4CCM111A Calculus I Lecture Notes

Dr. Paul P. Cook
Department of Mathematics, King's College London.

13th May 2020

"And what are these Fluxions? The Velocities of evanescent Increments? And what are these same evanescent Increments? They are neither finite Quantities nor Quantities infinitely small, nor yet nothing. May we not call them the ghosts of departed quantities?"

George Berkeley, in The Analyst, 1734.

0. Contents

1 Introduction **5**

1.1 Historic introduction 5

1.2 Course information and organisation 9

2 Functions of One Variable **15**

2.1 The Definition of a Function. 15

2.1.1 Domain and Range 17

2.1.2 Injections, Surjections and Bijections 20

2.2 Inverse functions 23

2.3 Some standard functions 27

2.3.1 Exponents, Logarithms and Polynomial Functions 27

2.3.2 The Logarithm 29

2.3.3 Polynomial Functions 31

2.3.4 The Exponential Function 33

2.3.5 The Trigonometric Functions. 36

2.3.6 The Hyperbolic Functions. 46

2.3.7 Trigonometric and Hyperbolic Functions with Complex Variables 52

2.3.8 Functions of Functions 58

2.4 A Zoo of Functions 59

3 The Limit **61**

3.1 Motivation 61

3.2 The Existence of the Limit. 64

3.2.1 Continuous Functions. 70

3.2.2 Limits Involving Infinity. 74

3.3 Working with Limits 75

3.3.1 Rules for Limits of Composite Expressions. 75

3.3.2 Multiple Limits. 80

3.3.3 Evaluating standard limits 81

4 The Derivative **93**
4.1 Differentiation from first principles 96
4.2 Differentiable Functions 104
4.3 Properties of the derivative 106
4.3.1 The Chain Rule 106
4.3.2 The Sum Rule 109
4.3.3 The Product Rule 109
4.3.4 The Quotient Rule 110
4.4 Derivatives of Implicit Functions 111
4.4.1 Inverse Functions 111
4.4.2 Curves 113
4.4.3 Parametric Functions 116
4.5 The Mean Value Theorem 119

5 Integration **123**
5.0.1 Motion at constant speed: the area under a straight line 124
5.1 The Riemann Integral 125
5.2 The Fundamental Theorem of Calculus 137
5.2.1 Indefinite and Definite Integrals. 142
5.3 Properties of the Integral and some Techniques for Integration 143
5.3.1 Solving Integrals by Substitution 145
5.3.2 Integration by Parts 152
5.3.3 Partial Fractions 155
5.3.4 Recursion Relations 159
5.4 Some Applications: Length, Area and Volume. 161
5.4.1 Areas of Circles and Ellipses 161
5.4.2 Volumes of Revolution 164
5.4.3 The Length of a Curve 166

6 Power Series **169**
6.1 Infinite sums 170
6.2 Convergence 170
6.2.1 Series Convergence Criteria. 173
6.3 Series as Functions of x. 175
6.4 Taylor's Theorem 180

6.4.1 Short-cuts for finding a Taylor Expansion 188
6.5 l'Hôpital's Rule 189

Appendix **193**
6.A The Ratio Test Implies the Root Test. 193

7 Tutorial Exercises **197**

Acknowledgements

These lecture notes were written in August-September 2015, with a major revision written in August-September 2018, and built upon the previous excellent set of notes written by Professor Ton Coolen, Professor Coolen's hand can be sensed throughout these notes and I am grateful to him for introducing me to this module in 2013. Each overhaul of a standard set of lecture notes builds its foundations on the previous courses and the work of Professor Peter Sollich and Dr Gérard Watts has had a profound impact on this iteration of the course. All typographical errors and errors of any other type are, of course, my fault and I will be grateful for all comments of any kind of mistake. Thanks are due to Jane Bennett-Rees, Robert Evans, Asuka Kumon, (Anthony) Peter Young, Michael Yiasemides, Haodong Sun, Shuo Huang, Pablo de Castro, Rishi Mouland, Alessio Sarti, Senan Sekhon, Keith Glennon, Veno Mramor, Véronique Fischer and many others for pointing out typos and making other comments on the text.

Paul P. Cook, 10^{th} September, 2018.

1. Introduction

In which we introduce and motivate the study of calculus; describe the course in outline; explain the organisation of the course and how it will be examined.

This will be covered during week 1 of lectures.

1.1 Historic introduction

Calculus is the mathematical study of change. Historically the subject was called the infinitesimal calculus or just the calculus and referred to the study of functions, limits, differentials (studying the gradients of curves) and integrals (studying the areas under curves). The major work in fundamental calculus was carried out in the 17^{th} century by Isaac Newton and Gottfried Leibniz and calculus marked the beginning of a new era of rigour in mathematics, it is the beginning of a part of mathematics known as analysis. To many, calculus is seen as the starting point of modern mathematics and it is apt to begin your university studies with it.

The fundamental questions that calculus and analysis are able to provide answers to were first posed thousands of years ago. Atomism, a branch of natural philosophy which developed in many ancient civilisations contends that physical reality is formed of atoms (this is not the same as the modern use of the word atom). The fundamental object is the atom which is considered to be indestructible and indivisible. For Plato, among others, who was an atomist and a geometer there arose the question of the atomism of mathematical objects. Is a line formed of a sequence of indivisible atoms? Plato thought yes although his student Aristotle was happy to discard the indivisibility of a continuous line. With the assumption of indivisible geometric atoms many paradoxes readily arise. For example, given a continuous line, we might ask how many atoms does it contain? Now suppose we posit that a continuous line is made of N atomic parts. A problem now arises as we can split a continuous line in geometrically in

two, forming two new continuous lines. Now if each continuous line contains N atoms then the pair of lines contain $2N$ atoms - we appear to have created indivisible atoms (and if we had joined lines together we would be apparently destroying indivisible atoms), which undermines the assumption of a finite number of geometric atoms making up any continuous line. There is another possibility: perhaps each continuous line contains an infinite number of indivisible parts, even lines of finite length - which gives a consistent picture but raises further questions. Especially a fundamental problem in constructing the calculus: how can an infinite number of infinitesimal quantities give a finite quantity? The most famous poser of such paradoxical problems was Zeno of Alea, and perhaps his most famous paradox is that of Achilles and the tortoise which is succinctly described by Aristotle:

> "In a race, the quickest runner can never overtake the slowest, since the pursuer must first reach the point whence the pursued started, so that the slower must always hold a lead."

It is really the same problem as that of indivisibles vs infinitesimals. The resolution is calculus which gives a rigorous way to deal with the algebra of infinitesimal quantities. Parts of the ancient world did make use of infinitesimal mathematics, although such methods were not trusted, for example Archimedes was able to find the volume of a sphere, a cone and a cylinder using his "method of indivisibles", but he verified his results using geometrical derivations as well. Zu Chongzhi and his son Zu Gengzhi were also able to make similar computations.

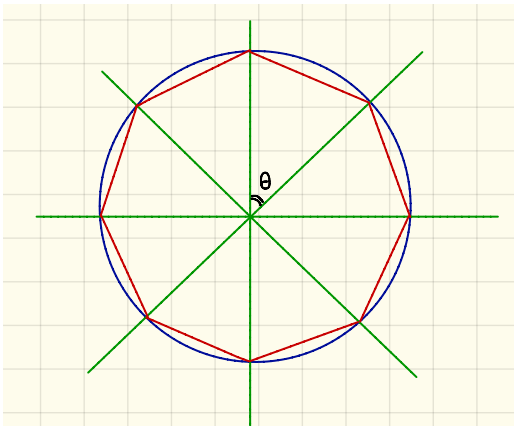
The question of the indivisibles or infinitesimals was not properly progressed for over a thousand years until Evangelista Torricelli and John Wallis's infinitesimals in the 17^{th} century. It is interesting to note how philosophically disturbing it was in that era to contemplate the idea that the mathematical world was underpinned by infinitesimal quantities rather than indivisible atoms. The solidity and surety of atoms of geometry sat easily with the idea of a stable universe, in which stable kingdoms are governed by unquestionable systems of law and religion. The idea of infinitesimals which were not mathematically well-understood was not welcome[1] and there are famous examples of respectable and historic figures who rejected the idea of an infinitesimal quantity, notably René Descartes (who first embraced and then turned away from infinitesimals) and George Berkeley who said that infinitesimals were

> "the ghosts of departed quantities."

The period that commenced with Newton and Leibniz saw the continual development of calculus and the branch of mathematics called analysis right up until 1910 when consensus was reached.

[1]There is an excellent book called Infinitesimal by Amir Alexander which discusses just how dangerous an idea the infinitesimal calculus was at the time.

The consensus was not concerned with whether a set of statements was true or false but focussed on the validity of the techniques used to make the arguments and their range of valid application. This was required in order to rigorously prove statements of calculus. It does seem peculiar that calculus can be founded properly in the 17^{th} century but discussions of its validity cannot be settled until the 20^{th} century. The reason is that calculus (and analysis) is concerned with the interplay between the very large (the infinities) and the very small (infinitesimal quantities) and as you will know infinities and zeroes can make even the most simple algebraic manipulation invalid or misleading. Consider the simple slicing up of disc into similar segments each of which subtends an angle θ at the centre of the circle:



The straight lines (in red above) can be summed to find the length of the circumference of the circle. As the angle θ is decreased the better the approximation of the circumference and in the limit where $\theta \to 0$ the approximation of circumference can be expected to become exact. However, in that limit, the length of the red lines goes to zero and we see we are arguing that an infinite sum of infinitesimal lengths is neither zero nor infinity but is finite. The central question is to understand under which circumstances this is the case.

Another similar example also highlights the need to resolve the question of how one works with infinitesimal quantities. Consider the square $ABCD$ whose edges are all of 1 metre long.

Imagine an ant travels the path along the edges AB and then BC, denote the path ABC. The distance travelled by the ant is 2 metres. Now consider a new path within the square $AB_1B_2B_3C$ whose edges all measure 0.5 metres as shown by the middle square in figure 1.1, i.e. $|AB_1| = |B_1B_2| = |B_2B_3| = |B_3C| = 0.5m$. The ant travels the path $AB_1B_2B_3C$ which has length $4 \times \frac{1}{2} = 2$ metres. Now repeat the process: half the length of each edge and form the staircase path whose edges all of length $\frac{1}{n}$ metres. There are now $2n$ edges of length $\frac{1}{n}$ giving a total path length of, again, 2 metres. No matter how small the edge length $\frac{1}{n}$ is made, the total path length remains 2 metres. However we recognise that as $n \to \infty$ the path travelled by the ant approaches the diagonal of the unit square, which has length $\sqrt{2}$ metres. So there is

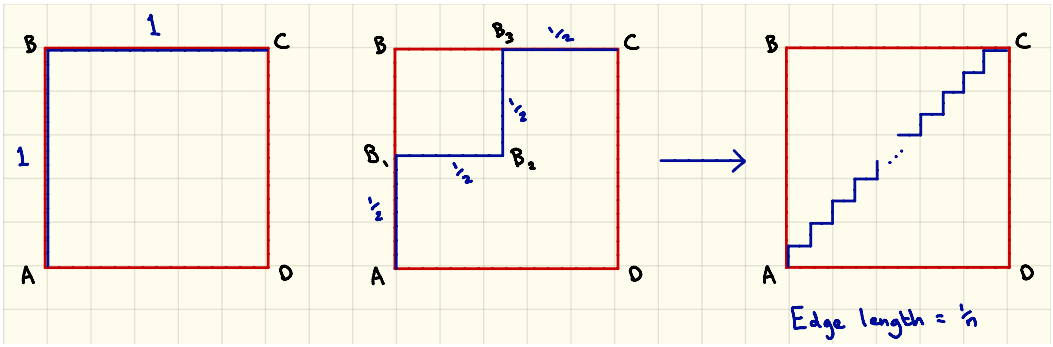


Figure 1.1: An ant travels along a staircase shaped path, the path always has length two metres, regardless of the number of steps.

a fundamental difference between the finite n case and the case where $n \to \infty$, and the process of taking the limit to infinity needs to be carefully constructed.

Consider a final example that was the subject of significant debate in the 18^{th} century, the series:

$$1 - 1 + 1 - 1 + 1 - 1 + 1 - 1 + \ldots .$$

The modern view is that this series is divergent as it does not tend to a limit, however you would be in the good company of Euler if you argued that it is an example of a geometric series of the form

$$1 + x + x^2 + x^3 + ...$$

(with $x = -1$) which sums to $\frac{1}{1-x} = \frac{1}{2}$. You would also be forgiven if, without a system of analysis, you argued that you could sum the series to zero or one.

A final confusion that we would like to emphasise here at the outset of our course was apparent already within the foundations of calculus and occurs when we consider two functions $f(x)$ and $g(x)$ which under a transformation $x \to x + dx$ such that $f \to f + df$ and $g \to g + dg$. The variation of a product of these functions is derived from the variations of f and g as

$$d(fg) = (f + df)(g + dg) - fg = f\,dg + g\,df + df\,dg$$

However Leibniz[2] argued that the infinitesimal variation of a product of the two functions is

$$d(fg) = f\,dg + g\,df$$

that is $df\,dg$ is treated as zero whereas df or dg is infinitesimally small but finite and not treated as zero - why is it valid to treat $df\,dg$ as zero? The answer to this question requires us to consider what we mean by a function and to develop a mathematical method that will

[2]This is Leibniz famous product rule, albeit written in a slightly unfamiliar notation.

allow us to work with finite but very small quantities rather than to attempt an algebra of zeros. The development of these techniques will pave the way for calculus. It will turn out that there are many mysteries which will need to be vanquished and even the most beguilingly simple expressions will acquire a clear definition in this quest. For example if you paused when reading the expression $f\,g$ and wondered just what is a product of functions - you would be asking a good question, one that we will answer as the course develops.

1.2 Course information and organisation

This course covers the basic methods of one variable calculus. Our major topic will be the definition of the function: we will pursue a purist's definition of a function which will ultimately lead us to the analytic function at the end of the course. Our emphasis in the first half of the course is on introducing fundamental concepts while at the same time developing our competence and fluency in a number of standard mathematical techniques. Although the material is largely quite elementary, it is fundamental to the development of calculus. An approximate indication of the programme and the material to be covered is shown in table 1.1.

Course tutorials

The tutorials will start in the second week of the teaching term.

Your tutorial group will be shown on your personal timetable where you will also find the the classroom number for your group. Attendance at tutorials is registered, if you fail to attend and/or fail to hand in solutions to tutorial exercises repeatedly you may **not** be allowed to sit the January examination. Of course, more importantly, by not trying the problems you will not get to enjoy many of the intricacies of analysis.

Tests and examinations

(i) There will be three class tests each lasting 30 minutes, taking place throughout the semester and beginning sometime after week 4 of the lectures (please check your personal timetable to find the precise dates and locations of your class tests). The three exams contribute 20% of your overall mark.

(ii) There will be a final (2 hour) examination in the early January, counting for 80% of the overall mark. Make sure you able to sit the exam in January (i.e. don't book travel tickets before the exam dates are announced - usually in November).

There will be no resits for class tests. See the KEATS page on class tests and course website for more details. Class test results will be announced on the class test KEATS page and solutions will also be published there.

Reading week

Reading week occurs in the middle of the first teaching semester - please check the current term dates to be certain which week is reading week. It is a week of no scheduled study and is intended as a time for catching up the material on all your courses: there are no lectures or tutorials classes or office hours in this week.

Tutorials

- Your tutorial group will normally consist of about 20 students. Tutorials are not like mini-lectures, but more like small workshops; they involve and demand active participation from the group.

- The idea for the tutorials is that at the start of the first tutorial you form small groups of (say) 3-4 people, and work together in these small groups for the rest of the semester. Each week you work during the tutorial hour on the problems set for that particular week. You will find these problems at the end of the course lecture notes.

- The teaching assistant will go around in the room and make herself/himself available to anyone who gets stuck. Their role, however, is not to give complete 'model solutions'; you learn more from twenty minutes struggling with a problem than from two minutes of being being told by someone else how it is done. Discussing problems is also one of the most efficient ways of putting new material into your long-term memory. The assistant will only go through model solutions on the board if it turns out that there are particular problems everyone got wrong (in her/his judgement).

- You will be expected to each present some solution or partial solution to a problem on the board to your classmates at some point in the tutorials. This is to give you some experience of presenting mathematics to a group, a skill which is high demand in the modern world (both academic and otherwise).

- You will usually not be able to do all exercises during the tutorial (they tend to be too hard and/or too many for that), but you will submit your solutions to those question marked with $\star$ (written and explained properly) to your tutor at the end of the tutorial. Normally you will need to commence writing your solutions in advance of the tutorial.

- After you have handed in your work for a particular week, the solutions will be made available for you to download from the Calculus I page on KEATS. The teaching assistant will inspect your solutions and return the work to you at your next tutorial.
- Experience shows that if you have understood the tutorial exercises (perhaps not in the week you first tried to do them, but upon further study), you will also be successful in the Calculus I class tests and in the January examination.
- You will know your tutorial group allocation from your personal timetable. But if you are uncertain which group you should be in please write to ug-mathematics@kcl.ac.uk. Always inspect regularly the Calculus I KEATS page to be updated on any changes.

Calculus I Course Materials and Reading List

There is **no required reading** beyond the lecture notes for this module.

This module does not follow any particular book but lecture notes will be are available via the KEATS page for Calculus I. You will be guided to read through the lecture notes in advance of the lectures. The major ideas will be covered in the lectures so it is important you have read ahead in order to be able to digest the lectures.

In addition, there are many books which you might find helpful for the module to provide background, alternative explanations, and generally supplement the lectures. The typical US-style introductory Calculus book has on the order of 1000 pages, is rather heavy to carry around, and goes a long way beyond this module (but will not necessarily cover all the material either). Having said that, such books can be very well written and also useful for Calculus II next semester and also possibly for Analysis I or Joint Honours Analysis. The standard calculus books tend to be expensive to buy. The library will keep reference copies of all major texts for the course, so please review any book in the library. There are many to choose from, mostly located in section QA303 in the library for example

- Calculus (one and several variables) by Salas, Hille and Etgen,
- Maths - A Student's Survival Guide by Jenny Olive,
- Calculus by Michael Spivak, or
- Calculus by James Stewart.

The book by Jenny Olive (which is available online from the library via the "Resource List" on the right-hand-side of the Calculus I KEATS page) is written in a very informal style, and intended for 'science' students, rather than mathematics students; in parts it is rather basic, as

well as missing out some of the more advanced topics, but it has well written sections on most of the material covered in this module and provides many examples - it also has hamsters. The book by Michael Spivak has yellow pigs. The book by James Stewart is extremely popular.

A fourth book which might also prove useful is

- Engineering Mathematics by K.A. Stroud.

Again this is not intended for mathematics students, but covers almost all the topics in the module in a very straightforward and clear way with many worked examples, and might prove useful if you find the other books too technical or need to see more examples.

There are a number of other very good books introducing more advanced topics which also cover in their introductory chapters some of the material in the Calculus 1 course. In particular complex numbers are not usually covered as part of a first course in calculus and you may not find complex numbers discussed in the books above. There are many excellent books on complex analysis which tend to begin with a clear introduction to complex numbers. I recommend

- Introduction to Complex Analysis by H. A. Priestley and
- Visual Complex Analysis by T. Needham

Introduction to Complex Analysis by H. A. Priestley is an excellent book whose first two chapters contain a review of the complex numbers and the complex plane. It contains a number of exercises which might prove a useful resource for those of you seeking extra questions concerning geometry in the complex plane. Visual Complex Analysis is a beautiful book which will give you an excellent introduction to complex analysis, a subject beyond the scope of our course, but you will find many common themes and topics which overlap with our course in this book.

As a complement to our course, you might wish to enrol on the Calculus One course on Coursera taught by Jim Fowler of the University of Utah. The course is available for free on Coursera at `https://www.coursera.org/learn/calculus1`, and is always open for registration. The content of this course overlaps and complements nicely many of the topics in our course - but of course the material in the coursera course will not be examined! Nor is it required watching for our course. Nevertheless I think it offers a good and fun way to keep up with and revise many topics in our course.

	Topics
Week 1	Historical introduction, course organisation, functions, domain, range, injections, surjections and bijections.
Week 2	Inverse functions, exponents, logarithms, polynomial functions and trigonometric functions.
Week 3	The inverse trigonometric functions, the hyperbolic functions, trigonometric functions with complex arguments, double-angle formulae and function composition.
Week 4	The limit, the $\epsilon - \delta$ definition of the limit, continuous functions, the intermediate value theorem, limits involving infinity and rules for working with sums and products of limits.
Week 5	More work with limits: limits of composite functions, multiple limits standard limits, the sandwich theorem and logarithmic vs polynomial vs exponential limits.
Week 6	**Reading week.**
Week 7	Differentiation from first principles, differentiable functions, properties of the derivative, the chain rule, the sum rule, the product rule and the quotient rule.
Week 8	Derivatives of implicit functions, derivatives of inverse functions, derivatives of parametric functions and the mean value theorem.
Week 9	Integration, the Riemann integral and the fundamental theorems of calculus.
Week 10	Indefinite and definite integrals, integration by substitution, integration by parts, recursion relations, partial fractions and surface areas.
Week 11	Volumes of revolution, the length of a curve, infinite sums, convergence criteria and power series.
Week 12	Taylor's theorem, l'Hôpital's rule and analytic functions.

Table 1.1: The approximate organisation of the course lectures.

2. Functions of One Variable

In which we introduce the most important character in the course: the function. We will define functions of one variable as a map and we will build up a catalogue of functions and study their properties. In our thinking we will come across the infinite sum which will foreshadow the idea of an analytic function which we will chase until we meet it formally at the culmination of the course.

This will be covered during weeks 1-3 of lectures.

2.1 The Definition of a Function.

Functions are some of the most simple objects in mathematics and yet remain the cornerstone of study in many areas of modern mathematics.

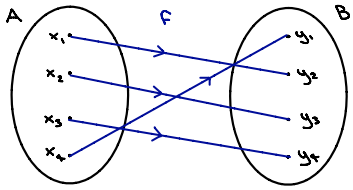
Definition 2.1.1. *A function $f : A \to B$ is a map from a set A to a set B that associates a unique element in B to each element in A.*

Comment(s). *(On functions...)*

1. *Practically the set A indicates the possible inputs to a function, while the set B specifies where the output resides.*

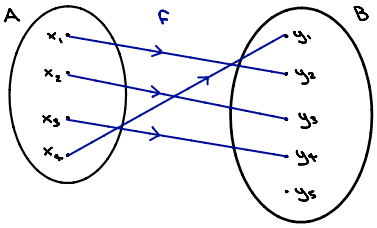
2. *The definition above defines the qualities of any function, but to define a particular function one must state how it acts on each element of A. Mathematically we write the action of a function which associates an element $x \in A$ to an element $y \in B$ as $f : x \mapsto y$. N.B. The symbol $\mapsto$ is used when a map is acting on a single element in a set, while the symbol $\to$ is used to indicate a map between sets of elements.*

3. *We might define a function $f : A \to B$ pictorially by drawing arrows between elements of A and elements of B indicating the action of f, e.g.*



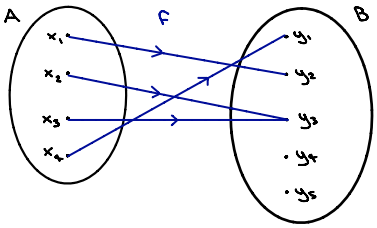
describes the function $f : A \to B$ where $A = \{x_1, x_2, x_3, x_4\}$ and $B = \{y_1, y_2, y_3, y_4\}$ given by $f : x_1 \mapsto y_2$, $f : x_2 \mapsto y_3$, $f : x_3 \mapsto y_4$ and $f : x_4 \mapsto y_1$. Defining functions in this way is only convenient when A and B are sets of low order.

4. *A function $f : A \to B$ acts on all elements in A to give an element in B, but it is not necessary that every element in B is associated with an element in A by a function. For example the map indicated by*

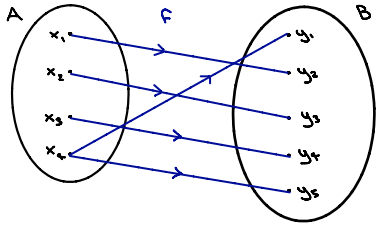


where now $B = \{y_1, y_2, y_3, y_4, y_5\}$ is a well-defined function even though there is no element in A which is mapped to y_5 in B.

5. *The definition of a function $f : A \to B$ requires that the output of f is a unique element in B, so for example the map defined by*



is a well-defined function, whereas the map defined by



is not a well-defined function as it does not map x_4 to a unique element in B. The definition implies that functions may be many-to-one maps, but they cannot be one-to-many maps.

2.1.1 Domain and Range

Definition 2.1.2. *For a function $f : A \to B$, the set A is called the domain of f and the set B is called the range of f.*

Comment(s). *(On the domain and range...)*

1. *There are many names which are commonly used for the domain and range of a function, the most common alternatives are the co-domain or image for the range and correspondingly the pre-image for the domain*[1]*. The word image has a more precise meaning as the set of all elements $f(x)$ for all $x \in A$, note that the image of a function must always be contained within the range of a well-defined function.*

2. *It is natural to define the range of f in an abstract way, as $f(x)$ for all elements $x \in A$. Hence we might define a function as $f : A \to B$ by $f : x \mapsto f(x)$ for all $x \in A$, where $y = f(x) \in B$.*

3. *Where the domain and range are known, we might swiftly refer to a function as $f(x)$ where x lies in the domain. This is a useful algebraic notation that will enable us to define functions whose domain and ranges are infinite sets, such as the number sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$.*

4. *The domain and range of a function might be unions of sets, for example*

$$f : \mathbb{R} \to (-\infty, -3] \cup [7, \infty), \qquad f(x) = \begin{cases} x^2 + 7 & x > 0 \\ x - 3 & x \leq 0 \end{cases}.$$

5. *The range B of a function $f : A \to B$ needs only to contain the elements $f(x)$ for each element $x \in A$. In particular, as we saw earlier, the range B may contain elements which are not mapped to by f. For example,*

$$f : \mathbb{R} \to \mathbb{R}, \qquad f(x) = \begin{cases} x^2 + 7 & x > 0 \\ x - 3 & x \leq 0 \end{cases}$$

[1]Note that conventions differ in the definition of the range, we have identified the range and co-domain in this course, however another common convention is to identify the range with the image. It is very common for fundamental definitions to differ in mathematics, we take the view of Shakespeare's Juliet when she said 'what's in a name? That which we call a rose by any other name would smell as sweet.'

is well-defined.

When defining a function it is necessary to specify the domain and range. The same operation applied over different ranges and domains gives a different function. As we will see in some of the following examples the same functional operation defined over different domains and ranges may not even give a well-defined function.

Example 2.1. *State whether the following function is well-defined:*

$$f : \mathbb{R} \to \mathbb{R}, \qquad f(x) = x + 2.$$

The function defined above is very simple: it translates the real number line by $+2$. It is because we have this picture in mind that we are immediately content that the output is unique, so this is a well-defined function. The specification of the domain tells us that $x \in \mathbb{R}$, and the definition of the range that $f(x) \in \mathbb{R}$. Evidently various parts of the definition of the function are independent, it is up to us as mathematicians to check that the domain and range are suitable for any explicitly defined function $f(x)$. There are many ways we could pick the domain and range so that this function is not well-defined, e.g. if we had changed the range so that $f : \mathbb{R} \to \mathbb{Z}$, for $f(x) = x + 2$ the range would simply be incorrect.

Example 2.2. *Check that the following function is well-defined:*

$$f : [0, \infty) \to [0, 12] \cup (14, 16) \cup \{9\}, \qquad f(x) = \begin{cases} 3x & \text{for } x \in [0, 4] \\ 2x + 6 & \text{for } x \in (4, 5) \\ 9 & \text{for } x \geq 5. \end{cases}$$

Here we have a complicated definition for a function, which is perfectly valid (as you should check). Notice that the notation for an open interval for all positive numbers is written $[0, \infty)$, it is an open set in the positive direction which never reaches ∞. As ∞ is not a real number [2].

Example 2.3. *Is*

$$f : \mathbb{R} \to \mathbb{R}, \qquad f(x) = \pm\sqrt{x}$$

a well-defined function?

---

[2] It is possible to extend the real number line to include $\pm\infty$. This is denoted $\mathbb{R} \cup \{-\infty, +\infty\}$ or sometimes $\overline{\mathbb{R}}$ and is called the affinely extended real number line.

It is not well-defined for two reasons. First $f(x)$ gives two outputs for every input. So let us correct this and redefine the function as

$$f : \mathbb{R} \to \mathbb{R}, \qquad f(x) = \sqrt{x}.$$

We emphasise that the notation $\sqrt{x}$ denotes the principal square root of a number, i.e. its positive square root. The function is still not well-defined for if $x < 0$ then $f(x) \notin \mathbb{R}$. However we may again amend the definition to

$$f : \mathbb{R}^+ \to \mathbb{R}, \qquad f(x) = \sqrt{x}.$$

This is well-defined, even though there is some redundancy in the range, we could have tightened the definition to

$$f : \mathbb{R}^+ \to \mathbb{R}^+, \qquad f(x) = \sqrt{x}$$

but it is not strictly necessary.

We have restricted our examples to simple algebraic functions but we could certainly have constructed well-defined functions of a more abstract and curious nature e.g.

$$f : \mathbb{R} \to \{0, 1\}, \qquad f(x) = \begin{cases} 1, & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases}.$$

This function, although horrible, is well-known in the development of mathematical analysis[3] as it is discontinuous at every point x. We will meet the definition of continuity later in the course, which will help us appreciate the horror of the above function. We could also have defined interesting functions in less mathematical language, for example the function which outputs the number of results when one googles an integer. The outputs of this function is shown in a log graph in figure 2.1. This graph was made by Brian Hayes and comes from his blog where there is a detailed and interesting discussion of the patterns in the output, see `http://bit-player.org/2014/600613`. By this point in our discussion of functions we may begin to feel that there are an overwhelming range of functions, and we would be right. Intuitively we have sense that certain functions are "nicer" than others - those that we can sketch a smooth graph of, those whose derivatives are well-defined, while there are many others which are less palatable such as the google hits function above. The desire to put functions into classes based on their mathematical behaviour will motivate much of our work in these lectures.

[3]It is called the Dirichlet function after Peter Gustav Lejeune Dirichlet.

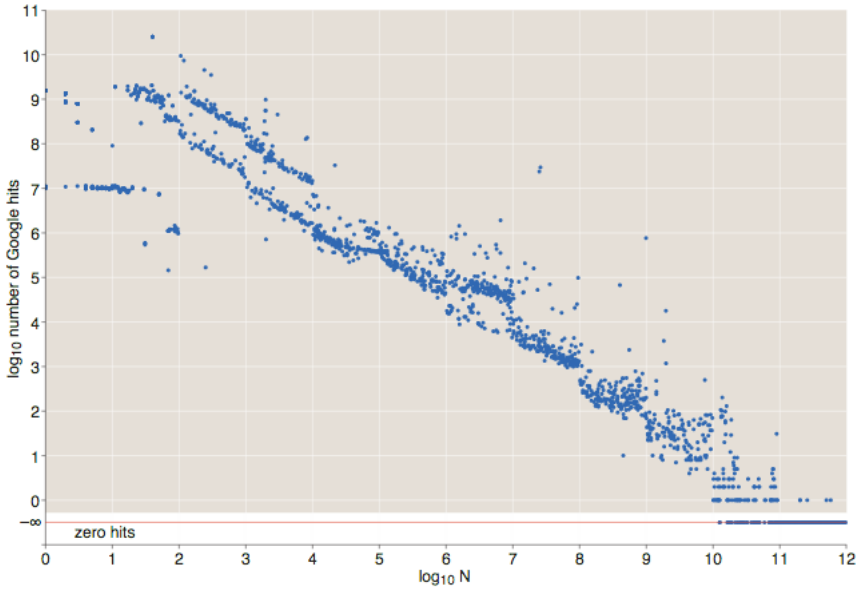


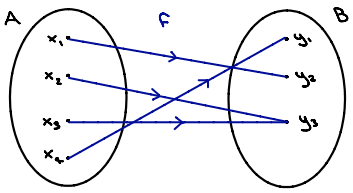
Figure 2.1: A log plot of the number of google hits of the integers from 0 to 10^{12}.

2.1.2 Injections, Surjections and Bijections

We will begin our classification of functions with the following definitions.

Definition 2.1.3. *A function $f : A \to B$ is surjective (or onto) if every element in B is mapped to by at least one element of A. A surjective function is called a surjection.*

For example the following defines a surjective function as every element of B is mapped to from A:

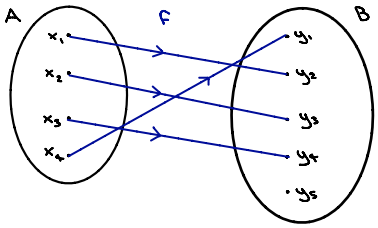


We can write the definition of a surjective function in notation as follows. If, for $f : A \to B$, $\forall y \in B \, \exists x \in A$ such that $y = f(x)$ then f is a surjection.

Definition 2.1.4. *A function $f : A \to B$ is injective (or one-to-one) if every element $f(x) \in B$ is mapped to by at most one element in the domain A. An injective function is called an injection.*

For example the following defines an injective function as each element of A is associated

with a unique element of B:



We can write the definition of an injection in mathematical notation as follows. If, for $f : A \to B$, $\forall\, x_1, x_2 \in A$, $f(x_1) = f(x_2) \implies x_1 = x_2$ then $f(x)$ is an injective function.

Example 2.4. *State which, if any, of the following functions are injective, surjective or both?*

(a) $f_1 : \mathbb{R} \to \mathbb{R}$ *given by* $f_1(x) = x^2$

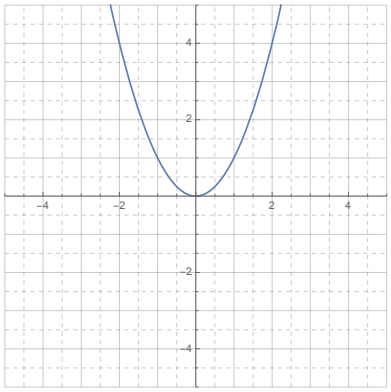
(b) $f_2 : \mathbb{R} \to \mathbb{R}$ *given by* $f_2(x) = x^3$

(c) $f_3 : \mathbb{R}^+ \to \mathbb{R}^+$ *given by* $f_3(x) = \sqrt{x}$

(d) $f_4 : \mathbb{R} \to \mathbb{R}$ *given by* $f_4(x) = 2^x$

Let's consider each function in turn.

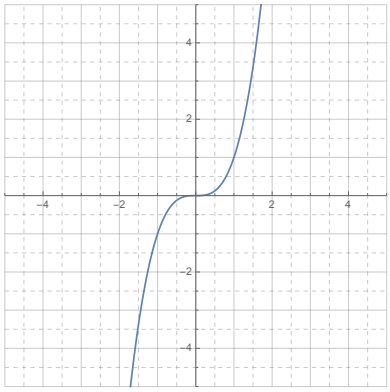
(a) The graph of f_1 *is*



The graph shows the function as a map from a point on the x-axis to a point on the y-axis and these points are paired together to give the coordinates of the curve we have sketched. If you wanted to draw an arrow from a real number on the x-axis to find its image in the real numbers on the y-axis you would draw a vertical line from the number on the x-axis until it meets the curve at a point and then draw a horizontal line from that point until it meets the y-axis. Now we can ask if f_1 *is surjective: is every number on the y-axis in*

the image of f_1? We can see from the graph, which is always positive, that none of the negative numbers on the y-axis are in the image of the function. To prove the function is not surjective we need only show that one element in $\mathbb{R}$ (which is the specified range of f_1) is not in the image of f_1, e.g. suppose that -1 was in the image of f_1 then there would exist some fixed value $x_0 \in \mathbb{R}$ such that $-1 \stackrel{!}{=} x_0^2$ as every real number squares to a positive real number we see that $x_0 \notin \mathbb{R}$. This contradicts the assumption that $x_0 \in \mathbb{R}$ and is an example of a proof by contradiction: we have made an assumption and then developed an expression logically using that assumption and shows that it leads to a contradiction - the only possibility is that our assumption was not correct. Hence there is no element in $\mathbb{R}$ that is mapped to -1 by f_1 hence f_1 does not map onto all of $\mathbb{R}$, so f_1 is not surjective. From the graph we can also quickly deduce that f_1 is not injective either, as both x and $-x$ are mapped to the same point on the y-axis as the curve is symmetric about the line $x = 0$. As two different numbers are mapped to the same number on the y-axis we see that the function is not one-to-one. If we had wanted to prove that f_1 is not injective algebraically we would need only to give one example where the map is not one-to-one (e.g. $f_1(2) = 4 = f_1(-2)$).

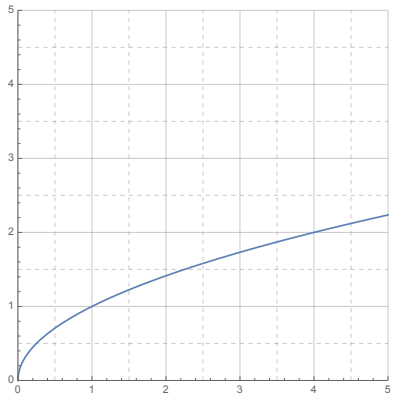
(b) The graph of f_2 is



which indicates that f_2 is both surjective and injective. To prove that f_2 is surjective then we need to show that for any value $y \in \mathbb{R}$ we can identify a value $x \in \mathbb{R}$ for which $f_2(x) = y$, i.e. we must solve $x^3 = y$ for x in terms of y. We readily obtain that $x = y^{1/3}$ which exists in $\mathbb{R}$. To prove injectivity, we may use a proof by contradiction. Suppose there exists $x_1 \neq x_2$ both in $\mathbb{R}$ such that $f_2(x_1) = f_2(x_2)$ then this implies that $x_1^3 = x_2^3$ and by cube-rooting both sides[4] we have $x_1 = x_2$, contradicting our assumption, hence f_2 must be an injective function.

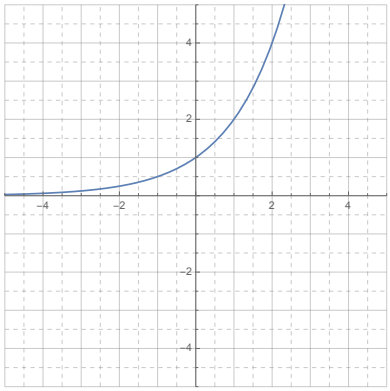
[4]Taking the cube-root, or raising to the power of one-third is a one-to-one map

(c) The graph of f_3 is



The square-root function is both injective and surjective (defined as a map from $\mathbb{R}^+ \to \mathbb{R}^+$).

(d) The graph of f_4 is



The function f_4 is injective but not surjective.

Definition 2.1.5. *A function $f : A \to B$ is bijective if it is both surjective and injective. A bijective function is called a bijection.*

In the examples above both f_2 and f_3 are bijections. Bijections are very useful functions because by being both injective and surjective between two sets A and B it means that every element in A is uniquely paired with an element in B and vice versa. This implies that a bijective function can be inverted.

2.2 Inverse functions

Definition 2.2.1. *The inverse function, denoted f^{-1}, of a bijective function $f : A \to B$ is a function defined by*

$$f^{-1} : B \to A, \qquad f^{-1}(f(a)) = a \qquad \forall\, a \in A.$$

Comment(s). *(On inverse functions...)*

1. *The notation is potentially confusing: f^{-1} is a special exception to the exponent notation and one should be careful to make sure its meaning is clear. In general, and throughout this course, f^{-1}, where f is a function, does not signify the reciprocal $\frac{1}{f}$ but the inverse function, e.g. if $f(x) = x^2$ then $f^{-1}(x) = \sqrt{x}$.*

2. *For a function to be invertible it must be both surjective and injective.*

   - *Consider the example of the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$, which is not injective as, for $x \neq 0$, the distinct values x and $-x$ are both mapped to x^2. Consequently if we attempted to invert the map we would not know whether x^2 originated from x or $-x$. For an injective function, its mapping is one-to-one by definition, so if $f(x) = y$ and f is injective then we can construct the inverse function as that function function f^{-1} which maps $y \mapsto x$, and so $f^{-1}(f(x)) = x$ for all x in the domain of f. Injectivity allows the inverse map to be constructed for those elements in the image of f.*

   - *Now consider an example of a function which is one-to-one but is not surjective: $f : \mathbb{R}^+ \to \mathbb{R}$ given by $f : x \mapsto \sqrt{x}$. This function is one-to-one, as suppose there exists $x_1 \neq x_2$ such that $f(x_1) = f(x_2)$ then this implies that $\sqrt{x_1} = \sqrt{x_2}$ and upon squaring this statement we find that $x_1 = x_2$, contradicting our assumption that f is not injective, so f is an injective function as we claimed. However f is not surjective as the image of f is $\mathbb{R}^+$, a subset of the range of f. Mathematically, we may prove that f is not surjective by identifying an element of the range of f, i.e. $\mathbb{R}$, which is not mapped to by f, in this example any negative number will do to prove that f is not surjective. Now what problems does this cause when we attempt to define the inverse function? To construct the inverse function we invert the map from elements of A to elements of B, so if the map is not surjective then there are some elements of B which are not associated with elements of A by the function. In this example the function does not map onto any of the negative numbers in $\mathbb{R}$ so the inverse function is not defined on any of the negative numbers, hence as a function $f^{-1} : \mathbb{R} \to \mathbb{R}^+$ is not well-defined.*

3. *On the other hand if a function $f : A \to B$ is bijective then it describes a one-to-one correspondence between elements of A and all elements of B, which we can use to construct the inverse function which undoes the work of f and takes us from B to A. The inverse function of a bijective function is also a bijection.*

Now if we are given a bijection how do we find the inverse function? We can do it in three steps:

(i) Write $y \equiv f(x)$.

(ii) Rearrange the equation to obtain $x = g(y)$.

(iii) Then as $y = f(x) = f(g(y))$ then $g(y) = f^{-1}(y)$, or changing variable label $g(x) = f^{-1}(x)$.

This procedure is straightforward, for example, if we were trying to find the inverse function to $f(x) = x + 4$ where $x \in \mathbb{R}$, we would write

$$y = x + 4 \implies x = y - 4 \equiv g(y) = f^{-1}(y)$$

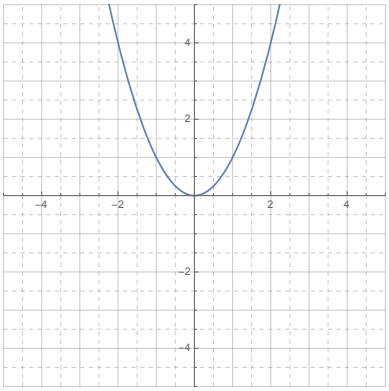
therefore $f^{-1}(x) = x - 4$ as expected. However it is worth practising further for yourself so you are happy that all the steps are sensible.

Exercise 2.1. *Find the inverse function $f^{-1}(x)$ for*

1. $f(x) = 6x + 4$ *for* $x \in \mathbb{R}$ *and*
2. $f(x) = \sqrt{x} + 5$ *for* $x \in \mathbb{R}^+$.

Many common functions are not bijections, but we can restrict the domain and range of their definition so as to construct a bijection which can then be inverted on the restricted domain and range. We will consider a simple example in this section and then in later sections we will look at more involved inverse functions.

Let us consider the quadratic function $f : \mathbb{R} \to \mathbb{R}$ given by $f : x \mapsto x^2$, whose graph is

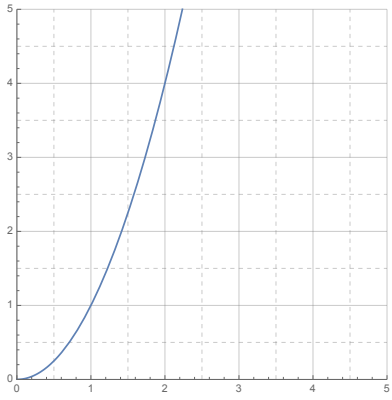


.

The claim that we cannot invert this function is equivalent to stating that $f^{-1} : \mathbb{R} \to \mathbb{R}$ given by $f^{-1}(x) = \sqrt{x}$ is not a well-defined function. We can see why, the (real) square-root is not defined on the negative real numbers. The function f fails to be a bijection because it is not

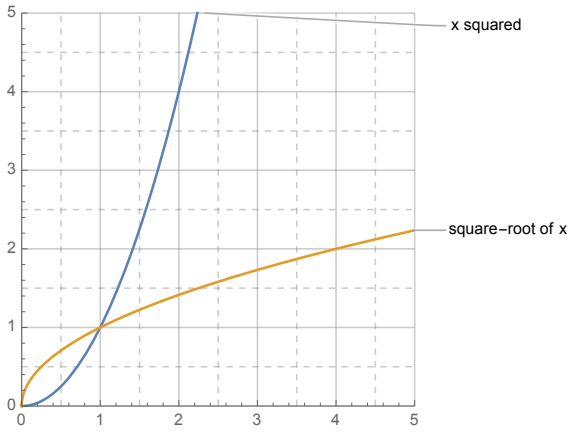
injective nor is it surjective onto $\mathbb{R}^+$. We can see both these observations from the graph. The image of f is $\mathbb{R}^+$, so that if we were to change the definition of f and restrict its range as $f : \mathbb{R} \to \mathbb{R}^+$ by $f : x \mapsto x^2$, then the new function would be surjective. Furthermore we can see from the graph that it is generally a two-to-one function (apart from at zero): if we drew a horizontal line across the graph for any non-zero, positive value on the y-axis, the line would meet the function at two points. However if we restricted the domain of the function's definition to, for example, just $\mathbb{R}^+$, it would become an injective function. In other words $f : \mathbb{R}^+ \to \mathbb{R}^+$ given by $f : x \mapsto x^2$ is both an injective and a surjective function, so it is a bijection and its inverse function exists and is given by $f^{-1} : \mathbb{R}^+ \to \mathbb{R}^+$ such that $f^{-1}(f(x)) = x$ for all $x \in \mathbb{R}^+$. This is a function we have a short-hand for, namely it is given by $f^{-1}(x) = \sqrt{x}$ - and this is the way the square-root function is constructed.

Before moving on, let's consider the graph of the bijective function $f : \mathbb{R}^+ \to \mathbb{R}^+$ given by $f : x \mapsto x^2$:



This is the restriction of the parabola $f(x) = x^2$ constructed over $\mathbb{R}$, and compared to the graph above we see that by restricting the domain we have effectively cut the function at its turning point and by stopping it turning we have ensured that what remains is a one-to-one function. This idea of restricting a function's domain to lie between its turning points so as to make it injective on the new domain is how we will set about constructing more complicated inverse functions. We will make one final remark about inverse functions here, as $f : A \to B$ and $f^{-1} : B \to A$ the graph of the inverse function is given by the graph of the original function but with the x-axis and y-axis interchanged. Hence if we to plot both f and f^{-1} on the same set of axes we would find that f^{-1} is the reflection of f in the line $y = x$. For the example we have considered here we plot both f and f^{-1} below so we can see that one is a reflection of the

other.



restricted the y-axis to the

Now we will turn our attention to trying to write down a large set of the better-behaved functions.

2.3 Some standard functions

2.3.1 Exponents, Logarithms and Polynomial Functions

Our aim in this section is to introduce the polynomial functions, as well as the exponential function and the logarithm. However we will first take some time to discuss the operation of raising a number to a power to illustrate a common technique in mathematics, that of developing a simple definition in a straightforward setting so that it can be applied in a more general setting. Our aim will be to start with the definition of a real number raised to the power of a positive integer and use this to arrive at a definition when the power is also a real number.

Definition 2.3.1. *In the expression x^n, n is called the power (or exponent) and x is called the base and it denotes the n-fold multiplicative product of x, where $x \in \mathbb{R}$ and $n \in \mathbb{Z}^+$, i.e.*

$$x^n = \underbrace{x \times x \times x \ldots \times x}_{n\text{-fold product of } x\text{'s}}$$

Comment(s). *The notation x^n,to denote the n'th power of x was first introduced by Descartes in his book Géométrie, published in 1637=$(1637)^1$ (1637 is a prime number so its prime factorisation, written in exponent notation, is 1637^1).*

It is clear from the definition of x^n that n must be a positive integer, what does it mean to multiply a number by itself a negative number of times? Or a fraction number of times? Or even zero times?

The definition allows us to develop some rules for multiplying powers of numbers, informed by our understanding of multiplication of the real numbers, but written in the short exponent notation.

$$x^n \times x^m = \underbrace{x \times x \times x \ldots \times x}_{n\text{-fold product of } x\text{'s}} \times \underbrace{x \times x \times x \ldots \times x}_{m\text{-fold product of } x\text{'s}} = \underbrace{x \times x \times x \ldots \times x}_{(n+m)\text{-fold product of } x\text{'s}} = x^{n+m}$$

where $n, m \in \mathbb{Z}^+$.

We may also use the exponent notation to divide powers of x:

$$\frac{x^n}{x^m} = \frac{\overbrace{x \times x \times x \ldots \times x}^{n\text{-fold product of } x\text{'s}}}{\underbrace{x \times x \times x \ldots \times x}_{m\text{-fold product of } x\text{'s}}} = \underbrace{x \times x \times x \ldots \times x}_{(n-m)\text{-fold product of } x\text{'s}} = x^{n-m} \quad \text{if } (n-m) \in \mathbb{Z}^+.$$

This last condition is a little restrictive and comes from our definition of x^n which is only defined when n is a positive integer. Using the rule for multiplication of powers we realise that we can give meaning to x^{-m} $(m \in \mathbb{Z}^+)$ as

$$x^{n-m} = x^n \times x^{-m} = \frac{x^n}{x^m}.$$

Hence we have constructed a logical definition of for the meaning of x^n when $n \in \mathbb{Z}^-$, namely:

$$x^{-n} \equiv \frac{1}{\underbrace{x \times x \times x \ldots \times x}_{n\text{-fold product of } x\text{'s}}}.$$

To recap, we have extended the definition of x^n from n being a positive integer, to n being a positive or a negative integer. From this the definition of x^0 also follows quickly as

$$x^0 \equiv x^{n-n} = x^n \times x^{-n} = \frac{\overbrace{x \times x \ldots \times x}^{n\text{-fold product of } x\text{'s}}}{\underbrace{x \times x \ldots \times x}_{n\text{-fold product of } x\text{'s}}} = 1.$$

Consequently we have extended our definition of x^n to any integer power, i.e. when $n \in \mathbb{Z}$.

Turning again to our knowledge and experience of multiplication of the real numbers we also understand that

$$(x^m)^n = \underbrace{x^m \times x^m \ldots \times x^m}_{n} = \overbrace{\underbrace{(x \times x \times x \ldots \times x)}_{m} \times \underbrace{(x \times x \times x \ldots \times x)}_{m} \times \ldots \underbrace{(x \times x \times x \ldots \times x)}_{m}}^{\text{n}} = x^{m \times n}.$$

Comment(s). *It is customary to drop the multiplication symbol algebraically and write $m \times n = mn$, where, as here, when the multiplication rule is obvious. Hence one writes $(x^m)^n = x^{mn} = x^{nm}$.*

We can now extend the notion of the exponent to any rational power ($\mathbb{Q}$) by noting that

$$(x^{p/q})^q = x^{qp/q} = x^p.$$

Therefore $(x^{p/q})$ is the q'th root of x^p, e.g. if $p = 1$ and $q = 2$ then $x^{1/2} = \sqrt[2]{x} \equiv \sqrt{x}$ is the square root, or if $p = 1$ and $q = 3$ then $x^{1/3} = \sqrt[3]{x}$ is the cube-root, or if $p = 2$ and $q = 5$ then $x^{2/5} = (x^2)^{1/5} = \sqrt[5]{x^2}$. With this interpretation we have extended the range of our definition of x^n to $n \in \mathbb{Q}$.

Exercise 2.2. *Check that the laws for the manipulation of powers (i.e.* $(x^n)(x^m) = x^{n+m}$ *and* $(x^m)^n = x^{nm}$*) are valid for* $n, m \in \mathbb{Q}$.

It is a little more challenging to go further and generalise to real powers of x, i.e. when $n \in \mathbb{R}$. We can use the definition of the exponent when $n \in \mathbb{Q}$ to identify a rational number that is arbitrarily close to any real number. Specifically given a degree of accuracy we can always identify a rational number p/q that is within that degree of accuracy of any irrational real number[5]. For example, suppose we wish to express π as a fraction to two decimal places then $3.14 = \frac{314}{100} = \frac{157}{50}$ and now we can give meaning to $x^{157/50}$ which is $x^{\pi(\text{up to 2.d.p.})}$. If we wanted to find a better approximation of x^π we can always find a more accurate rational approximation of π, e.g. to eight decimal places $3.14159265 = \frac{314159265}{10^8}$. As the rational approximations improve we approach the meaning of x^π, although it seems we never meet it. This may worry you, but such a definition involving a limiting procedure can be put on a definite footing and the subject of limits will be a major topic later in this course.

2.3.2 The Logarithm

Definition 2.3.2. *The logarithmic function, or logarithm, to the base* a*, written as* $\log_a(x)$*, is the inverse of the function* $f(x) = a^x$ *where* $x \in \mathbb{R}^+$ *and* $a \in \mathbb{R}^+ \setminus 1$.

Comment(s). *(On the logarithm...)*

1. *In other words, '*$\log_a(y)$ *gives the power to which I must raise* a *to get* y*' i.e.* $\log_a(a^x) = x$.

2. *The graph of* $y = \log_2 x$ *is shown in figure 2.2. This is the inverse function of* 2^x*. As* $f : A \to B$ *and* $f^{-1} : B \to A$*, then you would expect that the inverse function is the mirror image of the original function in the line* $y = x$ *as indicated in figure 2.2.*

3. *The natural logarithm is defined as the logarithmic function when the base is* e*, and is denoted* ln*, i.e.* $\ln(x) \equiv \log_e(x)$*. The graph of* $y = \ln(x)$ *is shown in figure 2.3. Notice*

[5]It is only when n is an irrational real number that we are lacking a definition of the exponent x^n.

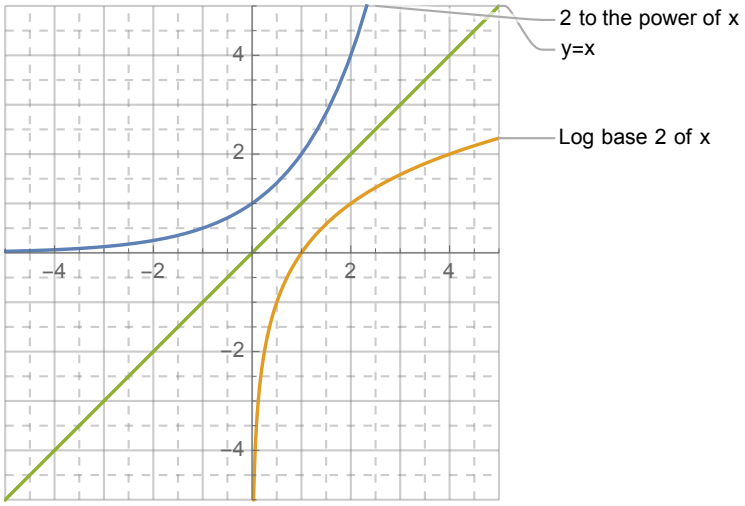


Figure 2.2: A sketch of $y = 2^x$, $y = x$ and $y = \log_2 x$. The curve $y = \log_2 x$ is the reflection in the line $y = x$ of $y = 2^x$.

that broadly there is little difference in the sketch of the two logarithm graphs $y = \ln x$ *and* $y = \log_2 x$.

The logarithm obeys several (very useful) identities:

(i) $\log_a x = \frac{\log_b x}{\log_b a}$.

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

(iii) $\log_a x^y = y \log_a x$.

Proof (of the logarithm identities):[6]

(i) We will prove that $\log_a x \log_b a = \log_b x$, by raising both sides of this equation as a power of base b. For the right-hand-side, by definition, we have

$$b^{\log_b x} = x$$

while from the left-hand-side we obtain

$$b^{\log_a x \log_b a} = (b^{\log_b a})^{\log_a x} = a^{\log_a x} = x.$$

Hence as the exponentiation of b gives a unique output for each input (it is a one-to-one function), we have shown the identity.

[6]Note that proving identities is always a matter of showing that the two sides of the given equation are equal - there is no need to derive the identity.

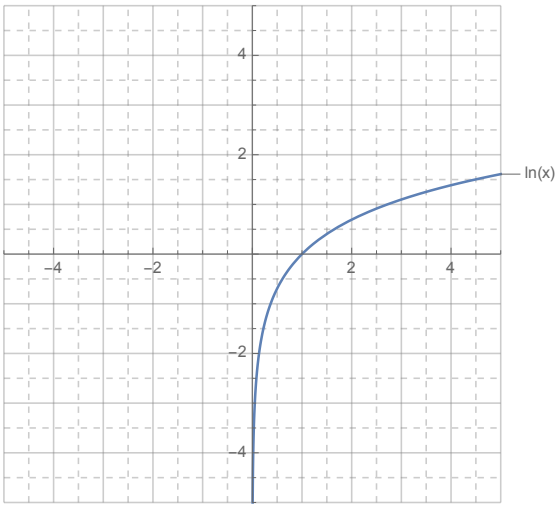


Figure 2.3: A sketch of $y = \ln(x)$. Notice that it is qualitatively similar to the graph of $y = \log_2(x)$ shown in figure 2.2.

(ii) We will raise both sides of the identity $\log_a xy = \log_a x + \log_a y$ as a power of a and show the result is identical. From the left-hand-side we have

$$a^{(\log_a xy)} = xy$$

while from the right-hand-side we find

$$a^{(\log_a x + \log_a y)} = a^{(\log_a x)} a^{(\log_a y)} = xy.$$

As required.

(iii) We will use the same device again and raise both sides of the identity $\log_a x^y = y \log_a x$ as a power of a. From the left-hand-side we have

$$a^{(\log_a x^y)} = x^y$$

while from the right-hand-side we find

$$a^{(y \log_a x)} = (a^{(\log_a x)})^y = x^y.$$

2.3.3 Polynomial Functions

Definition 2.3.3. *A polynomial function in x of degree n where $n \in \{0, 1, 2, 3, \ldots\}$ can be written in the form*

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

where $a_i \in \mathbb{R}$ for $i \in \{0, 1, 2, \ldots n\}$ and $a_n \neq 0$.

Comment(s). *On polynomial functions...*

1. *A constant function is trivially a polynomial of degree zero.*
2. *A linear equation $f(x) \equiv ax + b$ is a polynomial of degree one.*
3. *A quadratic equation $f(x) \equiv ax^2 + bx + c$ is a polynomial of degree two and so on....*
4. *We have not mentioned the domain and range of the polynomial functions as these functions might be real polynomial functions (mapping from $\mathbb{R}$ to $\mathbb{R}$) or complex polynomial functions (mapping from $\mathbb{C}$ to $\mathbb{C}$), or more generally from an abstract number field $\mathbb{F}$ to $\mathbb{F}$. The context of use of the polynomial functions will normally define clearly the range and domain of a particular polynomial function. In this course we will consider real and complex polynomial functions.*

We have not formally introduced the derivative yet in these lecture notes, but it will be useful to assume a simple result which we will prove later on, namely that the derivative of a monomial term ax^n is anx^{n-1}, i.e. $\frac{d}{dx}(ax^n) = anx^{n-1}$. Armed with this result (and with some experience of differentiation) we can see that a polynomial function of degree n which is defined by $n+1$ coefficients $\{a_0, a_1, a_2, \ldots a_n\}$ could be defined in an alternative way. Rather than defining the function globally and explicitly, we could reconstruct the function from knowledge of all its derivatives at a point - this is a very local way to define a function, and what is surprising is that it will be sufficient to reconstruct the function globally. Given the polynomial function:

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

we can compute the derivatives at the point $x = 0$ (any other point would do, but the expressions would become very complicated), as shown in the table. We see that from the values of the

m	$\frac{d^m}{dx^m} f(x)\|_{x=0}$
0	a_0
1	a_1
2	$2a_2$
3	$6a_3$
$\vdots$	$\vdots$
n	$n!a_n$

multiple derivatives at a point we have extracted all the data about the coefficients from which

we could reconstruct the polynomial function. This illustrates an alternative way to define a polynomial function: if we know the value of a function at a point and we know all its derivatives at that point we can define the polynomial function.

The polynomial functions are ubiquitous in mathematics, and we will meet many examples in this course, but their definition immediately suggests a way to generalise them: could we consider polynomial functions of infinite degree? This would mean the function would be an infinite sum of terms and each term dependent on x, and we will meet such functions, known as power series, later on in the course. However we raise the idea here because we will introduce such a function now and it will be defined most naturally in terms of its derivatives at a point.

2.3.4 The Exponential Function

The Exponential Function.

Definition 2.3.4. *The exponential function is denoted*

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

where e is Euler's number and e^x is defined as being the function which is its own derivative and equals 1 when $x = 0$.

Comment(s). *On the exponential function...*

1. *Euler's number is named after the prolific Swiss mathematician Leonhard Euler, and the number which is most commonly simply called 'e'. The value of the exponential function at $x = 1$ gives a value for e.*

2. *There are alternative and equivalent definitions of the exponential function such as*

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

or as the value of y which satisfies

$$x = \int_1^y \frac{dt}{t}.$$

These definitions make use the limit and the integral, concepts which we will define later in this course.

3. *$e^x > 0$ for all $x \in \mathbb{R}$.*

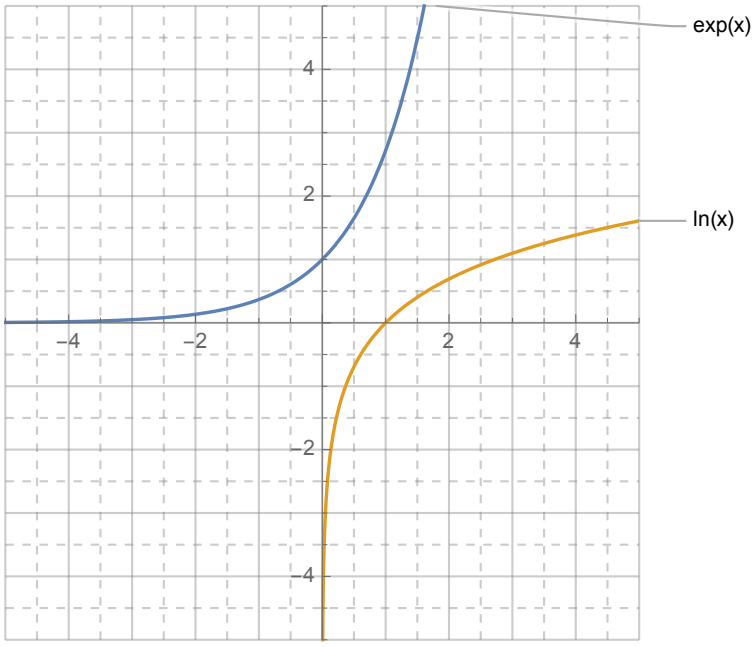


Figure 2.4: The exponential function e^x (or $exp(x)$) and its inverse the natural logarithm $\ln(x)$.

4. *e^x increases monotonically (a monotonically increasing function $f(x)$ is one which increases or stays constant - but does not decrease - as x increases).*

5. *For $x > 0$ e^x grows very quickly ('exponential growth') and is unbounded, consequently for $x < 0$ e^x decreases as x decreases and approaches zero as x approaches $-\infty$. See the sketch of e^x in figure 2.4.*

We motivated the exponential function as a polynomial function of infinite degree, we will introduce such power series later in the course, but let us develop an expression for e^x using the definition of the exponential function above. Recall that we commented earlier that two functions are identical if all their derivatives (including the zeroth derivative) at a point are equal. We will presume the following action for the derivative of a power of x: $\frac{d}{dx}(ax^n) = anx^{n-1}$, we will prove this later, but it will benefit us to make use of this fact here to find another expression for the exponential function. We will start by testing the assumption that e^x can be written as a polynomial in x of finite degree n (and we will see how this fails) so let us write

$$e^x \stackrel{!}{=} \sum_{k=0}^{n} a_k x^k$$

where $a_k \in \mathbb{R}$. Now we will take the derivative with respect to x:

$$\frac{d}{dx}(e^x) \stackrel{!}{=} \frac{d}{dx}(\sum_{k=0}^{n} a_k x^n) = \sum_{k=0}^{n} \frac{d}{dx}(a_k x^k) = \sum_{k=0}^{n} a_k k x^{k-1} = \sum_{k=1}^{n} a_k k x^{k-1} = \sum_{j=0}^{n-1} a_{j+1}(j+1)x^j$$

Now by the definition of the exponential function this derivative must equal our proposed polynomial expression for e^x i.e.

$$\sum_{k=0}^{n} a_k x^k \stackrel{!}{=} \sum_{k=0}^{n-1} a_{k+1}(k+1)x^k.$$

Notice that the left-hand-side is a polynomial of degree n while the right-hand-side is a polynomial of degree $n-1$. Of course this makes sense as the derivative has decreased the maximum power of x by one. We have hinted already how we can resolve this: we will allow the maximum power of x to tend to infinity, and the sum becomes infinite, i.e. our candidate exponential function becomes

$$e^x \stackrel{!}{=} \sum_{k=0}^{\infty} a_k x^k$$

and the condition that the first derivative equals e^x becomes

$$\sum_{k=0}^{\infty} a_k x^k \stackrel{!}{=} \sum_{k=0}^{\infty} a_{k+1}(k+1)x^k.$$

Can we pick coefficients a_k so that the left-hand-side equals the right-hand-side? If the expressions are equal then the coefficient of each power of x must be equal on both sides, i.e. we require $a_k = (k+1)a_{k+1}$ or $a_{k+1} = \frac{a_k}{k+1}$ for all k. We can repeatedly use this to find:

$$a_{k+1} = \frac{a_k}{k+1} = \frac{a_{k-1}}{(k+1)k} = \ldots = \frac{a_0}{(k+1)!}$$

where $(k+1)! = (k+1) \times k \times (k-1) \times (k-2) \times \ldots 2 \times 1$ is called the factorial of $(k+1)$ and is defined as indicated for all $k+1 \in \mathbb{Z}^+$, but note that $0! \equiv 1$ by definition. Putting together our findings we now have

$$e^x \stackrel{!}{=} \sum_{k=0}^{\infty} \frac{a_0}{k!} x^k.$$

Now as $e^0 = 1$ we see that $a_0 = 1$ hence we have the series form of the exponential function

$$e^x = \sum_{k=0}^{\infty} \frac{1}{k!} x^k = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots \tag{2.1}$$

We can use this expression to evaluate Euler's number as

$$e = e^1 = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} + \frac{1}{720} \ldots = 2.718\ldots$$

To find e with greater accuracy one can sum more of the first terms (convince yourself that the first terms are the largest terms in the sum) e.g. to 50 decimal places we have

$$e = 2.71828182845904523536028747135266249775724709369995\ldots.$$

e is an irrational, transcendental[7] number.

Let us pause and sound a note of alarm: what does an infinite sum mean? After all we can never write down the sum in completion, it does not end. How do we know that the sum does give a finite number? Shouldn't we be concerned that an infinite sum does not exist nor make sense? Well, yes we should and we are. It will be the subject of the final chapter of this course where we will address this concern and lose our worries. There is a useful name for such functions:

Definition 2.3.5. *A function which can be locally written as a convergent power series is called an analytic function.*

The word 'locally' in the above definition really means that function is well-defined in the vicinity of a certain point. We will understand the reason for including this in the definition when we study Taylor series. All polynomial functions, the exponential function, the trigonometric functions, power series, the logarithm function are all examples of analytic functions.

2.3.5 The Trigonometric Functions.

There are many equivalent definitions for the trigonometric functions and we will introduce both a geometric and an algebraic definition here. The natural introduction to the trigonometric functions is via geometry and the circle.

Definition 2.3.6. *The unit circle is the set of points (x, y) satisfying $x^2 + y^2 = 1$.*

Definition 2.3.7. *One radian is the angle subtended at the centre of a circle by an arc that is equal in length to the radius of the circle.*

In other words, the arc length on a circle of radius R is equal to $R\theta$ where θ is the angle (in radians) swept through by the radial vector for points along the arc. The radian is the standard unit of angular measurement and the circumference of the the circle $2\pi R$ for a circle of radius, R. Hence 2π is the length of the circumference of the unit circle. If one wishes to convert degrees to radians then using the conversion for the unit circle $360^\circ = 2\pi$ radians, or 1 radian $= \frac{360^\circ}{2\pi} = 57.295779513\ldots^\circ$.

Now we can define the trigonometric functions $\cos\theta$, $\sin\theta$ and $\tan\theta$.

Definition 2.3.8. *The points $(x, y) = (\cos\theta, \sin\theta)$ are the Cartesian coordinates of points on the unit circle whose radial vector (i.e. the straight line from the point $(0, 0)$ to (x, y)) is at an angle θ to the x-axis.*

[7] A transcendental number is any number which is not the solution of any non-constant polynomial equation with rational coefficients.

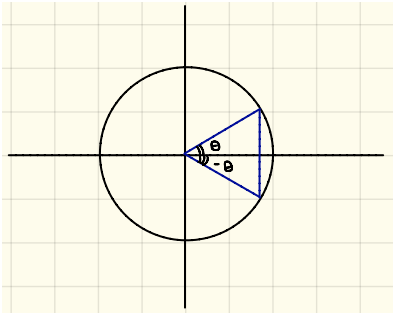
Definition 2.3.9. *The tangent function is given by* $\tan\theta \equiv \frac{\sin\theta}{\cos\theta}$.

Comment(s). *On the trigonometric functions...*
Due to the definition above we immediately see that

1. $\cos^2\theta + \sin^2\theta = 1$.
2. $\cos : \mathbb{R} \to [-1,1]$ *and* $\sin : \mathbb{R} \to [-1,1]$.
3. $\tan\theta$ *is not defined for* $\theta = \frac{\pi}{2} + n\pi$ *where* $n \in \mathbb{Z}$*: for these values of the* θ, $\cos\theta = 0$.
4. $\cos\theta$ *and* $\sin\theta$ *are both periodic, with period* 2π*, in other words* $\cos(\theta + 2\pi) = \cos\theta = \cos(\theta + n2\pi)$ *where* $n \in \mathbb{Z}$*, and similarly for* $\sin\theta$.
5. $\tan\theta$ *is periodic, but its period is* π *as* $\sin(\theta+\pi) = -\sin\theta$ *and* $\cos(\theta+\pi) = -cos(\theta)$ *(see below for proof of these statements).*
6. $\cos\theta$ *and* $\sin\theta$ *have simple values at* $\theta = 0, \frac{\pi}{2}, \pi, \ldots$ *and also at* $\theta = \frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \ldots$.

Sketches of the trigonometric functions can be derived quickly from the definitions and are shown in figure 2.5. In the comments above we made use of some fundamental properties of sine and cosine that originate from their definition as defining the coordinate $(x, y) = (\cos\theta, \sin\theta)$ on the unit circle. Since any point on the unit circle can be written in terms of sine and cosine, transformations which generate symmetries of the circle maps one point on the circle (x, y) to another point (x', y') also on the circle and so also expressible in terms of sine and cosine. That is the symmetry transformations of the circle give rise to interesting transformations of sine and cosine. We consider the effect of some simple symmetry transformations:

(i) Reflection in the x-axis, i.e. $(x, y) \to (x', y') = (x, -y)$



$$x = \cos\theta \to x' = \cos(-\theta) \equiv x = \cos\theta$$
$$y = \sin\theta \to y' = \sin(-\theta) \equiv -y = -\sin\theta$$

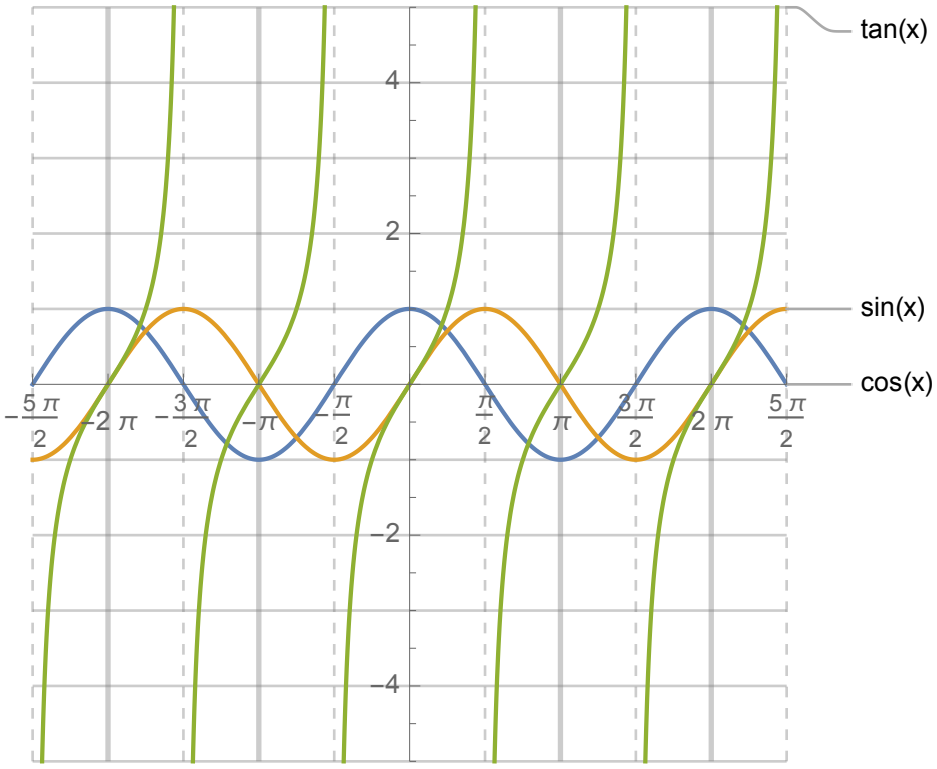
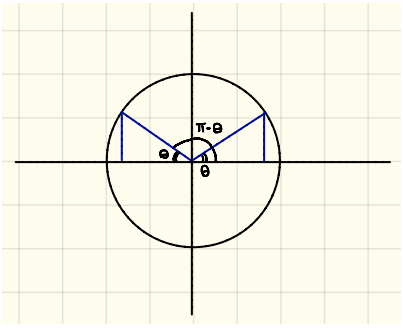


Figure 2.5: The trigonometric functions $\cos(x)$, $\sin(x)$ and $\tan(x)$.

In summary:

$$\cos(-\theta) = \cos\theta$$
$$\sin(-\theta) = -\sin\theta.$$

(ii) Reflection in the y-axis, i.e. $(x, y) \to (x', y') = (-x, y)$

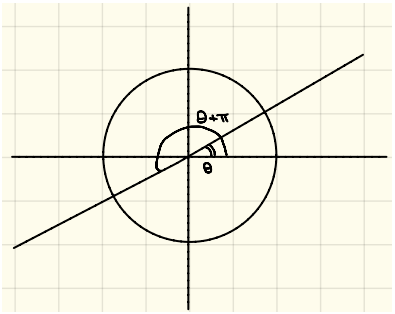


$$x = \cos\theta \to x' = \cos(\pi - \theta) \equiv -x = -\cos\theta$$
$$y = \sin\theta \to y' = \sin(\pi - \theta) \equiv y = \sin\theta$$

In summary:

$$\cos(\pi - \theta) = -\cos\theta$$
$$\sin(\pi - \theta) = \sin\theta.$$

(iii) Rotation by π i.e. $(x, y) \to (x', y') = (-x, -y)$

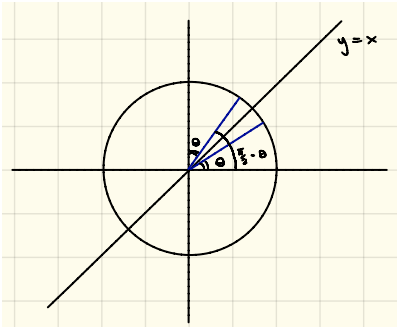


$$x = \cos\theta \to x' = \cos(\theta + \pi) \equiv -x = -\cos\theta$$
$$y = \sin\theta \to y' = \sin(\theta + \pi) \equiv -y = -\sin\theta$$

In summary:

$$\cos(\theta + \pi) = -\cos\theta$$
$$\sin(\theta + \pi) = -\sin\theta.$$

(iv) Reflection in the line $y = x$ i.e. $(x, y) \to (x', y') = (y, x)$



$$x = \cos\theta \to x' = \cos(\frac{\pi}{2} - \theta) \equiv y = \sin\theta$$
$$y = \sin\theta \to y' = \sin(\frac{\pi}{2} - \theta) \equiv x = \cos\theta$$

In summary:

$$\cos(\frac{\pi}{2} - \theta) = \sin\theta$$
$$\sin(\frac{\pi}{2} - \theta) = \cos\theta.$$

The identities above agree with the double-angle formulae for sine and cosine which are often taught without proof in pre-university calculus. We will return to the trigonometric functions and derive the double-angle formulae in a later chapter.

We will now give a second, equivalent, algebraic definition of the basic trigonometric functions in terms of a nice property of its derivatives.

Definition 2.3.10. *The solutions to the second order differential equation*

$$\frac{d^2 f}{dx^2} = -f$$

are

$$f = \begin{cases} \cos x & \text{if } f(0) = 1 \text{ and } \frac{df}{dx}(0) = 0 \\ \sin x & \text{if } f(0) = 0 \text{ and } \frac{df}{dx}(0) = 1. \end{cases}$$

We have asserted, in passing, that $\cos x$ and $\sin x$ are analytic functions. While we are not in a position to check convergence yet, we can certainly identify the correct power series.

Example 2.5. *Use the definition of* $\cos x$ *and* $\sin x$ *as solutions to* $\frac{d^2 f}{dx^2} = -f$ *with the appropriate boundary conditions to find their power series.*

Let $f(x) \equiv \sum_{n=0}^{\infty} a_n x^n$ *then*

$$\frac{d^2 f}{dx^2} = \sum_{n=0}^{\infty} a_n n(n-1)x^{n-2} = \sum_{n=0}^{\infty} a_{n+2}(n+2)(n+1)x^n \stackrel{!}{=} -\sum_{n=0}^{\infty} a_n x^n.$$

Equating coefficients of each power of x *gives* $a_{n+2}(n+2)(n+1) = -a_n$ *or*

$$a_{n+2} = -\frac{a_n}{(n+2)(n+1)}.$$

We can use this recursively but the outcome will depend on whether $n+2$ *is even or odd. After a little thought and taking care with the minus signs we find*

$$a_{n+2} = \begin{cases} (-1)^{(n+2)/2}\frac{a_0}{(n+2)!} & n \ \ even \\ (-1)^{(n+1)/2}\frac{a_1}{(n+2)!} & n \ \ odd \end{cases}$$

hence we have

$$f(x) = a_0 \sum_{n\,even} (-1)^{n/2}\frac{x^n}{n!} + a_1 \sum_{n\,odd} (-1)^{(n-1)/2}\frac{x^n}{n!}.$$

Applying the boundary conditions for $\sin(x)$*, where we have* $f(0) = 0$ *and* $f'(0) = 1$ *gives* $a_0 = 0$ *and* $a_1 = 1$*, therefore*

$$\sin(x) = \sum_{n\,odd} (-1)^{(n-1)/2}\frac{x^n}{n!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots. \tag{2.2}$$

While the boundary conditions for $\cos(x)$ *give* $a_0 = 1$ *and* $a_1 = 0$ *hence*

$$\cos(x) = \sum_{n\,even} (-1)^{n/2}\frac{x^n}{n!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots \tag{2.3}$$

where we have used $0! \equiv 1$*.*

We can take advantage of mathematical software to plot graphs of some of the partial sums of the first few terms of $\sin(x)$ and compare the result against the graph of $\sin(x)$. Bearing in mind that we expect the full sum to reproduce the sine function exactly, we may anticipate that the sum of the terms up to and including order n in x, given by $x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots + (-1)^{(n-1)/2}\frac{x^n}{n!}$ (for odd n) gives a better and better approximation to the sine function as n increases. Sketches of the partial sums, together with the sine function, are shown in figure 2.6.

At this stage you should correctly feel concerned about the infinite sums and infinite products. How do we know that they are finite? Addressing this concern is one of the major topics in analysis and which we will work on in this course.

In addition to the basic trigonometric functions there are a set of related and commonly used functions - they appear so frequently in mathematics that we will define them now.

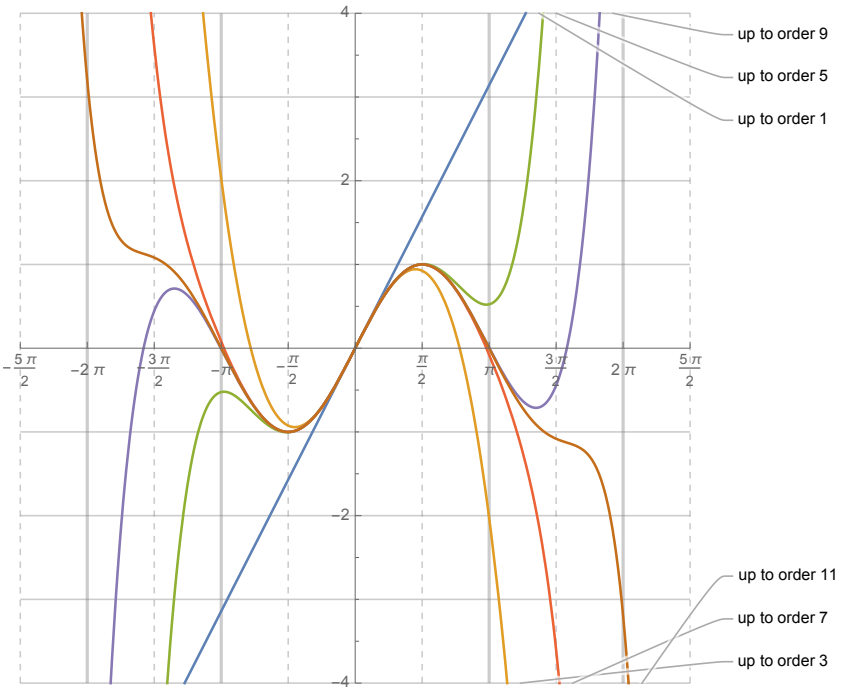


Figure 2.6: Building $\sin(x)$ as a power series, by taking more and more terms in the summation. Dashed: $\sin(x)$. Solid: $f(x) = \sum_{k\,odd}^{n} (-1)^{\frac{k-1}{2}} x^k/k!$ for different choices of n (values of n are indicated at the edges of the sketch by the order).

Definition 2.3.11. *The secant function is denoted* sec *and defined as*

$$\sec\theta = \frac{1}{\cos\theta}.$$

Comment(s). *On* $\sec\theta$*...*

1. *It is not defined for* $\theta = \frac{\pi}{2} + n\pi$ *where* $n \in \mathbb{Z}$*. These are the same points where* $\tan\theta$ *is not defined, so you might suspect there is a connection between* $\tan\theta$ *and* $\sec\theta$ *and you would be correct. . . .*

2. $\sec(\theta + 2\pi) = \sec\theta$ *and* $\sec(\theta + \pi) = -\sec\theta$*.*

Definition 2.3.12. *The cosecant function is denoted* cosec *and defined as*

$$\operatorname{cosec}\theta = \frac{1}{\sin\theta}.$$

Comment(s). *On* $\operatorname{cosec}\theta$*...*

1. *It is not defined for* $\theta = n\pi$ *where* $n \in \mathbb{Z}$*.*

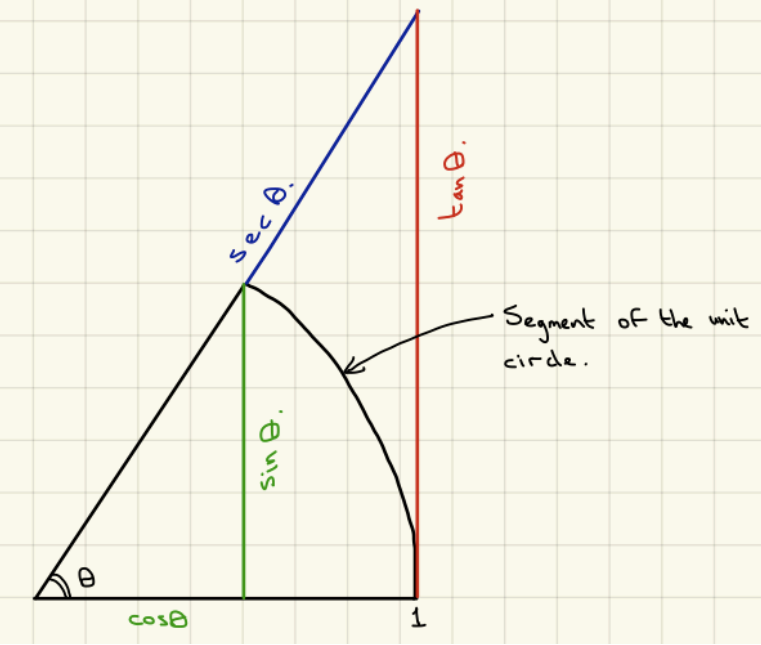


Figure 2.7: A segment of the unit circle, subtending an angle of θ. The scale factor to map the small right-angled triangle up to the large one is $\sec\theta$.

2. $\operatorname{cosec}(\theta + 2\pi) = \operatorname{cosec}\theta$ *and* $\operatorname{cosec}(\theta + \pi) = -\operatorname{cosec}\theta$.

Definition 2.3.13. *The cotangent function is denoted* cot *and defined as*

$$\cot\theta = \frac{1}{\tan\theta}.$$

Comment(s). *On* $\cot\theta$*...*

1. *It is not defined for* $\theta = n\pi$ *where* $n \in \mathbb{Z}$.

2. $\cot(\theta + \pi) = \cot\theta$.

But where do all these unusual names come from? Well, Latin is the short answer... the sine function is derived from 'sinus' meaning 'a bend', while cosine is short for 'complementary sine', the tangent originates from 'tangere' meaning 'touching' (see the sketch in figure 2.7) and secant from 'secare' meaning 'to cut' (again see the sketch in figure 2.7 to get the sense behind these names).

The Inverse Trigonometric Functions.

Definition 2.3.14. *The inverse trigonometric functions are known by particular names. The inverse sine function is denoted* arcsin *or* $\sin^{-1}$*, the inverse cosine function is denoted* arccos *of* $\cos^{-1}$ *and the inverse tangent function is denoted* arctan *or* $\tan^{-1}$.

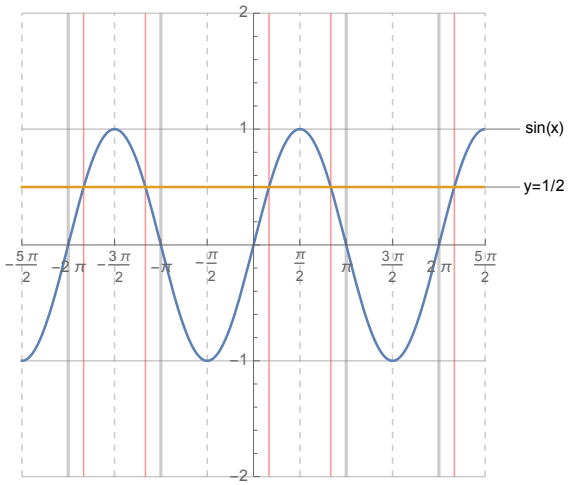


Figure 2.8: A sketch of $\sin x$ highlighting its many-to-one nature. The horizontal line indicates the line $y = \frac{1}{2}$ and its multiple intersections with $\sin x$ are shown over the domain $[-5\pi/2, 5\pi/2]$.

Comment(s). *(On the inverse trigonometric functions...) The trigonometric functions are all periodic, sine and cosine have period 2π while the tangent has period π, see the sketch in figure 2.8. Consequently for a given θ there are infinitely many values θ', such that $\theta \neq \theta'$ and $\sin(\theta') = \sin(\theta)$, $\cos(\theta') = \cos(\theta)$ and $\tan(\theta') = \tan(\theta)$. This means that the trigonometric functions are not bijective functions, so they are not invertible functions on their full domain. However by restricting the domain, one can then construct a well-defined inverse functions.*

From the graph in figure 2.8, and our knowledge of the continued periodicity of the trigonometric functions, we see that, for example, $\sin\theta = \frac{1}{2}$ has an infinite number of solutions for θ, i.e. $\theta = \frac{\pi}{6} + n2\pi$ and $\theta = \frac{5\pi}{6} + m2\pi$ for all $n, m \in \mathbb{Z}$. That is, there exist $\theta_1 \neq \theta_2$ such that $\sin\theta_1 = \sin\theta_2 = \ldots$, so that on the domain of the real numbers $\mathbb{R}$ the sine function is not an injective function and so the inverse sine function is not defined. The same is also true for the inverse cosine function and the inverse tangent function. However if we limit the domain and range we can find inverse trigonometric functions.

Our aim is to reduce the domain and range of the trigonometric function such that each input gives a different unique output. Graphically this corresponds to restricting the sine and cosine graphs to a domain between consecutive turning points. There are many ways to do this and we give the standard definitions here.

Definition 2.3.15. *The function* arcsin *(also denoted $\sin^{-1}$) is defined by*

$$\arcsin : [-1, 1] \to [-\pi/2, \pi/2] \qquad \arcsin(\sin x) = x.$$

The function $\arcsin x$ when sketched as a graph is a line segment and is a reflection in the line $y = x$ of $\sin x$ over the domain $[-\pi/2, \pi/2]$, this is shown in figure 2.9.

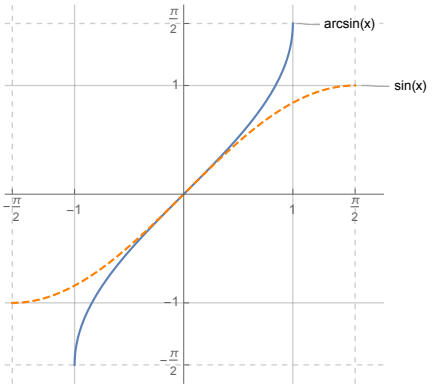


Figure 2.9: A sketch of $\arcsin x$ and $\sin x$ over the domains $[-1, 1]$ and $[-\pi/2, \pi/2]$ respectively. Notice how they are mirrored in the line $y = x$.

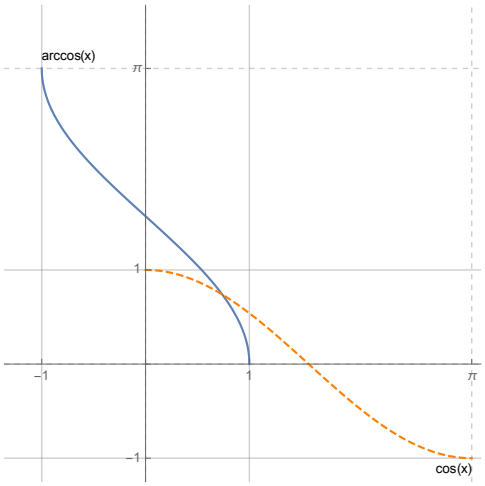


Figure 2.10: A sketch of $\arccos x$ and $\cos x$ over the domains $[-1, 1]$ and $[0, \pi]$ respectively.

Definition 2.3.16. *The function* arccos *(also denoted* $\cos^{-1}$*) is defined by*

$$\arccos : [-1, 1] \to [0, \pi] \qquad \arccos(\cos x) = x.$$

The sketch of $\cos(x)$ and $\arccos(x)$ are shown on the same graph in figure 2.10.

Definition 2.3.17. *The function* arctan *(also denoted* $\tan^{-1}$*) is defined by*

$$\arctan : \mathbb{R} \to (-\pi/2, \pi/2) \qquad \arctan(\tan x) = x.$$

A plot of $y = \arctan x$ and $\tan(x)$ is shown in figure 2.11.

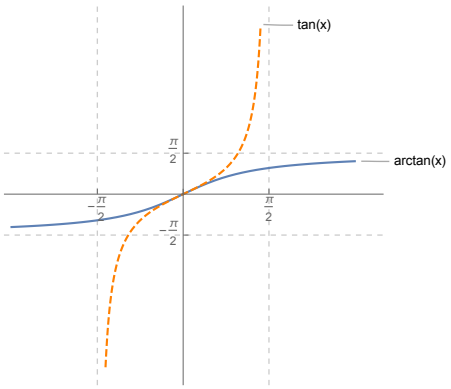


Figure 2.11: A sketch of $\arctan x$ and $\tan x$ over the domains $[-\pi, \pi]$ and $(-\pi/2, \pi/2)$ respectively.

2.3.6 The Hyperbolic Functions.

Recall that we geometrically defined the trigonometric functions $x = \cos\theta$ and $y = \sin\theta$ as the coordinates of points on the unit circle $x^2 + y^2 = 1$, we will now define the hyperbolic (trigonometric) functions in a similar manner.

Definition 2.3.18. *The hyperbolic functions $x = \cosh(a)$ and $y = \sinh(a)$ are the coordinates of points on the right-hand branch of the hyperbola $x^2 - y^2 = 1$, as depicted in figure 2.12.*

Comment(s). *On the hyperbolic functions...*

1. *The notation* cosh *and* sinh *are shorthand for 'hyperbolic cosine' and 'hyperbolic sine'.*

2. *There are multiple ways of pronouncing* sinh *(pronounced as either "shine" or "sinch") while* cosh *is pronounced as written ("kosh").*

3. *By definition we have the identity:*

$$\cosh^2(a) - \sinh^2(a) = 1.$$

4. *The parameter a used to define the point $(\cosh(a), \sinh(a))$ is not an angle but has the magnitude of an area and maybe positive or negative (if the point (x, y) is below the x-axis). Its magnitude is twice the area shaded in figure 2.12 i.e. up to a sign it is the area*[8] *enclosed between the x-axis, the right-hand branch of the hyperbola and the straight*

[8]The simplest way to prove this is by integration - try and show this once you have met the integral later in the course.

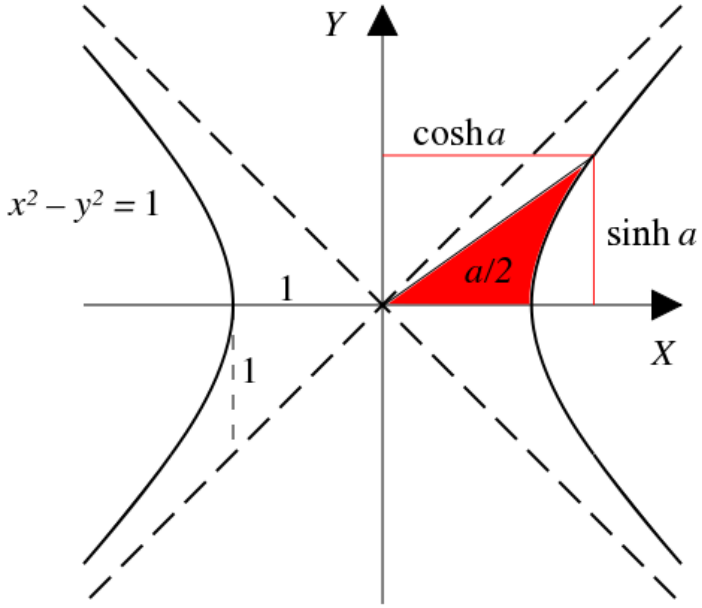


Figure 2.12: The hyperbolic functions $x = \cosh(a)$ and $y = \sinh(a)$ shown as points on the right-hand branch of the hyperbola $x^2 - y^2 = 1$.

line from the origin to the point $x = \cosh(a)$ *and* $y = \sinh(a)$*. It is worth comparing this definition of* a *with that of the argument of the standard trigonometric functions where the area of a segment of the unit circle subtending an angle* θ *at the origin is* $\frac{1}{2}\theta r^2 = \frac{1}{2}\theta$*.*

5. *The functions are defined on the domain and ranges:*

$$\sinh : \mathbb{R} \to \mathbb{R}$$

and

$$\cosh : \mathbb{R} \to [1, \infty).$$

Definition 2.3.19. *The hyperbolic tangent function is denoted* tanh *and defined by*

$$\tanh(a) = \frac{\sinh(a)}{\cosh(a)}.$$

Comment(s). *(On* tanh*)*

1. tanh *is pronounced "tanch" or "than".*

2. $\tanh(a)$ *is the gradient of the straight line through the origin and the point with coordinates* $(\cosh(a), \sinh(a))$ *on the hyperbola.*

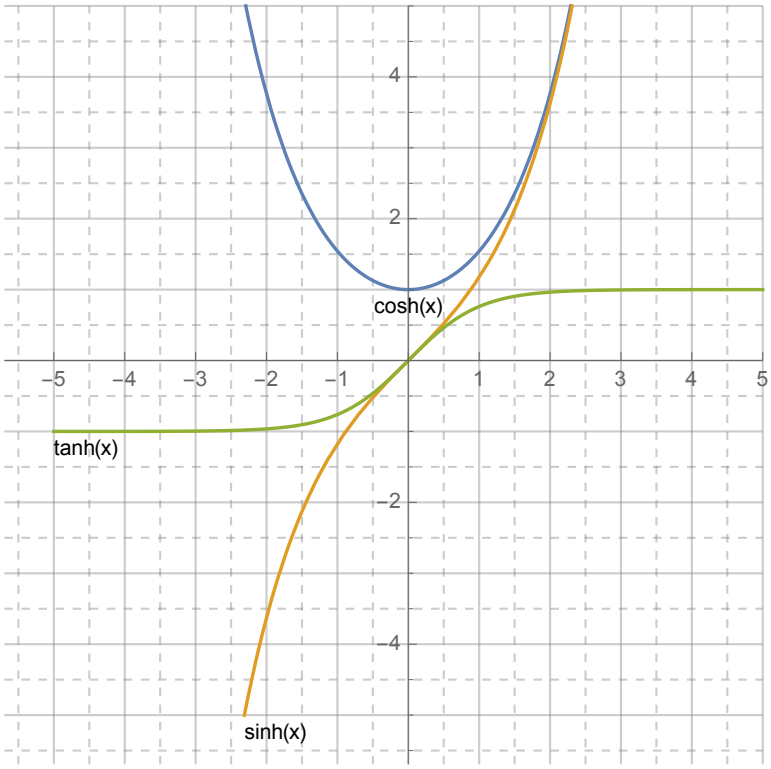


Figure 2.13: Sketches of the hyperbolic functions as the curves $y = \cosh(x)$, $y = \sinh(x)$ and $y = \tanh(x)$.

3. *As the hyperbola is tangent*[9] *to the lines* $y = x$ *and* $y = -x$ *then as* y *approaches* $\pm\infty$, $\tanh(a)$ *approaches* ± 1.

4. $\tanh : \mathbb{R} \to (-1, 1)$.

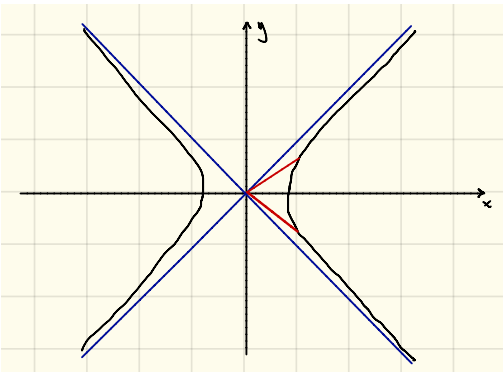
Sketches of the hyperbolic trigonometric functions are shown in figure 2.13. Related hyperbolic functions are given by

$$\begin{aligned} \text{sech}\,(x) &\equiv \frac{1}{\cosh(x)} \\ \text{cosech}\,(x) &\equiv \frac{1}{\sinh(x)} \\ \coth(x) &\equiv \frac{1}{\tanh(x)} \end{aligned}$$

One can derive some properties of the hyperbolic functions from the symmetries of the hyperbola:

[9]The equation defining the hyperbola can be rewritten as the pair of curves $y = \sqrt{x^2 - 1}$ and $y = -\sqrt{x^2 - 1}$ and for large x these equations approximate closely the lines $y = x$ and $y = -x$.

(i) Reflection in the x-axis, i.e. $(x, y) \to (x', y') = (x, -y)$.



$$x = \cosh(a) \to x' = \cosh(-a) \equiv x = \cosh(a)$$
$$y = \sinh(a) \to y' = \sinh(-a) \equiv -y = -\sinh(a)$$

In summary:

$$\cosh(-a) = \cosh(a)$$
$$\sinh(-a) = -\sinh(a).$$

The circle has an infinite number of symmetries, but the hyperbola only obviously has two: reflection in the x-axis and reflection in the y-axis. Why does reflection of the hyperbola in the y-axis not generate an identity of the hyperbolic functions?

One can also define the hyperbolic functions algebraically as solutions to differential equations.

Definition 2.3.20. *The solutions to the second order differential equation*

$$\frac{d^2 f}{dx^2} = f$$

are

$$f = \begin{cases} \cosh x & \text{if } f(0) = 1 \text{ and } \frac{df}{dx}(0) = 0 \\ \sinh x & \text{if } f(0) = 0 \text{ and } \frac{df}{dx}(0) = 1. \end{cases}$$

This definition, coming prior to the definition of the derivative is premature! However we can make use of it (again assuming only that $\frac{d}{dx}(ax^n) = anx^{(n-1)}$) to find power series.

Example 2.6. *Use the definition of* $\cosh(x)$ *and* $\sinh(x)$ *as solutions to* $\frac{d^2 f}{dx^2} = f$ *with the appropriate boundary conditions to find their power series.*

Let $f(x) \equiv \sum_{n=0}^{\infty} a_n x^n$ *then*

$$\frac{d^2 f}{dx^2} = \sum_{n=0}^{\infty} a_n n(n-1)x^{n-2} = \sum_{n=0}^{\infty} a_{n+2}(n+2)(n+1)x^n \stackrel{!}{=} \sum_{n=0}^{\infty} a_n x^n.$$

Equating coefficients of each power of x *gives* $a_{n+2}(n+2)(n+1) = a_n$ *or*

$$a_{n+2} = \frac{a_n}{(n+2)(n+1)}.$$

Up to a sign this is the same as the relation we found when deriving the power series for $\cos(x)$ *and* $\sin(x)$*, hence we have*

$$f(x) = a_0 \sum_{n\,even} \frac{x^n}{n!} + a_1 \sum_{n\,odd} \frac{x^n}{n!}.$$

Applying the boundary conditions for $\sinh(x)$*, where we have* $f(0) = 0$ *and* $f'(0) = 1$ *gives* $a_0 = 0$ *and* $a_1 = 1$*, therefore*

$$\sinh(x) = \sum_{n\,odd} \frac{x^n}{n!} = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!} + \ldots. \tag{2.4}$$

While the boundary conditions for $\cosh(x)$ *give* $a_0 = 1$ *and* $a_1 = 0$ *hence*

$$\cosh(x) = \sum_{n\,even} \frac{x^n}{n!} = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!} + \ldots \tag{2.5}$$

where we have used $0! \equiv 1$*.*

The Inverse Hyperbolic Functions

Definition 2.3.21. *The inverse hyperbolic functions are called* arcsinh *or* $\sinh^{-1}$*,* arccosh *or* $\cosh^{-1}$ *and* arctanh *or* $\tanh^{-1}$ *and are the inverse* sinh*,* cosh *and* tanh *functions respectively.*

Recall the graphs of the hyperbolic functions which are shown in figure 2.13. These functions are not periodic, and only cosh "doubles back on itself" and is the only one of the three which is not injective and whose inverse function will have a restricted domain. There are some further pleasant qualities of the hyperbolic functions which will allow us to find simple analytic expressions for their inverse functions. Our construction will rest upon the power series for the exponential function[10] which we recall is:

$$e^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \ldots$$

[10]We should note that we are building up a lot of debt here: we have assumed that the derivative of ax^n is anx^{n-1} and that the infinite sum for e^x converges - we will pay off our debts in due course.

and hence we also have

$$e^{-x} = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} x^n = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \frac{x^4}{4!} - \frac{x^5}{5!} + \dots .$$

While, earlier, in a similar analysis in equations (2.4) and (2.5) we found that

$$\sinh(x) = \sum_{n\ odd} \frac{x^n}{n!} = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!} + \dots \qquad \text{and}$$
$$\cosh(x) = \sum_{n\ even} \frac{x^n}{n!} = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!} + \dots .$$

Hence we notice that our power series for the two hyperbolic functions can be rewritten in terms of the exponential function, that is,

$$\sinh(x) = \frac{1}{2}(e^x - e^{-x}) \qquad \text{and} \tag{2.6}$$
$$\cosh(x) = \frac{1}{2}(e^x + e^{-x}). \tag{2.7}$$

Consequently we also have

$$\tanh(x) \equiv \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1}. \tag{2.8}$$

Using these expressions we can find simple analytic expressions for arcsinh , arccosh and arctanh .

Example 2.7. *Find an analytic expression for the inverse* sinh *function.*

Let us write $y \equiv \sinh x = \frac{1}{2}(e^x - e^{-x})$ *and rearrange this expression to find* x *as a function of* y*, i.e.* $x(y)$*. Putting all terms on the same side of the equation we find*

$$e^x - e^{-x} - 2y = 0$$

and multiplying this by e^x *gives*

$$e^{2x} - 1 - 2ye^x = 0.$$

We may complete the square to find

$$(e^x - y)^2 - y^2 - 1 = 0 \quad \Longrightarrow \; e^x - y = \sqrt{1 + y^2}$$

and therefore

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

That is

$$\operatorname{arcsinh} y = \ln(y + \sqrt{1 + y^2}) \qquad \forall\, y \in \mathbb{R}.$$

A sketch of the graph of arcsinh is given in in figure 2.14, you should take some time to confirm the plot agrees with your expectations of the analytic function we have derived in the previous example.

Exercise 2.3. *Repeat the procedure above to show that*

$$\operatorname{arccosh} x = \ln(\sqrt{x^2 - 1} + x) \qquad \forall\, x \in [1, \infty).$$

Exercise 2.4. *Repeat the procedure again to show that*

$$\operatorname{arctanh} x = \frac{1}{2} \ln\left(\frac{1 + x}{1 - x}\right) \qquad \forall\, x \in (-1, 1).$$

2.3.7 Trigonometric and Hyperbolic Functions with Complex Variables

What more do we have to say about the trigonometric and hyperbolic functions? There are a number of identities which will prove useful for integration and to develop them we will consider the useful role played by the complex numbers in the description of the trigonometric functions.

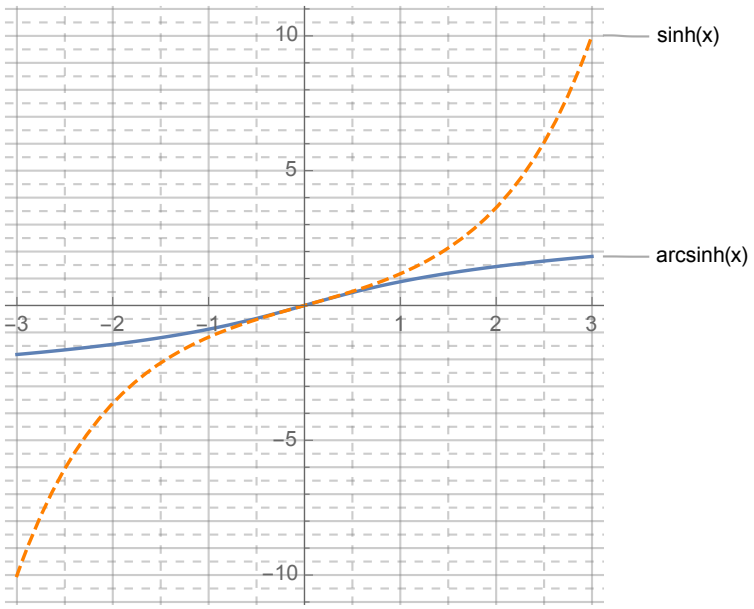


Figure 2.14: A sketch of $\operatorname{arcsinh}(x)$ and $\sinh(x)$.

Euler's Formula

The complex numbers $\mathbb{C}$ introduces the imaginary number $i = \sqrt{-1}$ to the real numbers and consists of all numbers of the form $x + iy$ where $x, y \in \mathbb{R}$. The complex numbers $\mathbb{C}$ is an extension of $\mathbb{R}$, and we may wonder if we can we extend the range of validity of the exponent a^n to include the case where $n \in \mathbb{C}$. For example just what could an expression such as a^i mean? What is the sense in saying it means we multiply a by itself i times? Not much, but this is just the kind of loose end we might have expected when we investigating the complex numbers, and we must hope that we can give such expressions some logical meaning.

Recall that we can write e^x, $\cosh x$ and $\sinh x$ as infinite sums, and even relate the formulae to find exponential expressions for the hyperbolic functions. Now that we have introduced the complex number i we can also re-express our power series for sine in equation (2.2) and for cosine in equation (2.3) in terms of the exponential in a similar way. Recalling our power series for e^x given in equation (2.1) we can now add in the imaginary number to find:

$$e^{ix} = \sum_{k=0}^{\infty} \frac{1}{k!}(ix)^k = 1 + ix - \frac{x^2}{2!} - i\frac{x^3}{3!} + \frac{x^4}{4!} + i\frac{x^5}{5!} - \frac{x^6}{6!} - i\frac{x^7}{7!} + \dots$$

and

$$e^{-ix} = \sum_{k=0}^{\infty} \frac{1}{k!}(-ix)^k = 1 - ix - \frac{x^2}{2!} + i\frac{x^3}{3!} + \frac{x^4}{4!} - i\frac{x^5}{5!} - \frac{x^6}{6!} + i\frac{x^7}{7!} + \dots$$

where we have used our definition of i, that $i^2 = -1$ to expand out and simplify the powers of

i. By comparing with the power series proposed for $\sin(x)$ and $\cos(x)$ in equations (2.2) and (2.3), we deduce that

$$\sin x = \frac{1}{2i}(e^{ix} - e^{-ix}) \tag{2.9}$$

$$\cos x = \frac{1}{2}(e^{ix} + e^{-ix}). \tag{2.10}$$

We may now invert these equations to find an expression for e^{ix} in terms of $\cos x$ and $\sin x$. This gives us Euler's formula.

Definition 2.3.22. *Euler's formula* *is*

$$e^{i\theta} \equiv \cos\theta + i\sin\theta$$

where $\theta \in \mathbb{R}$.

Leonhard Euler (1707-1783) was a Swiss mathematician who played a pivotal role in the development of analysis. Euler, pronounced "Oiler" rather than "U-ler"[12], is widely celebrated as one of the greatest mathematicians to have ever lived. While measures of the greatness of mathematicians are subjective, what is not subjective is how prolific Euler was, he had 13 children[13] and still managed to write over 800 papers! It was said of Euler that "he calculated just as men breathe, as eagles sustain themselves in the air".



Figure 2.15
Leonhard Euler[11]

Our development of Euler's formula has been based on a number of assumptions which we have yet to prove, so let us at this stage use some of our prior knowledge of calculus to check that the formula is, at least, sensible. So forgoing the idea that we have yet to discover the derivative we note that:

$$\frac{d}{d\theta}(e^{i\theta}) = ie^{i\theta} \qquad \text{while} \qquad \frac{d}{d\theta}(\cos\theta + i\sin\theta) = -\sin\theta + i\cos\theta = i(i\sin\theta + \cos\theta) = ie^{i\theta}.$$

Thus this remarkable formula of Euler's passes a first test - why did we do this? We have no good handle on an imaginary power, nothing in our experience with which to check the logic of the formula and we have developed our power series on the back of a set of ideas which we have yet to convince ourselves of. So it is reassuring to check that this formula agrees with our pre-knowledge of calculus. We will make much use of it in the following.

[11]This portrait is by Jakob Handmann. Euler is winking at you. No he isn't, in later life he had problems with his eye, although it is understood he developed a cataract in his left eye, so his closed right eye in this portrait is puzzling.

[12]See the 2014 film "The Imitation Game" for a good example of how not to pronounce Euler.

[13]This is not what I mean by prolific!

Comment(s). *(On Euler's Formula)*

1. *When* $\theta = 0$ *we have* $e^0 = 1 = \cos(0) + i\sin(0)$ *as required.*

2. *When* $\theta = \pi/2$ *we have* $e^{i\pi/2} = \cos(\pi/2) + i\sin(\pi/2) = i$.

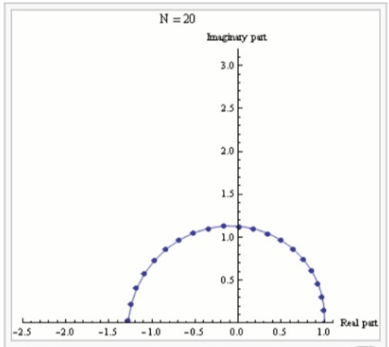
3. *When* $\theta = \pi$ *we have* $e^{i\pi} = -1$. *This remarkable statement is known as Euler's identity and is frequently celebrated as it relates* e, i, π *and* -1. *See figure 2.16 for many rapturous comments on this identity.*

4. $e^{i\theta}$ *has a period of* 2π *inherited from the trigonometric functions, and indeed* $e^{i2\pi} = 1$. *This will be of concern in defining its inverse function, the natural logarithm, over the complex numbers.*

On "Euler's identity" [edit]

In mathematical analysis, **Euler's identity** is the equation "$e^{i\pi} + 1 = 0$."

- **One of the most frequently mentioned equations was Euler's equation, $e^{i\pi} + 1 = 0$. Respondents called it "the most profound mathematical statement ever written"; "uncanny and sublime"; "filled with cosmic beauty"; and "mind-blowing".** Another asked: **"What could be more mystical than an imaginary number interacting with real numbers to produce nothing?" The equation contains nine basic concepts of mathematics — once and only once — in a single expression. These are: e (the base of natural logarithms); the exponent operation; π; plus (or minus, depending on how you write it); multiplication; imaginary numbers; equals; one; and zero.**
  - Robert P. Crease, in "The greatest equations ever" at PhysicsWeb (October 2004)
- Like a Shakespearean sonnet that captures the very essence of love, or a painting that brings out the beauty of the human form that is far more than just skin deep, **Euler's equation reaches down into the very depths of existence.**
  - Keith Devlin, as quoted in *Dr. Euler's Fabulous Formula : Cures Many Mathematical Ills* (2006) ISBN 978-0691118222
- **Our jewel ... one of the most remarkable, almost astounding, formulas in all of mathematics.**
  - Richard Feynman on Euler's formula, of which Euler's identity is a special case, in *The Feynman Lectures on Physics*, Vol. I, p. 22
- There is a famous formula, perhaps the most compact and famous of all formulas — developed by Euler from a discovery of de Moivre: $e^{i\pi} + 1 = 0$. **It appeals equally to the mystic, the scientist, the philosopher, the mathematician.**
  - Edward Kasner and James R. Newman in *Mathematics and the Imagination* (1940)
- **Gentlemen, that is surely true, it is absolutely paradoxical; we cannot understand it, and we don't know what it means. But we have proved it, and therefore we know it must be the truth.**
  - Benjamin Peirce, as quoted in notes by W. E. Byerly, published in *Benjamin Peirce, 1809-1880 : Biographical Sketch and Bibliography* (1925) by R. C. Archibald; also in *Mathematics and the Imagination* (1940) by Edward Kasner and James Newman



Gentlemen, that is surely true, it is absolutely paradoxical; we cannot understand it, and we don't know what it means. But we have proved it, and therefore we know it must be the truth. ~ Benjamin Peirce

Figure 2.16: Quotes on the beauty of Euler's identity and formula.

Double-Angle Formulae

It will prove very useful to us to develop the double-angle formulae for sine and cosine (i.e. to rewrite the expressions $\cos(\theta+\phi)$ and $\sin(\theta+\phi)$ in terms of $\cos\theta$, $\cos\phi$, $\sin\theta$ and $\sin\phi$). Due to Euler's formula we now have a very simple way to quickly write these formulae down.

Let us multiply two complex numbers, both of modulus 1:

$$e^{i\theta}e^{i\phi} = e^{i(\theta+\phi)}.$$

There appears to be nothing to this, but notice that if we use Euler's formula on the left we will have trigonometric functions of single angles, while on the right it will give us trigonometric functions of the sum of the angles:

$$\begin{aligned}\cos(\theta+\phi) + i\sin(\theta+\phi) &= e^{i(\theta+\phi)}\\ &= e^{i\theta}e^{i\phi}\\ &= (\cos\theta + i\sin\theta)(\cos\phi + i\sin\phi)\\ &= \cos\theta\cos\phi - \sin\theta\sin\phi + i(\cos\theta\sin\phi + \sin\theta\cos\phi).\end{aligned}$$

If we equate the real part on each side and the imaginary part on each side of the above we arrive at the double-angle formulae for cosine and sine:

$$\cos(\theta+\phi) = \cos\theta\cos\phi - \sin\theta\sin\phi \tag{2.11}$$

$$\sin(\theta+\phi) = \cos\theta\sin\phi + \sin\theta\cos\phi. \tag{2.12}$$

Exercise 2.5. *Use the double angle formulae for cosine and sine to show that*

$$\tan(\theta+\phi) = \frac{\tan\theta + \tan\phi}{1 - \tan\theta\tan\phi}.$$

Comment(s). *(On the double angle formulae...)*

1. *When $\phi = \theta$ we have*

$$\begin{aligned}\cos(2\theta) &= \cos^2\theta - \sin^2\theta\\ \sin(2\theta) &= 2\cos\theta\sin\theta\\ \tan(2\theta) &= \frac{2\tan\theta}{1-\tan^2\theta}\end{aligned}$$

2. *Adding $\cos^2\theta + \sin^2\theta = 1$ and $\cos^2\theta - \sin^2\theta = \cos(2\theta)$ gives*

$$\cos^2\theta = \frac{1}{2}(1 + \cos(2\theta))$$

while the difference gives:

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

These formula reduce the power of the trigonometric function at the expense of doubling the argument. This can be very useful when solving integrals.

3. *The difference formula are quick to derive (using* $\sin(-x) = -sin(x)$ *and* $\cos(-x) = \cos(x)$*):*

$$\begin{aligned}\cos(\theta-\phi) &= \cos\theta\cos\phi + \sin\theta\sin\phi\\ \sin(\theta-\phi) &= -\cos\theta\sin\phi + \sin\theta\cos\phi\\ \tan(\theta-\phi) &= \frac{\tan\theta-\tan\phi}{1+\tan\theta\tan\phi}\end{aligned}$$

4. *One can convert sums of trigonometric functions into products and vice-versa:*

$$\begin{aligned}\frac{1}{2}(\cos(\theta+\phi)+\cos(\theta-\phi)) &= \cos\theta\cos\phi\\ \frac{1}{2}(\cos(\theta-\phi)-\cos(\theta+\phi)) &= \sin\theta\sin\phi\\ \frac{1}{2}(\sin(\theta+\phi)+\sin(\theta-\phi)) &= \sin\theta\cos\phi\\ \frac{1}{2}(\sin(\theta+\phi)-\sin(\theta-\phi)) &= \sin\phi\cos\theta\end{aligned}$$

5. *By writing* $\alpha = \theta+\phi$ *and* $\beta = \theta-\phi$*, so that* $\alpha+\beta = 2\theta$ *and* $\alpha-\beta = 2\phi$ *one can use the product formulae above to find formulae relating half-angles to angles:*

$$\begin{aligned}\cos(\alpha)+\cos(\beta) &= 2\cos(\frac{\alpha+\beta}{2})\cos(\frac{\alpha-\beta}{2})\\ \sin(\alpha)+\sin(\beta) &= 2\sin(\frac{\alpha+\beta}{2})\cos(\frac{\alpha-\beta}{2})\\ \cos(\alpha)-\cos(\beta) &= -2\sin(\frac{\alpha+\beta}{2})\sin(\frac{\alpha-\beta}{2})\\ \sin(\alpha)-\sin(\beta) &= 2\sin(\frac{\alpha-\beta}{2})\cos(\frac{\alpha+\beta}{2})\end{aligned}$$

Exercise 2.6. *Write* $\cos\theta$*,* $\sin\theta$ *and* $\tan\theta$ *in terms of* $t = \tan\frac{\theta}{2}$*. (Motivation: these formulae are very useful for solving otherwise difficult integrals.)*

Trigonometric to Hyperbolic Functions and Back Again

There are some simple relations between the trigonometric functions and the hyperbolic functions which are evident in the similarity of the analytic expressions. By allowing the arguments of cos, sin, cosh and sinh to be extended to the complex numbers we can use the analytic expressions to find direct relations between the trigonometric functions with argument θ and the hyperbolic functions with argument $i\theta$:

$$\begin{aligned}\cos\theta &= \frac{1}{2}(e^{i\theta} + e^{-i\theta}) = \cosh(i\theta)\\ \sin\theta &= \frac{1}{2i}(e^{i\theta} - e^{-i\theta}) = -i\sinh(i\theta)\\ \tan\theta &= \frac{-i\sinh(i\theta)}{\cosh(i\theta)} = -i\tanh(i\theta).\end{aligned}$$

Notice that we can use our relations above to convert the fundamental trigonometric identity

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

into

$$\cosh^2(i\theta) - \sinh^2(i\theta) = 1.$$

Similarly we can also rewrite hyperbolic functions with argument θ as trigonometric functions of $i\theta$ with complex coefficients:

$$\begin{aligned}\cosh\theta &= \cos(-i\theta) = \cos(i\theta)\\ \sinh\theta &= i\sin(-i\theta) = -i\sin(i\theta)\\ \tanh\theta &= \frac{-i\sin(i\theta)}{\cos(i\theta)} = -i\tan(i\theta).\end{aligned}$$

2.3.8 Functions of Functions

How do we construct more functions? Given two functions $f(x)$ and $g(x)$ there are some obvious ways to combine them to form a new function. We may add the functions

$$f(x) + g(x)$$

or we might multiply them

$$f(x)g(x).$$

Furthermore since the output of both $f(x)$ and $g(x)$ are real numbers for any $x \in \mathbb{R}$ then we may even raise one function as a power of the other:

$$f(x)^{g(x)}.$$

This gives us a clue to consider a more abstract way to combine functions: functions of functions. It is evidently simple for us to combine functions abstractly as

$$f(g(x)).$$

Definition 2.3.23. *Function composition is the application of one function f to the output of another function g which is defined point-wise for $x \in \mathbb{R}$ (for each value in the domain at a time) and is denoted*

$$(f \circ g)(x) \equiv f(g(x)) \qquad \forall\, x \in \mathbb{R}.$$

Note the use of the common mathematical shorthand $\forall$ which means "for all". By composing two functions we link them together in the sense that the output of one function ($g(x)$ above) becomes the input of another function ($f(x)$ above).

We note that function composition is very common, consider classical mechanics which was a major motivation for the development of the calculus, where one might expect a speed v to be defined as a function of a position x, which in turn may be defined as a function of time, t, i.e. $v(x(t))$ this function v is naturally written in terms of x, when it is called an implicit function of t but an explicit function of x. Of course if we know $x(t)$ then we can write the composition $v(x(t))$ as an explicit function of t denoted $v(t)$. For example if $v = ax + b$ and $x = t^2$ then we can also define $v(t) = at^2 + b$.

2.4 A Zoo of Functions

How many functions are there? Obviously there are an infinite number of linear functions (i.e. $f(x) = ax + b$ where $a, b \in \mathbb{R}$). So rather than try to count functions, can we classify them and write down the classes? This is a difficult question, which we will make some inroads to answering in this course. The first problem in classifying functions is how do we know whether two functions $f(x)$ and $g(x)$ are identical? The answer is that we check all the values of each function and see if they are the same for each identical input! Alternatively, as we have seen, we could check one point is the same for both functions (i.e. check that $f(x_0) = g(x_0)$) and also check that all the derivatives at the same point are the same (i.e. $f'(x_0) = g'(x_0)$, $f''(x_0) = g''(x_0)$ etc.. Notice that this second check uses the derivative (which we have not formally introduced yet, and are denoting here by a prime on the function i.e. $f' \equiv \frac{df}{dx}$) and may not require an infinite set of checks, as for example for any polynomial equation of degree n the derivatives will vanish after taking the $(n+1)$'th derivative, e.g. for two linear functions $f(x) \equiv ax + b$ and $g(x) \equiv cx + d$ we can evaluate them both at $x = 0$ and find all their

derivatives:

$$f(0) = b \qquad g(0) = d$$
$$f'(0) = a \qquad g'(0) = c.$$

So if $b = d$ and $a = c$ then the two linear functions are the same. Since we have not introduced the derivative yet we will leave this idea on the back-burner, but note that if we cannot find the derivative of a function (if there are functions for which it does not exist) then we will not be able to invoke this check of whether two functions are the same.

One is not often tempted to compare all values of two functions but, instead, to manipulate one algebraic expression for a function $f(x)$ into another $g(x)$ but there are some traps with this, which we should be aware of.

Example 2.8. *Is $f(x) \equiv \frac{x^2}{x}$ the same function as $g(x) = x$?*

We are tempted to simplify $f(x)$ as follows

$$f(x) = \frac{x^2}{x} = x = g(x) \tag{2.13}$$

and argue that the two functions are identical. But what is $f(0) = \frac{0}{0}$, such expressions are called undefined, whereas $g(0) = 0$. So the two functions are not identical. The message is that one must be careful when simplifying algebraic expressions where one might be implicitly dividing by zero.

---

3. The Limit

In which we encounter the most important idea in the course: the limit. We develop its formal definition, gain experience in computing limits and use it to define continuous functions.

This will be covered during weeks 4 and 5 of the lectures.

3.1 Motivation

A major problem in mathematics occurs whenever we encounter an expression of the form $\frac{0}{0}$ - this expression is not a number. We know this because we can see the evident (and ill-defined) division by zero, however we are nevertheless tempted to try to draw some conclusions about it. We might take any of the following stances in opposition to the idea that it is not a number:

(a.) it is zero, as any number multiplied by zero gives zero, or

(b.) it is one as any number of the form $\epsilon/\epsilon = 1$, or even

(c.) it tends to either $\pm\infty$, as any number divided by a small, positive number blows up towards plus infinity, while division by a small, negative number blows up towards negative infinity - but zero is equally close to a small positive and a small negative number, so we are even confused over which number $\frac{1}{\epsilon}$ approaches as $\epsilon \to 0$ is it plus or minus infinity?

By way of example consider the function

$$f(x) = \frac{x^2 - 9}{x - 3}$$

and resist the strong inclination to algebraically simplify the function to $(x + 3)$ using the difference of two squares as follows

$$\frac{x^2 - 9}{x - 3} = \frac{(x + 3)(x - 3)}{x - 3} \stackrel{!}{=} x + 3 \equiv g(x).$$

Why should we not do this? The function $f(x)$ is almost but not quite the same function as $g(x) = x + 3$, the functions are the same apart from at the point $x = 3$. To rewrite the line above without the exclamation mark we should add a condition so that it now reads

$$f(x) = \frac{(x+3)(x-3)}{x-3} = x + 3 = g(x) \qquad \text{for } x \neq 3$$

otherwise we would be implicitly dividing by zero. The function $f(x)$ is identical to $g(x)$ accept at the point $x = 3$ where it involves a division by zero and so becomes undefined. If one plotted all the points of $f(x)$ for $x \in \mathbb{R}$ it would be identical to the straight line $g(x) = x + 3$ apart from missing a single point at $x = 3$. If we plugged in some numbers to see what happens to $f(x)$ as x gets close to 3 we could build up a table of data to help us understand what happens before and after $f(3)$ and compare it with $g(x)$.

x	$f(x) = \frac{x^2-9}{x-3}$
2.9	5.9
2.99	5.99
2.999	5.999
2.9999	5.9999
3	Not defined.
3.0001	6.0001
3.001	6.001
3.01	6.01
3.1	6.1

x	$g(x) = x + 3$
2.9	5.9
2.99	5.99
2.999	5.999
2.9999	5.9999
3	6
3.0001	6.0001
3.001	6.001
3.01	6.01
3.1	6.1

In the table above we have evaluated $f(x)$ and $g(x)$ at and near $x = 3$. We could have picked a value for x infinitesimally close to 3, but not quite 3, e.g. $x = 3 + \epsilon$ such that $\epsilon \neq 0$ and we would have found that $f(3 + \epsilon) = 6 + \epsilon = g(3 + \epsilon)$. As $\epsilon \to 0$, $f(3 + \epsilon) \to 6$, in words, as x approaches 3, $f(x)$ approaches 6.

As mathematicians we are interested in fixing problems and a resolution to the problem for $f(x)$ at $x = 3$ that we are tempted towards is to discard the point $x = 3$ and to treat $f(x)$ as we did algebraically. We would be able to work with the function $f(x)$ to an arbitrary degree of accuracy as close to $x = 3$ as we would care to choose. So long as we do not have to do any computations at exactly $x = 3$ then we could replace $f(x)$ in all our calculations with the better behaved function $g(x) = x + 3$.

This reasoning is very disturbing. Of course it would be nice to replace a function which is an ill-defined function on $\mathbb{R}$ with one which is well-defined everywhere by adding in a point but what would we gain? And would it be sufficiently interesting to justify the seemingly illogical manoeuvre of replacing $f(x)$ with $g(x)$? Now even supposing that we could convince ourselves that we would learn enough about $f(x)$ to justify the transition to work with $g(x)$, what about other functions which have ill-defined, singular points[1], we would like to know how to go about finding the well-behaved function that we could work with.

Let us think about this last problem. If we had been presented with $f(x)$ how would we have known that we could add in a point and it would become $g(x)$? In the case we considered the algebra was simple to do and we could immediately find a way to get rid of the singularity in $f(x)$ and find $g(x)$, but what about more complicated examples. Suppose instead we considered the function

$$f(x) \equiv \frac{\sin x}{x}$$

whose graph looks like an oscillating $\sin x$ function but whose amplitude varies as $\frac{1}{x}$ (see figure 3.1). With the aid of the computer-generated graph, we can guess that approaching the singular

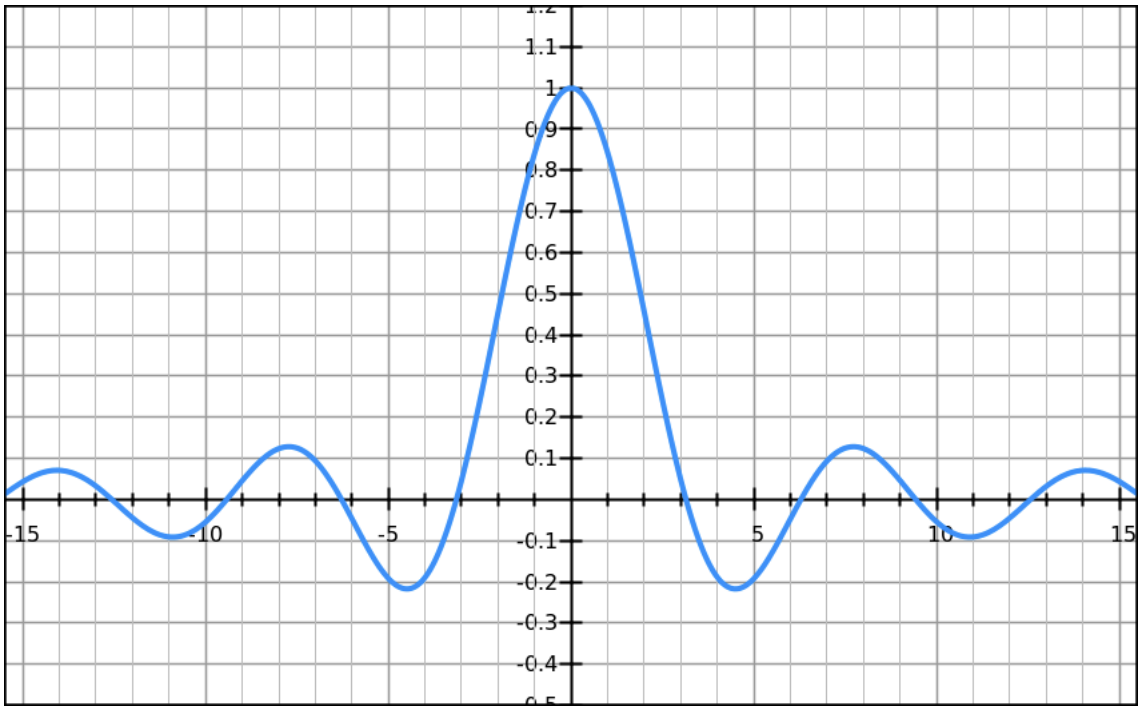


Figure 3.1: The graph of $f(x) = \frac{\sin x}{x}$.

point (as $x \to 0$) then $\frac{\sin x}{x} \to 1$. With this example we guess that the well-behaved function is

[1]Points which become ill-defined are called singularities or singular points.

found by adding the point 1 to the function to obtain:

$$g(x) \equiv \begin{cases} \frac{\sin x}{x} & x \in \mathbb{R} \setminus 0 \\ 1 & x = 0 \end{cases}$$

This is all very well, but we have been entirely loose with our method and our mathematical language. Let's try and rectify this[2].

3.2 The Existence of the Limit.

In terms of the language we have been speaking of "well-behaved" functions but we have not spelled out exactly what we mean by this. Suppose a function $f(x)$ is singular at a single point $x = x_0$. We have seen an example where we added a single point (to the definition of the function $f(x)$) - and we could determine the value of that point because it was the value the function was approaching as its argument approached arbitrarily close to the singular point x_0. The value that the function approaches is called the limit of $f(x)$ as x approaches x_0 and we denote this as

$$\lim_{x \to x_0} [f(x)].$$

We will develop this terminology to give its full definition shortly. However before doing so let's address any doubts we have about whether what we are proposing has any mathematical value. We may think it is sufficient to be able to sketch a graph and read off its tendencies as it approaches any singular points but can we trust our intuition about the output data from our computers and calculators to work out the value of a limit? Let us look at a slightly different example of an ill-defined function which will trouble us on this point. Consider now

$$f(x) \equiv \sin\left(\frac{\pi}{x}\right)$$

whose argument π/x is singular at $x = 0$. Near the singularity at $x = 0$ we may compute the value of the function to try to deduce the limit of the function as x approaches 0. We show some data in the table below:

[2]We will return later to $\frac{\sin x}{x}$ to see if our intuition from the graph was correct, but first we need to develop our mathematical technology.

x	$f(x) = \sin(\frac{\pi}{x})$
0.1	0.0000000000020665
0.01	0.0000000000206930
0.001	0.0000000002065886
0.0001	0.0000000020658863
0	Not defined.
-0.0001	-0.0000000020658863
-0.001	-0.0000000002065886
-0.01	-0.0000000000206930
-0.1	-0.0000000000020665

For our computations shown in the table we have approximated $\pi \approx 3.14159265359$ and we realise that if our calculator had only shown eight decimal places we would have thought this function to be zero in the domain above. This would have led us to conclude that as x approaches zero $f(x)$ also approaches zero. But we'd have been wrong. With a little thought (and without a calculator) we realise that the values of x we chose to evaluate $f(x)$ are special points for the sine function, as for $x = \pm\left(\frac{1}{10}\right)^n$ for $n = 1, 2, 3, 4$ we see that we were evaluating $\sin(\frac{\pi}{x}) = \sin(\frac{\pm\pi}{10^{-n}}) = \sin(\pm 10^n \pi) = 0$.

If we had chosen different values of x we would have found wildly different answers as we approached $x = 0$:

x	$f(x) = \sin(\frac{\pi}{x})$
0.7	-0.9749279121818890
0.07	0.7818314824698670
0.007	0.4338837390909680
0.0007	0.9749279121161520
0	Not defined.
-0.0007	-0.9749279121161520
-0.007	-0.4338837390909680
-0.07	-0.7818314824698670
-0.7	0.9749279121818890

Again we have used the approximation $\pi \approx 3.14159265359$ to evaluate the function in

the table. Notice how the output of the function now varies wildly between -1 and 1. This behaviour of the function is (slightly) better appreciated by considering the graph of the function which is shown in figure 3.2.

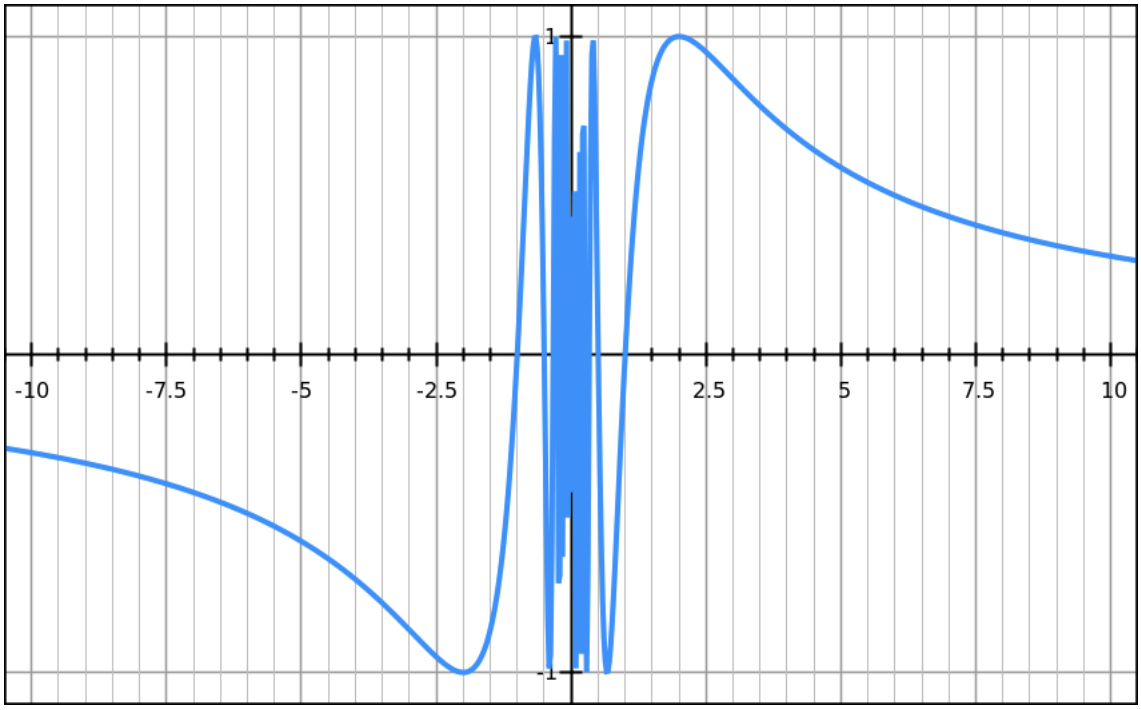


Figure 3.2: The graph of $f(x) = \sin\left(\frac{\pi}{x}\right)$.

Around $x = 0$ the graph becomes a solid blob - we might think this is a consequence of the thick line of the graph, but if we zoom in, the dense lines of the graph only persist (see figure 3.3). We are left to conclude that neither our calculator nor our graphing software is helping us to find the limit of $\sin\left(\frac{\pi}{x}\right)$ as x approaches zero.

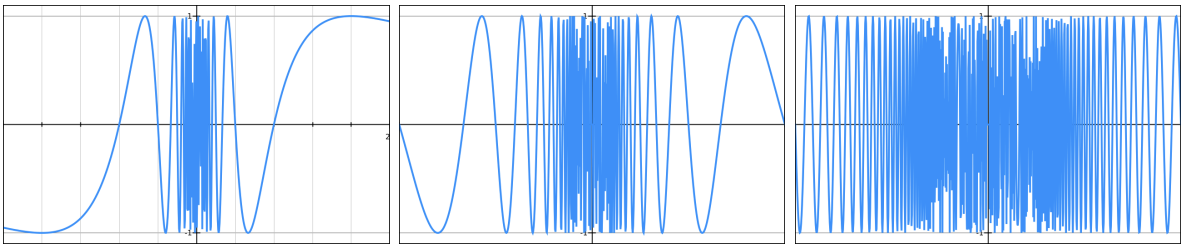


Figure 3.3: Zooming in on the graph of $f(x) = \sin\left(\frac{\pi}{x}\right)$ around $x = 0$: the blob persists!

In fact in this function $\sin\left(\frac{\pi}{x}\right)$ our scheme of discovering a limit has hit a big problem: this function does not have a limit as x approaches zero! Recall our rough definition of $\lim_{x\to x_0}[f(x)]$ as the value $f(x)$ approaches as its argument approaches arbitrarily close to x_0. In making this

definition we imagined that functions do always have such limit points - we were picturing a graph given by a continuous curve with one point missing. How can we tell if no such limit exists? We must bear in mind that we will not be able to learn much from the graph - but we could refine our definition of a limit further by saying that as x approaches closer to x_0 so $f(x)$ must approach closer to $\lim_{x\to x_0}[f(x)] \equiv L$. In other words as we restrict the argument x to only a small neighbourhood of x_0, if $f(x)$ is converging towards a value then that number is the limit, L. Mathematically we are saying that a limit exists if for $|x - x_0| < \delta$ that $|f(x) - L| < \epsilon$ for all values of $\epsilon \in \mathbb{R}^+$ and where $\delta \in \mathbb{R}^+$ (in figure 3.4 we illustrate these ranges for the graph of a simple function).

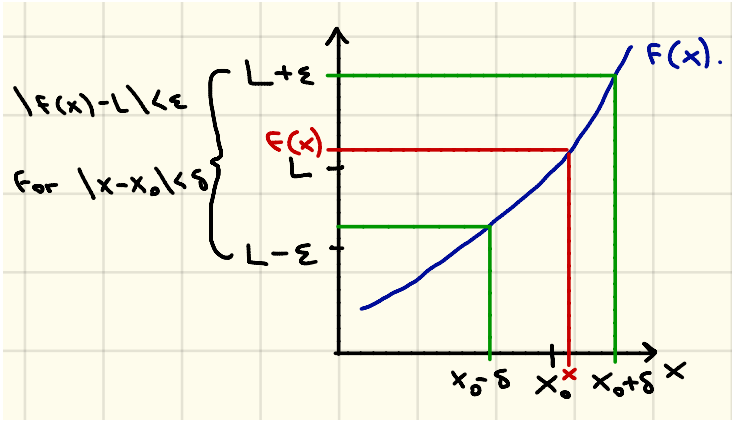


Figure 3.4: For any given $\epsilon > 0$ then if $f(x)$ lies in the range $(L - \epsilon, L + \epsilon)$ when x lies in the range $(x_0 - \delta, x_0 + \delta)$ then we say the $\lim_{x\to x_0}(f(x)) = L$.

Consider the limit of two simple functions at $x_0 = 0$: $y = 2$ and $x = 0$. In the first case the function is a horizontal line with coordinates $(x, 2)$ for all $x \in \mathbb{R}$ - by visualising the function we can immediately see the limit exists: the function can always be trapped between $2+\epsilon$ and $2-\epsilon$ for values of x near $x_0 = 0$. More algebraically as $x \to 0$ and the limit is 2: $|f(x) - 2| = 0 < \epsilon$ for $x \in (0 - \delta, 0 + \delta)$ for all $\delta > 0$. The second function $x = 0$ is a vertical line consisting of the points $(0, y)$ for all $y \in \mathbb{R}$, one cannot find a bound on the range of the function when x approaches $x = 0$, i.e. the output of the function at $x = 0$ is $\mathbb{R}$ and so we cannot find a limit that the function approaches in this second example.

Returning our thoughts to the function $f(x) \equiv \sin\left(\frac{\pi}{x}\right)$ we realise that we can always find any output value in the range $[-1, 1]$ no matter how close x is to zero, i.e. this function does not have a limit. Let's convince ourselves of this by using our mathematical definition of the limit. If there is no limit then we claim that

$$|\sin\left(\frac{\pi}{x}\right) - L| \overset{!}{<} \epsilon \qquad \forall\, \epsilon \in \mathbb{R}^+$$

is violated for a restricted range of x around zero, i.e. $|x| < \delta$. To see this pick, for example, $\epsilon = \frac{1}{2}$, then we must check

$$|\sin\left(\frac{\pi}{x}\right) - L| \stackrel{!}{<} \frac{1}{2}$$

for all x close to zero. Before we can check this we need to specify the limit L we think the function approaches. For simplicity in presenting the ideas we will limit our proof to show only that the function does not limit to $L = 0$. We wish to check whether

$$|\sin\left(\frac{\pi}{x}\right) - 0| = |\sin\left(\frac{\pi}{x}\right)| \stackrel{!}{<} \frac{1}{2}$$

for all x near $x = 0$. If we choose

$$\frac{\pi}{x} = \frac{\pi}{2} + 2n\pi$$

then $x = (\frac{1}{2} + 2n)^{-1} = \frac{2}{1+4n}$ and so for large n x can be made as close to $x = 0$ as we like and all the while at these points $\sin\left(\frac{\pi}{x}\right) = 1$, i.e.

$$|\sin\left(\frac{\pi}{x}\right)| = 1 > \frac{1}{2} = \epsilon \qquad \text{for } x = \frac{2}{1+4n}, \quad n \in \mathbb{Z}.$$

The above argument gives a specific proof that $\lim_{x\to 0}[\sin\left(\frac{\pi}{x}\right)] \neq 0$ - we do not have to repeat the proof for any other values of ϵ as the inequality has already been violated for $\epsilon = \frac{1}{2}$ and if the limit exists it must be true for all $\epsilon > 0$. What about other limits besides $L = 0$? We leave it as an exercise for the curious reader to pursue for we have some other intricate points to develop about the limit.

Thus far we have obscured an important point. In the notation $\lim_{x\to x_0}(f(x))$ we have presumed that the meaning of $x \to x_0$ is unambiguous, but of course it is not. We could approach $x = x_0$ from either above x_0 or below and, if they exist, the limits found in approaching from two different directions may be different.

Definition 3.2.1. *The left limit (or limit from below) of $f(x)$ is denoted $\lim_{x\to x_0^-}[f(x)]$ or $\lim_{x\uparrow x_0}[f(x)]$, if it exists it is the value that $f(x)$ approaches as x approaches x_0 from below i.e. values of $x < x_0$. Formally the limit $\lim_{x\to x_0^-}[f(x)] = L$ exists if for all $\epsilon > 0$ there exists $\delta > 0$ such that $|f(x) - L| < \epsilon$ for $x \in (x_0 - \delta, x_0)$.*

Definition 3.2.2. *The right limit (or limit from above) of $f(x)$ is denoted $\lim_{x\to x_0^+}[f(x)]$ or $\lim_{x\downarrow x_0}[f(x)]$, if it exists it is the value that $f(x)$ approaches as x approaches x_0 from above i.e. values of $x > x_0$. Formally the limit $\lim_{x\to x_0^+}[f(x)] = L$ exists if for all $\epsilon > 0$ there exists $\delta > 0$ such that $|f(x) - L| < \epsilon$ for $x \in (x_0, x_0 + \delta)$.*

We'll consider two examples.

Example 3.1. *Show that the limits from above and below as x approaches zero for the function $f(x) = \frac{1}{x}$ are different.*

Now 1 divided by any small number is large, however if the small number is negative then its reciprocal will be a large negative number, but if the small number is positive it will be a large positive number, i.e.

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

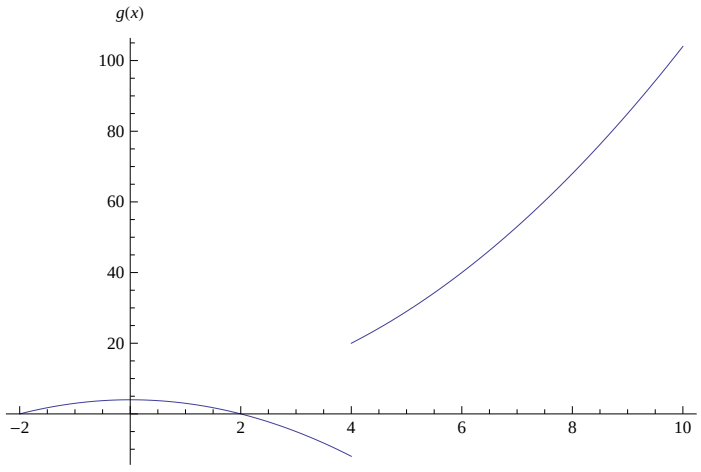
You might feel uncomfortable about setting a limit equal to ∞, after all ∞ is not a number. But this is one of the advantages of the limit, it allows us to consider values which were not in the range of the original function. In writing the limit $\lim_{x\to x_0}(f(x)) = \infty$ means only that the function's output approaches ∞ as $x \to x_0$ - we need never get to the point where we claim that the function's output is ∞.

Example 3.2. *Show that the limits from above and below as x approaches 4 for the function*

$$g(x) = \begin{cases} x^2 + 4 & x > 4 \\ -x^2 + 4 & x \leq 4 \end{cases}$$

are different.

We can sketch the graph around $x = 4$:



Notice the jump that occurs at $x = 4$. Consequently the limit will differ if we approach $x =$ from above or below. Specifically we have

$$\lim_{x\to 4^-}(g(x)) = -12 \qquad \text{and} \qquad \lim_{x\to 4^+}(g(x)) = 20.$$

This is an example of a function which is not continuous and we will soon use the limit to give a rigorous definition to the continuity of a function.

When the limit from above is equal to the limit from below then only a single notation is needed - this is the notation we had adopted initially. Let us now give the formal definition of the limit.

Definition 3.2.3. *The limit of a function $f(x)$ exists and is equal to L as $x \to x_0$ is denoted*

$$\lim_{x \to x_0} [f(x)] = L$$

if and only if

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

for all $\epsilon > 0$ where $0 < |x - x_0| < \delta$ for $\delta > 0$.

Comment(s). *On the existence of a limit...*

1. *The definition of the limit as stated above implicitly means that the limit from above is equal to the limit from below on any open interval (a, b), i.e.*

$$\lim_{x \to x_0^-} [f(x)] = \lim_{x \to x_0^+} [f(x)] \equiv \lim_{x \to x_0} [f(x)].$$

2. *This so called $\epsilon - \delta$ definition of the limit was first given by Bolzano in 1817, but is most closely associated with Augustin-Louis Cauchy[3] (1789-1857) of whom it is claimed more mathematical concepts are named after than for any other mathematician.*

We will now take a short digression from our topic of limits to discuss one important use of the limit: to define continuous functions.

3.2.1 Continuous Functions.

Definition 3.2.4. *A function $f(x)$ is continuous at the point $x = x_0$ if $\lim_{x \to x_0}[f(x)]$ exists and is equal to $f(x_0)$.*

Definition 3.2.5. *A function $f(x)$ is continuous if it is continuous at all points in its domain.*

Comment(s). *On continuous functions...*

1. *Continuity implies that a small change in the argument of a function gives a small change in its output, so for example, the function $g(x)$ used in example 3.2 is not continuous.*

[3] Even though it was only expressed in this modern form by Karl Weierstrass.

2. *Colloquially it is often said that continuous functions are those for which the graph can be drawn without lifting the pen from the paper (however be wary of functions such as* $\sin\left(\frac{\pi}{x}\right)$ *which you might draw without lifting your pen by forming a blob around* $x = 0$*).*

3. *Some functions which we might think of as very well-behaved are not continuous, for example* $f(x) = \frac{1}{x}$ *on* $\mathbb{R}$*.*

4. *Some functions which we might not think of as being so well-behaved are continuous, for example* $f(x) = |x|$*. See the following example.*

Example 3.3. *The function* $f(x) = |x|$ *is a continuous function.*

The graph of the function is shown in figure 3.5.

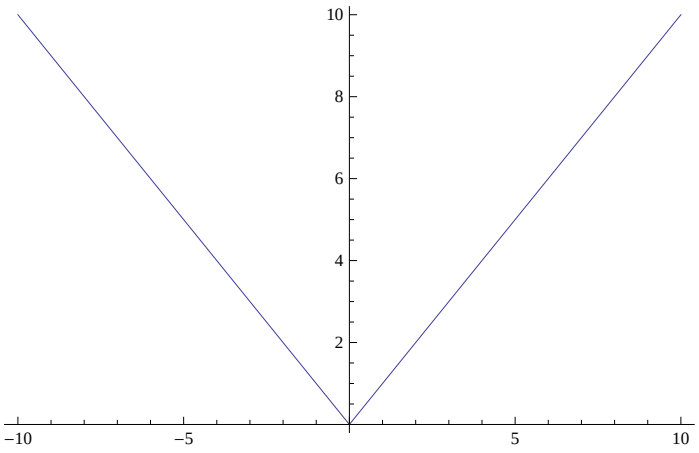


Figure 3.5: The graph of $f(x) = |x|$.

It should be obvious that it is continuous when $x \neq 0$ *and the point* $x = 0$ *is our main concern. If we take the limit from above we have*

$$\lim_{x\to 0^+}[|x|] = \lim_{x\to 0^+}[x] = 0$$

and from below we have

$$\lim_{x\to 0^-}[|x|] = \lim_{x\to 0^-}[-x] = 0.$$

Hence the limits from above and from below are both equal to zero and hence the limit $\lim_{x\to 0}[|x|]$ *exists and is also equal to zero. Since this is also the value of the function at zero (i.e.* $|0| = 0$*) then we have*

$$\lim_{x\to 0}[|x|] = |0|.$$

Hence the modulus function is a continuous function, despite its nasty-looking right-angle at $x = 0$.

Exercise 3.1. *Show that the function*

$$f(x) \equiv \begin{cases} e^x & x \leq 2 \\ e^{4-x} & x > 2 \end{cases}$$

is continuous. The graph of the function is shown in figure 3.6.

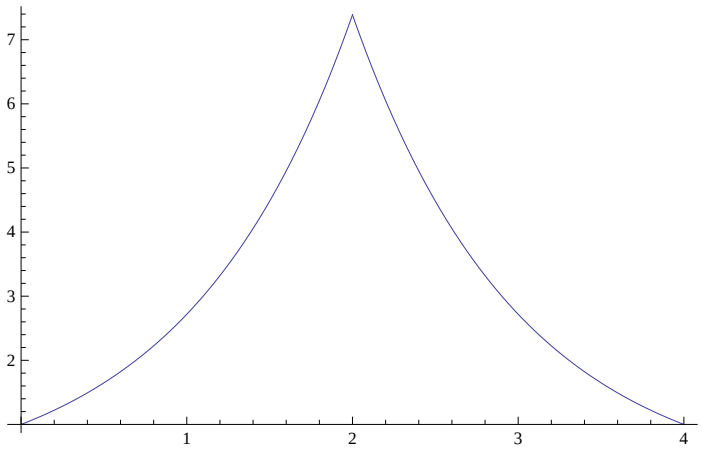


Figure 3.6: The function $f(x)$ defined in exercise 3.1 about $x = 2$.

The Intermediate Value Theorem

Knowing that a function is continuous has some interesting consequences. One of the seemingly straightforward consequences is called the intermediate value theorem and it can be used to make some fascinating observations. Let's define the theorem here.

Theorem 3.1. *(The Intermediate Value Theorem) Let* $f(x)$ *be a continuous function on* $[a, b]$ *and suppose* $f(a) < y < f(b)$ *then there exists an* $x = c$ *with* $c \in (a, b)$ *such that* $f(c) = y$.

We emphasise the theorem on a graph in figure 3.7.

This theorem appears almost unworthy of the dignified title of a theorem. It does after all seem obvious (once the mathematical jargon has been decoded). It is saying that if a continuous function is defined at $f(a)$ and at $f(b)$ then as it is continuous it has no gaps and so must pass through all output values between $f(a)$ and $f(b)$. It may even pass through certain values many times (as in figure 3.7 the only constraint is that it moves continuously from $f(a)$ to $f(b)$).

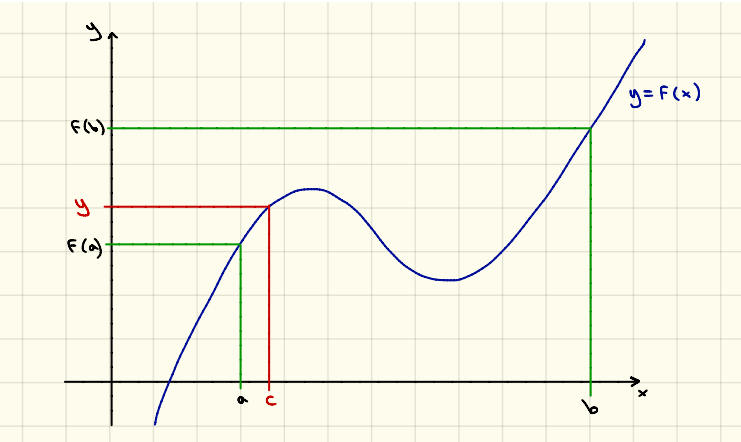


Figure 3.7: An illustration of the intermediate value theorem.

One can use the intermediate value theorem to find approximate expressions for irrational numbers. For example one might consider the continuous function $f(x) = x^2 - 2$ to evaluate $\sqrt{2}$[4]. Now we know that when $f(x) = 0$ then $x = \sqrt{2}$, but why can we be sure that $f(x)$ does equal zero for some value of x. The intermediate value theorem tells us that such a value exists if we can identify a and b such that $f(a) < 0 < f(b)$. Of course we can do this, for example we can check that $f(1) = -1$ and $f(2) = 2$, i.e. $a = 1$ and $b = 2$ tells us that there exists $1 < c < 2$ such that $f(c) = 0$. Hence we have observed that $\sqrt{2}$ lies in the interval $(1, 2)$. For a better approximation one can reduce the range (a, b). In principle this method for other polynomial functions gives us some certainty that roots exist, where the function is zero.

A neat observation that follows from a particular version of the intermediate value theorem. Assume we have a continuous function $f(x)$ on the interval $[0, 1]$ such that $0 \leq f(x) \leq 1$, then there exists $c \in [0, 1]$ such that $f(c) = c$, i.e. at one point the function behaves like the identity function! Now $f : [0, 1] \to [0, 1]$ so you can picture f as a set of reassignments of all the numbers in the interval $[0, 1]$, and we are arguing that one number c is mapped to itself - hence it is called a fixed point. Let's use the intermediate value theorem to prove our claim. To do so, consider the function $g(x) \equiv f(x) - x$ which is continuous (as $f(x)$ is continuous by assumption and the identity function $f(x) = x$ is continuous too). Noting that $g(1) \leq g(0)$ as $g(0) = f(0) - 0 = f(0) \geq 0$ and $g(1) = f(1) - 1 \leq 0$ then by the intermediate value theorem we know that there exist some value $x = c$ such that $g(c) = 0$ where $c \in [0, 1]$. Therefore we have that $g(c) = f(c) - c = 0$ hence $f(c) = c$, as claimed. For example consider the sine function or the cosine function on the interval $[0, 1]$ - this theorem gives us the insight that there are values $\sin(c) = c$ and $\cos(\hat{c}) = \hat{c}$. Can you find a way to estimate these fixed points of sine and cosine?

[4]See the University of Utah video by Jim Fowler on Coursera course "Calculus 1" to see this done in detail.

The intermediate value theorem is seemingly simple but its applications are manifold. In particular it is the intermediate value theorem that lies behind Brouwer's fixed point theorem which has applications in the investigation of differential equations, differential geometry and even game theory. From small acorns, mighty oak trees grow.

3.2.2 Limits Involving Infinity.

The interplay between the notation used for limits and the notion of infinity is interesting and can be confusing. Earlier we remarked how it was sensible to write that $\lim_{x\to x_0}[f(x)] = \infty$ even though infinity is not a number. The analogy that is useful is that of a function with a single point missing for example $\frac{\sin x}{x}$ is undefined at $x = 0$ while we claimed earlier (and still have to prove) that $\lim_{x\to 0}[\frac{\sin x}{x}] = 1$. It does not trouble us that 1 is not in the range of $\frac{\sin x}{x}$ and yet its limit is 1 - the limit tells us only that the function is approaching a value, not that it attains that value. In the same way although a function maps to $\mathbb{R}$ (and not to $\mathbb{R} \cup \pm\infty$) it does not present a problem to say that its value tends towards infinity.

It is also possible to consider the value of a function as its argument tends towards infinity, i.e. as $x \to \infty$ or similarly as $x \to -\infty$ as the 'tending to' is not the same as x reaching infinity. Hence the following "asymptotic" limit is well-defined.

Definition 3.2.6.

$$\lim_{x\to\infty}[f(x)] = L \qquad \Longleftrightarrow \qquad \forall\, \epsilon > 0, \quad \exists\, X > 0 \quad \textit{such that } |f(x) - L| < \epsilon \quad \textit{for } x > X.$$

Comment(s). *(On limits to infinity...)*

1. *Compare this definition with the definition for approaching a limit from below, i.e. contrast $x \in (x_0 - \delta, x_0)$ with $x \in (X, \infty)$ used here. Of course ∞ can only be approached from below which explains the change of form for the definition above.*

2. *Similarly one can take the limit $x \to -\infty$, which takes the form of a limit from above:*

$$\lim_{x\to-\infty}[f(x)] = L \qquad \Longleftrightarrow \qquad \forall\, \epsilon > 0, \quad \exists\, X < 0 \quad \textit{such that } |f(x) - L| < \epsilon \quad \textit{for } x < X.$$

Example 3.4. *Consider the function*

$$f(\alpha) \equiv \tanh(\alpha)$$

and show that

$$\lim_{\alpha\to\infty}[f(\alpha)] = 1.$$

We have to show that for all $\epsilon > 0$ there exists an $X > 0$ such that $|f(\alpha) - 1| < \epsilon$ when $\alpha > X$. We will first give a sketch proof using the graph of $\tanh\alpha$ shown in figure 2.13. Our aim is to show that $\tanh\alpha$ approaches 1 to a specified degree of accuracy encoded in ϵ so for example if we started with $\epsilon = \frac{1}{100}$ then we aim to show that we can identify a range of $\alpha > X$ such that $|\tanh x - 1| < \frac{1}{100}$. From the graph we understand this is possible immediately as $\tanh\alpha$ asymptotes to $y = 1$ at infinity, hence it is only a small matter of computation to identify when $|\tanh X - 1| = \frac{1}{100}$, i.e. we aim to solve $\tanh X = \frac{99}{100}$ which gives, roughly $X = 2.65$ radians. This is not a proof, but by this stage we should be convinced that we could repeat the computation of X for any ϵ.

Let us consider a more challenging version of the same question, let us try to show that the limit is 1 directly. Recalling the definition of the hyperbolic functions as points on the right hand branch of the hyperbola $x^2 - y^2 = 1$, where $x = \cosh\alpha$ and $y = \sinh\alpha$ then we have

$$\tanh\alpha = \frac{\sinh\alpha}{\cosh\alpha} = \frac{y}{x} = \frac{y}{\sqrt{1+y^2}}$$

As $y = \sinh\alpha$ as $\alpha \to \infty$ then so $y \to \infty$. We have recast the limit as

$$\lim_{\alpha\to\infty}[\tanh\alpha] = \lim_{y\to\infty}[\frac{y}{\sqrt{1+y^2}}] = \lim_{y\to\infty}[\frac{y}{\sqrt{1+y^2}} \times \frac{\left(\frac{1}{y}\right)}{\left(\frac{1}{y}\right)}] = \lim_{y\to\infty}[\frac{1}{\sqrt{\frac{1}{y^2}+1}}] = 1$$

as $\lim_{y\to\infty}[\frac{1}{y^2}] = 0$. In the above line we have made a number of standard manipulations of limits and while it should certainly be possible to read and understand the line of working above, it should also make us a little worried as we have not yet discussed the fundamentals of working with limits. The working above is correct, but what we will discuss next is why these lines of working are valid.

3.3 Working with Limits

In this section we will develop the standard set of tools for being able to compute limits.

3.3.1 Rules for Limits of Composite Expressions.

Our first rules concern the limits of sums, products and quotients of functions. Before giving the rules, let us think about the about the sum of two functions $h(x) \equiv f(x) + g(x)$. Now as a definition of a new function $h(x)$ this makes sense so long as both $f(x) \in \mathbb{R}$ and $g(x) \in \mathbb{R}$ as then we can use the addition of numbers in $\mathbb{R}$ to build up the function $h(x)$ for all points $x \in \mathbb{R}$.

Now if it happens that the functions, rather than be well-defined in $\mathbb{R}$, are slightly weaker but do still have well-defined limits which lie in $\mathbb{R}$, then we can quickly say that the limit of $h(x)$ is equal to the sum of the limits $f(x)$ and $g(x)$: "the limit of the sum is the sum of the limits", so long as the limits both lie in $\mathbb{R}$. With the same constraint that the limit must be a well-defined number we can convince ourselves that also the limit of a product is the product of the limits. But what about division, well then we must take care as although zero is a well-defined number the operation of division by zero is ill-defined. Let's list the rules.

Sums, Product and Quotients of Limits

If $\lim_{x\to x_0}[f(x)] = a$ and $\lim_{x\to x_0}[g(x)] = b$ where $a, b \in \mathbb{R}$ then

1. $$\lim_{x\to x_0}[f(x)+g(x)] = \lim_{x\to x_0}[f(x)] + \lim_{x\to x_0}[g(x)] = a+b$$

2. $$\lim_{x\to x_0}[f(x)g(x)] = \lim_{x\to x_0}[f(x)] \lim_{x\to x_0}[g(x)] = ab$$

3. $$\lim_{x\to x_0}[\frac{f(x)}{g(x)}] = \frac{\lim_{x\to x_0}[f(x)]}{\lim_{x\to x_0}[g(x)]} = \frac{a}{b} \text{ only if } \lim_{x\to x_0}[g(x)] \neq 0.$$

We have laboured the point that these rules for limits are not generally valid if $\lim_{x\to x_0}[f(x)] = \infty$. Why? As infinity is not a real number we do not, for example, have a way to define its addition and subtraction in $\mathbb{R}$. Let's convince ourselves that this is a problem. Let $\lim_{x\to x_0}[f(x)] = \infty$ and $\lim_{x\to x_0}[g(x)] = 1$, then if we could apply the rules above for the sum of limits we would find

$$\lim_{x\to x_0}[f(x)+g(x)] \stackrel{!}{=} \infty + 1 = \infty \tag{3.1}$$

$$\lim_{x\to x_0}[f(x)-f(x)] \stackrel{!}{=} \infty - \infty = \lim_{x\to x_0}[0] = 0 \tag{3.2}$$

In fact the first of the above equations presents no problems, but it is the second one where we have subtracted one infinite limit from another infinite limit that will cause all manner of illogical deductions, for example, if this second rule were correct we could write

$$\lim_{x\to x_0}[f(x)+g(x)-f(x)] \stackrel{!}{=} \infty + 1 - \infty \stackrel{!}{=} 0$$

i.e. that $1 \stackrel{!}{=} 0$ which is a problem. Hence we must take more care when manipulating sums of infinite limits and so these general rules above are not generally valid when the limit is infinite.

Example 3.5. *Evaluate*

$$\lim_{x\to 3}\left[\frac{x^2-9}{x-3}\right].$$

$$\lim_{x\to 3}\left[\frac{x^2-9}{x-3}\right]=\lim_{x\to 3}\left[\frac{(x-3)(x+3)}{x-3}\right]=\lim_{x\to 3}\big[(x+3)\big]=\lim_{x\to 3}\big[x\big]+\lim_{x\to 3}\big[3\big]=3+3=6.$$

Example 3.6. *Evaluate*

$$\lim_{x\to -1}\left[\frac{x^2-4x-5}{x(x+1)}\right].$$

$$\lim_{x\to -1}\left[\frac{x^2-4x-5}{x(x+1)}\right]=\lim_{x\to -1}\left[\frac{(x+1)(x-5)}{x(x+1)}\right]=\lim_{x\to -1}\left[\frac{(x-5)}{x}\right]=\frac{\lim_{x\to -1}\big[(x-5)\big]}{\lim_{x\to -1}\big[x\big]}=\frac{-6}{-1}=6.$$

Example 3.7. *Evaluate*

$$\lim_{x\to\infty}\left[\frac{x^2+7x-3}{3x^2-18x+24}\right].$$

$$\begin{aligned}
\lim_{x\to\infty}\left[\frac{x^2+7x-3}{3x^2-18x+24}\right] &= \lim_{x\to\infty}\left[\frac{x^2+7x-3}{3x^2-18x+24}\times\frac{1/x^2}{1/x^2}\right]\\
&= \lim_{x\to\infty}\left[\frac{1+\frac{7}{x}-\frac{3}{x^2}}{3-\frac{18}{x}+\frac{24}{x^2}}\right]\\
&= \frac{\lim_{x\to\infty}\left[1+\frac{7}{x}-\frac{3}{x^2}\right]}{\lim_{x\to\infty}\left[3-\frac{18}{x}+\frac{24}{x^2}\right]}\\
&= \frac{\lim_{x\to\infty}\left[1\right]+\lim_{x\to\infty}\left[\frac{7}{x}\right]-\lim_{x\to\infty}\left[\frac{3}{x^2}\right]}{\lim_{x\to\infty}\left[3\right]-\lim_{x\to\infty}\left[\frac{18}{x}\right]+\lim_{x\to\infty}\left[\frac{24}{x^2}\right]}\\
&= \frac{1+0-0}{3-0+0}\\
&= \frac{1}{3}.
\end{aligned}$$

Notice in the first line that we have multiplied the function inside the limit by 1 *written in the form* $\frac{1/x^2}{1/x^2}$ *- doing this to a function would normally result in changing the function so it is no longer defined at* $x=0$*, but doing so inside the limit where* $x\to\infty$ *means we are not considering the function near* $x=0$ *so the non-trivial change to the function does not affect the value of the limit.*

We will introduce a final rule that can be very useful in evaluating limits, this is a rule for the limit of a composite function.

If $\lim_{x\to x_0}[g(x)] = b$, $\lim_{x\to b}[f(x)] = a$ and $f(x)$ is a continuous function then for $x_0 \in \mathbb{R}$ we have

4.
$$\lim_{x\to x_0}[f(g(x))] = f(\lim_{x\to x_0}[g(x)]) = a.$$

You should convince yourself that this rule is numerically plausible for well-behaved functions and for such functions it is a consequence of the rules for the limits of sums, products and quotients for all functions which can be written as a sum, product or quotient.

Comment(s). *(On limits of composite functions)*

1. *The conditions surrounding this limit are more constraining than for the other rules, note in particular that as $x_0 \in \mathbb{R}$ the asymptotic limit $x_0 \to \infty$ is not necessarily covered by the rule. However when facing such a limit one may always make a change of variables $x \equiv \frac{1}{y}$ such that the limit becomes $y \to 0$ and is then covered by the rule. We will discuss changing the limit variable in more detail in the upcoming section on the evaluation of limits.*

2. *Notice that functions $f(x)$ which are not continuous are not covered by this rule. Why is this? Consider the following functions:*
$$f(x) = \begin{cases} x^2+1 & x \neq 0 \\ 7 & x = 0 \end{cases} \qquad \text{and} \qquad g(x) = 0$$
As $\lim_{x\to 0}[f(x)] = 1 \neq f(0) = 7$ then $f(x)$ is not a continuous function. Now let us note that the composite function is $f(g(x)) = f(0) = 7$ is a continuous function with a well-defined limit, i.e. we know that $\lim_{x\to 0}[f(g(x))] = \lim_{x\to 0}[7] = 7$. However if we had attempted (recklessly) to use the rule above we would have needed $\lim_{x\to 0^+}[g(x)] = 0$ and $\lim_{x\to 0}[f(x)] = 1$ and so if we could use the rule we would find
$$\lim_{x\to 0}[f(g(x))] \stackrel{!}{=} f(\lim_{x\to 0}[g(x)]) \stackrel{!}{=} 1.$$
Note in the line above the rule for limits of composite functions has been applied (in order to check if it is valid for a function with a discontinuity - we see that it is not). We draw attention to that the constraint on using the rule for taking limits of composite functions is only on the function $f(x)$ being continuous, there is no constraint that $g(x)$ be a continuous function.

We will illustrate its utility with two examples below.

Example 3.8. *Evaluate*

$$\lim_{x\to\frac{\pi}{2}}\left[e^{\left(\cos^2 x\right)}\right].$$

$$\lim_{x\to\frac{\pi}{2}}\left[e^{\left(\cos^2 x\right)}\right] = e^{\lim_{x\to\frac{\pi}{2}}\left[\cos^2 x\right]} = e^0 = 1.$$

If we wished to compare this with the abstract formulation of the limit rule we would define

$$f(x) \equiv e^x \qquad \textit{and} \qquad g(x) = \cos^2 x$$

and then note that $\lim_{x\to\frac{\pi}{2}}[g(x)] = 0$ *and* $\lim_{x\to 0}[f(x)] = 1$.

Example 3.9. *Evaluate*

$$\lim_{x\to 1}\left[\sin\left(\frac{\pi(x^2-1)}{4(x-1)}\right)\right]$$

where $a \in \mathbb{R}^+$.

$$\lim_{x\to 1}\left[\sin\left(\frac{\pi(x^2-1)}{4(x-1)}\right)\right] = \sin\left(\lim_{x\to 1}\left[\left(\frac{\pi(x^2-1)}{4(x-1)}\right)\right]\right) = \sin\left(\lim_{x\to 1}\left[\frac{\pi}{4}(x+1)\right]\right) = 1.$$

To further convince ourselves that this is correct we show the graph in figure 3.8

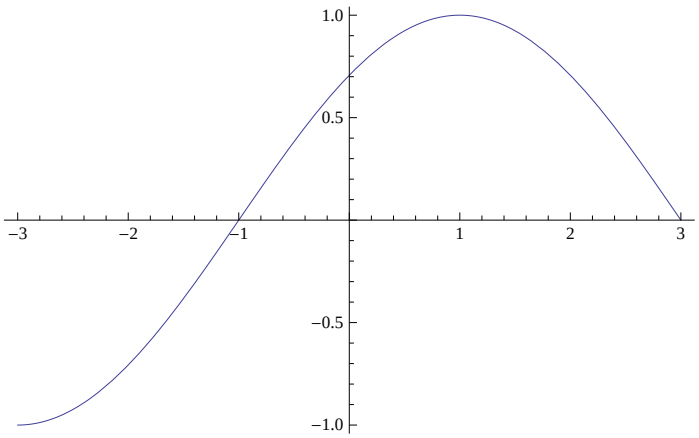


Figure 3.8: The graph of $f(x) = \sin\left(\frac{\pi(x^2-1)}{4(x-1)}\right)$ in the vicinity of $x = 1$.

A very common trick that can be used to manipulate expressions is to replace a function with the exponential function acting on the natural logarithm, e.g.

$$\lim_{x\to x_0}[f(x)^{g(x)}] = \lim_{x\to x_0}[e^{\ln(f(x)^{g(x)})}] = \lim_{x\to x_0}[e^{g(x)\ln(f(x))}] = e^{\lim_{x\to x_0}[g(x)\ln(f(x))]}.$$

It's not clear at this stage that we have much call for such a rearrangement, but as the functions we consider become more complicated it will be useful for us to try such operations in order to split complicated limits into simpler ones. The logarithm allows powers to be re-cast as coefficients which can prove very useful, for example a limit that you will evaluate in the tutorial exercises can be rearranged as

$$\lim_{x\to 0}[(1+x)^{\frac{1}{x}}] = e^{\lim_{x\to 0}[\frac{1}{x}\ln(1+x)]}.$$

Since we do not know how to evaluate $\lim_{x\to 0}[\frac{1}{x}\ln(1+x)]$ yet we pause here and leave the development of this limit for later after we have shown that $\lim_{x\to 0}[\frac{1}{x}\ln(1+x)] = 1$.

3.3.2 Multiple Limits.

Functions of multiple variables may have their limit taken for each variable. Of course when one takes a limit the global structure of the function is lost and only some of the local information remains about the limit point. Consequently for functions of multiple variables it will in general make a difference in which order limits are taken, there is an implicit ordering in which the notation indicates the limits will be taken:

$$\lim_{x\to x_0}\lim_{y\to y_0} f(x,y) \equiv \lim_{x\to x_0}\left[\lim_{y\to y_0}\big[f(x,y)\big]\right].$$

Example 3.10. *Show that changing the order the limits are taken in the following double-limit changes its value:*

$$\lim_{x\to\infty}\lim_{y\to-\infty}(1+\tanh(x+y)).$$

$$\lim_{x\to\infty}\lim_{y\to-\infty}(1+\tanh(x+y)) = \lim_{x\to\infty}(1-1) = 0$$

while

$$\lim_{y\to-\infty}\lim_{x\to\infty}(1+\tanh(x+y)) = \lim_{y\to-\infty}(1+1) = 2.$$

We include a sketch of $(1+tanh(x+y))$ *in figure 3.9 to help us visualise the limits.*

However there are many functions for which the order of the limits does not change the evaluation.

Example 3.11. *Show that changing the order the limits are taken in the following double-limit does not change its value:*

$$\lim_{x\to 0}\lim_{y\to 1}(x^2y - e^{-x-y}).$$

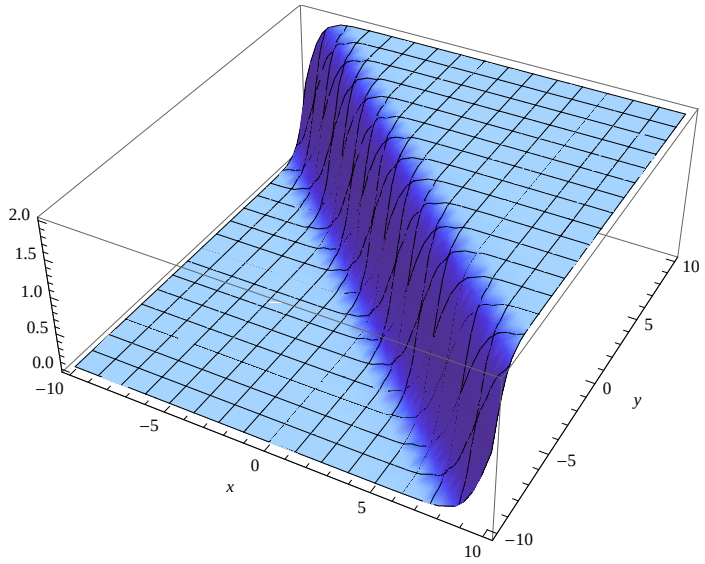


Figure 3.9: A sketch of the function $f(x, y) = 1 + \tanh(x + y)$.

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

and

$$\lim_{y\to 1}\lim_{x\to 0}(x^2y - e^{-x-y}) = \lim_{y\to 1}(-e^{-y}) = -\frac{1}{e}.$$

We include a sketch of $x^2y - e^{-x-y}$ in figure 3.10 to help us visualise the limits.

We do not intend to draw any conclusions about the class of functions and double-limits for which it is possible to change the order of the limits. However you can begin to see that the order of the limits specifies the path one takes in approaching a limit on a surface, or on a more general multi-variable function. Consequently the condition for a a multi-variable function to be continuous (i.e. the limit is independent of the path) is constraining. This will be a subject of study in some of the following courses in your degree.

3.3.3 Evaluating standard limits

So far we have spent some time establishing how one might tell if a limit exists, but we have only found limits for a small class of functions. Many limits can be evaluated by using the rules for manipulating sums and products of limits. The interesting cases are those which cannot be investigated this way. We have met only a few of these interesting cases so far, including those of the form

$$\lim_{x\to x_0}\left[\frac{f(x)}{g(x)}\right]$$

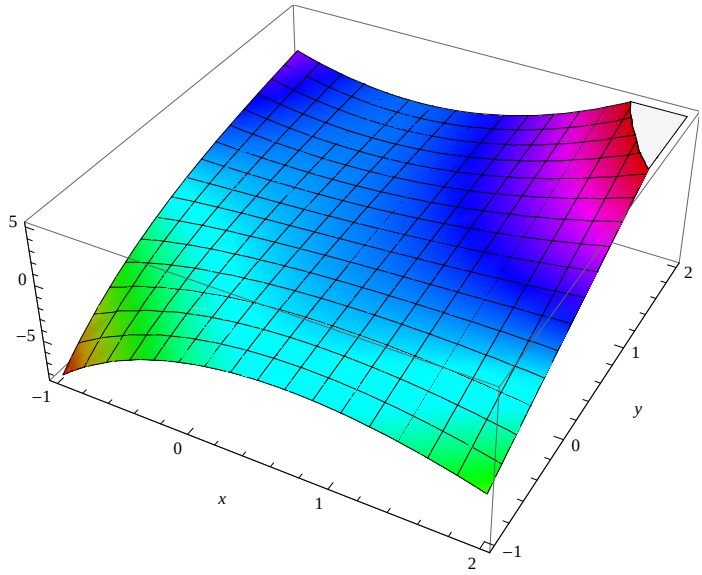


Figure 3.10: A sketch of the function $f(x,y) = x^2 y - e^{-x-y}$.

where $\lim_{x\to x_0}[f(x)] \in \mathbb{R}$, $\lim_{x\to x_0}[g(x)] = 0$, and $g(x)$ is a factor of $f(x)$, then if $h(x) \equiv \frac{f(x)}{g(x)}$ inside the limit and $\lim_{x\to x_0}[h(x)]$ exists we can evaluate the limit. We saw this method used in examples 3.5 and 3.6. When the limit is taken as $x \to \pm\infty$ on ratios of polynomial functions of the same degree we can evaluate the limit by looking at the terms of highest order, i.e.

$$\lim_{x\to\infty}\left[\frac{\sum_{j=0}^{n} a_j x^j}{\sum_{k=0}^{n} b_k x^k}\right] = \lim_{x\to\infty}\left[\frac{a_n x^n}{b_n x^n}\right] = \lim_{x\to\infty}\left[\frac{a_n}{b_n}\right] = \frac{a_n}{b_n}.$$

We saw this method used in example 3.7.

Inspired by the graph, shown in figure 3.1 we claimed earlier that $\lim_{x\to 0}\left[\frac{\sin x}{x}\right] = 1$. Now we could at this stage invoke the infinite power series for x and convince ourselves that our guess was correct, i.e.

$$\begin{aligned}
\lim_{x\to 0}\left[\frac{\sin x}{x}\right] &= \lim_{x\to 0}\left[\frac{\sum_{n,\,\text{odd}}(-1)^n \frac{x^n}{n!}}{x}\right] \\
&= \lim_{x\to 0}\left[\frac{x - \frac{x^3}{3!} + \frac{x^5}{5!} - \mathcal{O}(x^7)}{x}\right] \\
&= \lim_{x\to 0}\left[1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \mathcal{O}(x^6)\right] \\
&= \lim_{x\to 0}\left[1\right] - \lim_{x\to 0}\left[\frac{x^2}{3!}\right] + \lim_{x\to 0}\left[\frac{x^4}{5!}\right] - \lim_{x\to 0}\left[\mathcal{O}(x^6)\right] \\
&= 1
\end{aligned}$$

Notice the use of the $\mathcal{O}(x^7)$ notation above - in this context it simply means that the size of the next largest terms in the infinite series, after those we have written out, is proportional to

x^7. This notation is a shorthand and allows us to write only the leading terms in an infinite power series.

However our formula for the sine function as an infinite power series rests on some conjectures we have yet to prove. Fortunately there is another common method used to evaluate limits, which we can use to confirm this function limits to one as x approaches zero.

The Sandwich Theorem

This theorem is sometimes called the Squeeze Theorem or the Pinching Theorem.

Theorem 3.2. *The Sandwich Theorem. If*

$$f(x) \leq g(x) \leq h(x)$$

for x such that $|x - x_0| < \delta$ and if $\lim_{x\to x_0}[f(x)] = \lim_{x\to x_0}[h(x)] = L$ where $L \in \mathbb{R}$ then $\lim_{x\to x_0}[g(x)] = L$.

That is if we can identify two functions $f(x)$ and $h(x)$ that sandwich the function whose limit we are interested in $g(x)$ in the vicinity of x in the limit, and that they are equal in the limit then as $g(x)$ lies between them, then it too must equal the same limit. This theorem is very helpful if we can find two sandwiching functions and their limit is simpler to evaluate than the function in the middle of the sandwich. Let us look at some examples.

Example 3.12. *Use the sandwich theorem to prove that*

$$\lim_{x\to 0}\left[\frac{\sin x}{x}\right] = 1.$$

The trickiest part in applying the sandwich theorem is identifying the bounding functions which lie above and below the function we are interested in, in the vicinity of the limit. To find some useful inequalities we will look again at the two right-angled triangles (with sides of length $\{\cos\theta, \sin\theta, 1\}$ and the second with edges of length $\{1, \tan\theta, \sec\theta\}$) which bound the segment of the unit circle subtending an angle θ at the centre of the circle - this is hard to comprehend without a diagram, fortunately we have drawn the central idea in an earlier image - see figure 2.7. The length of the arc shown in the figure in radians is θ, so from the diagram we immediately have the following trigonometric inequality when $0 \leq \theta \leq \frac{\pi}{2}$ (note θ is used as a positive length in constructing the inequality):

$$\sin\theta \leq \theta \leq \tan\theta \qquad \text{for } 0 \leq \theta \leq \frac{\pi}{2}$$

We may usefully split this inequality in two and make use of both

$$\sin\theta \leq \theta \qquad \text{and} \qquad \theta \leq \tan\theta \qquad \text{for } 0 \leq \theta \leq \frac{\pi}{2}.$$

From the first inequality we may divide through (recall $\theta > 0$ here) to obtain

$$\frac{\sin\theta}{\theta} \leq 1 \qquad \text{for } 0 \leq \theta \leq \frac{\pi}{2}$$

which will give the upper bound we will use when we invoke the sandwich theorem. While from $\theta \leq \tan\theta = \frac{\sin\theta}{\cos\theta}$ we have

$$\cos\theta \leq \frac{\sin\theta}{\theta} \qquad \text{for } 0 \leq \theta \leq \frac{\pi}{2}.$$

Altogether we have

$$\cos\theta \leq \frac{\sin\theta}{\theta} \leq 1 \qquad \text{for } 0 \leq \theta \leq \frac{\pi}{2}$$

which we may extend to negative θ as $\cos(-\theta) = \cos(\theta)$ and $\frac{\sin(-\theta)}{-\theta} = \frac{\sin(\theta)}{\theta}$. So we have,

$$\cos\theta \leq \frac{\sin\theta}{\theta} \leq 1 \qquad \text{for } -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$$

which we may use in the sandwich theorem as by taking the limit to zero we find:

$$1 = \lim_{\theta\to 0}(\cos\theta) \leq \lim_{\theta\to 0}(\frac{\sin\theta}{\theta}) \leq \lim_{\theta\to 0}(1) = 1$$

hence as the upper bound and the lower bound both limit to one as x approaches zero, by the sandwich theorem we have that

$$\lim_{\theta\to 0}(\frac{\sin\theta}{\theta}) = 1$$

as required.

Example 3.13. *Use the sandwich theorem to prove that*

$$\lim_{x\to 0}\left[x\sin\left(\frac{1}{x}\right)\right] = 0.$$

In this example we are at an advantage as it features a trigonometric function and our first thought might be to try to use

$$-1 \leq \sin\left(\frac{1}{x}\right) \leq 1.$$

To turn this into something close to our expression, without losing the order of the inequality, we must multiply by a positive number. Hence let's multiply by $|x|$ rather than x to obtain:

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

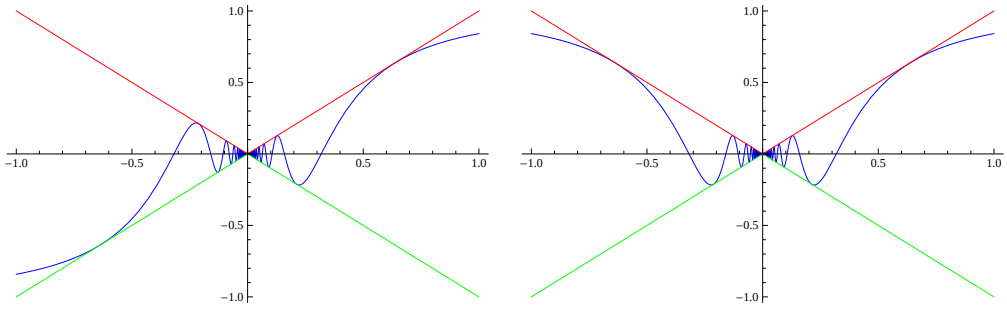


Figure 3.11: In the left-hand graph we show $-|x| \leq |x|\sin(1/x) \leq |x|$, while on the right we show $-|x| \leq x\sin(1/x) \leq |x|$

In the graph on the left of figure 3.11 we sketch the graphs of these functions to see the inequality in terms of the curves. How do we arrive at an inequality useful for the problem at hand? Now we notice that $-|x| \leq x \leq |x|$*, hence we now can deduce*

$$-|x| \leq x\sin\left(\frac{1}{x}\right) \leq |x|$$

and this inequality is shown graphically on the right in figure 3.11. Now the function sandwiched in the middle of the inequality is the one we are interested in, and furthermore the inequalities are of the form of the sandwich theorem as in the limit we have $\lim_{x\to 0}(|x|) = \lim_{x\to 0}(-|x|) = 0$*. Hence if take the limit on the inequality we find:*

$$0 = \lim_{x\to 0}(-|x|) \leq \lim_{x\to 0}((x\sin\left(\frac{1}{x}\right)) \leq \lim_{x\to 0}(|x|) = 0.$$

Therefore we surmise that

$$\lim_{x\to 0}((x\sin\left(\frac{1}{x}\right)) = 0$$

as required.

We did a lot of work to find these limits, can we learn anything else from these results? The two limits in the example do look rather similar and we can modify them to show that they tell us about two different limits of the same function.

Let us commence with the result of the second limit

$$\lim_{x\to 0}((x\sin\left(\frac{1}{x}\right)) = 0.$$

If we change the variable used in the limit to $y \equiv \frac{1}{x}$ then the function in the limit becomes $\frac{1}{y}\sin y$, the same function, albeit written in terms of a different variable, as in the first example. We also have to change the limit $x \to 0$ into a limit in terms of y, using $y = \frac{1}{x}$ as x approaches

a very small number so y approaches a very large number, but there is a small concern for as $x \to 0^+$ then $y \to \infty$, while as $x \to 0^-$, $y \to -\infty$. So in translating the limit from one in terms of x we in fact generate two limits in terms of y:

$$\lim_{y\to-\infty}(\frac{\sin y}{y}) = 0 \quad \text{and} \quad \lim_{y\to\infty}(\frac{\sin y}{y}) = 0.$$

Given that these are limits to infinity it is in fact necessary that the limits are one-sided (i.e. from above and from below) so what seemed like a complexity is rather satisfying and we can be pleased with this two-for-one result[5]. We may also use the same change of variables to show that the limit $\lim_{x\to0}(\frac{\sin x}{x}) = 1$ can be rewritten as the two asymptotic limits:

$$\lim_{y\to-\infty}(y\sin\left(\frac{1}{y}\right)) = 1 \quad \text{and} \quad \lim_{y\to\infty}(y\sin\left(\frac{1}{y}\right)) = 1.$$

Furthermore we may even consider changing to imaginary variables if the limit remains well-defined, for example,

$$1 = \lim_{x\to0}(\frac{\sin x}{x}) = \lim_{x\to0}\left(\frac{\frac{1}{2i}(e^{ix} - e^{-ix})}{x}\right) = \lim_{y\to0}\left(\frac{(e^{y} - e^{-y})}{2y}\right) = \lim_{y\to0}(\frac{\sinh y}{y})$$

where we have substituted $y \equiv ix$, in order to find another useful limit.

Exercise 3.2. *Prove that* $\lim_{x\to0}[\frac{\tan x}{x}] = 1$. *Hint: you may assume, without further proof, that* $\lim_{x\to0}[\frac{\sin x}{x}] = 1$ *in your answer.*

Exercise 3.3. *Find the following limit:*

$$\lim_{x\to\frac{\pi}{2}}[\frac{2\cos x}{2x-\pi}].$$

Hint: Try a change of the limit variable.

The result in example 3.13 above may seem rather exciting to us, as previously we argued that $\sin\left(\frac{\pi}{x}\right)$ had no limit as $x \to 0$. Now if we rescale the variable by defining $y = \frac{x}{\pi}$ then as $x \to 0$ we have $y \to 0$ and so we claim that $\lim_{y\to0}\left(\sin\left(\frac{1}{y}\right)\right)$ does not exist, where from the results above we see that just by multiplying the function by y we discover a well-defined limit i.e. $\lim_{y\to0}\left(y\sin\left(\frac{1}{y}\right)\right) = 0$. Let us emphasise that we cannot use the fact that the limit of a product is the product of the limits here, i.e.

$$\lim_{y\to0}\left(y\sin\left(\frac{1}{y}\right)\right) \neq \lim_{y\to0}(y)\lim_{y\to0}(\sin\left(\frac{1}{y}\right))$$

[5]Given the symmetries of the sine function in the numerator and the linear function in the denominator then even if we had only one of these results we could have determined the second by substituting $z = -y$.

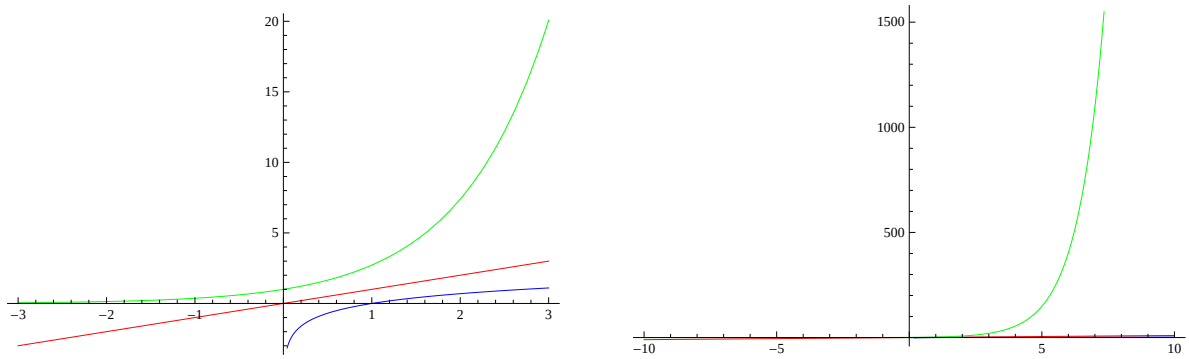


Figure 3.12: The graphs of $f(x) = \ln(x)$ (blue), $g(x) = x$ (red) and $h(x) = e^x$ (green) above, near zero on the left and for a slightly larger domain on the right. Notice that for positive x $f(x) < x < e^x$ and that exponential growth is really very fast indeed!

as the limit $\lim_{y\to 0}(\sin\left(\frac{1}{y}\right))$ is undefined and so the rule for limits of products is not valid here. What has happened is that the linear function y has gone to zero "faster" than $(\sin\left(\frac{1}{y}\right))$ has oscillated about zero as $y \to 0$. While we cannot split the product of limits up in this case it is useful to adopt this kind of thinking and ask yourself how parts of functions grow or shrink as the limit is taken and which types of function dominate. This thinking does not replace a detailed analysis of a function but it can prove a helpful guide. We will now compare some other common types of function in the limit to get a sense of which dominate in specified limits.

Logarithmic vs. Polynomial vs. Exponential Functions: Which grows faster?

The answer to the question in the title is given away by these terms use in modern language where it is common to speak of exponential growth (or decay) to refer to something which grows incredibly fast (or shrinks very quickly). You are also probably familiar with logarithmic graph paper which is used to plot functions whose outputs vary very quickly - the logarithm is applied to slow down the variation of the function. So as $\ln(x) = \ln(\ln(e^x))$ we may guess that $ln(x)$ grows more slowly as x increases than x (an example of a polynomial function), and in turn as $x = \ln(e^x)$ that polynomial functions grow more slowly than exponential functions as x grows. In each case it is common to speak of logarithmic growth, polynomial growth and exponential growth. Of course the linguistic meaning of the terms is confirmed by plotting the graphs of $f(x) = \ln(x)$, $g(x) = x$ and $h(x) = e^x$. The result shown in figure 3.12 is what we should expect: $f(x)$ is the mirror image of $h(x)$ (where it is defined) in $g(x)$ as $h(x)$ is the inverse function of $f(x)$.

Let us now investigate how these types of functions behave in the limit by way of the following examples.

Example 3.14. *Show that*

$$\lim_{x\to 0}\left[\frac{\ln(1+x)}{x}\right]=1.$$

Let us reflect and notice that we are comparing the speed with which $\ln(1+x)$ *approaches zero - this is a shifted version of the logarithm function sketched in figure 3.12. Hence as* $x\to 0$, $\ln(1+x)\to\ln(1)=0$, *and in all the limit in question is approaching the dreaded* $0/0$ *and we cannot obviously split the function as a product. Instead let us try making a substitution to change the variable, we will substitute* $1+x=e^y$ *so that the logarithm is simplified, we end up with*

$$\lim_{y\to 0}\left[\frac{y}{e^y-1}\right]=\lim_{y\to 0}\left[\frac{y}{e^{\frac{y}{2}}(e^{\frac{y}{2}}-e^{-\frac{y}{2}})}\right]=\lim_{y\to 0}\left[\frac{ye^{-\frac{y}{2}}}{2\sinh\left(\frac{y}{2}\right)}\right]=\lim_{y\to 0}\left[\frac{y/2}{\sinh\left(\frac{y}{2}\right)}\right]\lim_{y\to 0}\left[e^{-\frac{y}{2}}\right]=1$$

where we have used the fact that $\lim_{x\to 0}[\frac{\sinh x}{x}]=1$ *which was shown earlier.*

If you did not find this proof instinctive, do not worry: there are often many ways to analyse a limit. For example we could have used the power series for e^y *above to simplify the limit in a different way:*

$$\lim_{y\to 0}\left[\frac{y}{e^y-1}\right]=\lim_{y\to 0}\left[\frac{y}{y+\frac{y^2}{2!}+\mathcal{O}(y^3)}\right]=\lim_{y\to 0}\left[\frac{1}{1+\frac{y}{2!}+\mathcal{O}(y^2)}\right]=\frac{\lim_{y\to 0}\left[1\right]}{\lim_{y\to 0}\left[1+\frac{y}{2!}+\mathcal{O}(y^2)\right]}=1.$$

There is an artistry to analysis and which route you prefer to take will depend upon your tastes.

Exercise 3.4. *Prove that*

$$\lim_{x\to\infty}\left[x\ln(1+\frac{1}{x})\right]=1.$$

Hint 1: You may assume all the limits proven in the lecture notes this far and try a substitution.
Hint 2: This should be a very quick exercise!

Example 3.15. *Show that*

$$\lim_{x\to 0^+}\left[x\ln(x)\right]=0.$$

This limit is telling us that the polynomial function x *which goes to zero dominates the logarithmic function* $\ln(x)$ *which approaches* $-\infty$ *as* $x\to 0$. *We give the sketch of the graph near zero in figure 3.13. Notice that the limit is taken from above as* $ln(x)$ *is undefined for* $x\leq 0$.

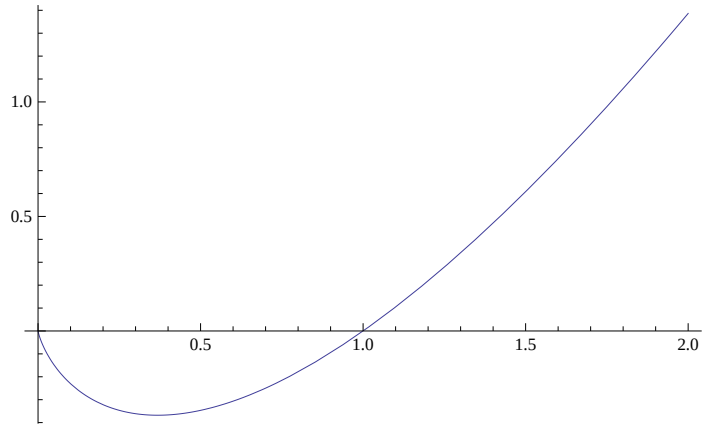


Figure 3.13: The graph of $f(x) = x \ln x$ in the vicinity of $x = 0$.

This will be our most intricate proof of a limit and we will set it out in three steps below. However our first instinct might be to change variables to remove the logarithm, so let us substitute $x = e^{-y}$ *so that the limit becomes*[6]

$$\lim_{y\to\infty} \left[-y\, e^{-y} \right] = 0.$$

Again we note if it is true then we learn that the behaviour of the exponential near $-\infty$ *suppresses the growth of* y*, as expected. Our aim is to use the sandwich theorem and so we must identify some useful inequalities. At its root we are trying to compare* e^y *with* y *which will motivate the first of our three steps:*

(i) We claim that $e^w \geq w$ *for all* $w \geq 0$*. Let's prove this algebraically. To do so we will consider the difference function* $f(w) = e^w - w$*, and we will have proved our claim if we can show that* $f(w) \geq 0$ *for all* $w \geq 0$*. First we note that* $f(0) = 1$*, so the claim is true at* $w = 0$*. Now using the definition of the exponential function (that* $\frac{d}{dx}(e^x) = e^x$*) we can show that*

$$\frac{df}{dw} = e^w - 1 \geq 0 \qquad \text{for all } w \geq 0.$$

Hence the difference function is a function $f(w)$ *with an ever increasing gradient as* w *increases from zero towards* ∞*. As it is positive at* $w = 0$ *it must (due to its positive gradient) therefore be positive for all* $y > 0$ *too. Hence we have argued that* $f(w) = e^w - w \geq 0$ *for all* $w \geq 0$ *as claimed.*

[6]There is no reason at this stage to substitute $x = e^{-y}$ rather than $x = e^y$ but it will prove very useful later to make this choice, the other choice will not work with the following method. Why not try it, after you have worked through the three steps and find out where the substitution $x = e^y$ runs into difficulty with this method?

(ii) *Now let* $w \equiv \frac{z}{2}$ *in the above inequality. It is perhaps surprising that such a substitution will prove valuable but it is only the first prong of this two-pronged manoeuvre. From step (i) we have*

$$e^{\frac{z}{2}} \geq \frac{z}{2} \qquad \textit{for all } z \geq 0.$$

The second part here is to square the above inequality (as both sides are positive this does not present any ambiguity) and we find

$$e^{z} \geq \frac{z^2}{4} \qquad \textit{for all } z \geq 0$$

hence after rearranging (and taking care not to divide by zero implicitly) we have

$$ze^{-z} \leq \frac{4}{z} \qquad \textit{for all } z > 0.$$

Note that $\lim_{z\to\infty}[\frac{4}{z}] = 0$.

(iii) *We are interested in the limit* $\lim_{y\to\infty}[-ye^{y}]$. *From step (ii) we have found one useful inequality whose bounding function tends to* 0 *in the limit, if we can identify another function* $g(y)$ *such that* $g(y) \leq ye^{-y}$ *and satisfying* $\lim_{y\to 0}[g(y)] = 0$ *for large* y *we will be in a position to use the sandwich theorem. The simplest function to try is the constant function* $g(y) = 0$, *now for* $y > 0$ *then* $ye^{-y} > 0$ *hence we may use the constant function and we have the inequalities:*

$$0 \leq ye^{-y} \leq \frac{4}{y} \qquad \forall\, y > 0.$$

You might be concerned about the sign of the function in the above compared to our function, in truth if the sandwiching is correct it does not matter, but for the fastidious we may change the sign and exchange the inequality signs:

$$0 \geq -ye^{-y} \geq -\frac{4}{y} \qquad \forall\, y > 0.$$

Now taking the limit we have

$$0 = \lim_{y\to\infty}[0] \geq \lim_{y\to\infty}[-ye^{-y}] \geq \lim_{y\to\infty}[-\frac{4}{y}] = 0$$

hence by the sandwiching theorem

$$\lim_{y\to\infty}[-ye^{-y}] = 0$$

as required. This was, perhaps, a tedious proof, but its method was interesting and you can be assured that there are simpler methods for finding this limit which we will meet

in this course. For example if we had been willing to make use of the power series for $e^y = \sum_{n=0}^{\infty} \frac{y^n}{n!}$ *then we have the straightforward inequality that* $e^y \geq \frac{y^m}{m!}$ *for any* $y > 1$ *and* $m \in \mathbb{Z}^+$*, hence rearranging gives* $\frac{m!}{y^m} \geq e^{-y}$ *and from which we obtain* $ye^{-y} \leq \frac{m!}{y^{m-1}}$ *which gives an upper bound to use in the sandwich theorem.*

From this last example we also readily obtain another useful inequality by substitution:

$$0 = \lim_{x \to 0^+} \big[x \ln(x)\big] \overset{(x \equiv \frac{1}{y})}{=} \lim_{y \to \infty} \big[\frac{\ln(y^{-1})}{y}\big] = \lim_{y \to \infty} \big[\frac{-\ln(y)}{y}\big].$$

Finding intelligent ways to discover limits of functions is an art-form and requires practise: do try out all the limit questions you can find. We now turn our attention to the a fundamental object in calculus: the derivative and the limit will play a central role.

4. The Derivative

In which we will meet the derivative function, the function which gives the slope of the tangent to a point on function at any point. We will see that the limit will play a central role in the definition of the derivative and we will derive the derivatives of many of the most common functions from first principles.

The material in this chapter will be covered in weeks 7 and 8 of the course.

Limits are clearly very interesting in the detailed study of a function, but are they of practical importance? Is it ever important to know the value a function is approaching rather than the actual value of the function? One of the motivating factors in developing the calculus was the desire to understand physical properties in the natural world. It is intrinsically interesting to enquire what is the speed of an object, but it can be surprisingly difficult to define. Let us explain why such a simple concept proved challenging to pin down until the calculus was developed.

The average speed of an object is defined as

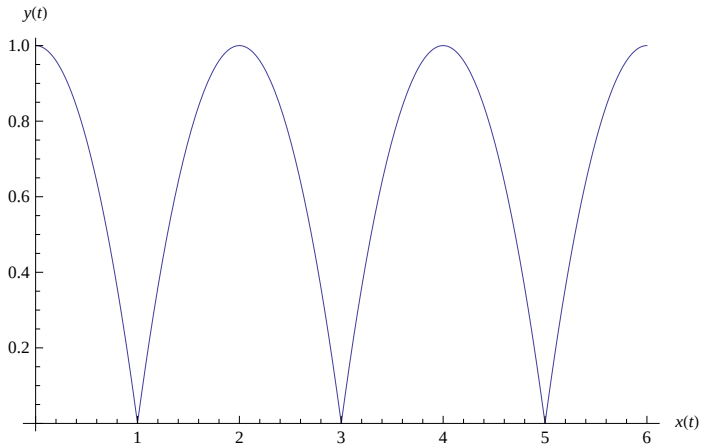
$$\text{Average speed} \equiv \frac{\text{Change in the object's position.}}{\text{Change in time.}}$$

If an object moves from $\vec{r}(t_1)$ to $\vec{r}(t_2)$ over the time interval $[t_1, t_2]$ then we can use this definition to find that its average speed is

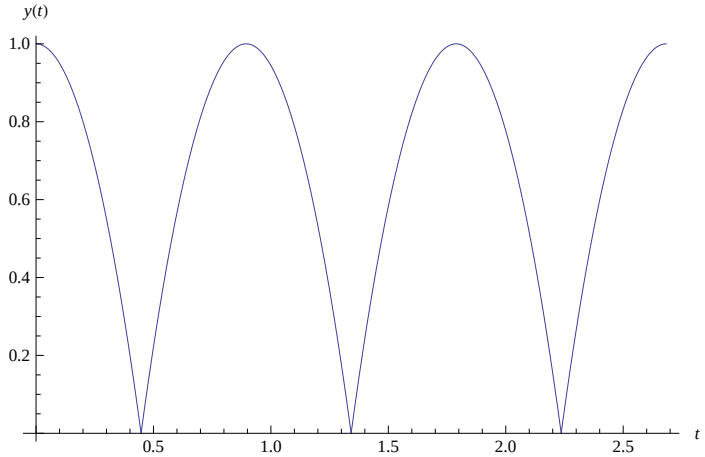
$$\frac{|\vec{r}(t_2) - \vec{r}(t_1)|}{t_2 - t_1}.$$

Obviously this presents some problems for objects whose speed changes rapidly, the average speed loses a lot of information. For example consider an idealised bouncing ball moving in the x-direction with constant speed [1]:

[1] In case you were wondering, for a simple illustration we have neglected resistance in the ball's motion, and presumed gravity is $10ms^{-2}$ acting vertically downwards, started the ball at a height of $1m$ with zero initial



The horizontal speed is constant by design but the vertical speed is varying from zero at the top of the bounce and its maximum speed when it hits the floor. Let's look at the plot of its vertical position y against time t for its first bounces.



As $x(t) = \sqrt{5}t$ this graph looks like a squashed version of the last sketch.

Let us make a computation of the ball's average speed for the time interval $[0, \frac{1}{\sqrt{5}}]$ which from the graph (or better from the equation $y(t) = -5t^2 + 1$) we can read off $y(0) = 1$ and $y(\frac{1}{\sqrt{5}}) = 0$. In figure 4.1 we have annotated the graph of the first bounce (in a plot of $y(t)$ against t), to help compute the average speed. Reading off from the graph for $t_1 = 0$ and $t_2 = 1$ (so that $t_2 - t_1 = 1$), we have $y(t_2) - y(t_1) = -1$. The average vertical speed (in ms^{-1}) is

$$\frac{|y(t_2) - y(t_1)|}{t_2 - t_1} = \frac{|-1|}{\frac{1}{\sqrt{5}}} = \sqrt{5}.$$

vertical speed and a horizontal speed of $\sqrt{5}ms^{-1}$. The equations of motion give us $y = -x^2 + 1$, before any bounce occurs. In addition we have assumed a perfect bounce takes place, i.e. an instantaneous reflection of the speed of the ball in the floor-line at the moment of impact. The situation is not very realistic but is good enough for our purpose.

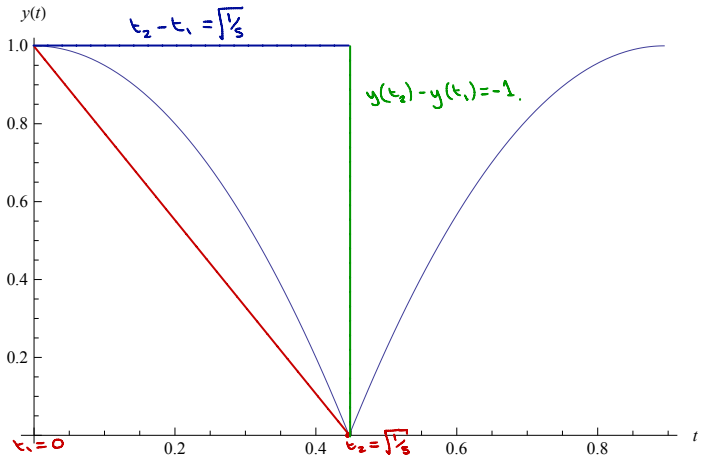


Figure 4.1: The computation of the average speed from $t = 0$ to the first bounce at $t = \sqrt{\frac{1}{5}}$.

From the graph we can see this is the speed a ball which moved with constant speed from $y = -1$ to $y = 0$ would move at. Our ball actually has this speed for only one moment during its first descent: we can surmise this only because we are actually experts in this kind of motion and we know that the vertical speed starts at $0ms^{-1}$ and continually increases due to the gravitational acceleration to its maximum speed. In other words, the straight line (in red on figure 4.1) is tangent to the path $y(t)$ at just one point.

Of course if we were to compute the average speed for smaller and smaller time intervals we would be able to find an average velocity for each time interval, i.e. we could build up a set of average speeds $\hat{v}_1, \hat{v}_2, \hat{v}_3, \ldots \hat{v}_n$ for the time intervals $[t_1, t_2], [t_2, t_3], [t_3, t_4], \ldots [t_n, t_{n+1}]$ respectively. We attempt to illustrate the improved accuracy in finding in figure 4.2, where one can see that for sufficiently short time intervals the sequence of straight lines whose slopes are the average speed for the interval approaches and becomes almost indistinguishable from the curve $y(t)$. If we consider the limit in which a time interval $[t_n, t_{n+1}]$ becomes vanishingly

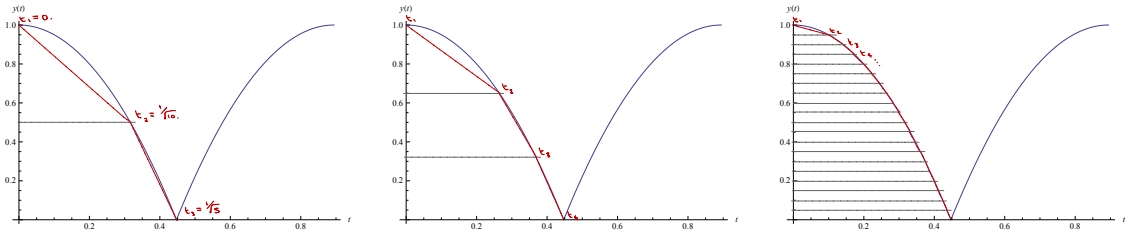


Figure 4.2: Approaching a speed function by finding the average speed for smaller and smaller time intervals.

small (i.e. $t_{n+1} \to t_n$) then the average speed for that interval approached the instantaneous speed: it is the slope of the tangent to the curve as $t_{n+1} \to t_n$. This gives us a definition of

the instantaneous speed at time t_n, as the slope of the tangent to the curve $y(t)$ at time $t = t_n$ plotted against t. Since this curve is relatively well-behaved (apart from at the bounces) this gives a good practical definition for constructing the speed function, namely the speed in the y-direction at time $t = t_n$ is given by evaluating

$$\frac{|y(t_{n+1}) - y(t_n)|}{t_{n+1} - t_n}.$$

when $t_{n+1} = t_n$. But of course when $t_2 = t_1$ the denominator becomes zero and the function is not defined. Earlier we saw the example of the function $\frac{x^2}{x}$ which was not defined at $x = 0$ and the limit was introduced as a method to make sense of the algebraic simplification of this function as $\lim_{x\to 0}[\frac{x^2}{x}] = \lim_{x\to 0}[x] = 0$ which is valid arbitrarily close to $x = 0$. In exactly the same manner we may avoid contemplating division by zero in the speed function for $y(t)$ by extending the definition to include a limit as:

$$\dot{y}(t_n) \equiv \lim_{t_{n+1}\to t_n} \left[\frac{y(t_{n+1}) - y(t_n)}{t_{n+1} - t_n}\right].$$

The dot on the y is Newton's notation for a time-derivative, hence $\dot{y}$ is the standard notation for the velocity in the y-direction[2].

4.1 Differentiation from first principles

Historically, Newton's early ideas in calculus came directly from 'Fermat's[3] way of drawing tangents'. Fermat had proposed to find the instantaneous slope by simply evaluating the change in a function $f(x)$ as x is varied and divide it by the corresponding change in x and then to let this change be put equal to zero. Precisely, Fermat proposed to find the slope of a function for the interval $[x, x + h]$ by evaluating the fraction

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

and then setting $h = 0$ in the result. While this is evidently undefined as a function, the development of Newton to consider not the slope function but the limit of the slope function gives the derivative.

Definition 4.1.1. *The derivative of a function $f(x)$ with respect to x is*

$$\frac{df}{dx} \equiv \lim_{h\to 0} \left[\frac{f(x + h) - f(x)}{h}\right]$$

[2]Leibniz used the dominant notation for derivatives featuring the letter d to indicate an infinitesimal change in a quantity and would have written $\dot{y} = \frac{dy}{dt}$. We will introduce this notation soon.

[3]You can discover truly marvellous facts about Fermat elsewhere, which this margin is too narrow to contain.

where the limit exists.

Comment(s). *(On the derivative...)*

1. *There are different notations in common use for the derivative of a function $f(x)$, the most common notation is due to Leibniz and is $\frac{df}{dx}$, but one may also commonly write $f'(x)$ for the derivative of $f(x)$ with respect to x. This second 'primed' notation is due to Lagrange and means that the derivative is taken with respect to the argument of the function, hence $f'(y) = \frac{df}{dy}$, when the argument is a complicated function this notation is very useful, e.g. the derivative of $f(x - ct)$, a travelling wave function, with respect to $(x - ct)$ is written $f'(x - ct)$ and frequently the argument is dropped in the notation so one may simply write f' for this same derivative - use of this notation obviously requires the function and its argument to be clearly defined. As mentioned earlier time derivatives have a special "dotted" notation reserved for them, which was first used by Newton who was studying dynamics, so that $\dot{x} = \frac{dx}{dt}$.*

2. *The notation for derivatives of derivatives is written*

$$\frac{d^n f}{dx^n}$$

where $n \in \mathbb{Z}^+$. When $n = 1$ we do not write the number 1, this denotes the first derivative of a function. When $n \geq 2$ we write the value of n explicitly, hence the second derivative of $f(x)$ with respect to x is written

$$\frac{d^2 f}{dx^2}.$$

Other notations for derivatives employ repeated use of the prime or the dot to indicate further derivatives. For example one might write

$$f''(x) \equiv \frac{d^2 f}{dx^2}$$

or

$$\ddot{x} \equiv \frac{d^2 x}{dt^2}.$$

3. *Geometrically the definition of the derivative is delightful and instructive. Suppose you wish to evaluate the derivative of a function $f(x)$ at $x = x_0$, the definition above is equivalent to drawing the straight line that connects $f(x_0)$ to any other point on the curve $f(x_0 + h)$. This straight line (so long as it exists) is called a secant and has a well-defined slope, which we may compute. Now in taking the limit one allows the two separate points to become close, so that $f(x_0 + h)$ approaches $f(x_0)$. If the slope is tending to a well-defined limit then the derivative exists: the secant becomes a tangent. As $h \to 0$ we are*

able to construct the straight line through two points arbitrarily close to each other (but not identical) in this way we may have a well-defined definition of the tangent line to a function: a line apparently given meaning by only a single point on $f(x)$.

4. *In this definition we have defined a derivative function - the result is a function of x and in doing so we have been efficient but have side-stepped an important notion. The derivative at a point $x = x_0$ is given by*

$$\left[\frac{df}{dx}\right]_{x=x_0} \equiv \lim_{h\to 0}\left[\frac{f(x_0+h)-f(x_0)}{h}\right].$$

This gives a value for the slope of the tangent to $f(x)$ at $x = x_0$, it is not a function of x. To construct the derivative function, which we have defined above we should evaluate the derivative at all points x_0 in the domain of the function, the resulting set of values can be used to then define the derivative as given above. One understands why we have taken the short cut we have in our definition, but we must wonder what happens if the derivative at a point $x = x_0$ does not exist. Such a function is said to not be differentiable at x_0 and we will return to this idea in the following section.

5. *Note that the limit in the derivative will only exist on an open interval $(a, b) \in \mathbb{R}$ when the limit from above is equal to the limit from below.*

Finding the derivative by evaluating the limiting value of the slope is often referred to as 'differentiation from first principles' as it does not rely upon knowing the derivative of any other function. We will now find from first principles some standard derivatives.

Example 4.1. *Find the derivative of the linear function $f(x) = ax + b$ with respect to x where $a, b \in \mathbb{R}$ at the point $x = x_0$.*

Commencing with the definition of the derivative, we compute:

$$\frac{df}{dx} = \lim_{h\to 0}\left[\frac{f(x_0+h)-f(x_0)}{h}\right] = \lim_{h\to 0}\left[\frac{a(x_0+h)+b-a(x_0)-b}{h}\right] = \lim_{h\to 0}\left[\frac{ah}{h}\right] = a.$$

As expected the derivative of the straight line is constant.

Example 4.2. *Find the derivative of the quadratic function*

$$f(x) = ax^2 + bx + c$$

with respect to x where $a, b, c \in \mathbb{R}$ as a function of x.

Commencing with the definition of the derivative, we compute:

$$\begin{aligned}
\frac{df}{dx} &= \lim_{h\to 0}\left[\frac{f(x+h)-f(x)}{h}\right] \\
&= \lim_{h\to 0}\left[\frac{a(x+h)^2+b(x+h)+c-a(x)^2-bx-c}{h}\right] \\
&= \lim_{h\to 0}\left[\frac{ax^2+2axh+ah^2+bx+bh+c-ax^2-bx-c}{h}\right] \\
&= \lim_{h\to 0}\left[\frac{2axh+ah^2+bh}{h}\right] \\
&= \lim_{h\to 0}\left[2ax+ah+b\right] \\
&= 2ax+b
\end{aligned}$$

Example 4.3. *Find the derivative of the function*

$$f(x) = ax^n$$

with respect to x where $a, \in \mathbb{R}$ and $n \in \mathbb{Z}^+$ as a function of x.

Note that this derivative was assumed in earlier chapters when defining the exponential function, the trigonometric functions and the hyperbolic functions as infinite series. This was one key pillar that we built much of the course so far upon, the other was that the infinite sum for e^x converges. From first principles we have:

$$\begin{aligned}
\frac{df}{dx} &= \lim_{h\to 0}\left[\frac{f(x+h)-f(x)}{h}\right] \\
&= \lim_{h\to 0}\left[\frac{a(x+h)^n-ax^n}{h}\right] \\
&= \lim_{h\to 0}\left[\frac{a}{h}\left(\sum_{k=0}^{n}\left(\binom{n}{k}x^{n-k}h^k\right)-x^n\right)\right] \\
&= \lim_{h\to 0}\left[\frac{a}{h}\sum_{k=1}^{n}\binom{n}{k}x^{n-k}h^k\right] \\
&= \lim_{h\to 0}\left[a\sum_{k=1}^{n}\binom{n}{k}x^{n-k}h^{k-1}\right] \\
&= \lim_{h\to 0}\left[a\binom{n}{1}x^{n-1}+a\binom{n}{2}x^{n-2}h+\mathcal{O}(h^2)\right] \\
&= anx^{n-1}
\end{aligned}$$

Or, alternatively, without using the binomial expansion (so that the following becomes a proof for all $n \in \mathbb{R}$):

$$\begin{aligned}
\frac{df}{dx} &= \lim_{h\to 0}\left[\frac{f(x+h)-f(x)}{h}\right] \\
&= \lim_{h\to 0}\left[\frac{a(x+h)^n - ax^n}{h}\right] \\
&= ax^n \lim_{h\to 0}\left[\frac{(1+\frac{h}{x})^n - 1}{h}\right] \\
&= ax^n \lim_{h\to 0}\left[\frac{e^{\ln[(1+\frac{h}{x})^n]} - 1}{h}\right] \\
&= ax^n \lim_{h\to 0}\left[\frac{e^{n\ln(1+\frac{h}{x})} - 1}{h}\right].
\end{aligned}$$

Let us now change the limit variable using $h \equiv xg$ so we have:

$$\begin{aligned}
\frac{df}{dx} &= ax^n \lim_{g\to 0}\left[\frac{e^{n\ln(1+g)} - 1}{xg}\right] \\
&= ax^{n-1} \lim_{g\to 0}\left[\frac{e^{n\ln(1+g)} - 1}{g}\right].
\end{aligned}$$

Here we now choose to insert $\ln(1+g)/\ln(1+g) = 1$, so that we may make use of the standard limits (evaluated in the previous chapter): $\lim_{x\to 0}[\frac{\ln(1+x)}{x}] = 1$ and $\lim_{x\to 0}[\frac{e^x-1}{x}] = 1$ - see example 3.14 for the proof of both limits. So, upon inserting $\ln(1+g)/\ln(1+g)$ we have

$$\begin{aligned}
\frac{df}{dx} &= ax^{n-1} \lim_{g\to 0}\left[\left(\frac{e^{n\ln(1+g)} - 1}{\ln(1+g)}\right)\left(\frac{\ln(1+g)}{g}\right)\right] \\
&= ax^{n-1} \lim_{g\to 0}\left[\frac{e^{n\ln(1+g)} - 1}{\ln(1+g)}\right] \lim_{g\to 0}\left[\frac{\ln(1+g)}{g}\right] \\
&= ax^{n-1} \lim_{k\to 0}\left[\frac{e^k - 1}{k/n}\right] \\
&= anx^{n-1}
\end{aligned}$$

where we substituted $k \equiv n\ln(1+g)$. This second method is much longer than using the binomial expansion, but it used two nice and common "tricks" to manipulate the limit into the form of some standard limits.

Example 4.4. *Find the derivative of the function*

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

with respect to x where $a, \in \mathbb{R}$ as a function of x.

$$\begin{aligned}
\frac{df}{dx} &= \lim_{h\to 0}\left[\frac{f(x+h)-f(x)}{h}\right] \\
&= \lim_{h\to 0}\left[\frac{a^{x+h}-a^x}{h}\right] \\
&= \lim_{h\to 0}\left[a^x\left(\frac{a^h-1}{h}\right)\right] \\
&= \lim_{h\to 0}[a^x]\lim_{h\to 0}\left[\frac{a^h-1}{h}\right] \\
&= a^x\lim_{h\to 0}\left[\frac{e^{\ln(a^h)}-1}{h}\right] \\
&= a^x\lim_{k\to 0}\left[\frac{e^k-1}{k/\ln(a)}\right] \\
&= a^x\ln(a)\lim_{k\to 0}\left[\frac{(1+k+\frac{k^2}{2!}+\mathcal{O}(k^3))-1}{k}\right] \\
&= a^x\ln(a)\lim_{k\to 0}\left[1+\mathcal{O}(k)\right] \\
&= a^x\ln(a)
\end{aligned}$$

Where we changed the limit variable using $k = \ln a^h = h\ln a$. Note that had we chosen $a = e$, Euler's number, then we would have found $\frac{df}{dx} = e^x\ln(e) = e^x$ which was part of our defining relation for the exponential in earlier chapters.

Example 4.5. *Find the derivative of the function*

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

with respect to x as a function of x.

$$\begin{aligned}
\frac{df}{dx} &= \lim_{h\to 0}\left[\frac{f(x+h)-f(x)}{h}\right] \\
&= \lim_{h\to 0}\left[\frac{\ln(x+h)-\ln(x)}{h}\right] \\
&= \lim_{h\to 0}\left[\frac{\ln(x(1+\frac{h}{x}))-\ln(x)}{h}\right] \\
&= \lim_{h\to 0}\left[\frac{\ln(1+\frac{h}{x})+\ln(x)-\ln(x)}{h}\right] \\
&= \lim_{h\to 0}\left[\frac{\ln(1+\frac{h}{x})}{h}\right] \\
&= \lim_{g\to 0}\left[\frac{\ln(1+g)}{gx}\right] \\
&= \frac{1}{x}\lim_{g\to 0}\left[\frac{\ln(1+g)}{g}\right] \\
&= \frac{1}{x}
\end{aligned}$$

where we have used $h \equiv gx$ *and the standard limit* $\lim_{x\to 0}[\frac{\ln(1+x)}{x}] = 1$ *(see example 3.14).*

Example 4.6. *Find the derivative of the function*

$$f(x) = \sin(x)$$

with respect to x *as a function of* x*.*

$$\begin{aligned}
\frac{df}{dx} &= \lim_{h\to 0}\left[\frac{f(x+h)-f(x)}{h}\right] \\
&= \lim_{h\to 0}\left[\frac{\sin(x+h)-\sin(x)}{h}\right] \\
&= \lim_{h\to 0}\left[\frac{\sin(x)\cos(h)+\cos(x)\sin(h)-\sin(x)}{h}\right] \\
&= \lim_{h\to 0}\left[\frac{\sin(x)(\cos(h)-1)+\cos(x)\sin(h)}{h}\right] \\
&= \sin(x)\lim_{h\to 0}\left[\frac{\cos(h)-1}{h}\right] + \cos(x)\lim_{h\to 0}\left[\frac{\sin(h)}{h}\right]
\end{aligned}$$

Let us comment at this stage that we do have recourse to using the series expansion for $\cos(h) = 1 - \mathcal{O}(h^2)$ *to rapidly show that*

$$\lim_{h\to 0}\left[\frac{\cos(h)-1}{h}\right] = \lim_{h\to 0}\left[\frac{\mathcal{O}(h^2)}{h}\right] = \lim_{h\to 0}\left[\mathcal{O}(h)\right] = 0$$

and thence after substituting the limit $\lim_{h\to 0}\left[\frac{\sin(h)}{h}\right] = 1$ *we would find* $\frac{df}{dx} = \cos(x)$*. However since this is an analysis course we are going to think of another proof just to practise using the sandwich theorem again, and of course it will take slightly longer to get to the same answer but will be interesting. As* $-1 \leq \cos(h) \leq 1$ *then we may deduce that* $(\cos(h) - 1) \leq 0$*, this gives the upper bound for use in the sandwich theorem. For the lower bound we turn to geometry and the sector shown in figure 2.5. The base of the large right-angled-triangle has length* 1*, and we have embedded a similar triangle within it whose base is* $\cos(\theta)$*. Hence the remaining part of the base line (whose length isn't indicated on the diagram) has length* $1 - \cos(\theta)$*. Now consider the right-angled-triangle with base* $1 - \cos(\theta)$ *and height* $\sin\theta$*, its hypotenuse has length-squared:*

$$(1 - \cos\theta)^2 + \sin^2\theta = 2 - 2\cos\theta.$$

As this triangle would be embedded within the sector of the unit circle shown in figure 2.5, its hypotenuse must have length less than or equal to the arc length shown, which is θ*, i.e. we have*

$$2(1 - \cos\theta) \leq \theta^2$$

or

$$(\cos\theta - 1) \geq -\frac{\theta^2}{2}$$

which gives our lower bound. Returning to our limit, we are now able to sandwich the limit as follows

$$0 = \lim_{h\to 0}\left[-\frac{h}{2}\right] = \lim_{h\to 0}\left[-\frac{h^2}{2h}\right] \leq \lim_{h\to 0}\left[\frac{\cos(h)-1}{h}\right] \leq 0.$$

Hence by the sandwich theorem we have

$$\lim_{h\to 0}\left[\frac{\cos(h)-1}{h}\right] = 0$$

which we may use to show[4] *that* $\frac{d}{dx}(\sin(x)) = \cos(x)$*.*

Exercise 4.1. *Find the derivative of* $f(x) \equiv \cos(x)$ *with respect to* x*.*

[4] Once again!

4.2 Differentiable Functions

But do all functions have a derivative at all points? The answer is emphatically no. There were some caveats involved in defining the derivative and we did not draw attention to them in the previous section, so we will look again now at the definition of the derivative in definition 4.1.1. It is a good idea now to turn back and re-read the definition. Where does the definition of the derivative encounter problems? There are two situations which spring to mind:

(i) if the function $f(x)$ is not continuous at $x = x_0$ and

(ii) if the limit

$$\lim_{h\to 0}\left[\frac{f(x+h)-f(x)}{h}\right]$$

does not exist for some particular value of $x = x_0$.

If either of these cases occur, the derivative at $x = x_0$ does not exist and the function is said to be 'not differentiable' at $x = x_0$.

Definition 4.2.1. *A function $f(x)$ is differentiable at the point $x = x_0$ if the derivative $\frac{df}{dx}$ exists at $x = x_0$. A function for which the derivative exists for all points in its domain is called a differentiable function.*

Theorem 4.1. *All differentiable functions are continuous functions.*

Proof: Suppose that $f(x)$ is a differentiable function. Now recall the definition for $f(x)$ to be continuous at the point $x = x_0$, namely that

$$\lim_{x\to x_0}(f(x)) = f(x_0).$$

This is what we are aiming to show, we will nevertheless start with this statement and see how it relates to the limit which defines the derivative of $f(x)$. We begin by rearranging the statement of the continuity of $f(x)$ at $x = x_0$:

$$\begin{aligned}
\lim_{x\to x_0}(f(x)) - f(x_0) &= \lim_{x\to x_0}[f(x) - f(x_0)] \\
&= \lim_{x\to x_0}\left[\frac{f(x)-f(x_0)}{(x-x_0)}(x-x_0)\right] \\
&= \lim_{h\to 0}\left[\frac{f(x_0+h)-f(x_0)}{h}(h)\right] \\
&= \left[\frac{df}{dx}\right]_{x=x_0} \times \lim_{h\to 0}[h] \\
&= 0
\end{aligned}$$

where we redefined the limit variable $h \equiv x - x_0$ and in the penultimate line we were able to split the limit of the product into the product of the limits as we know both limits exist. We only know that both limits exist because we assumed that the function is differentiable, and hence it is differentiable at the point $x = x_0$. Hence by assuming that $f(x)$ is differentiable we have shown that it is also continuous.

At this stage you might be wondering if there are any continuous functions which are not differentiable. Indeed there are, and this point tackles our second problem with the derivative: when the limit of the slope does not exist. The most famous example used to show that a continuous function is not necessarily differentiable is the modulus function (which we defined in definition 2.5.3) and is sketched in figure 4.3.

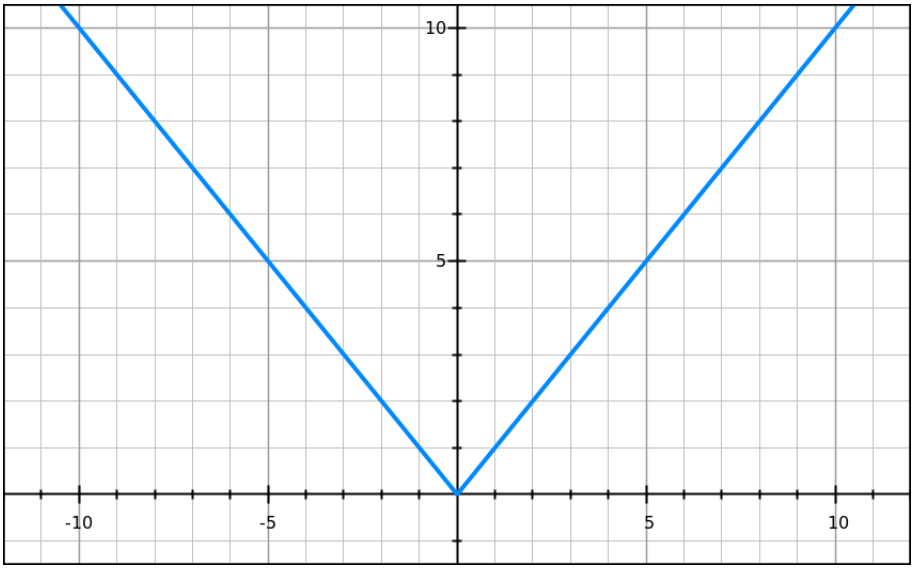


Figure 4.3: The modulus function is a continuous function, but is not a differentiable function.

One can immediately guess where the problem point will be from the graph: the graph of the function although continuous has a right-angle in it at $x = 0$. Our instincts will prove to be correct but let us check it carefully. First we ought to confirm that the modulus function is continuous, namely we check that

$$\lim_{x \to x_0} [|x|] = |x_0|$$

this function is simple enough that we may be confident enough to assert the above from the sketch of the graph: it is a continuous function. But is it differentiable? The derivative can be readily checked to give (for $f(x) = |x|$)

$$\left[\frac{df}{dx}\right]_{x=x_0} = \begin{cases} +1 & x_0 > 0 \\ -1 & x_0 < 0. \end{cases}$$

Hence now we see that at $x_0 = 0$ the limit from above in the definition of the derivative does not equal the limit from below. Above zero the limit in the derivative is $+1$, while from below it is -1, hence

$$\lim_{h\to 0}\left[\frac{f(0+h)-f(0)}{h}\right]$$

does not exist and so the modulus function is an example of a continuous function which is not differentiable at $x = 0$, and hence is not a differentiable function.

These thoughts above about the conditions necessary for a function to have a well-defined derivative, highlight a very simple idea behind the derivative, namely that if one zooms in sufficiently close to the graph of a differentiable function, its graph approaches a straight line. The straight line it approaches is part of the tangent to the function at a point, and the slope of the tangent is the derivative of the function at the point. No matter how much one zoomed into the microscopic detail of the modulus function at $x = 0$, the V shape of the curve would not smooth out into a straight line - so it is not differentiable. Of course when we face questions of differentiability we will rarely be in the position of being able to quickly sketch the function and understand what happens to the curve as we zoom in on a point.

4.3 Properties of the derivative

We have thought about and computed derivatives for many common functions. As we have seen it is possible to create new functions from a basic set by adding the functions, multiplying the function, dividing the functions or composing the functions (functions of functions). Do we have to compute the derivative from first principles for all functions formed in this way? No, fortunately we can rely on properties of the derivative to determine a set of rules which can be used to break down the derivatives of complicated functions into derivatives of simpler functions which we can quickly compute.

4.3.1 The Chain Rule

Given a composite function $f(g(x))$ where both f and g are differentiable functions we can find the derivative $\frac{d}{dx}(f(g(x))$ using our knowledge of the derivatives of f and g. Let us elaborate a little on this idea using some dynamical quantities. Suppose we were interested in computing the acceleration function from the velocity function $v(x)$ written as a function of the displacement functions $x(t)$, by taking its time derivative. In other words we wish to compute:

$$\frac{d}{dt}\Big(v(x(t))\Big).$$

At this point it is useful to emphasise the difference between the function $v(x(t))$ which is the velocity written as an explicit function of $x(t)$ (i.e. when one writes $v(x(t))$ it is written as a function of x, it is the same as writing $v(x)$ where the notation indicating that x is a function of t has been suppressed) and the function $v(t)$ which is the velocity function written as an explicit function of t. We can obtain $v(t)$ from $v(x(t))$ by substituting the expression for $x(t)$ into $v(x(t))$ so we obtain a function of t rather than x. No matter which variable we use to express v the velocity is still the same for any particular value of $t = t_0$ or $x = x_0 = x(t_0)$. Hence while it is a simple matter to compute from first principles the derivative $\frac{d(v(t))}{dt}$ we know that it must also be possible to compute $\frac{d(v(x(t)))}{dt}$ - the only change is that we use the variable x to express v before taking the derivative. This is the aim of the chain rule: to give a method to compute $\frac{dv}{dt}$ starting from $v(x)$, without having to substitute $x(t)$ immediately. It is worth commenting that if we begin with $v(x)$ and $x(t)$ and we wish to compute derivatives there are two simple derivatives we can compute without much thought, namely $\frac{dv}{dx}$ and $\frac{dx}{dt}$. The Chain rule is a simple rule for combining these two derivatives to obtain $\frac{dv}{dt}$.

Let us now state and prove the chain rule formula in terms of a pair of differentiable functions $f(x)$ and $g(x)$. The chain rule is:

$$\frac{d(f(g))}{dx} = \left(\frac{df}{dg}\right)\left(\frac{dg}{dx}\right)$$

Proof:

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

where we have made a change of the limit variable so that $\epsilon \equiv g(x+h) - g(x)$, hence $g(x+h) = g(x) + \epsilon$ and

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

Exercise 4.2. *Show that*

$$\frac{d}{dx}\Big(\sin(x^2)\Big) = 2x\cos(x^2)$$

both from first principles and by using the chain rule.

Example 4.7. *Use the chain rule to find the derivative of $f(ax)$ with respect to x where a is a constant.*

Let us write $g(x) = ax$ then we have

$$\frac{d(f(g))}{dx} = \left(\frac{df}{dg}\right)\left(\frac{d(ax)}{dx}\right) = af'(ax)$$

e.g. if $f(x) = x^3$ then $f(ax) = a^3x^3$ and $\frac{df(ax)}{dx} = \frac{d(a^3x^3)}{dx} = 3a^3x^2$ while $f'(ax)$ can be found by writing $y \equiv ax$ and computing $f'(ax) = f'(y) = \frac{df(y)}{dy} = \frac{d(y^3)}{dy} = 3y^2 = 3a^2x^2$, hence $af'(ax) = 3a^3x^2$, in agreement with the earlier computation.

Example 4.8. *Find the derivative with respect to x of the function e^{x^n} as a function of x.*

Using the chain rule (with $f(x) = e^x$ and $g(x) = x^n$, so that $f(g(x)) = e^{x^n}$) we have

$$\frac{d}{dx}\left(e^{x^n}\right) = \frac{d(f(g))}{dx} = \left(\frac{df}{dg}\right)\left(\frac{dg}{dx}\right) = nx^{n-1}e^{x^n}.$$

Example 4.9. *Find the derivative with respect to x of the function $\ln(\cosh(x))$ as a function of x.*

Using the chain rule (with $f(x) = \ln(x)$ and $g(x) = \cosh(x)$, so that $\frac{df}{dx} = \frac{1}{x}$ and $\frac{dg}{dx} = \sinh(x)$) we have

$$\frac{d}{dx}\Big(\ln(\cosh(x))\Big) = \frac{d(f(g))}{dx} = \left(\frac{df}{dg}\right)\left(\frac{dg}{dx}\right) = \left(\frac{1}{g}\right)\Big(\sinh(x)\Big) = \frac{\sinh(x)}{\cosh(x)} = \tanh(x).$$

Example 4.10. *Find the derivative with respect to x of the function $f(g(h(x)))$ as a function of x.*

Using the chain rule repeatedly we find:

$$\frac{d}{dx}\Big(f(g(h(x)))\Big) = \left(\frac{df}{dg}\right)\left(\frac{dg}{dx}\right) = \left(\frac{df}{dg}\right)\left(\frac{dg}{dh}\right)\left(\frac{dh}{dx}\right).$$

4.3.2 The Sum Rule

The derivative acts linearly on a sum of differentiable functions $f(x)$ and $g(x)$:

$$\frac{d}{dx}\Big(f(x)+g(x)\Big)=\frac{df}{dx}+\frac{dg}{dx}.$$

Proof: (relies on the properties of limits)

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

Exercise 4.3. *Employ the sum rule and, subsequently, differentiation from first principles to show that*

$$\frac{d}{dx}\Big(\cos^2 x+\sin^2 x\Big)=0$$

i.e. prove this without using the identity $\cos^2 x+\sin^2 x=1$.

4.3.3 The Product Rule

This is also known as the Leibniz product rule, and was mentioned (albeit in an alternative notation) in the introduction to our course as one of the puzzles we would aim to demystify. The product rule is defined for the derivative of a product of differentiable functions $f(x)$ and $g(x)$:

$$\frac{d}{dx}\Big(f(x)g(x)\Big)=\frac{df}{dx}g(x)+f(x)\frac{dg}{dx}.$$

Proof:

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

Exercise 4.4. *Use the product rule to find the derivatives of $f(x) = \sin^2 x$ and $g(x) = \cos^2 x$. Make sure your results agree with the calculation of the derivatives from first principles found in answering the previous exercise.*

Example 4.11. *Find the derivative with respect to x of the function $x^2 \sin x$ as a function of x.*

Using the product rule (on the product of functions x^2 and $\sin x$) we have

$$\frac{d}{dx}\left(x^2 \sin x\right) = \frac{d(x^2)}{dx}\sin x + x^2 \frac{d(\sin x)}{dx} = 2x \sin x + x^2 \cos x.$$

Example 4.12. *Find the derivative with respect to x of the function x^x as a function of x.*

We will insert a natural logarithm and an exponential to manipulate x^x into a form upon which we may employ the chain rule:

$$\frac{d}{dx}\left(x^x\right) = \frac{d}{dx}\left(e^{\ln(x^x)}\right) = \frac{d}{dx}\left(e^{x\ln(x)}\right).$$

Now we may use the chain rule (and the product rule) with $f(x) = e^x$ and $g(x) = x\ln(x)$, so that $f(g(x)) = e^{x\ln x}$. Hence we have:

$$\frac{d}{dx}\left(e^{x\ln(x)}\right) = e^{x\ln x}\frac{d}{dx}\left(x\ln x\right) = e^{x\ln x}(\ln x + 1) = x^x(1 + \ln x).$$

4.3.4 The Quotient Rule

The quotient rule is a method for finding the derivative of a pair of differentiable functions $f(x)$ and $g(x)$ written as a quotient, or fraction, i.e. it is a method for computing

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right).$$

The quotient rule can be derived from the product rule for all points where $\frac{1}{g(x)}$ is well-defined. Instead of treating $f(x)/g(x)$ as a quotient we instead consider it as a product $f(x)\times(1/(g(x)))$, and employ the product rule and the chain rule. Explicitly we have:

$$\begin{aligned}
\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) &= \frac{df}{dx}\frac{1}{g(x)} + f(x)\frac{d}{dx}\left(\frac{1}{g(x)}\right) \\
&= \frac{df}{dx}\frac{1}{g(x)} + f(x)\left(-\frac{1}{(g(x))^2}\right)\frac{dg}{dx} \\
&= \frac{1}{(g(x))^2}\left(\frac{df}{dx}g(x) - f(x)\frac{dg}{dx}\right).
\end{aligned}$$

This is not so simple a formula to remember, so it is good news that it can be derived from the product rule and the chain rule together.

Exercise 4.5. *Check the validity of the quotient rule by first computing the derivative of* $\tan(x)$ *from first principles and comparing your result with the derivative* $\frac{d}{dx}\left(\frac{\sin x}{\cos x}\right)$ *computed using the quotient rule.*

4.4 Derivatives of Implicit Functions

The methods we have used so far to find derivatives are effective for functions which are defined, or can be defined, explicitly in terms of a variable, i.e. functions which may be written as $f(x)$. However it is often interesting to be able to find derivatives of functions whose definition are given in terms of their properties without an explicit function being available, or even possible in principle. The prime example of such a function that we have met earlier in the course is the abstract definition of the inverse function, defined by $f(f^{-1}(x)) = x$, but there are other common implicit functions defined as curves or by parametric equations. The common quality that all implicit functions share is that they are defined by a relation rather than given an explicit functional definition in terms of a parameter. Let us consider these examples in turn and evaluate their derivatives.

4.4.1 Inverse Functions

Earlier in the course we saw that the inverse function $f^{-1}(x)$ of a function $f(x)$ is defined via $f(f^{-1}(x)) = x$. We will invoke the inverse function theorem without proof here (it is not part of this course, but it will prove invaluable in this section):

Theorem 4.2. *(Inverse Function Theorem) If* f *is a differentiable function with a non-zero derivative at the point* x_0 *then* f *is invertible in any neighbourhood of* x_0 *and* f^{-1} *is a differentiable function such that*

$$(f^{-1})'(x_0) = \frac{1}{f'(f^{-1}(x_0))}.$$

We will only make use of part of the theorem, but that part is the most interesting part, namely that the inverse function f^{-1} is a differentiable function if f is a differentiable function. By assuming these properties we will be able to find the derivative of the inverse function using the chain rule.

Assume that $f(x)$ is a differentiable function (hence by the inverse function theorem $f^{-1}(x)$ is also a differentiable function) then starting from the definition of the inverse function, and

acting with the derivative on both sides of the equation we obtain:

$$\frac{d}{dx}\bigg(f(f^{-1}(x))\bigg) = \frac{d}{dx}\bigg(x\bigg) = 1.$$

Now, since both f and f^{-1} are differentiable functions we may use the chain rule to obtain:

$$\frac{d}{dx}\bigg(f(f^{-1}(x))\bigg) = \bigg(\frac{df}{df^{-1}}\bigg)\bigg(\frac{d(f^{-1}(x))}{dx}\bigg) = 1.$$

Hence,

$$\bigg(\frac{d(f^{-1})}{dx}\bigg) = \frac{1}{\bigg(\frac{df}{df^{-1}}\bigg)}$$

or using the primed notation to denote the derivative with respect to the argument we have

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

which is the statement in the last part of the inverse function theorem. The meaning of the above is best seen through some examples.

Example 4.13. *Find the derivative of* $\ln(x)$.

The natural logarithm is the inverse function of the exponential function $f(x) = e^x$ *and is defined for* $x > 0$. *The defining relation is:*

$$e^{\ln(x)} = x.$$

By taking the derivative of both sides of the equation above we have:

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

Hence,

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

Example 4.14. *Find the derivative of* $\arcsin(x)$.

$\arcsin(x)$ *is the inverse function for* $f(x) = \sin(x)$ *and is defined for* $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. *The defining relation is:*

$$\sin\left(\arcsin(x)\right) = x.$$

By taking the derivative of both sides of the equation above we have:

$$\frac{d}{dx}\Big(\sin\left(\arcsin(x)\right)\Big) = \cos\left(\arcsin(x)\right)\frac{d}{dx}\Big(\arcsin(x)\Big) = 1.$$

Hence,

$$\frac{d}{dx}\Big(\arcsin(x)\Big) = \frac{1}{\cos\left(\arcsin(x)\right)} = \frac{1}{\sqrt{1-\sin^2(\arcsin(x))}} = \frac{1}{\sqrt{1-x^2}}.$$

Where we have made use of the identity $\cos^2(x) + \sin^2(x) = 1$.

Example 4.15. *Find the derivative of* $\arctan(x)$.

$\arctan(x)$ *is the inverse function for* $f(x) = \tan(x)$ *and is defined for* $x \in (-\frac{\pi}{2}, \frac{\pi}{2})$. *The defining relation is:*

$$\tan\left(\arctan(x)\right) = x.$$

By taking the derivative of both sides of the equation above we have:

$$\frac{d}{dx}\Big(\tan\left(\arctan(x)\right)\Big) = \left(1+\tan^2\left(\arctan(x)\right)\right)\frac{d}{dx}\Big(\arctan(x)\Big) = (1+x^2)\frac{d}{dx}\Big(\arctan(x)\Big) = 1.$$

Hence,

$$\frac{d}{dx}\Big(\arctan(x)\Big) = \frac{1}{1+x^2}.$$

4.4.2 Curves

A curve, for the purpose of this course, means the set of solutions to equations defined in terms of x and y which are related by

$$R(x, y) = 0.$$

Typically the sets of points which satisfy the above relation form lines which may be curved (as the name suggests) or straight lines (as the name does not suggest). There is a difference between a relation being used to implicitly define a function as above and the idea that one can find an explicit function $f(x)$ to define the curve as $y = f(x)$. Usually one can only locally find a function $y = f(x)$ which will relate the y-coordinate of a point on a curve to its x-coordinate, but one can cover the curve with these local functions. Consider the example of the unit circle which is defined by the relation:

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

One can rearrange this equation to find two functions which cover the curve for different values of the coordinates, namely

$$y = \begin{cases} +\sqrt{1-x^2} & y \geq 0 \\ -\sqrt{1-x^2} & y < 0 \end{cases}$$

For more interesting curves the principle remains the same: one can locally find functions of the form $y = f(x)$ that are the same shape as the curve (locally). These local functions are the functions which are defined implicitly by $R(x, y) = 0$.

To find the gradient of a tangent to a curve, one can employ the chain rule and the defining relation, together with the notion that $y = f(x)$:

$$\frac{d}{dx}\Big(R(x,y)\Big) = \frac{d}{dx}\Big(R(x,f(x))\Big) = 0.$$

Let us consider some examples.

Example 4.16. *Find the derivative $\frac{dy}{dx}$ for the function $y(x)$ defined implicitly by the relation:*

$$R(x,y) \equiv y + \sin y - x = 0.$$

We treat y as a function of x, i.e. $y = y(x)$ and use the chain rule when taking the derivative of the relation with respect to x:

$$\frac{d}{dx}\Big(R(x,y)\Big) = \frac{d}{dx}\Big(y + \sin y - x\Big) = \frac{dy}{dx} + \cos y \frac{dy}{dx} - 1 = 0$$

which we may rearrange to find

$$\frac{dy}{dx} = \frac{1}{1+\cos y}.$$

This is an unusual result in our experience of derivative functions so far, the derivative function is given as a function of y itself - whereas we are used to obtaining the derivative function as a function of x. However it presents no serious problems for us: to find the derivative of y at $x = x_0$, we first must identify the corresponding point $y = y_0$ from the relation

$$y_0 + \sin y_0 - x_0 = 0$$

armed with the explicit point (x_0, y_0) which lies on $R(x, y) = 0$ we can then compute the gradient of the the tangent line to the curve at that point, i.e.

$$\frac{dy}{dx}(x_0) = \frac{1}{1+\cos y_0}.$$

For example at $x = 0$ we have $y = -sin(y)$ which is satisfied when $y = 0$, so the point $(0, 0)$ lies on the curve and at that point the gradient of the tangent line is $\frac{dy}{dx}(0) = \frac{1}{1+\cos 0} = \frac{1}{2}$. As y increases we see that the gradient oscillates between being undefined (i.e. the curve becomes vertical) and $\frac{1}{2}$. So we expect this curve to wiggle its way through the point $(0, 0)$ and a plot of the graph can confirm this. The sketch of the graph shown here in figure 4.4 was found by solving the differential equation for the gradient (and setting the constant to zero) to obtain $y(x)$.

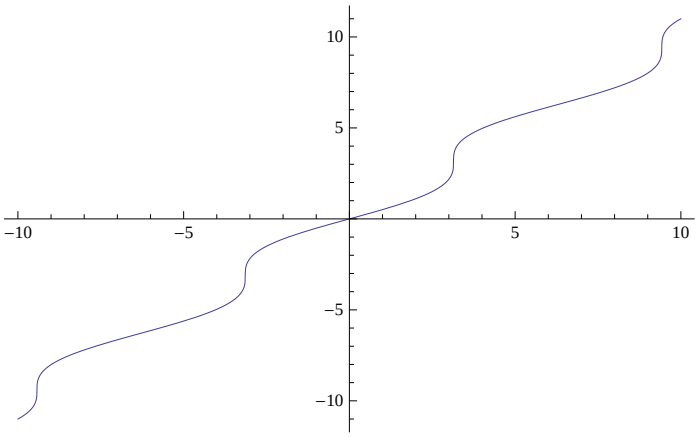


Figure 4.4: The graph of the curve $y(x)$ defined as the solution set of $y + \sin y - x = 0$.

Example 4.17. *Find the derivative $\frac{dy}{dx}$ for the function $y(x)$ defined implicitly by the relation:*

$$R(x, y) \equiv y - \tanh(xy) = 0.$$

We have,

$$\frac{d}{dx}\Big(R(x,y)\Big) = \frac{d}{dx}\Big(y - \tanh(xy)\Big) = \frac{dy}{dx} - \frac{1}{\cosh^2(xy)}\frac{d(xy)}{dx} = \frac{dy}{dx} - \frac{1}{\cosh^2(xy)}(y + x\frac{dy}{dx}) = 0$$

which we may rearrange to find

$$\frac{dy}{dx} = \frac{y}{\cosh^2(xy) - x}.$$

This is a challenging graph to sketch! First we notice that when $y = 0$ we have $\frac{dy}{dx}(y = 0) = \frac{0}{1-x}$ which equals zero apart from at the point $x = 1$ where it becomes undefined. For which values of x does $y(x) = 0$? We must solve $0 = \tanh 0 \times x$ which is true for all x. Hence we find one set of solutions is given by the $x-axis$ whose gradient is zero. Let us turn our thoughts to the point $x = 1$ where the derivative was undefined: evidently $y = \tanh(y)$ is solved when $y = 0$, but this we already knew. So we may (correctly) wonder why the derivative is undefined

there. Now as $y = \tanh(xy)$ *then* $y \in (-1, +1)$*. Let us see if there are any points on the curve where* $y(x) = 1$*, this corresponds to* $\tanh(x) = 1$*, which is only true in the limit, i.e.* $\lim_{x\to\infty}(\tanh(x)) = 1$ *where we find*

$$\lim_{x\to\infty, y\to 1}\left(\frac{dy}{dx}\right) = \lim_{x\to\infty, y\to 1}\left(\frac{y}{\cosh^2(xy) - x}\right) = \lim_{x\to\infty}\left(\frac{1}{\cosh^2(x) - x}\right) = 0.$$

So the curve is tangential to $y = 1$ *as* $x \to +\infty$*. We can make a similar argument for* $y = -1$ *when* $x \to \infty$*, so* $y = -1$ *is also a tangent to the curve as* $x \to \infty$*. We have now three tangent lines to the curve as* $x \to \infty$*, given by* $y = \{-1, 0, 1\}$*, we can conclude that the curve has at least three parts to it as* x *grows. We may wonder if there are more branches to this curve? Consider* $y = \delta$ *where* $0 < \delta < 1$*, then we are aiming to solve the equation* $\delta = \tanh(x\delta)$ *which has a unique solution* $x = \frac{1}{\delta}\text{arctanh}(\delta)$*. Hence we now have the picture that there are three parts to the curve: a curve when* $y > 0$ *and* $x > 1$*, the line* $y = 0$ *and a curve when* $y < 0$ *and* $x > 0$ *(for the last curve consider* $y = \delta'$ *where* $-1 < \delta' < 0$*). Finally consider the gradient for the part of the curve where* $y > 0$ *and* $x > 1$*, at* $x \to \infty$ *the gradient approaches zero. We may consider the points near* $x = 1$ *by substituting* $x = 1 + \epsilon$ *where* $\epsilon \geq 0$ *to find:*

$$\begin{aligned}
\lim_{y\to 0^+, x\to 1^+}\left(\frac{dy}{dx}\right) &= \lim_{y\to 0^+, \epsilon\to 0^+}\left(\frac{y}{\cosh^2((1+\epsilon)y) - (1+\epsilon)}\right) \\
&= \lim_{y\to 0^+, \epsilon\to 0^+}\left(\frac{y}{(1 + \frac{(1+\epsilon)^2y^2}{2!} + \mathcal{O}(y^4))^2) - (1+\epsilon)}\right) \\
&= \lim_{y\to 0^+}\left(\frac{y}{y^2 + \mathcal{O}(y^4)}\right) \\
&= \infty.
\end{aligned}$$

Hence the upper curve becomes vertical as it approaches the point $(x = 1, y = 0)$*. A similar argument can be made for the curve when* $y < 0$*. Eventually these lengthy observations can allow us to sketch the curve and we show the graph in figure 4.5.*

4.4.3 Parametric Functions

Given two functions $x(t)$ and $y(t)$ for $t \in \mathbb{R}$ we can define a curve as the set of points $(x(t), y(t))$ which are sketched out as t varies. In other words, t is a single coordinate that defines a point on a curve: one might like to think of t as a distance along a curve as measured from the point on the curve given by $t = 0$. As one moves along the curve the value of t changes, and t is said to parameterise the curve.

One can easily rewrite common curves with a parametric definition, e.g. the parabola $y = x^2$ consists of the points (t, t^2) parameterised by t, or the straight line $y = mx+c$ is the set of points

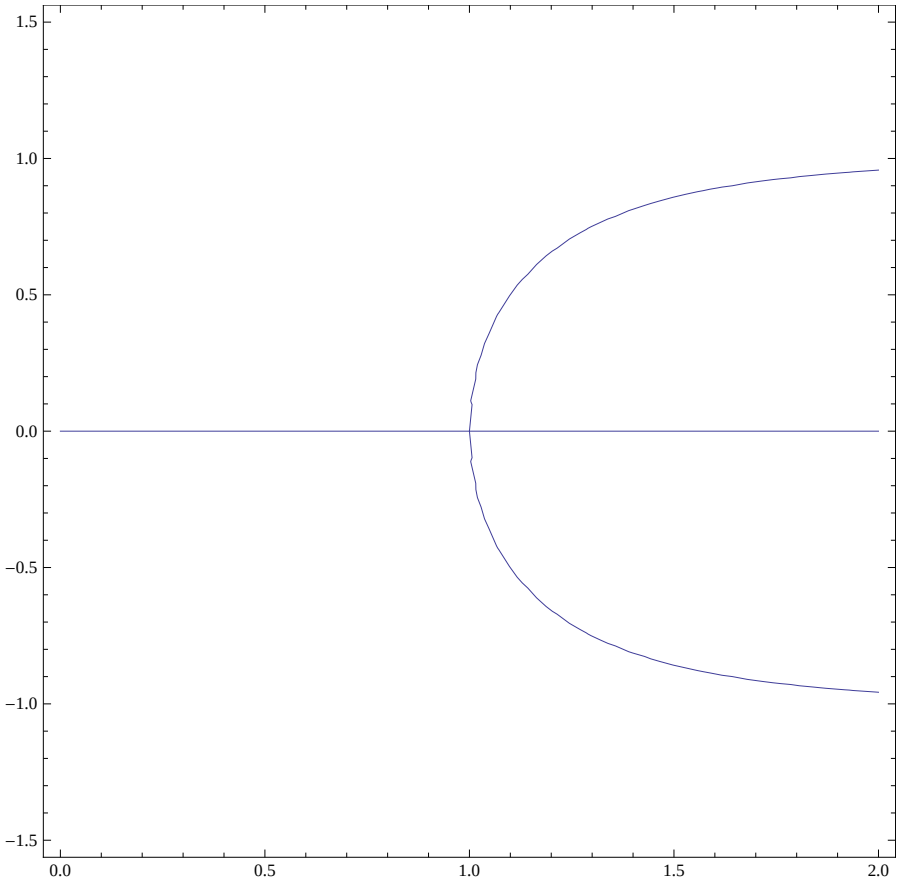


Figure 4.5: The graph of the curve $y(x)$ defined as the solution set of $y - \tanh(yx) = 0$.

$(t, mt + c)$. The choice of parameterisation is not unique: one can always shift t by a constant t_0 so as to move the point corresponding to $t = 0$ along the curve: if $t' \equiv t + t_0$ then $t = t' - t_0$ and the parameterisation of the straight line becomes $(t, mt + c) = (t' - t_0, mt' - mt_0 + c)$, and if we picked the value of $t_0 = \frac{c}{m}$ then the line becomes the points $(t' - \frac{c}{m}, mt')$ now parameterised by t': one can move the origin in t ($t = 0$) to wherever one wishes on a parametric curve in this way, which can be very useful.

Now, given $(x(t), y(t))$ we may wonder if $y(x)$ exists. Suppose that $x = f(t)$, if f^{-1} exists and is well-defined we have $t = f^{-1}(x)$, hence we would then have $y(t) = y(f^{-1}(x)) \equiv y(x(t))$. So if $y(x(t))$ exists then by the chain rule we have:

$$\frac{dy}{dt} = \left(\frac{dy}{dx}\right)\left(\frac{dx}{dt}\right)$$

and hence

$$\frac{dy}{dx} = \frac{\left(\frac{dy}{dt}\right)}{\left(\frac{dx}{dt}\right)}.$$

Now as $x(t)$ and $y(t)$ are differentiable functions we may compute directly $\frac{dx}{dt}$ and $\frac{dy}{dt}$, and combine to obtain $\frac{dy}{dx}$, so long as $\frac{dx}{dt} \neq 0$.

Example 4.18. *Find the derivative function $\frac{dy}{dx}$ for the curve defined parametrically by*

$$x(t) = t + \cos(t)$$
$$y(t) = \ln(\cosh(\sin(t)))$$

where $t \in \mathbb{R}$.

We may compute the following derivatives immediately:

$$\frac{dx}{dt} = 1 - \sin(t)$$
$$\frac{dy}{dt} = \frac{1}{\cosh(\sin(t))}\frac{d}{dt}\Big(\cosh(\sin(t))\Big) = \frac{\sinh(\sin(t))}{\cosh(\sin(t))}\frac{d}{dt}\Big(\sin(t)\Big) = \tanh(\sin(t))\cos(t)$$

Hence we see that the derivative is not defined when $t = \frac{\pi}{2} + n2\pi$ for $n \in \mathbb{Z}$, but is defined elsewhere. We have:

$$\frac{dy}{dx} = \frac{\left(\frac{dy}{dt}\right)}{\left(\frac{dx}{dt}\right)} = \frac{\tanh(\sin(t))\cos(t)}{1 - \sin(t)}.$$

A sketch of the parametric curve is shown in figure 4.6 where it can be observed that the derivative is not well-defined at $x = \frac{\pi}{2} + n2\pi$.

Example 4.19. *Find the derivative function $\frac{dy}{dx}$ for the curve defined parametrically by*

$$x(t) = e^t$$
$$y(t) = \tan(t)$$

where $t \in \mathbb{R}$.

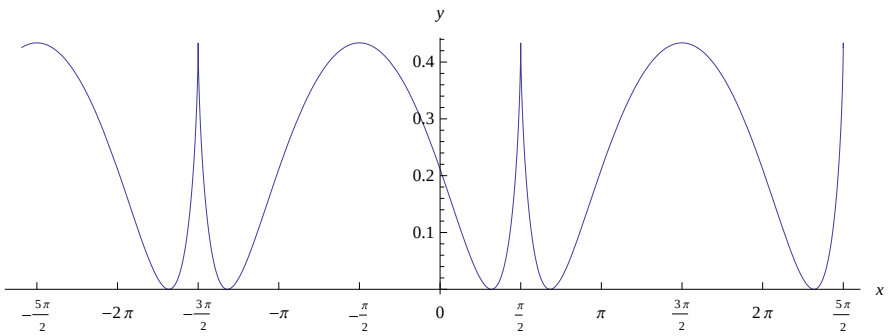


Figure 4.6: The graph of the curve $y(x)$ defined parametrically as the points $(t + \cos(t), \ln(\cosh(\sin(t))))$ for $t \in \mathbb{R}$.

Our preliminary observation is that $x(t) > 0$ so the curve is defined only for $x \in \mathbb{R}^+$. We may compute the following derivatives immediately:

$$\frac{dx}{dt} = e^t = x$$
$$\frac{dy}{dt} = 1 + \tan^2(t) = 1 + y^2$$

Hence we see that the derivative is not defined when $x = 0$ but is defined elsewhere. We have:

$$\frac{dy}{dx} = \frac{\left(\frac{dy}{dt}\right)}{\left(\frac{dx}{dt}\right)} = \frac{1 + y^2}{x}.$$

A sketch of this parametric curve is shown in figure 4.7.

4.5 The Mean Value Theorem

Suppose that we know a lot of information about the derivatives of a function, but we don't know much about the function at all. In the last few sections we saw some examples of how this may come about by defining curves via an equation or defining them parametrically, but in these examples we did in principle have a definition of the function whose derivative we were trying to compute and it just happened to be simpler to work with the derivative due to the way the function was defined. Here we are imagining that we have a table of data defining the derivative function in detail and we are wondering if we can use this data to reconstruct the function. A real world example of this occurs in speed traps for cars, where a car's position is measured at two times and from this information it can be deduced whether the car broke the

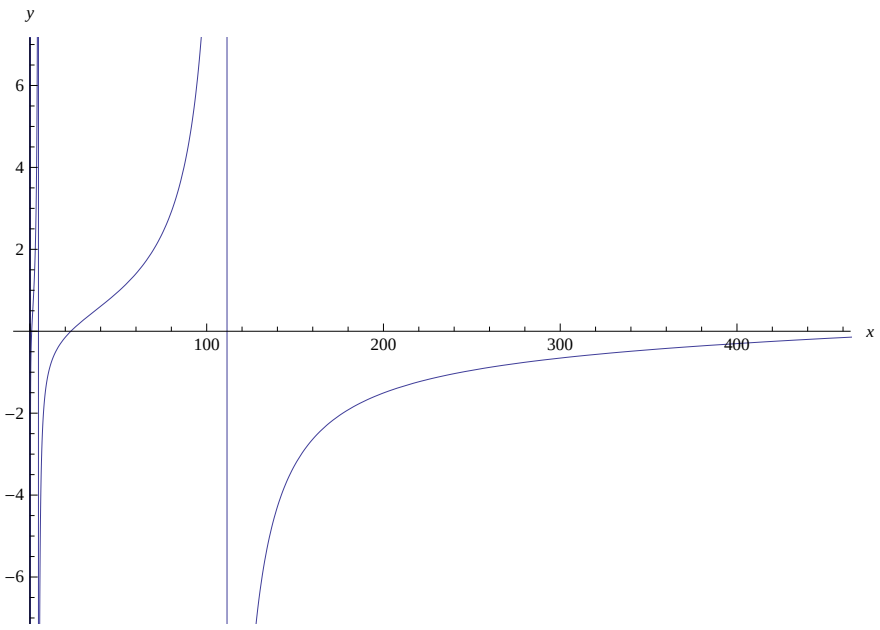


Figure 4.7: The graph of the curve $y(x)$ defined parametrically as the points $(e^t, \tan(t))$ for $t \in [-8, 8]$.

speed limit at any point between the two measurement points. All this rests upon the mean value theorem.

Theorem 4.3. *(The Mean Value Theorem) Let f be a continuous function on $[a, b]$ and a differentiable function on (a, b), then there exists a point $c \in (a, b)$ where*

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

The mean value theorem says that at some point c the slope $\frac{f(b)-f(a)}{b-a}$ of the straight line connecting $f(a)$ to $f(b)$ is actually equal to the derivative of f at c. It is helpful to think what this means when the function is the position function and the derivative is taken with respect to time. In this case the mean value theorem tells us that for a journey from $f(a)$ to $f(b)$ then your average speed $\frac{f(b)-f(a)}{b-a}$ is equal to your actual instantaneous speed at least one moment, time c. Let us try some numbers and stick with the setting of the distance, time and speed. Suppose that it takes you 1 hour to travel 60 miles, then your average speed for the journey is 60 miles per hour (mph). The mean value theorem tells you that at least for one moments in the journey you were travelling at 60 mph. Of course you might have travelled the route in a number of different ways, e.g. you may have travelled at 60 mph for the entire journey, or you may have travelled part of the journey at a speed greater than 60 mph and another at a speed less than 60 mph (so that the average speed is still 60 mph), but in that case your speed

would have to pass through 60 mph at some point. This is obviously an unpleasant realisation for an apprentice analyst wishing to break the speed limit. Such a mathematician might have thought to argue in court against the evidence of a speed trap by imagining that there existed a position function whose derivative nowhere exceeded the speed limit, but the mean value theorem tells her or him that no such function exists.

Let us return to abstract functions and illustrate the mean value theorem on a graph. In figure 4.8 we see a curve $y = f(x)$ and construct the straight line which passes through the

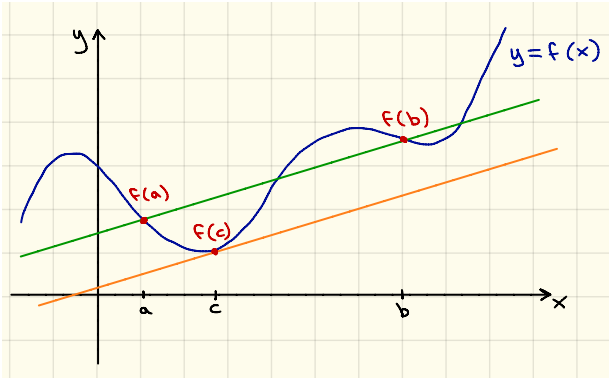


Figure 4.8: The mean value theorem illustrated on the curve $y = f(x)$.

points $f(a)$ and $f(b)$. The slope of this straight line is $\frac{f(b)-f(a)}{b-a}$ and the mean value theorem tells us that this straight line is tangent to the curve $y = f(x)$ for at least one point $x = c$ where $c \in (a, b)$, in figure 4.8 we illustrate this by translating the line passing through $f(a)$ and $f(b)$ until it is tangent to $y = f(x)$. For the curve we have sketched there is at least one other point on $y = f(x)$ whose tangent has the same slope.

A very useful example of the mean value theorem occurs when $f(a) = f(b)$ as in this case we learn that there exists some c such that

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

i.e. there exists a stationary point at $x = c$. This example of the mean value theorem is important enough to have its own name, it is called Rolle's theorem after Michel Rolle who first proved this in 1691.

We commenced this section by wondering whether we can construct a function from knowledge of its derivatives. The mean value theorem tells us about at least one derivative on the open interval (a, b), so how might we use this to construct the function on the open interval? There is a very useful theorem that allows us to reconstruct certain functions: the constant functions.

Theorem 4.4. *If $f(x)$ is a differentiable function on an open interval such that*

$$f'(x) = 0$$

for all x then $f(x)$ is constant on the interval.

Proof: We may consider two points a and b defined on the open interval and then we have, by the mean value theorem, that

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

where $c \in (a, b)$. However we know (from the statement of the theorem) that the function's derivatives are all zero on the open interval, which includes (a, b) hence $f'(c) = 0$. Therefore we have

$$f(b) - f(a) = 0$$

for all a and b, hence $f(a) = f(b)$ for any choice of a and b in the open interval where the derivative vanishes, hence it is a constant function.

In this chapter we developed a definition of the derivative function and studied its properties. The derivative is the instantaneous slope of a function and is defined using the limit. We developed some properties of the derivative, most importantly the chain rule and the product rule. It was proved that every differentiable function is also continuous and we saw how we can use the derivative on functions defined implicitly, via an inverse or curves defined parametrically and we had some practise sketching functions with the aid of the derivative function. The final section of this chapter presented the mean value theorem and we saw how the vanishing derivative on an open interval implied that the function must be constant on the open interval. We began our final comments by speculating whether we could commence with knowledge of a derivative function and use it to reconstruct the function. Such an operation would be the inverse of the derivative, it would be the process of finding the antiderivative. In the following chapter we will see how the antiderivative is related to the integral which gives a method for finding the area under a curve.

5. Integration

In which we meet the integral: a way of turning a sum of infinitely many infinitesimally thin parallelograms into finite number which gives the area under a curve. Astonishingly this will be related by the fundamental theorem of calculus to the antiderivative of a function!

The material in this chapter will be covered in weeks 9-11.



Figure 5.1: Isaac Newton (top) in 1689 and Gottfried Leibniz (below).

Integration was developed almost independently by both Isaac Newton (1643-1726) and Gottfried Leibniz (1646-1716) in the late 17$^{\text{th}}$ century. The argument over who had priority in the invention of calculus is one of the most famous disputes over the ownership of an idea in all of science: the passion shown by these two intellectual giants over priority rather than joy in their shared legacy of a great discovery leaves a bitter aftertaste. While Newton made his discoveries first he had kept them secret, on the other hand it seems indisputable that Leibniz had found partial inspiration from Newton's wider work. It is put succinctly in James Gleick's biography "Isaac Newton":

> "Newton had made his discoveries first, and he had discovered more, but Leibniz had done what Newton had not: published his work for the world to use and to judge."

It is widely accepted that both men discovered the calculus independently but Newton did his reputation no favours during the dispute. When a committee of the Royal Society gave its report on the argument it said, in the words of Gleick, that:

> "It judged Newton's method to be not only the first - "by many years" - but also more elegant, more natural, more geometrical, more useful and more certain."

The President of the Royal Society at the time the report was made and also the secret author of the report was... Isaac Newton.

5.0.1 Motion at constant speed: the area under a straight line

To intuit that the area under a curve and the operation giving the gradient may be related it is worth calling to mind the dynamical examples that inspired Newton. If we sketch the graph of the position of a particle $x(t)$ which moves at constant speed $\frac{dx}{dt}$ we have a simple straight line (its tangent line at every point has the same, constant slope): The velocity function for this

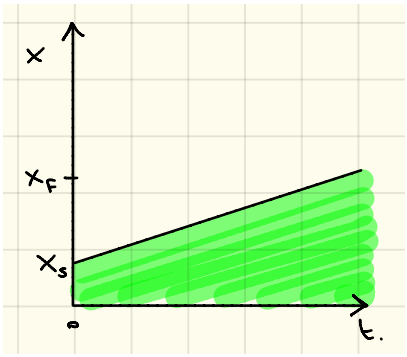


Figure 5.2: The function $x(t)$ where $x'(t)$ is constant. The area under the graph as the position changes from x_S to x_F is shaded, badly, but not particularly dynamically interesting.

motion is a constant function and we sketch it below. The area under the velocity function is

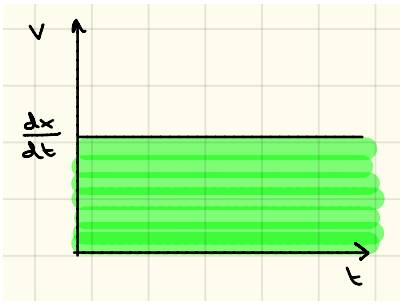


Figure 5.3: The function $v(t)$ is constant. The area under the graph as the position changes from x_S to x_F is shaded.

dynamically interesting, as the speed for this example is constant, we can use the formula for the average speed as the particle moves from x_S at time t_S to x_F at time t_F is

$$\frac{dx}{dt} \times (t_F - t_S) = x_F - x_S.$$

For this simple motion we see that from the graph of the velocity, $\frac{dx}{dt}$, we can obtain information related to the graph of the position function $x(t)$ rather than just $v(t)$.

We have considered a very simple straight line function, but for infinitesimal domains any continuous function is approximately a straight line, so we may expect this feature to continue, that integration (the process of computing the area under a graph) is roughly the inverse procedure to taking the derivative, that is,

$$x(t) \xrightarrow{\frac{d}{dt}} v(t) \qquad \text{and} \qquad x(t) \overset{\text{Integration}}{\longleftarrow} v(t).$$

That integration is the inverse of differentiation is remarkable and is the fundamental theorem of calculus - we will return to this later in this chapter. First we must formalise and develop integration.

5.1 The Riemann Integral

The integral of a function is up to a sign the area "under" a graph of the function. Let us look at the fine-print in the definition of the integral.

Definition 5.1.1. *The integral, denoted $\int_a^b f(x)dx$ is the total area between the graph of $y = f(x)$ and the x-axis in the x, y plane between $x = a$ and $x = b$, counted positively for $f(x) > 0$ (area above the x-axis) and negatively for $f(x) < 0$ (area below the x-axis).*



Figure 5.4: Bernhard Riemann in 1863.

But how does one practically go about computing $\int_a^b f(x)dx$? Perhaps surprisingly, a rigorous definition of the integral was not given until much later than the fundamental theorem of calculus and the surrounding mathematics had been established by Newton and Leibniz. It was not until 1854 that the first rigorous definition of the integral was written down by Bernhard Riemann (1826-1866)[1]. The idea behind the integral is not challenging, namely that one can split the area underneath a curve into a sequence of rectangles and that one can approximate the height of the rectangles to the curve, although the widths of the rectangles will typically become infinitesimal... The integral of a function calculated in this way is called the Riemann integral.

First let us consider a graph which is precisely a sequence of rectangles. Such a graph is called (for obvious reasons) a 'staircase graph' or 'staircase function', let us consider one such arbitrary staircase function as shown in figure 5.5. The staircase function is defined by

[1]Note how young Riemann was at the time of his death. Mathematics surely lost one of its finest minds too young. After his death his housekeeper threw out many of his papers, which Riemann was not content to publish, but which probably contained many correct insights and results.

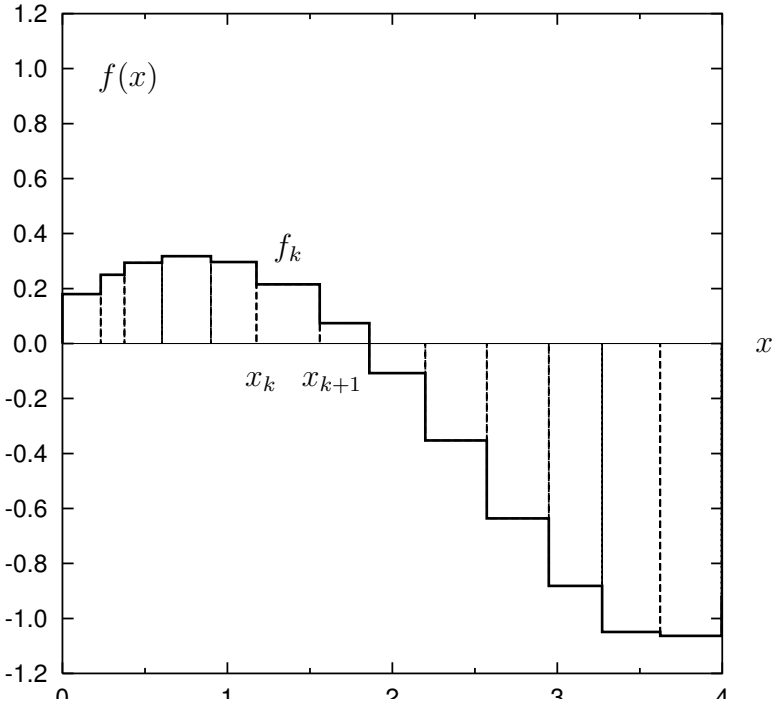


Figure 5.5: A staircase function: let the vertical sides of the rectangles sit at $x_1 < x_2 < x_3 < \ldots < x_n < x_{n+1}$ and let the height of the k'th rectangle by f_k.

$$f(x) = f_k \quad \text{if } x \in [x_k, x_{k+1})$$

where f_k is a sequence of real numbers (not specified here, although it is possible to read off some values of f_k from the graph). Now the area of the k'th rectangle is its width times its height or

$$(x_{k+1} - x_k) f_i$$

and notice that the sign convention agrees with that of the integral: if $f_k < 0$ then the "area" is counted negatively. This unusual function is chosen solely because we can compute its integral exactly:

$$\int_{x_1}^{x_{n+1}} f(x)dx = \sum_{k=1}^{n} (x_{k+1} - x_k) f_k.$$

On the left-hand-side we have the notation denoting the integral, while on the right-hand-side we have been able to write an expression for the sum of the areas of the n rectangles under the staircase function between $x = x_1$ and $x = x_{n+1}$. This exact integral is the foundation upon which all of our Riemann integrals will be constructed.

In the previous example we considered a staircase function and showed that we could compute its integral exactly. This was because the staircase function's special shape allowed it to be partitioned into rectangles whose areas could be rapidly evaluated. We would like to be able to find the integral of an arbitrary curve, not only staircase functions, and what we will do is to try to come up with a method of approximating curves to staircase functions. What we will achieve ultimately will be a way of sandwiching the integral of any function between two staircase functions, such that in a limit we can compute the integral. We will worry what this means soon enough, first let us try to see how we might construct staircase functions which closely approximate an arbitrary continuous function.

So how would we approximate a continuous function to a staircase function? What we would like is a staircase whose steps are the same height as the function. But as the function is continuous it cannot have any jumps in values as occurs for the staircase function when one moves between steps. But what if we made the width of the steps infinitesimally thin? If we could do this then we could approximate a continuous function by a staircase function whose steps have infinitesimal width and whose heights are equal to the function. This will be our goal, but first let us think a little about some staircase functions which begin to approximate an arbitrary continuous function.

Imagine writing an algorithm to associate a staircase function to a curve: it would be difficult to know where to begin, indeed there are many possible beginnings. There are a number of decisions to make: how wide should the steps be? Should the step width vary or be constant? How do decide the height of each step so that it is related to the function? For example should the step height be the value of the function at the mid-point of the step width (as depicted in figure 5.6), or it could be given by the top left-hand-corner of each step, or the right-hand corner or indeed any point on the step could be put equal to the function at the same point.

How will we find a unique way to associate a staircase function to a continuous curve? Well, the answer is that for steps of finite width we will not find a unique staircase function, but once we take the limit so that the steps go to infinitesimal width then all staircase functions will approach the continuous function. The algorithm we will follow is:

(i) consider all possible staircase functions where each step touches the curve we are interested in (typically this will be an infinite set of staircase functions),

(ii) find the area under each staircase graph in the limit where the width of all steps goes to zero and

(iii) check whether all these areas for the limits of the variety of staircase functions give the same area.

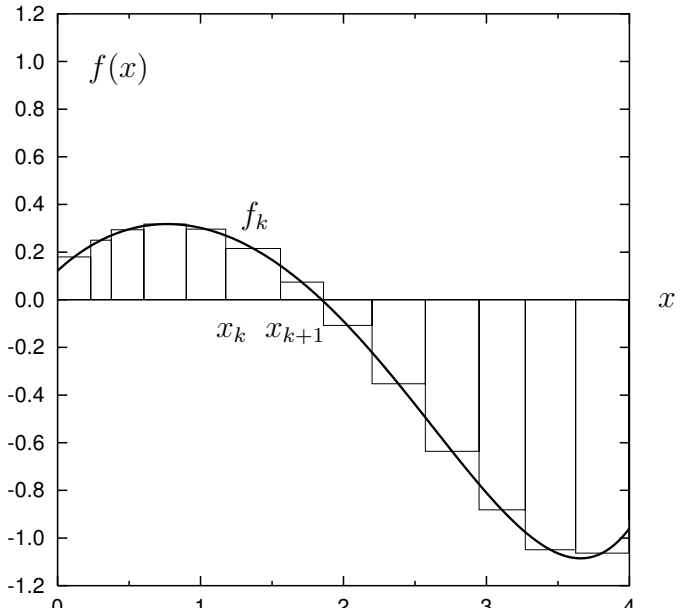


Figure 5.6: An example of a staircase function (split into rectangular strips) for which the mid-point of the step is equal to the function, note the varying width of the steps.

The method proposed is tedious and, potentially, will take an infinite time to actually carry out but since such a method may exist it gives us courage that, for a continuous function $f(x)$ defined on the interval $[a, b]$ the integral $\int_a^b f(x)\,dx$ exists. This encapsulates the formal definition of the Riemann integral:

Definition 5.1.2. *The (Riemann) integral of a function $f(x)$ on the interval $[a, b]$, if it exists, is equal to the sum of the areas of the steps of any staircase function $S(x)$ in the limit that the step width, w, approaches zero so long as $\lim_{w\to 0}(S(x)) = f(x)$.*

Comment(s). *(On the Riemann integral...) Notice that we have generalised the discussion and the Riemann integral is defined for any function, not just continuous functions in this definition and we have introduced the idea that the integral might not exist for general functions. As we have seen earlier in the course the limit does not always exist and it is therefore possible for the integral of a function to not be defined. For example, think of integrating $1/x$ on the interval $[0, b]$ where $b > 0$, from the graph we know that the integral will be infinite and so not exist. The short story (but not the full story) is that if a function is continuous on the interval then the Riemann integral exists and the function is said to be integrable on the interval.*

To construct the Riemann integral we will first notice that out of the infinite set of staircase

functions which we may associate with a function we are trying to integrate, there are two classes of staircase function which stand out:

- $S^-(x)$ those which have the minimum area (for a given, not necessarily constant, set of step-widths) and
- $S^+(x)$ those which have the maximum area (for a given, not necessarily constant, set of step-widths).

Before defining these functions let us consider the graphs in figure 5.7 showing examples of $S^-(x)$ and $S^+(x)$ for an arbitrary continuous function $f(x)$. Notice that, when $f(x) > 0$, the

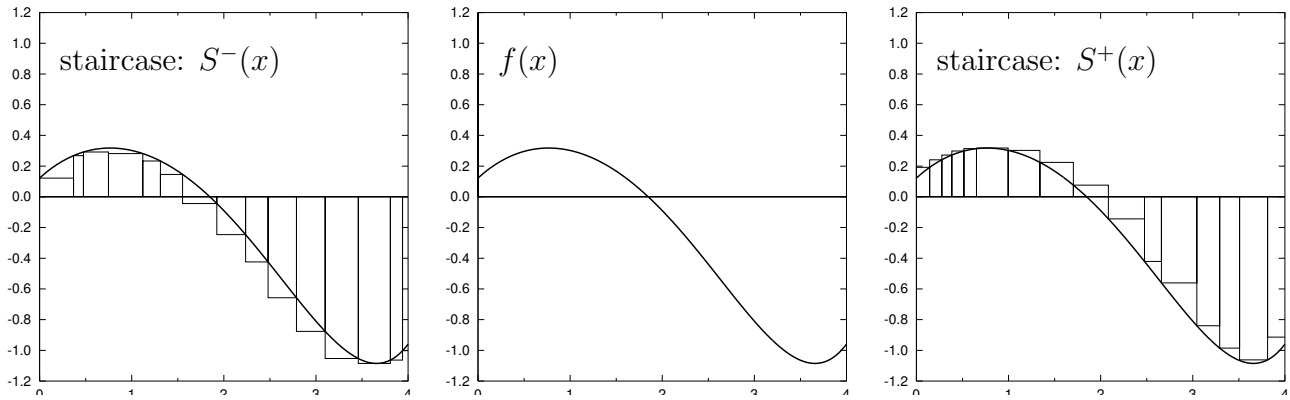


Figure 5.7: Staircase functions constructed for the function $f(x)$ (shown in the centre and repeated in the diagrams on the left and on the right). The area of the staircase function $S^-(x)$ underestimates the integral of $f(x)$, while the area of $S^+(x)$ overestimates the same integral.

area of each rectangular step in the staircase function $S^-(x)$ is less than the area under the graph evaluated on the interval covered by the step's width. When $f(x) < 0$ the integral of each step in $S^-(x)$ is more negative (less than) that of $f(x)$ on the same interval. Now consider $S^+(x)$ and notice that the integral of each step in the staircase function is always greater than the integral of $f(x)$. In other words the staircase functions satisfy the relation

$$S^-(x) \leq f(x) \leq S^+(x) \qquad \forall\, x \text{ in the domain of } f(x).$$

Practically how do we define these staircase functions? It is a simple matter of finding a definition that agrees with the inequality given above. Recall from our first use of the staircase function that a staircase function $S(x)$ is defined by specifying the height of its steps on the intervals $[x_k, x_{k+1})$, such that $x_1 < x_2 < x_3 < \ldots < x_n < x_{n+1}$ are the x coordinates of each

side of the n steps. We will consider the integral over the interval $x \in [a, b]$ so $x_1 = a$ and $x_{n+1} = b$. Hence we can define $S^-(x)$ and $S^+(x)$ for a given function $f(x)$ as follows:

$$S^-(x) = S_k^- \equiv \min[f(x)] \qquad \forall\, x \in [x_k, x_{k+1})$$
$$S^+(x) = S_k^+ \equiv \max[f(x)] \qquad \forall\, x \in [x_k, x_{k+1}).$$

In words the $k'th$ step of the staircase function has constant height over the range $x \in [x_k, x_{k+1})$. We pick the height of each step of $S^-(x)$ to just touch $f(x)$ at its minimum point, so it will always have an integral less than that of $f(x)$, while $S^+(x)$ has steps which just touch $f(x)$ at its maximum point in the width of the step, so the integral of $S^+(x)$ is always greater than the integral $f(x)$. Notice that the width of each step has not been fixed in any way, the steps have arbitrary width, so there still remain an infinite set of choices for $S^-(x)$ and $S^+(x)$. However we are now sure that:

$$A^- \equiv \sum_{k=1}^{n} S_k^-(x_{k+1} - x_k) = \int_a^b S^-(x)dx \leq \int_a^b f(x)dx \leq \int_a^b S^+(x)dx = \sum_{k=1}^{n} S_k^+(x_{k+1} - x_k) \equiv A^+.$$

We have "sandwiched" the integral of $f(x)$ between the integrals for the two staircase functions which we are able to evaluate as summations. The idea now is to use the sandwich theorem. If we can take a limit so that $S^+(x)$ and $S^-(x)$ both approach $f(x)$ then if the limits of their integrals exist and are equal then that limiting value is the value of the integral of $f(x)$. The real question is what is the limit that we should take? By now we should expect to take a limit such that the width of the steps in $S^{\pm}(x)$ become infinitesimal. But of course if we simply make the step widths thinner without increasing the number of steps then the staircase functions will no longer cover the domain of the function. So we must do two things at once when we take the limit:

- the width of each step must become infinitesimal i.e. $(x_{k+1} - x_k) \to 0$ and
- the number of steps must approach infinity i.e. $n \to \infty$.

How do we know that the number of steps must grow infinitely large? Well we are interested in the integral over $x \in [a, b]$, hence the width of the staircase functions in total must remain $b - a$ as we take the limit, i.e. we need $b - a = \sum_{k=1}^{n+1}(x_{k+1} - x_k)$ and so as $(x_{k+1} - x_k)$ shrinks, so the number of steps n must increase as $b - a$ is constant.

You might think of this limit as moving through the space of staircase functions $S^{\pm}(x)$ to arrive at the pair of staircase functions whose steps are infinitesimally thin. For example pictorially for $S^-(x)$ the process of taking this limit and moving through the staircase functions is shown in figure 5.8. If we now take this (abstract) limit on the inequality which sandwiches

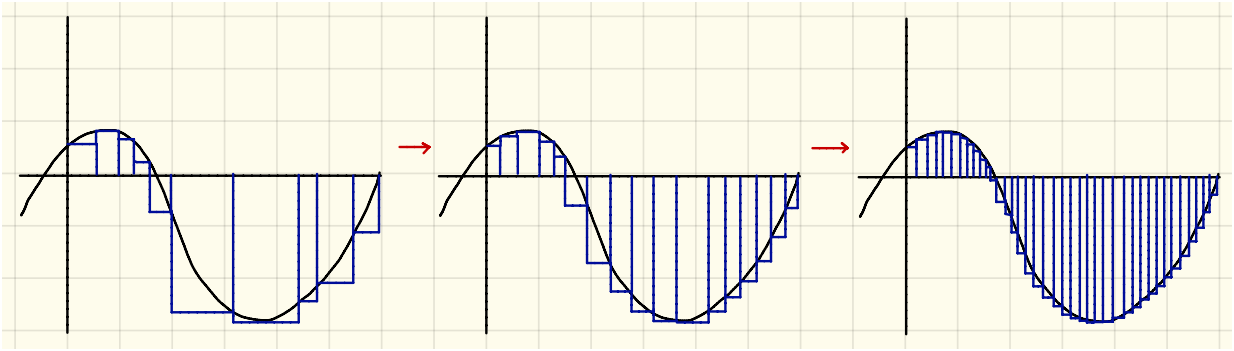


Figure 5.8: The width of each step in the staircase function $S^-(x)$ is reduced and the number of steps is increased, so that the integral of $S^-(x)$ approaches from below that of $f(x)$.

$\int_a^b f(x)dx$ we have:

$$\lim_{(x_{k+1}-x_k)\to 0, n\to\infty}[A^-] \leq \int_a^b f(x)dx \leq \lim_{(x_{k+1}-x_k)\to 0, n\to\infty}[A^+].$$

Now if the two limits tend to the same value, i.e.

$$\lim_{(x_{k+1}-x_k)\to 0, n\to\infty}[A^-] = A \qquad \text{and} \qquad \lim_{(x_{k+1}-x_k)\to 0, n\to\infty}[A^+] = A$$

then we conclude that $\int_a^b f(x)dx = A$. This is the process by which we can construct and evaluate the Riemann integral of a function, if it exists, but how hard is it to apply this construction in practise?

Example 5.1. *Let*

$$\int_0^b \cos(x)dx = A \qquad \text{for } b \leq \pi.$$

Construct the staircase functions and take the appropriate limit to evaluate the Riemann integral and find A.

For $x \in [0, b]$, the cosine function decreases monotonically (i.e. if $x, x' \in [0, b]$ and $x' > x$ then $\cos(x') < \cos(x)$). Now we may construct the staircase functions $S^{\pm}(x)$ via

$$S_k^- = \min[\cos(x)] \qquad \text{for } x \in [x_k, x_{k+1}) \implies S_k^- = \cos(x_{k+1})$$
$$S_k^+ = \max[\cos(x)] \qquad \text{for } x \in [x_k, x_{k+1}) \implies S_k^+ = \cos(x_k).$$

Now, for convenience, we may choose steps of equal width, w, with $x_1 = 0$ and $x_{n+1} = b$ so that as

$$b = x_{n+1} = nw \qquad \text{then} \qquad w = \frac{b}{n}.$$

We can also find expressions for each of the x-coordinates of the sides of the steps, x_k,

$$x_k = x_1 + (k-1)w = (k-1)w = \frac{b(k-1)}{n}$$

i.e. $x_1 = 0$, $x_2 = w$, $x_3 = 2w, \ldots$, $x_n = (n-1)w$ *and* $x_{n+1} = nw = b$. *Hence in this notation we have,*

$$A^- = \sum_{k=1}^{n} S_k^-(x_{k+1} - x_k) = \sum_{k=1}^{n} w\cos(x_{k+1}) = w\sum_{k=1}^{n}\cos(kw).$$

The limit we wish to take involves simultaneously taking $n \to \infty$ and $w \to 0$, which will be problematic unless we can rewrite the summation, in order to move n from the summation and into an expression where we can take the limit.

Exercise 5.1. *Show that*

$$\sum_{k=0}^{n}\cos(k\theta) = \frac{1 - \cos((n+1)\theta) - \cos\theta + \cos(n\theta)}{2 - 2\cos\theta}.$$

[Hint: First prove that $\sum_{k=0}^{n} z^k = \frac{1-z^{n+1}}{1-z}$ then substitute $z = e^{i\theta} = \cos(\theta) + i\sin(\theta)$ into this result.]

Hence we have,

$$\begin{aligned}
A^- &= w\sum_{k=1}^{n}\cos(kw) \\
&= w(\sum_{k=0}^{n}\cos(kw) - 1) \\
&= w\left(\frac{1 - \cos((n+1)w) - \cos(w) + \cos(nw)}{2 - 2\cos(w)} - 1\right) \\
&= w\left(\frac{-1 - \cos((n+1)w) + \cos(w) + \cos(nw)}{2 - 2\cos(w)}\right) \\
&= w\left(\frac{-1 - \cos(b+w) + \cos(w) + \cos(b)}{2 - 2\cos(w)}\right) \\
&= w\left(\frac{-1 - \cos(b)\cos(w) + \sin(b)\sin(w) + \cos(w) + \cos(b)}{2 - 2\cos(w)}\right)
\end{aligned}$$

where we have used $b = nw$ to eliminate n from the expression. Hence we may now attempt to

take the limit $w \to 0$*:*

$$\begin{aligned}
\lim_{w\to 0}[A^-] &= \lim_{w\to 0}\left[w\left(\frac{-1-\cos(b)\cos(w)+\sin(b)\sin(w)+\cos(w)+\cos(b)}{2-2\cos(w)}\right)\right] \\
&= \frac{1}{2}\lim_{w\to 0}\left[w\left(\frac{(-1+\cos(b))(1-\cos(w))+\sin(b)\sin(w)}{1-\cos(w)}\right)\right] \\
&= \frac{1}{2}\lim_{w\to 0}\left[w(-1+\cos(b))\right] + \frac{\sin(b)}{2}\lim_{w\to 0}\left[\frac{w\sin(w)}{1-\cos(w)}\right] \\
&= \frac{\sin(b)}{2}\lim_{w\to 0}\left[\frac{w(w-\mathcal{O}(w^3))}{1-(1-\frac{w^2}{2!}+\mathcal{O}(w^4))}\right] \\
&= \frac{\sin(b)}{2}\lim_{w\to 0}\left[\frac{1-\mathcal{O}(w))}{\frac{1}{2!}-\mathcal{O}(w^2))}\right] \\
&= \sin(b).
\end{aligned}$$

For A^+ *we have*

$$A^+ = \sum_{k=1}^{n} S_k^+(x_{k+1}-x_k) = \sum_{k=1}^{n} w\cos(x_k) = w\sum_{k=1}^{n}\cos((k-1)w).$$

Hence

$$\begin{aligned}
A^+ &= w\sum_{k=1}^{n}\cos((k-1)w) \\
&= w\sum_{\ell=0}^{n-1}\cos(\ell w) \\
&= w(1+\sum_{\ell=1}^{n}\cos(\ell w)-\cos(nw)) \\
&= w(1-\cos(b))+A^-
\end{aligned}$$

where we used $\ell \equiv k-1$ *in writing the second line above. Hence we have,*

$$\lim_{w\to 0}\left[A^+\right] = \lim_{w\to 0}\left[w(1-\cos(b))\right] + \lim_{w\to 0}\left[A^-\right] = \lim_{w\to 0}\left[A^-\right] = \sin(b).$$

Therefore as we have

$$\sin(b) = \lim_{w\to 0}\left[A^-\right] \le \int_0^b \cos(x)dx \le \lim_{w\to 0}\left[A^+\right] = \sin(b)$$

then

$$\int_0^b \cos(x)dx = \sin(b).$$

Exercise 5.2. *The construction of the staircase functions in the example above rested on the observation that* $\cos(x)$ *is monotonically decreasing for* $x \in [0, b]$ *where* $0 \leq b \leq \pi$*. Outline the procedure for showing that*

$$\int_0^b \cos(x)dx = \sin(b)$$

for $b > \pi$*.*

Exercise 5.3. *Construct the staircase functions one would use to evaluate*

$$\int_0^b \sin(x)dx$$

first for $0 \leq b \leq \pi$ *and more generally when* $b > 0$*.*

Example 5.2. *Use staircase functions and the sandwich theorem to evaluate*

$$A = \int_a^b x^m dx$$

where $0 < a < b$ *and* $m \in \mathbb{Z}^+$*.*

We choose the bounding staircase functions as follows:

$$\begin{aligned} S_k^- &= \min[x^m] \qquad \text{for } x \in [x_k, x_{k+1}) \implies S_i^- = x_k^m \\ S_k^+ &= \max[x^m] \qquad \text{for } x \in [x_k, x_{k+1}) \implies S_k^+ = x_{k+1}^m. \end{aligned}$$

In this example we will follow a seemingly more flamboyant path to evaluate the integral, namely we will consider steps of varying size which grow as a geometric progression in w*. Given that* $x_1 = a$ *and* $x_{n+1} = b$ *then we define the step sides to lie at*

$$x_1 = a,\ x_2 = aw\ x_3 = aw^2,\ \ldots x_k = aw^{k-1},\ \ldots x_{n+1} = aw^n = b.$$

Therefore we have

$$w = \left(\frac{b}{a}\right)^{\frac{1}{n}} \qquad \implies \qquad x_k = a\left(\frac{b}{a}\right)^{\frac{k-1}{n}}.$$

Let us make a few mathematical comments about this choice and a few consequences that will prove useful.

(i) $b > a$ *then* $w > 1$*,*

(ii) $\ln(b/a) = \ln(w^n) = n\ln(w)$ *therefore* $n = \frac{\ln(b)-\ln(a)}{\ln(w)}$*,*

(iii) $x_{k+1} - x_k = aw^k - aw^{k-1} = aw^{k-1}(w-1)$ *is the k'th step's width,*

(iv) the largest step size is the final one which has width $x_{n+1} - x_n = aw^{n-1}(w-1) = b(1-\frac{1}{w})$, *and*

(v) the limit in which the largest step's width is reduced to zero is the limit when $w \to 1^+$ *(the limit is taken from above as by point (i)* $w > 1$*).*

An upper bound on the integral A is given by

$$\begin{aligned} A^+ &= \sum_{k=1}^{n} S_k^+(x_{k+1} - x_k) \\ &= \sum_{k=1}^{n} x_{k+1}^m(x_{k+1} - x_k) \\ &= \sum_{k=1}^{n} (aw^k)^m(aw^{k-1}(w-1)) \\ &= a^{m+1}\frac{(w-1)}{w}\sum_{k=1}^{n} w^{(m+1)k} \\ &= a^{m+1}\frac{(w-1)}{w}\sum_{k=1}^{n} (w^{m+1})^k \end{aligned}$$

We can rewrite the geometric sum to find

$$\begin{aligned} A^+ &= a^{m+1}\frac{(w-1)}{w}\left(\frac{w^{m+1}((w^{m+1})^n - 1)}{w^{m+1} - 1}\right) \\ &= a^{m+1}\frac{(w-1)w^m((w^n)^{m+1} - 1)}{w^{m+1} - 1}. \end{aligned}$$

We now eliminate n using $w^n = \frac{b}{a}$,

$$\begin{aligned} A^+ &= a^{m+1}w^m(w-1)\frac{((b/a)^{m+1} - 1)}{w^{m+1} - 1} \\ &= a^{m+1}w^m(w-1)\frac{((b/a)^{m+1} - 1)}{w^{m+1} - 1}. \end{aligned}$$

Turning to the lower bound and noting that

$$S_k^- = x_k^m = (aw^{k-1})^m = (aw^k)^m w^{-m} = S_k^+ w^{-m}$$

we see that the lower bound on A is related to the upper bound by

$$A^- = \sum_{k=1}^{n} S_k^-(x_{k+1} - x_k) = \sum_{k=1}^{n} S_k^+ w^{-m}(x_{k+1} - x_k) = w^{-m}A^+.$$

We have,

$$A^- = w^{-m}A^+ \leq \int_a^b x^m dx \leq A^+ = a^{m+1}w^m(w-1)\frac{((b/a)^{m+1}-1)}{w^{m+1}-1}.$$

This looks simultaneously promising and horrendous: the good news is that in the limit $w \to 1$ it is clear that we will have $\lim_{w\to 1^+}[A^+] \leq A \leq \lim_{w\to 1^+}[A^+]$ so that A will be sandwiched by the limit if it exists, the bad news is that we will need to evaluate the $\lim_{w\to 1^+}[A^+]$. So,

$$\begin{aligned}
\lim_{w\to 1^+}[A^+] &= \lim_{w\to 1^+}\left[a^{m+1}w^m(w-1)\frac{((b/a)^{m+1}-1)}{w^{m+1}-1}\right] \\
&= \lim_{w\to 1^+}\left[(b^{m+1}-a^{m+1})\frac{w^m(w-1)}{w^{m+1}-1}\right] \\
&= (b^{m+1}-a^{m+1})\lim_{w\to 1^+}\left[w^m\right]\lim_{w\to 1^+}\left[\frac{(w-1)}{w^{m+1}-1}\right] \\
&= (a^{m+1}-b^{m+1})\lim_{w\to 1^+}\left[\frac{1-w}{w^{m+1}-1}\right] \\
&= -(a^{m+1}-b^{m+1})\lim_{w\to 1^+}\left[\left(\frac{1-w^{m+1}}{1-w}\right)^{-1}\right] \\
&= -(a^{m+1}-b^{m+1})\left(\lim_{w\to 1^+}\left[\frac{1-w^{m+1}}{1-w}\right]\right)^{-1} \\
&= -(a^{m+1}-b^{m+1})\left(\lim_{w\to 1^+}\left[\sum_{j=0}^{m} w^j\right]\right)^{-1} \\
&= -(a^{m+1}-b^{m+1})(m+1)^{-1} \\
&= \frac{b^{m+1}-a^{m+1}}{m+1}.
\end{aligned}$$

We also have (trivially),

$$\lim_{w\to 1^+}\left[A^-\right] = \lim_{w\to 1^+}\left[w^{-m}A^+\right] = \lim_{w\to 1^+}\left[w^{-m}\right]\lim_{w\to 1^+}\left[A^+\right] = \lim_{w\to 1^+}\left[A^+\right].$$

Consequently as $\lim_{w\to 1^+}\left[A^-\right] = \lim_{w\to 1^+}\left[A^+\right]$ then we conclude that

$$\int_a^b x^m dx = \frac{b^{m+1}-a^{m+1}}{m+1}.$$

It is worth noting a small point about generalising this result. If we had considered $x < 0$ then the defining relations of S_k^- and S_k^+ would have depended crucially on whether n is odd or even, e.g. for $x < 0$ and odd n then

$$\begin{aligned}
S_k^- &= \min[x^n] \qquad \text{for } x \in [x_k, x_{k+1}) \implies S_i^- = x_k^n \\
S_k^+ &= \max[x^n] \qquad \text{for } x \in [x_k, x_{k+1}) \implies S_k^+ = x_{k+1}^n.
\end{aligned}$$

while for $x < 0$ and even n we would have defined

$$S_k^- = \min[x^n] \qquad \text{for } x \in [x_k, x_{k+1}) \implies S_i^- = x_{k+1}^n$$
$$S_k^+ = \max[x^n] \qquad \text{for } x \in [x_k, x_{k+1}) \implies S_k^+ = x_k^n.$$

But notice that in the long run this fact would not cause us much concern as the change of n from even to odd amounts to interchanging the definition of S_k^+ and S_k^-. Hence if the integral exists and the pair of staircase functions limit to the same sum, then it will not change the result if the bounding functions are swapped depending upon even or odd n.

Of course it is very satisfying that meaning can be given to an integral and that we can construct the Riemann integral from first principles in this way. However it is a tedious and lengthy computation and it will come as a relief to learn that we do not need to repeat this process for each and every function we have met in the course so far. Instead we can rely on the fundamental theorem of calculus to give us a quick way to integrate functions.

5.2 The Fundamental Theorem of Calculus

The fundamental theorem of calculus was understood in various forms by John Gregory (1638-1675) and Isaac Barrow (1630-1677) before Isaac Newton and Gottfried Leibnitz formalised the calculus. The idea was understood long before Riemann put the integral on a rigorous foundation. The basic idea is one that you will already be familiar with, namely, that the integral is the antiderivative, i.e. that the integral of the derivative of a function gives back the function, the integral "undoes" the derivative. This is quite incredible and we are a little anaesthetised to how remarkable an idea it is because we are very familiar with the idea. Let us comment on why it is a surprising idea.

Differentiation and integration are the major operations within calculus and if we did not have a practical experience of working with these operations we would not think that they would be closely related. After all, differentiation gives the *local* value of the gradient of a function at a point and is determined using infinitesimal data while integration involves finding the area under a function as it runs the finite or even infinite distance between the limits of the integration and it is a *global* measure related to the curve (i.e. it requires more than local data from one point). Infinitesimal vs. finite (or infinite), local vs. global: it seems unlikely that the two operations will be related but yet, incredibly, integration and differentiation are inverse operations - this fact is called the fundamental theorem of calculus. The theorem is stated in two parts called the first and second fundamental theorems of calculus.

Theorem 5.1. *(The first fundamental theorem of calculus.) Let $f : [a, b] \to \mathbb{R}$ be a continuous function and let $F : [a, b] \to \mathbb{R}$ be defined by*

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

Then, F is a continuous function on $[a, b]$, differentiable on the open interval (a, b) and

$$\frac{dF}{dx} = f(x).$$

Theorem 5.2. *(The second fundamental theorem of calculus.) Let $f : [a, b] \to \mathbb{R}$ be a continuous function and let $F(x)$ be defined by*

$$\frac{dF}{dx} = f(x) \qquad \forall\, x \in [a, b].$$

Then, if $f(x)$ is integrable on $[a, b]$,

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

Comment(s). *(On the fundamental theorem of calculus...)*

1. *The function $F(x)$ whose derivative gives $f(x)$, i.e. $\frac{dF}{dx} = f(x)$, is called an antiderivative or a primitive of $f(x)$. For example if $f(x) = x^2$ then the antiderivative (think: integral) is $F(x) = x^3/3$ as $\frac{d}{dx}(x^3/3) = x^2 = f(x)$.*

2. *The first fundamental theorem states that the integral of a function (from a to x) is the antiderivative $F(x)$. At this stage it should be clear that*

$$\int_a^x f(t)dt$$

   is a function of x, but it should not yet be clear that this function will be the antiderivative and the appearance of a as a limit of the integral is not yet explained.

3. *The second fundamental theorem tells us that if we know the antiderivative of a function, then we can use this to rapidly evaluate the integral (rather than resort to the time-consuming construction of staircase functions and the use of the sandwich theorem, as we have seen in the previous section).*

We will now sketch some proofs of these theorems.

Proof: (the first fundamental theorem of calculus) In this proof we will limit ourselves to the case where $f(x)$ is a differentiable function. Now,

$$\begin{aligned}\frac{dF}{dx} &= \lim_{h\to 0}\left[\frac{F(x+h)-F(x)}{h}\right] \\ &= \lim_{h\to 0}\left[\frac{1}{h}\left(\int_a^{x+h} f(t)dt - \int_a^x f(t)dt\right)\right] \\ &= \lim_{h\to 0}\left[\frac{1}{h}\int_x^{x+h} f(t)dt\right].\end{aligned}$$

Next we will introduce an upper and lower bound on $f(t)$ where $t \in [x, x+h]$ and make use of the sandwich theorem to in order to take the limit above. We will make use of

$$C \equiv \max_{t\in[x,x+h]}\left(\left|\frac{df}{dt}\right|\right)$$

this is the maximum of the modulus of the derivative in the range $[x, x+h]$. Now we don't need to evaluate this number, it will prove very useful in evaluating the limit, but will vanish when we take the limit. As C is the maximum rate of increase (or decrease) of the function $f(t)$ on the interval $[x, x+h]$ we know that

$$f(x) - Ch \le f(t) \le f(x) + Ch \qquad \forall\, t \in [x, x+h].$$

Where did this inequality come from, the sketch in figure 5.9 will help us to understand it. In figure 5.9 we have marked in the straight lines with gradient $\pm C$ which pass through $f(x)$. As, by the definition of C, $\frac{df}{dt} \le C$ and $\frac{df}{dt} \ge -C$ for all $t \in [x, x+h]$, then the straight lines with the gradients $\pm C$ always are greater/less than or equal to $f(t)$ for $t \in [x, x+h]$. The maximum/minimum point on these lines lie at $t = x+h$ and hence we have the inequality (which we repeat here):

$$f(x) - Ch \le f(t) \le f(x) + Ch \qquad \forall\, t \in [x, x+h].$$

Returning to our proof, we now have

$$\lim_{h\to 0}\left[\frac{1}{h}\int_x^{x+h}(f(x)-Ch)dt\right] \le \lim_{h\to 0}\left[\frac{1}{h}\int_x^{x+h} f(t)dt\right] \le \lim_{h\to 0}\left[\frac{1}{h}\int_x^{x+h}(f(x)+Ch)dt\right].$$

To take the limit is a simple matter as

$$\lim_{h\to 0}\left[\frac{1}{h}\int_x^{x+h}(f(x)+Ch)dt\right] = \lim_{h\to 0}\left[\frac{1}{h}(f(x)+Ch)h\right] = \lim_{h\to 0}\left[f(x)+Ch\right] = f(x)$$

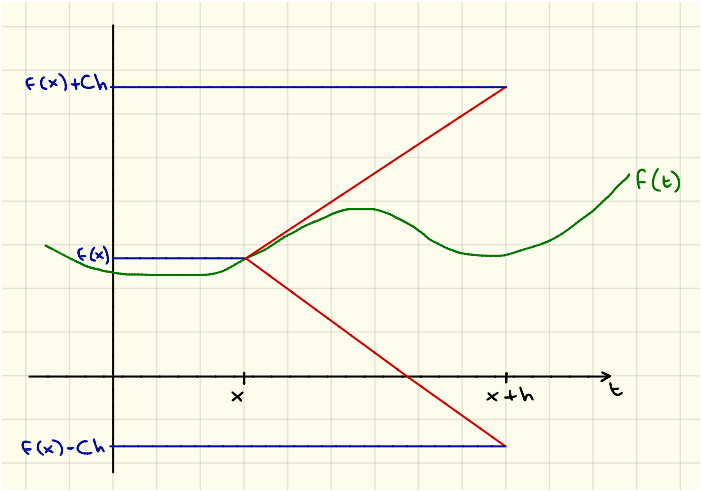


Figure 5.9: An arbitrary differentiable function $f(t)$ and we are interested in its values when $t \in [x, x+h]$. C is the absolute value of the maximum gradient of $f(x)$ and the straight lines marked in are those with gradient $\pm C$ passing through $f(x)$.

as $f(x) + Ch$ are constants in the integral with respect to t and $\int_x^{x+h} dt = h$ (we can evaluate the integral of the constant function: in this case it evaluates the area of a rectangle of width h and height 1). We can also evaluate the lower bound in the limit to find:

$$\lim_{h\to 0}\left[\frac{1}{h}\int_x^{x+h}(f(x)-Ch)dt\right] = \lim_{h\to 0}\left[\frac{1}{h}(f(x)-Ch)h\right] = \lim_{h\to 0}\left[f(x)-Ch\right] = f(x).$$

Hence, by the sandwich theorem, we have:

$$\frac{dF}{dx} = f(x)$$

as required.

Proof: (the second fundamental theorem of calculus) If $f(x) = \frac{dF}{dx}$ is integrable on $[a, b]$ then we can construct the Riemann integral as the limit of a sum of n rectangular areas of width $w_i \equiv x_{i+1} - x_i$ such that $x_0 = a$ and $x_{n+1} = b$. That is we know that,

$$\begin{aligned}\int_a^b f(x)dx &= \lim_{w_i\to 0, n\to\infty}\left[\sum_{i=1}^{n} f(x)\Big|_{x\in[x_i,x_{i+1}]}(x_{i+1}-x_i)\right]\\ &= \lim_{w_i\to 0, n\to\infty}\left[\sum_{i=1}^{n} \frac{dF}{dx}\Big|_{x\in[x_i,x_{i+1}]}(x_{i+1}-x_i)\right]\end{aligned}$$

Now $\frac{dF}{dx}\Big|_{x\in[x_i,x_{i+1}]}$ is some value of the function $\frac{dF}{dx}$ evaluated on the interval $[x_i, x_{i+1}]$ and by

the mean value theorem we know one special value of $\frac{dF}{dx}$ on the interval, namely,

$$\frac{F(x_{i+1}) - F(x_i)}{x_{i+1} - x_i} = \frac{dF}{dx}(c_i) \qquad \text{for some } c_i \in [x_i, x_{i+1}]$$

and we may use this value for the derivative of $F(x)$ in the interval $[x_i, x_{i+1}]$ in the Riemann sum above. We have:

$$\begin{aligned}
\int_a^b f(x)dx &= \lim_{w_i \to 0, n \to \infty} \left[\sum_{i=1}^{n} \frac{F(x_{i+1}) - F(x_i)}{(x_{i+1} - x_i)} (x_{i+1} - x_i) \right] \\
&= \lim_{w_i \to 0, n \to \infty} \left[\sum_{i=1}^{n} (F(x_{i+1}) - F(x_i)) \right] \\
&= \lim_{w_i \to 0, n \to \infty} \Big[(F(x_1) - F(x_0)) + (F(x_2) - F(x_1) + \ldots + (F(x_{n+1}) - F(x_n)) \Big] \\
&= \lim_{w_i \to 0, n \to \infty} \Big[F(x_{n+1}) - F(x_0) \Big] \\
&= \lim_{w_i \to 0, n \to \infty} \Big[F(b) - F(a) \Big] \\
&= F(b) - F(a)
\end{aligned}$$

as required.

Comment(s). *(On another proof of the second fundamental theorem of calculus...) If we knew that $f(x)$ had an antiderivative (and not just that it is integrable[2]) we could have proved the second fundamental theorem by using the first fundamental theorem. As $f(x)$ is a continuous function on $[a, b]$ then the integral $\int_a^x f(t)dt$ exists for $x \in [a, b]$. Let us give this useful function a name*

$$G(x) \equiv \int_a^x f(t)dt$$

and by the first fundamental theorem we have

$$\frac{dG}{dx} = f(x) \qquad \forall\, x \in [a, b].$$

Notice that $G(a) = 0$ and $G(b) = \int_a^b f(t)dt$, i.e. $G(b)$ is the integral of interest in the statement of the second fundamental theorem of calculus. Now consider a second antiderivative of $f(x)$, denoted $F(x)$, such that

$$\frac{dF}{dx} = f(x) \qquad \forall\, x \in [a, b].$$

[2]There exist functions which are integrable, but which do not have an antiderivative e.g. Thomae's function ($f(x) = 0$ if $x \in \mathbb{R} \setminus \mathbb{Q}$ and $f(x) = 1/q$ if $x = p/q \in \mathbb{Q}$ and p and q have no common factors), and there also exist functions who have an antiderivative but which are not Riemann integrable e.g. Volterra's function, see `https://en.wikipedia.org/wiki/Volterra%27s_function`.

In this proof we aim to show that $G(x) = F(x) - F(a)$, so let us introduce a third function, the difference between $F(x)$ and $G(x)$:

$$H(x) \equiv F(x) - G(x)$$

then

$$\frac{dH}{dx} = f(x) - f(x) = 0 \qquad \forall\, x \in [a, b].$$

So we have learnt that $H(x)$ has a gradient of zero for $x \in [a, b]$, therefore $H(x) = H(a)$ which implies that

$$F(x) - G(x) = F(a) - G(a) \qquad \forall\, x \in [a, b].$$

Now, as $G(a) = 0$ then we have $G(x) = F(x) - F(a)$, i.e.

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

as required.

5.2.1 Indefinite and Definite Integrals.

Definition 5.2.1. *The definite integral $A \equiv \int_a^b f(x)dx$ has specific limits $x = \{a, b\}$.*

Comment(s). *(On definite integrals...)*

1. *If $f(x)$ is integrable, then, by the (second) fundamental theorem of calculus:*

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

   where $\frac{dF}{dx} = f(x)$ and A is a number.

2. *So far we have only defined $\int_a^b f(x)dx$ for $a \leq b$ but we can generalise this to include $b \leq a$ via the (second) fundamental theorem of calculus:*

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

Definition 5.2.2. *The indefinite integral $\int f(x)dx$ is a function: it has no specified limits.*

Comment(s). *(On indefinite integrals...) The indefinite integral at first sight is an abuse of notation: the integral is defined (up to a sign) as an area via its limits, but the indefinite integral*

has no limits - so what does it mean? It is another notation for the antiderivatives or primitives of $f(x)$. Recall that by the (first) fundamental theorem of calculus that the antiderivative is

$$F(x) = \int_0^x f(t)dt = \int_0^x \frac{dF}{dt}dt = F(x) - F(0)$$

and note that $F(0) = 0$, hence

$$F(x) - F(a) = \int_a^x f(t)dt$$

so that $\frac{d}{dx}(F(x) - F(a)) = \frac{dF}{dx} = f(x)$, in other words there are multiple antiderivatives (antiderivatives are specified only up to a constant). Now it is useful to think of the class of antiderivative functions for $f(x)$ and while one could write $F(x) - F(a)$ to denote the antiderivatives it is useful to have a briefer notation, hence the notation for the indefinite integral, instead of writing $\int_a^x f(t)dt$ we write $\int f(x)dx$ to denote the class of antiderivative functions and emphasise that it is a function of x, hence the indefinite integral is

$$\int f(x)dx \equiv F(x) + K \qquad K \in \mathbb{R}$$

where K is called the constant of integration and $F(x)$ is an antiderivative of $f(x)$.

Using the fundamental theorem(s) of calculus we can now try to solve

$$\int_a^b f(x)dx$$

by identifying the antiderivative $F(x)$, i.e. by solving $\frac{dF}{dx} = f(x)$.

5.3 Properties of the Integral and some Techniques for Integration

Our knowledge of the derivatives of standard functions will prove invaluable when trying to identify a primitive, but it will also prove necessary to understand the basic properties of the integral, so that we may manipulate it into a form where we may identify the antiderivative or primitive.

The integral is a linear operation, i.e. $\int(af(x) + bg(x))dx = a\int f(x)dx + b\int g(x)dx$ where $a, b \in R$ are constants with respect to x (i.e. they do not depend on the integration variable x). We can prove the linearity of the integral by using our knowledge of derivatives and the fundamental theorem of calculus. In the same manner we can also show how we can transform

an integral by a substitution or by using the technique known as integration by parts. Both these methods will open up new ways to find primitives or antiderivatives and evaluate definite integrals. Let $F(x) \equiv \int f(x)dx$ and $G(x) \equiv \int g(x)dx$ then:

(i) Multiplication by a constant:

$$\int (af(x))dx = a \int f(x)dx$$

where a is a constant.

Proof:

$$\frac{d}{dx}(aF(x)) = a\frac{dF}{dx} = af(x).$$

Hence,

$$\int \frac{d}{dx}(aF(x))dx = aF(x) + K = a \int f(x)dx.$$

(ii) Integrals of sums of functions are sums of integrals:

$$\int (f(x) + g(x))dx = F(x) + G(x).$$

Proof:

$$\frac{d}{dx}(F(x) + G(x)) = \frac{dF}{dx} + \frac{dG}{dx} = f(x) + g(x).$$

By points (i) and (ii), integration is a linear operation.

(iii) Transforming an integral with respect to x into an integral with respect to t by substituting $x = x(t)$ to find:

$$\int_{x_1}^{x_2} f(x)dx = \int_{t_1}^{t_2} f(x(t))\frac{dx}{dt}dt$$

where $x(t_1) = x_1$ and $x(t_2) = x_2$.

Proof: We know that

$$\int_{x_1}^{x_2} f(x)dx = \int_{x_1}^{x_2} \frac{dF}{dx}dx = F(x_2) - F(x_1),$$

and if $x_1 = x(t_1)$ and $x_2 = x(t_2)$ then we have

$$\begin{aligned} F(x_2) - F(x_1) &= F(x(t_2)) - F(x(t_1)) \\ &= \Big[F(x(t))\Big]_{t_1}^{t_2} \\ &= \int_{t_1}^{t_2} \frac{d}{dt}(F(x(t)))dt \\ &= \int_{t_1}^{t_2} \frac{dF}{dx}\frac{dx}{dt}dt \end{aligned}$$

where we have used the chain rule $\frac{dF(x(t))}{dt} = \frac{dF}{dx}\frac{dx}{dt}$.

(iv) Integration by parts:

$$\int \frac{df}{dx} g(x)dx = f(x)g(x) - \int f(x)\frac{dg}{dx}dx + K.$$

Proof:

$$\frac{d}{dx}(f(x)g(x)) = \frac{df}{dx}g + f\frac{dg}{dx},$$

hence,

$$\int \frac{d}{dx}(f(x)g(x))dx = f(x)g(x) + K = \int \frac{df}{dx}g(x)dx + \int f(x)\frac{dg}{dx}dx$$

as required.

We will now practise these techniques to evaluate definite integrals or identify the antiderivative for an indefinite integral with multiple examples.

5.3.1 Solving Integrals by Substitution

Example 5.3. *Express the indefinite integral*

$$\int \sin^m(x)\cos(x)dx$$

as a function of x by substituting $sin(x) = t$.

Here $x(t) = \arcsin(t)$, but it is simpler to find $\frac{dx}{dt}$ by acting with $\frac{d}{dt}$ on $sin(x) = t$ to obtain $\frac{d}{dt}(\sin(x)) = \cos(x)\frac{dx}{dt} = 1$, hence,

$$\begin{aligned}
\int \sin^m(x)\cos(x)dx &= \int \sin^m(x(t))\cos(x(t))\frac{dx}{dt}dt \\
&= \int t^m \cos(x(t))\frac{1}{\cos(x(t))}dt \\
&= \int t^m dt \\
&= \frac{t^{m+1}}{m+1} + K \\
&= \frac{\sin^{m+1}(x)}{m+1} + K
\end{aligned}$$

It is rather lengthy to present the substitution as above, it is perfectly acceptable instead to observe that $\sin(x) = t \implies \cos(x)dx = dt$.

Example 5.4. *Express the indefinite integral*

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

as a function of x by using an appropriate substitution.

There are many options, but we are guided to simplify the terms in the integral and so choose the substitution

$$x^2 = t$$

hence $2xdx = dt$. Therefore we have upon substituting into the integral

$$\begin{aligned}\int e^{x^2} x dx &= \int e^t \frac{1}{2} dt \\ &= \frac{1}{2} e^t + K \\ &= \frac{1}{2} e^{x^2} + K.\end{aligned}$$

Example 5.5. *Express the indefinite integral*

$$\int (1 - x^2)^{-\frac{1}{2}} dx$$

as a function of x by using an appropriate substitution.

Let

$$x = \sin\theta$$

hence $dx = \cos\theta d\theta$. Our motivation for choosing this substitution is the trigonometric identity $1 - \sin^2\theta = \cos^2\theta$, i.e. $(1 - x^2)^{-\frac{1}{2}} = 1/\cos\theta$, hence we have

$$\begin{aligned}\int (1 - x^2)^{-\frac{1}{2}} dx &= \int \frac{1}{\cos\theta} (\cos\theta d\theta) \\ &= \int d\theta \\ &= \theta + K \\ &= \arcsin(x) + K.\end{aligned}$$

Of course we may have recognised immediately the antiderivative by recalling that $\frac{d}{dx}(\arcsin(x)) = \frac{1}{\sqrt{1-x^2}}$.

Example 5.6. *Express the indefinite integral*

$$\int (1+x^2)^{-\frac{1}{2}} dx$$

as a function of x by using an appropriate substitution.

Let

$$x = \sinh\theta$$

hence $dx = \cosh\theta d\theta$. Hence we have

$$\begin{aligned}\int (1+x^2)^{-\frac{1}{2}} dx &= \int (1+\sinh^2\theta)^{-\frac{1}{2}} (\cosh\theta) d\theta \\ &= \int d\theta \\ &= \theta + K \\ &= \operatorname{arcsinh}(x) + K.\end{aligned}$$

Where we have used the identity $1+\sinh^2\theta = \cosh^2\theta$.

Example 5.7. *Express the indefinite integral*

$$\int (1+x^2)^{-1} dx$$

as a function of x by using an appropriate substitution.

Let

$$x = \tan\theta$$

hence $dx = (1+\tan^2\theta)d\theta$. Hence we have

$$\begin{aligned}\int (1+x^2)^{-1} dx &= \int (1+\tan^2\theta)^{-1}(1+\tan^2\theta) d\theta \\ &= \int d\theta \\ &= \theta + K \\ &= \arctan(x) + K.\end{aligned}$$

Example 5.8. *Express the indefinite integral*

$$\int (1-x^2)^{-1} dx$$

as a function of x by using an appropriate substitution.

Let

$$x = \tanh\theta$$

hence $dx = (1 - \tanh^2\theta)d\theta$. *Hence we have*

$$\begin{aligned}\int (1-x^2)^{-1}dx &= \int (1-\tanh^2\theta)^{-1}(1-\tanh^2\theta)d\theta \\ &= \int d\theta \\ &= \theta + K \\ &= \operatorname{arctanh}(x) + K.\end{aligned}$$

Example 5.9. *Evaluate the definite integral*

$$\int_2^3 (x^2+2x)^{-\frac{1}{2}}dx$$

by using an appropriate substitution.

First note that

$$\int_2^3 (x^2+2x)^{-\frac{1}{2}}dx = \int_2^3 ((x+1)^2-1)^{-\frac{1}{2}}dx.$$

Now the integral suggests the substitution

$$x + 1 = \cosh\theta$$

so that $dx = \sinh\theta d\theta$. *Hence we have*

$$\begin{aligned}\int_2^3 ((x+1)^2-1)^{-\frac{1}{2}}dx &= \int_{\operatorname{arccosh}(3)}^{\operatorname{arccosh}(4)} (\cosh^2\theta - 1)^{-\frac{1}{2}}\sinh\theta d\theta \\ &= \int_{\operatorname{arccosh}(3)}^{\operatorname{arccosh}(4)} d\theta \\ &= \Big[\theta\Big]_{\operatorname{arccosh}(3)}^{\operatorname{arccosh}(4)} \\ &= \operatorname{arccosh}(4) - \operatorname{arccosh}(3)\end{aligned}$$

Example 5.10. *Evaluate the definite integral*

$$\int_{-\frac{3}{2}}^{-\frac{1}{2}} (-x^2 - 2x)^{-\frac{1}{2}} dx$$

by using an appropriate substitution.

First note that

$$\int_{-\frac{3}{2}}^{-\frac{1}{2}} (-x^2 - 2x)^{-\frac{1}{2}} dx = \int_{-\frac{3}{2}}^{-\frac{1}{2}} (-(x+1)^2 + 1)^{-\frac{1}{2}} dx.$$

Now the integral suggests the substitution

$$x + 1 = \sin\theta$$

so that $dx = \cos\theta d\theta$. *Hence we have*

$$\begin{aligned}
\int_{-\frac{3}{2}}^{-\frac{1}{2}} (-(x+1)^2 + 1)^{-\frac{1}{2}} dx &= \int_{\arcsin(-1/2)}^{\arcsin(1/2)} (-\sin^2\theta + 1)^{-\frac{1}{2}} \cos\theta d\theta \\
&= \int_{\arcsin(-1/2)}^{\arcsin(1/2)} d\theta \\
&= \arcsin(1/2) - \arcsin(-1/2) \\
&= \frac{\pi}{6} - \frac{-\pi}{6} \\
&= \frac{\pi}{3}.
\end{aligned}$$

Example 5.11. *Express the indefinite integral*

$$\int (6 + 4x - 2x^2)^{-\frac{1}{2}} dx$$

as a function of x *by using an appropriate substitution.*

First note that

$$\int (6 + 4x - 2x^2)^{-\frac{1}{2}} dx = \int (-2(x-1)^2 + 8)^{-\frac{1}{2}} dx = \frac{1}{\sqrt{8}} \int (-\frac{1}{4}(x-1)^2 + 1)^{-\frac{1}{2}} dx$$

Now the form of the integral suggests the substitution

$$\frac{1}{2}(x-1) = \sin\theta$$

so that $\frac{1}{2}dx = \cos\theta d\theta$. *Hence we have*

$$\begin{aligned}\frac{1}{\sqrt{8}}\int(-\frac{1}{4}(x-1)^2+1)^{-\frac{1}{2}}dx &= \frac{1}{\sqrt{8}}\int(-\sin^2\theta+1)^{-\frac{1}{2}}(2\cos\theta d\theta)\\ &= \frac{2}{\sqrt{8}}\int d\theta\\ &= \frac{1}{\sqrt{2}}(\arcsin(\frac{x-1}{2}))+K\end{aligned}$$

Example 5.12. *Express the indefinite integral*

$$\int \tan(x)dx$$

as a function of x *by using an appropriate substitution.*

First note that

$$\int \tan(x)dx = \int \frac{\sin(x)}{\cos(x)}dx.$$

With a little thought we realise that the numerator is the derivative of the denominator and so the substitution

$$\cos(x) = t$$

so that $-\sin(x)dx = dt$, *will greatly simplify the integral. We have*

$$\begin{aligned}\int \frac{\sin(x)}{\cos(x)}dx &= \int \frac{-dt}{t}\\ &= \int \frac{d}{dt}(-\ln(|t|))dt\\ &= -\ln(|t|)+K\\ &= -\ln(|\cos(x)|)+K\end{aligned}$$

Exercise 5.4. *Find the integral*

$$\int \frac{3x^2+2x}{x^3+x^2+5}dx$$

as a function of x *by making an appropriate substitution. [Hint: compare the numerator and the denominator.]*

Example 5.13. *Express the indefinite integral*

$$\int \frac{dx}{\sin(x)}$$

as a function of x by using an appropriate substitution.

There are many correct ways to approach this integral. We will first try to simplify the integral with the substitution

$$\frac{1}{\sin(x)} = t$$

so that $dt = -\frac{\cos(x)}{\sin^2(x)}dx = -\frac{\sqrt{1-\frac{1}{t^2}}}{\frac{1}{t^2}}dx = -t^2\sqrt{1-\frac{1}{t^2}}dx = -t\sqrt{t^2-1}dx$. *Upon substitution we find*

$$\begin{aligned}\int \frac{dx}{\sin(x)} &= \int t\left(\frac{-dt}{t\sqrt{t^2-1}}\right) \\ &= -\int \frac{dt}{\sqrt{t^2-1}}.\end{aligned}$$

Now we will use another substitution:

$$t = \cosh(u)$$

so that $dt = \sinh(u)du$ *and we have*

$$\begin{aligned}-\int \frac{dt}{\sqrt{t^2-1}} &= -\int \frac{\sinh(u)du}{\sqrt{\cosh^2(u)-1}} \\ &= -\int du \\ &= -u + K \\ &= -\text{arccosh}(t) + K \\ &= -\text{arccosh}(\frac{1}{\sin(x)}) + K.\end{aligned}$$

This answer can be written in many different forms using the various identities we have seen for the trigonometric and hyperbolic functions. Similarly if we had pursued a different path to find the antiderivative, by making different substitutions for example then we would have ended up with the answer expressed in a different form. The solution here is frequently written in

terms of $\tan(x/2)$ *and the logarithm, let us convert our answer into this common format:*

$$\begin{aligned}
-\operatorname{arccosh}(\frac{1}{\sin(x)}) &= -\ln(\left|\sqrt{\frac{1}{\sin^2 x}-1}+\frac{1}{\sin x}\right|)\\
&= -\ln(\left|\sqrt{\frac{1-\sin^2 x}{\sin^2 x}}+\frac{1}{\sin x}\right|)\\
&= -\ln(\left|\frac{\cos x+1}{\sin x}\right|)\\
&= \ln(\left|\frac{\sin x}{\cos x+1}\right|)\\
&= \ln(\left|\frac{2\sin(x/2)\cos(x/2)}{\cos^2(x/2)-\sin^2(x/2)+1}\right|)\\
&= \ln(\left|\frac{2\sin(x/2)\cos(x/2)}{2\cos^2(x/2)}\right|)\\
&= \ln(\left|\frac{\sin(x/2)}{\cos(x/2)}\right|)\\
&= \ln(|\tan(x/2)|).
\end{aligned}$$

5.3.2 Integration by Parts

This method can be very useful when integrating product. Let us rewrite the statement of an integration by parts but this time for a definite integral:

$$\int_a^b f(x)\frac{dg}{dx}dx = -\int_a^b \frac{df}{dx}g(x)dx + \Big[f(x)g(x)\Big]_a^b.$$

The technique allows us to move the derivative from $g(x)$ to its coefficient $f(x)$ (and introduce a minus sign plus a boundary term). This will be very useful if $\frac{df}{dx}$ is a simple function, e.g. a constant. Let us practise using the method by looking at some examples.

Example 5.14. *Find, up to a constant, the function of* x *given by the indefinite integral*

$$\int \arcsin(x)dx.$$

First we will make a substitution aimed at making the integral more palatable, we substitute $x = \sin\theta$*, so that* $dx = \cos\theta d\theta$ *hence*

$$\int \arcsin(x)dx = \int \theta\cos\theta d\theta = \int \theta\frac{d}{d\theta}(\sin\theta)d\theta.$$

In the last line we have rewritten it so that a derivative is applied to one term under the integral in order to emphasise that we are now in a situation where using integration by parts is a good idea. Why is it a good idea, because integration by parts gives us a way to move the derivative from $\sin\theta$ *onto* θ *and* $\frac{d\theta}{d\theta} = 1$ *is simple. Let us be careful and identify the functions* f *and* g *given in our abstract derivative with the functions in our present example*[3]*: we take* $f(\theta) = \theta$, $g(\theta) = \sin\theta$, *so that*

$$\int f(\theta)\frac{dg}{d\theta}d\theta = -\int \frac{df}{d\theta}g(\theta)d\theta + f(\theta)g(\theta)$$

hence

$$\begin{aligned}\int \theta\frac{d}{d\theta}(\sin\theta)d\theta &= -\int \frac{d\theta}{d\theta}\sin\theta d\theta + \theta\sin\theta \\ &= -\int \sin\theta d\theta + \theta\sin\theta \\ &= -\int \frac{d}{d\theta}(-\cos\theta)d\theta + \theta\sin\theta \\ &= \cos\theta + \theta\sin\theta + K \\ &= \cos(\arcsin(x)) + x\arcsin x + K \\ &= \sqrt{1-x^2} + x\arcsin x + K\end{aligned}$$

Example 5.15. *Find, up to a constant, the function of* x *given by the indefinite integral*

$$\int x^2\cos(2x)dx.$$

First we note that

$$\int x^2\cos(2x)dx = \int x^2\frac{d}{dx}(\frac{1}{2}\sin(2x))dx.$$

Now we integrate by parts. It is a good idea to practise integrating by parts without the need to identify $f(x)$ *and* $g(x)$ *before carrying out the procedure. Specifically notice that integration by parts tells you to add a minus sign and shift the derivative from one term onto its coefficients in the integral, and of course do not forget to add the boundary term - this one should practise*

[3]Note that in the abstract definition the variable was x but in our example it is now θ.

doing immediately almost as an algebraic manoeuvre on the integral, so that:

$$\begin{aligned}
\int x^2 \frac{d}{dx}(\frac{1}{2}\sin(2x))dx &= -\int \frac{d}{dx}(x^2)(\frac{1}{2}\sin(2x))dx + \frac{x^2}{2}\sin(2x) \\
&= -\int x\sin(2x)dx + \frac{x^2}{2}\sin(2x) \\
&= \int x\frac{d}{dx}(\frac{1}{2}\cos(2x))dx + \frac{x^2}{2}\sin(2x) \\
&= -\int \frac{dx}{dx}(\frac{1}{2}\cos(2x))dx + \frac{x}{2}\cos(2x) + \frac{x^2}{2}\sin(2x) \\
&= -\int \frac{1}{2}\cos(2x)dx + \frac{x}{2}\cos(2x) + \frac{x^2}{2}\sin(2x) \\
&= -\int \frac{1}{2}\frac{d}{dx}(\frac{1}{2}\sin(2x))dx + \frac{x}{2}\cos(2x) + \frac{x^2}{2}\sin(2x) \\
&= -\frac{1}{4}\sin(2x) + \frac{x}{2}\cos(2x) + \frac{x^2}{2}\sin(2x) + K
\end{aligned}$$

Example 5.16. *Find, up to a constant, the function of x given by the indefinite integral*

$$\int xe^{-x}dx.$$

$$\begin{aligned}
\int xe^{-x}dx &= \int x\frac{d}{dx}(-e^{-x})dx \\
&= -\int x\frac{d}{dx}(e^{-x})dx \\
&= \int e^{-x}dx - xe^{-x} \\
&= -e^{-x} - xe^{-x} + K.
\end{aligned}$$

Example 5.17. *Find, up to a constant, the function of x given by the indefinite integral*

$$\int \ln(x)dx.$$

This example involves a good trick: to use integration by parts we need a derivative under the integral, now since $\frac{dx}{dx} = 1$ we can always insert this trivial derivative without changing the

integral, so that,

$$\begin{aligned}\int \ln(x)dx &= \int \ln(x)\frac{dx}{dx}dx \\ &= -\int \frac{d}{dx}(\ln(x))xdx + x\ln(x) \\ &= -\int \frac{1}{x}xdx + x\ln(x) \\ &= -\int dx + x\ln(x) \\ &= -x + x\ln(x) + K\end{aligned}$$

Exercise 5.5. *Use integration by parts to find*

$$\int \arctan x dx$$

as a function of x, up to a constant. [Hint: insert $\frac{dx}{dx} = 1$ as in the last example.]

5.3.3 Partial Fractions

We saw among our examples of solving integrals by substitution, that if we were integrating a fraction where the numerator was the derivative of the denominator then a substitution could greatly simplify the integral, i.e.

$$\int \frac{f'(x)}{f(x)}dx = \int \frac{du}{u}$$

where we have made the substitution $u = f(x)$, so that $du = f'(x)dx$. But how do we go about integrating fractions which do not have this form? If we are working with fractions of polynomial functions, we can try to split the fraction up into a sum of simpler fractions. This process is known as finding the partial fractions. We will investigate the use of partial fractions through a few simple examples to illustrate the central idea.

Example 5.18. *Find the function of x given by the indefinite integral*

$$\int \frac{3x+1}{x^2-3x+2}dx.$$

As mentioned our aim is to split the fraction into a sum of fractions, now we know that,

$$\frac{A}{f(x)} + \frac{B}{g(x)} = \frac{Ag(x) + Bf(x)}{f(x)g(x)}$$

and our aim is to reverse this process. Therefore it will aid us to identify the factorisation of the denominator in the problem. We have:

$$\int \frac{3x+1}{x^2-3x+2}dx = \int \frac{3x+1}{(x-1)(x-2)}dx \equiv \int \left(\frac{A}{x-1}+\frac{B}{x-2}\right)dx.$$

Our aim now is to find A and B, we know (from summing the fractions) that

$$A(x-2)+B(x-1) = 3x+1$$

hence we may now identify A and B by equating the coefficients of x and the constants on each side of the above and solving the simultaneous equations in A and B that doing so gives us. We find:

$$\begin{aligned} A+B &= 3 \\ -2A-B &= 1 \end{aligned}$$

Hence $A=-4$ and $B=7$, so we have:

$$\int \frac{3x+1}{(x-1)(x-2)}dx = \int \left(\frac{-4}{x-1}+\frac{7}{x-2}\right)dx = -4\ln(|x-1|)+7\ln(|x-2|)+K.$$

Most integrals which can be quickly solved using partial fractions are similar to the above example: one attempts to factorise the denominator, to find the partial fractions and then integrate these simpler fractions. The simplest complication emerges when the denominator has a repeated factor as in the following example.

Example 5.19. *Find the function of x given by the indefinite integral*

$$\int \frac{2x+3}{(x-1)^2}dx.$$

If we commence by trying to split the fraction into partial fractions with the standard procedure we find

$$\frac{2x+3}{(x-1)^2} = \frac{A}{x-1}+\frac{B}{x-1} = \frac{(A+B)(x-1)}{(x-1)^2}$$

and so we have the contradiction that $A+B=2$ and $A+B=-3$, something has gone astray here, and the source of the problem is the repeated factor in the denominator. Notice that if it had been possible to find the partial fractions as above it would have meant that a factor of $(x-1)$ could be removed from both the numerator and the denominator. From the form of the fraction that we start with we see that this is not possible (otherwise we would have simplified

the fraction). Hence our first move was amiss, instead of splitting the fraction into partial fractions whose denominators are identical we seek A and B such that:

$$\frac{2x+3}{(x-1)^2} = \frac{A}{x-1} + \frac{B}{(x-1)^2} = \frac{Ax - A + B}{(x-1)^2}.$$

Now we find that $A = 2$ and $B = 5$, hence,

$$\int \frac{2x+3}{(x-1)^2}dx = \int \frac{2}{x-1}dx + \int \frac{5}{(x-1)^2}dx = 2\ln(|x-1|) - \frac{5}{x-1} + K.$$

One further common complication concerns what happens when the factor in the denominator is a polynomial of order greater than one. Consider the following example.

Example 5.20. *Find the function of x given by the indefinite integral*

$$\int \frac{2x^2 - 5x + 7}{x(x^2+3)}dx.$$

If we commence naïvely by trying to split the fraction with constant numerators we find

$$\frac{2x^2 - 5x + 7}{x(x^2+3)} = \frac{A}{x} + \frac{B}{x^2+3} = \frac{Ax^2 + 3A + Bx}{x(x^2+3)}$$

and we see that the simultaneous equations for A and B are contradictory: $A = 2$, $B = -5$ and $3A = 7$, our choice for the partition was too tightly constraining. Notice also that in every other example we have chosen partial fractions such that the numerator was of order one less than the denominator, hence we are motivated to try the following:

$$\frac{2x^2 - 5x + 7}{x(x^2+3)} = \frac{A}{x} + \frac{Bx + C}{x^2+3} = \frac{Ax^2 + 3A + Bx^2 + Cx}{x(x^2+3)}$$

so that the simultaneous equations to solve are

$$\begin{aligned} A + B &= 2 \\ C &= -5 \\ 3A &= 7. \end{aligned}$$

These equations are not contradictory and we can solve them to find $A = \frac{7}{3}$, $B = -\frac{1}{3}$ and $C = -5$. Hence,

$$\begin{aligned} \int \frac{2x^2 - 5x + 7}{x(x^2+3)}dx &= \frac{7}{3}\int \frac{1}{x}dx + \int \frac{-\frac{1}{3}x - 5}{x^2+3}dx \\ &= \frac{7}{3}\int \frac{1}{x}dx - \frac{1}{3}\int \frac{x}{x^2+3}dx - \int \frac{5}{x^2+3}dx \\ &= \frac{7}{3}\ln(|x|) - \frac{1}{6}\ln(|x^2+3|) - \frac{5}{\sqrt{3}}\arctan\left(\frac{x}{\sqrt{3}}\right) + K. \end{aligned}$$

There are some further variations which we will highlight here, namely, what happens when the degree of the polynomial in the numerator is the same as that in the denominator, or even greater? We will consider two simple examples in theses classes and also make some comments on what to do if one cannot factorise the denominator.

Example 5.21. *Find the function of x given by the indefinite integral*

$$\int \frac{x+2}{x-2} dx.$$

Here the denominator has the same power as the numerator, our aim will be to write the numerator as a multiple of the denominator plus some remainder to simplify the problem. If we can do this, we may then cancel out common factors and hope to find a simpler problem. Hence we must first find $f(x)$ and the remainder R such that

$$(x-2)f(x) + R = x + 2$$

we may carry out the long division or alternatively we may write $f(x) = ax + b$ and find a and b, i.e. from $(x-2)(ax+b) + R = ax^2 + bx - 2ax - 2b + R = x + 2$ we have $a = 0$, $b = 1$ and so $R = 4$. Hence,

$$\int \frac{x+2}{x-2} dx = \int \frac{(x-2)+4}{x-2} dx = \int \frac{x-2}{x-2} dx + \int \frac{4}{x-2} dx = x + 4\ln(|x-2|) + K.$$

The factorisation of the numerator allowed us to split the fraction into a constant part and a term in which the order of the numerator was less than the denominator.

Example 5.22. *Find the function of x given by the indefinite integral*

$$\int \frac{(x+2)^3}{(x-2)^2} dx.$$

We will aim to repeat the tactic of the previous example, but now there is some increased complication. Recall that the success of the previous example rested upon removing factors in the denominator from the numerator, so for this example we will aim to solve

$$(ax+b)(x-2)^2 + c(x-2) + d = (x+2)^3$$

or, after multiplying out,

$$ax^3 + (-4a+b)x^2 + (4a-4b+c)x + 4b - 2c + d = x^3 + 6x^2 + 12x + 8.$$

From which we may find that $a = 1$, $b = 10$, $c = 48$ and $d = 64$ and so we have,

$$\begin{aligned}\int \frac{(x+2)^3}{(x-2)^2}dx &= \int \frac{(x+10)(x-2)^2 + 48(x-2) + 64}{(x-2)^2}dx \\ &= \frac{1}{2}x^2 + 10x + 48\ln(|x-2|) - \frac{64}{x-2} + K.\end{aligned}$$

5.3.4 Recursion Relations

Integration by parts is useful when there is a product appearing in an integral and integration by parts can be used to simplify one of the products. Frequently it is necessary to apply the integration by parts many times before an integral becomes simple enough to be computed directly. The repeated procedure can be lengthy and it is often useful to derive a formula which encodes the action of integration by parts which can be used repeatedly to simplify an integral. Such a relation, which describes one integral in terms of another (simpler) integral is called a recursion relation or also a reduction formula.

For example, consider the integral

$$I_n \equiv \int_0^\infty x^n e^{-x} dx$$

where $n \in \mathbb{Z}^+$. Integrating by parts we find a recursion relation,

$$\begin{aligned}I_n &= \int_0^\infty x^n e^{-x} dx \\ &= \int_0^\infty x^n \frac{d}{dx}(-e^{-x})dx \\ &= -\int_0^\infty \frac{d}{dx}(x^n)(-e^{-x})dx + \left[(x^n)(-e^{-x})\right]_0^\infty \\ &= \int_0^\infty n x^{n-1} e^{-x} dx \\ &= nI_{n-1}.\end{aligned}$$

This is a recursion formula for the integral I_n: it expresses I_n as a function of I_{n-1} and more general recursion formulae may involve other integrals of lower order in n, e.g. $\hat{I}_n = f(\hat{I}_{n_1}, \hat{I}_{n_2}, \ldots \hat{I}_0)$ where $n > n_1 > n_2 > \ldots > 0$. A recursion formula allows one to replace more complex integrals (where n is large) with integrals (where n is smaller) which one hopes will be simpler to compute. We can repeatedly use $I_n = nI_{n-1}$ for successively smaller values of n until we find a simple integral that we can solve:

$$I_n = nI_{n-1} = n(n-1)I_{n-2} = n(n-1)(n-2)I_{n-3} = \ldots = n(n-1)(n-2)\ldots 2I_1 = n!I_0.$$

Now we can solve I_0 as

$$I_0 = \int_0^\infty x^0 e^{-x} dx = \int_0^\infty e^{-x} dx = \big[-e^{-x} \big]_0^\infty = 0 - (-1) = 1.$$

Therefore,

$$I_n \equiv \int_0^\infty x^n e^{-x} dx = n!$$

This is a beautiful expression, but notice how simple it is to compute all I_n (i.e. for any $n \in \mathbb{Z}^+$) once a recursion formula has been identified. What is odd about this particular integral is that $\int_0^\infty x^n e^{-x} dx$ is well-defined for any real value of n, not only positive integer values of n, while the computation of the recursion relation is left unchanged. In fact mathematicians use this integral to define what is meant by $n!$ when $n \in \mathbb{R}^+$. For this reason, this integral occurs frequently in mathematics and is called the gamma function and is defined as follows:

$$\Gamma(n+1) \equiv \int_0^\infty x^n e^{-x} dx \equiv n! \qquad \forall\, n \in \mathbb{R}^+.$$

By repeating the integration by parts, one can find the recursion relation for the gamma function is the same as for the original integral, namely,

$$\Gamma(n+1) = n\Gamma(n) \qquad \forall\, n \in \mathbb{R}^+.$$

Of course if one begins with a non-integer value for n then by repeated use of the recursion relation the value of n (appearing in the integral) can be lowered until it lies between 0 and 1, at which point one can only try to compute the remaining integral. A very interesting set of examples occur when n is half-integer, then the final integral to compute (after using the recursion relation) is $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, so, example:

$$\Gamma(\frac{9}{2}) = \frac{7}{2}\Gamma(\frac{7}{2}) = \frac{7\times 5}{2^2}\Gamma(\frac{5}{2}) = \frac{7\times 5\times 3}{2^3}\Gamma(\frac{3}{2}) = \frac{7\times 5\times 3\times 1}{2^4}\Gamma(\frac{1}{2}) = \frac{105}{16}\sqrt{\pi}.$$

The appearance of π through $\Gamma(\frac{1}{2})$ is interesting, and so this is the subject of the following (optional) exercise.

Exercise 5.6. *Compute the integral*

$$\Gamma(\frac{1}{2}) = \int_0^\infty x^{-\frac{1}{2}} e^{-x} dx$$

by using the substitution $x = y^2$ and the integral of (half of) the Gaussian function given by

$$\int_0^\infty e^{-y^2} dy = \frac{\sqrt{\pi}}{2}.$$

Exercise 5.7. *Show that the integral defined by*

$$I_n = \int \cos^n x dx \qquad \forall n \in \mathbb{Z}^+$$

has the following recursion relation:

$$I_n = \frac{n-1}{n} I_{n-2} + \frac{1}{n} \cos^{n-1} x \sin x.$$

Hence find I_5 as a function of x. [Hint: commence by writing $\cos^n x = \cos^{n-1} x \cos x$.]

5.4 Some Applications: Length, Area and Volume.

5.4.1 Areas of Circles and Ellipses

We defined the integral to measure (up to a sign) the area under a curve $f(x)$. It produces the area under a curve if the curve lies above the x-axis, but it produces minus the area between the curve and the x-axis where $f(x) < 0$. This seems to present a problem to anyone trying to use integration to find for example an expression for the area of a circle using integration. If the circle has its centre located on the x-axis then the integral of the circle will be zero, as the positive half-area above the axis cancels out the negative half-area found by integrating the part of the circle beneath the x-axis. Hence we need a trick to use integration to evaluate the area of a circle, or indeed the area of any shape which is left unchanged by a reflection in the x-axis. However the trick is very simple, since we know that half the area is found by integrating under the positive part of the curve which defines the shape, we may simply evaluate this integral and then double the result to find the area of the full shape.

Let us consider the example of the circle of radius R, centred at the origin. It is defined as the set of points (x, y) satisfying

$$x^2 + y^2 = R^2.$$

Hence writing the curve $y(x)$ we have $y(x) = \pm\sqrt{R^2 - x^2}$ for $x \in [-R, R]$. There are a pair of curves which are mapped into each other by reflection in the x-axis. Our aim will be to find the area under the positive curve, $y(x) = \sqrt{R^2 - x^2}$, this is the area of the semi-circle as shaded in figure 5.10. Now we may compute the area of the semicircle by integration:

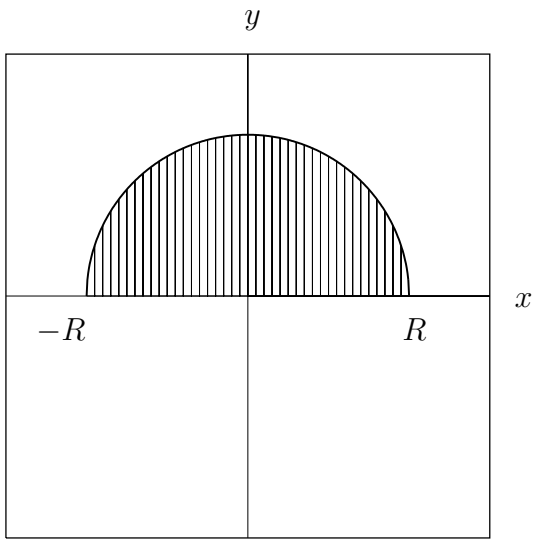


Figure 5.10: The integral of $y(x) = \sqrt{R^2 - x^2}$ from $x = -R$ to $x = R$ gives the area of the semicircle.

$$\begin{aligned}
A_{\text{semicircle}} &= \int_{-R}^{R} y(x)dx \\
&= \int_{-R}^{R} \sqrt{R^2 - x^2}dx \\
&= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sqrt{R^2 - R^2\sin^2\theta}(R\cos\theta d\theta) \\
&= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} R^2\cos^2\theta d\theta \\
&= R^2\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{2}(1+\cos(2\theta))d\theta \\
&= \frac{R^2}{2}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{d}{d\theta}(\theta + \frac{1}{2}\sin(2\theta))d\theta \\
&= \frac{R^2}{2}\left[\theta + \frac{1}{2}\sin(2\theta)\right]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \\
&= \frac{R^2}{2}((\frac{\pi}{2}+0) - (-\frac{\pi}{2}+0)) \\
&= \frac{1}{2}\pi R^2
\end{aligned}$$

where we made the substitution $x = R\sin\theta$, hence $dx = R\cos\theta d\theta$. This is the area of half the

circle, hence the area of the circle of radius R is

$$A_{\text{circle}} = 2A_{\text{semicircle}} = \pi R^2.$$

One can carry out a similar computation to find the area of an ellipse, this is the subject of the following example.

Example 5.23. *Find the area of the ellipse with foci located at* $(\pm a, 0)$ *where* $a \in \mathbb{R}^+$.

The ellipse is defined by the equation

$$\sqrt{(x-a)^2+y^2} + \sqrt{(x+a)^2+y^2} = 2R$$

where $R \in \mathbb{R}^+$ *and* $R > a$. *We would like to work with an equation of the form* $y = f(x)$, *so we rearrange the equation to obtain*[4]

$$y^2 = R^2 - x^2 - a^2 + \frac{a^2x^2}{R^2}.$$

Hence the upper edge of the ellipse is given by

$$y = \sqrt{R^2 - x^2 - a^2 + \frac{a^2x^2}{R^2}} = \sqrt{(R^2-a^2) - x^2(1-\frac{a^2}{R^2})}$$

. We would like to use an integral to find the area of the shaded region shown in figure 5.11 The area of the ellipse will be twice the area of upper half of the ellipse, i.e. twice the integral of $y(x)$, *hence,*

$$\begin{aligned}
A_{ellipse} &= 2\int_{-R}^{R} y(x)dx \\
&= 2\int_{-R}^{R} \sqrt{(R^2-a^2) - x^2(\frac{R^2-a^2}{R^2})}dx \\
&= 2\int_{-R}^{R} \sqrt{R^2-a^2}\sqrt{1-\frac{x^2}{R^2}}dx \\
&= 2\sqrt{R^2-a^2}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sqrt{1-\sin^2\theta}(R\cos\theta d\theta) \\
&= 2\sqrt{R^2-a^2}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} R\cos^2\theta d\theta \\
&= 2R\sqrt{R^2-a^2}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{2}(1+\cos(2\theta))d\theta \\
&= \pi R\sqrt{R^2-a^2}
\end{aligned}$$

[4]This is a reasonably lengthy algebraic manipulation, there is no trick used here to simplify the work but we are only presenting the result

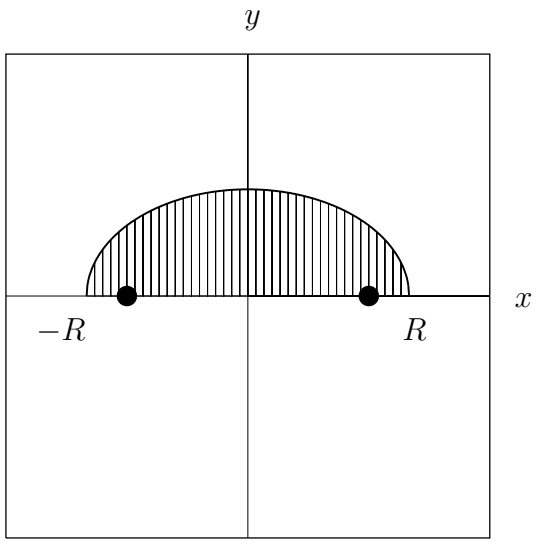


Figure 5.11: The integral of $y(x) = \sqrt{(R^2 - a^2) - x^2(1 - \frac{a^2}{R^2})}$ from $x = -R$ to $x = R$ gives the area of half of the ellipse.

where we have used the substitution $x = R\sin\theta$. Notice that this formula can be used to find the formula for the circle in the limit $a = 0$. Also note that the y-intercept of $y(x)$ is given by $y(0) = \sqrt{R^2 - a^2}$, which is a helpful way to recall the formula for the ellipse.

5.4.2 Volumes of Revolution

By dragging an area along a line we fill out a volume and we may use integration to compute volumes constructed in this way. Some of the most simple volumes (which remain interesting) are constructed by rotating an area about the x-axis: these are highly symmetric and are known as volumes of revolution. We may commence with a continuous function $f(x)$ on $x \in [a, b]$ and then consider slicing it into cross-sectional circles of width w and varying radius as depicted in figure 5.12. Let the cross-sectional area of the solid at $x = x_i$ be denoted $A(x_i)$ then the volume is approximated by

$$V = \sum_{i=1}^{n} wA(x_i) = \sum_{i=1}^{n} w\pi(f(x_i))^2$$

as $A(x_i) = \pi(f(x_i))^2$, the area of a circle of radius $f(x_i)$. If we split the volume into n cylinders of equal width w then we have $x_i = a + (i-1)w$ and as $w = \frac{b-a}{n}$ then $x_i = a + (i-1)\frac{b-a}{n}$, so that $x_{n+1} = b$ and $x_1 = a$. Now in the limit $w \to 0$ we have (by the definition of the Riemann

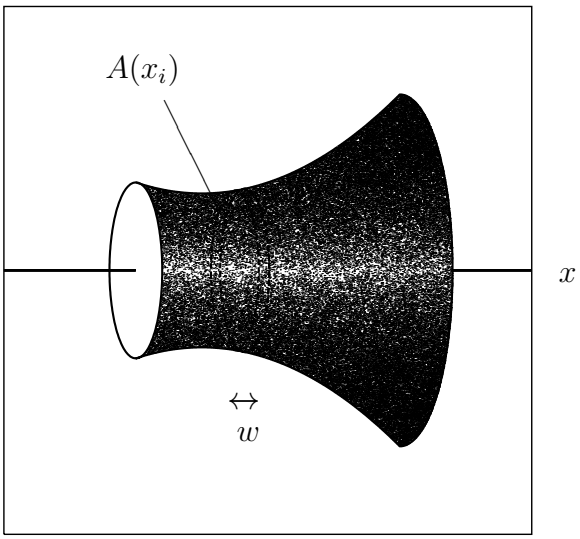


Figure 5.12: A volume whose surface is formed by rotating a continuous curve $y = f(x)$ about the x-axis may be sliced into thin circles of width w and radius $f(x_i)$ at $x = x_i$. In the limit when $w \to 0$ the circle has area $A(x_i) = \pi f(x_i)^2$.

integral):

$$\lim_{w\to 0, n\to\infty}(V) = \lim_{w\to 0, n\to\infty}(\sum_{i=1}^{n} w\pi(f(x_i))^2) = \pi\int_a^b f(x)^2 dx.$$

This integral above is the definition of the volume of revolution of the continuous function $f(x)$ about the x-axis from $x = a$ to $x = b$.

Example 5.24. *Use the formula for the volume of revolution to find the volume of the sphere of radius R.*

By rotating the circle defined by $x^2 + y^2 = R^2$ around the x-axis we find the volume of the

sphere of radius R:

$$\begin{aligned} V &= \pi \int_{-R}^{R} (f(x))^2 dx \\ &= \pi \int_{-R}^{R} (R^2 - x^2) dx \\ &= \pi \int_{-R}^{R} \frac{d}{dx}(R^2 x - \frac{1}{3}x^3) dx \\ &= \pi \left[R^2 x - \frac{1}{3}x^3 \right]_{-R}^{R} \\ &= \pi((R^3 - \frac{1}{3}R^3) - (-R^3 + \frac{1}{3}R^3)) \\ &= \pi(\frac{2}{3}R^3 + \frac{2}{3}R^3) \\ &= \frac{4}{3}\pi R^3. \end{aligned}$$

Exercise 5.8. *The volume bounded by the ellipse when it is rotated about the x-axis is called an ellipsoid or sometimes a cigar. Use the formula for the volume of revolution to show that the volume formed by rotating the ellipse defined by*

$$\sqrt{(x-a)^2 + y^2} + \sqrt{(x+a)^2 + y^2} = 2R$$

about the x-axis is

$$\frac{4}{3}\pi R(R^2 - a^2).$$

5.4.3 The Length of a Curve

A seemingly simple question can now be answered using the integral: how does one measure the length of a curve $f(x)$ from $x = a$ to $x = b$? The answer is that one can use Pythagorus' theorem to compute the straight line distance between points (x_i, y_i) and (x_{i+1}, y_{i+1}) on a curve. If we follow our usual construction and split the curve into n segments of equal width w for the length between $f(a)$ and $f(b)$ we have $w = \frac{b-a}{n}$, and $x_i = a + (i-1)w$. Now the straight line distance from $(x_i, f(x_i))$ to $(x_{i+1}, f(x_{i+1}))$ is given by

$$l_i = \sqrt{(x_{i+1} - x_i)^2 + (f(x_{i+1}) - f(x_i))^2} = w\sqrt{1 + \left(\frac{f(x_{i+1}) - f(x_i)}{w}\right)^2}$$

as $w = x_{i+1} - x_i$. Now the full length of the curve is found by summing the l_i and taking the limit as $w \to 0$ while $x_1 = a$ and $x_{n+1} = b$ are held fixed (i.e. this is the usual prescription for

constructing the Riemann integral), so we find

$$\begin{aligned}\text{Length, L} &\equiv \lim_{w\to 0, n\to\infty}\left(\sum_{i=1}^{n} l_i\right)\\ &= \lim_{w\to 0, n\to\infty}\left(\sum_{i=1}^{n} w\sqrt{1+\left(\frac{f(x_{i+1})-f(x_i)}{w}\right)^2}\right)\\ &= \int_a^b dx\sqrt{1+\left(\frac{df}{dx}\right)^2}.\end{aligned}$$

Example 5.25. *Find the length of the circumference of a circle of radius R.*

The circumference is twice the length of the curve $y=\sqrt{R^2-x^2}$, from $x=-R$ to $x=R$. We compute the ingredients used to find the length of the curve, namely, by taking the derivative with respect to x on both sides of $y^2=R^2-x^2$:

$$2y\frac{dy}{dx} = -2x \qquad \Longrightarrow \qquad \frac{dy}{dx} = -\frac{x}{y}.$$

Therefore,

$$1+\left(\frac{dy}{dx}\right)^2 = 1+\frac{x^2}{y^2} = \frac{x^2+y^2}{y^2} = \frac{R^2}{y^2} = \frac{R^2}{R^2-x^2}.$$

Hence,

$$\begin{aligned}L &= 2\int_{-R}^{R}\sqrt{\frac{R^2}{R^2-x^2}}dx\\ &= 2\int_{-R}^{R}\frac{R}{\sqrt{R^2-x^2}}dx\\ &= 2\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\frac{1}{\sqrt{R^2-R^2\sin^2\theta}}(R\cos\theta d\theta)\\ &= 2R\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}d\theta\\ &= 2\pi R.\end{aligned}$$

Of course one can use this technique to evaluate the lengths of more interesting curves, for example the length of a spiral. However we will leave the examples of greater complexity to the problem sheets. Instead here we will make a small digression to make an interesting comment that is contained in one of the more seemingly paradoxical mathematical results of the twentieth century which was commented on in a famous mathematical paper.

The famous paper was called "How Long is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension." It was written by Benoît Mandelbrot and published in 1967 and involved self-similar curves called fractals. One of his motivations was an observation of Lewis Fry Richardson[5] that the length of a coastline depends on the measurement scale that is used to measure it. Empirical measurement confirms that the smaller the ruler used the longer the coastline is measured to be! Our integral expressions for the length of a curve allow us to understand the idea as $w \to 0$ (our length scale is decreased) the curve length increases to a maximum (but not infinite) length. The curves we are considering are well-behaved, continuous curves so we might conclude that result concerning the length of the coastline of Britain is perfectly sensible, however we might also wonder whether the coastline of Britain is not given by a continuous function. The paradox contained in the measurement leads us to ponder the question: what is the smallest unit of measurement in the natural world and how long is a coastline measured using this length as a basic unit. We might speculate that this basic unit is of sub-nuclear length and then we might wonder further where exactly is the boundary of the nucleus - i.e. what boundary should we measure along at these small scales? We might take heart if we know that most fundamental physics involves descriptions of the world (often) in terms of continuous functions. So there may be hope yet of measuring the coastline of Britain in the 21^{st} century.

[5]Not the Hollyoaks character but a real English mathematician who lived from 1881 to 1953.

6. Power Series

In which we return to study the infinite sum and give mathematically sound criteria when such a sum exists. We meet the radius of convergence and show that the function e^x and the trigonometric and hyperbolic functions are well-defined. We state Taylor's theorem which gives a way to write a function locally as a power series and finally we show that an application of Taylor's theorem helps in the computation of many limits.

The material in this chapter will be covered in weeks 11-12.

At the start of this course we asked the question: how many functions are there? Our aim was not to give an answer to this question but to guide us to thinking about classes of functions. In the course we have met many large classes of functions, for example polynomial functions, trigonometric functions, inverse functions and so on. Throughout this tour of functions we were haunted by the notion that some functions were more fundamental than others (i.e. they could be used to form the other functions through addition, multiplication or composition) and indeed that some functions were better behaved than other well-defined functions (for example continuous functions and differentiable functions). In this chapter we will combine some of these thoughts to define a purist's notion of a well-behaved function: this will be the class of analytic functions. To define an analytic function we will also face another puzzle that has been with us in the course since we introduced the definition of the exponential function as a series: how do we know when an infinite sum is well-defined and converges? We will use these ideas to give the definition of a Taylor series expansion of a function about a point and then be able to say precisely what the properties of an analytic function are.

6.1 Infinite sums

Definition 6.1.1. *An infinite series is an expression of the form* $\sum_{n=n_0}^{\infty} a_n$ *where* $n, n_0 \in \mathbb{Z}$. *For a real series* $a_n \in \mathbb{R}$.

Comment(s). *(On infinite series...)*

1. *Typically* $n_0 = 0$ *or* 1, *hence a series is simply a sum with an infinite number of terms, for example*
$$\sum_{n=0}^{\infty} \frac{1}{2^n} = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots$$
where here $a_n = \frac{1}{2^n}$.

2. *A partial sum of an infinite series* $S \equiv \sum_{n=n_0}^{\infty} a_n$ *has the form* $S_N \equiv \sum_{n=n_0}^{N} a_n$. *This will be a useful tool to investigate properties of a series* S.

3. *We can immediately think of infinite series which do not sum to a finite number*[1] *(e.g.* $\sum_{n=0}^{\infty} n$*), so there are infinite series which do not sum to a finite number and are not well-defined. As mathematicians we would like to know when this is sensible.*

6.2 Convergence

As mathematicians we are expected to know whether or not we can find the sum of a series (infinite or not). It is not common in everyday life that sums do not converge but in mathematics this is very common, and such a sum or series is called divergent. Let us formally define these terms.

Definition 6.2.1. *A series* $S \equiv \sum_{n=n_0}^{\infty} a_n$ *is called convergent if the limit* $S \equiv \lim_{N\to\infty}(S_N)$ *exists, where* $S_N \equiv \sum_{n=n_0}^{N} a_n$. *If a series is not convergent it is called divergent.*

What can we say if we are given a series and asked to determine whether it is convergent or divergent? There is something we can deduce immediately about any convergent series, namely

[1] The example used here has made it into popular culture because it is a very important divergent sum that appears, among other places, in string theory, and because the oft-quoted claim that it equals $-\frac{1}{12}$ is disturbing. One can show that this sum can be split in a sensible way into a finite part (equal to $-\frac{1}{12}$) and an infinite part, and there is a method called ζ-regularisation which is used to throw away the infinite part, however sometimes you will see it claimed erroneously that $\sum_{n=0}^{\infty} n = -\frac{1}{12}$, where in reality it is only after the regularisation which removes an infinite part that this is the case. Without any regularisation it is a divergent sum.

that if an infinite series $S \equiv \sum_{n=n_0}^{\infty} a_n$ is convergent then we can say that:

$$\lim_{n\to\infty}(a_n) = 0.$$

This property of a convergent series is known as "the vanishing condition." One can see it is true by considering the difference between two convergent sums and taking a limit:

$$\lim_{n\to\infty}(S_n - S_{n-1}) = \lim_{n\to\infty}(S_n) - \lim_{n\to\infty}(S_{n-1}) = S - S = 0$$

but notice that $S_n - S_{n-1} = a_n$, hence we have the vanishing condition for any convergent series. For example consider the convergent series:

$$\sum_{n=0}^{\infty} \frac{1}{2^n} = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots$$

where $a_n = \frac{1}{2^n}$. It is clear that the vanishing condition is satisfied on a_n as

$$\lim_{n\to\infty} \frac{1}{2^n} = 0$$

and on the other hand we can compute the partial sum using the geometric identity to find

$$S_N = \sum_{n=0}^{N} \frac{1}{2^n} = \frac{1 - \frac{1}{2^{N+1}}}{1 - \frac{1}{2}} = \frac{2 - \frac{1}{2^N}}{2 - 1} = 2 - \frac{1}{2^N}$$

and hence

$$\lim_{N\to\infty}(S_N) = \lim_{N\to\infty}(2 - \frac{1}{2^N}) = 2.$$

So this infinite sum is convergent and it satisfies the vanishing condition. However the vanishing condition is a *necessary condition for a convergent series but it is not a* sufficient one. That is, the vanishing condition is true for all convergent series, but it is not a criterion to base a test for convergence on, as there are some divergent series which pass the vanishing condition test as well. For example the series

$$\sum_{n=0}^{\infty} \frac{1}{n}$$

does not converge, but does satisfy the vanishing condition as $\lim_{n\to\infty}(\frac{1}{n}) = 0$. There is a cunning argument to show that this sum does not converge. First consider the partial sum for this series:

$$S_N \equiv \sum_{n=1}^{N} \frac{1}{n}$$

and consider the difference between a pair of its partial sums

$$\begin{aligned} S_{2N} - S_N &= (1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{N} + \frac{1}{N+1} + \ldots + \frac{1}{2N}) - (1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{N}) \\ &= \frac{1}{N+1} + \ldots + \frac{1}{2N}. \end{aligned}$$

Now we observe that the smallest term in this difference of the two partial sums is the final one $\frac{1}{2N}$ and as there are N terms in the expression we have:

$$S_{2N} - S_N = \frac{1}{N+1} + \ldots + \frac{1}{2N} \geq \frac{N}{2N} = \frac{1}{2}.$$

But if a series T converges then the limit $\lim_{N\to\infty} T_N = T$ exists for all partial sums and hence if a series converges then we find

$$\lim_{N\to\infty} (T_{2N} - T_N) = T - T = 0$$

but for the series $\sum_{n=1}^{\infty} \frac{1}{n}$ we have found that

$$\lim_{N\to\infty} (S_{2N} - S_N) \geq \frac{1}{2}$$

hence we conclude that the series diverges even though it satisfies the vanishing condition.

The series $\sum_{n=n_0}^{\infty} \frac{1}{n}$ diverges as the terms in the sum do not become small enough at a fast enough rate. Whereas the series $\sum_{n=n_0}^{\infty} \frac{1}{n(n+1)}$ has terms which diminish in size at faster rate than those in the series $\sum_{n=n_0}^{\infty} \frac{1}{n}$ and does converge. Let us convince ourselves of this. The partial sum of this series is

$$\begin{aligned} S_N &\equiv \sum_{n=1}^{N} \frac{1}{n(n+1)} \\ &= \sum_{n=1}^{N} \left(\frac{1}{n} - \frac{1}{(n+1)} \right) \\ &= \left(\frac{1}{1} - \frac{1}{2} \right) + \left(\frac{1}{2} - \frac{1}{3} \right) + \left(\frac{1}{3} - \frac{1}{4} \right) + \ldots \left(\frac{1}{N} - \frac{1}{N+1} \right) \\ &= 1 - \frac{1}{N+1}. \end{aligned}$$

Hence,

$$S \equiv \lim_{N\to\infty} (S_N) = \lim_{N\to\infty} \left(1 - \frac{1}{N+1} \right) = 1.$$

To recap the series $\sum_{n=1}^{\infty} \frac{1}{n}$ is divergent, while the series $\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1$ is convergent. It will be an interesting challenge to understand the relation between the exponent of n in the series term a_n and the convergence of the series. However our more immediate challenge is to develop some criteria which we can use to determine whether a series converges or not.

6.2.1 Series Convergence Criteria.

We now state three tests which can be used to determine whether a series $\sum_{n=n_0}^{\infty} a_n$, (we will adopt the shorthand notation $\sum_n a_n \equiv \sum_{n=n_0}^{\infty} a_n$ for the series):

(C1) (*The Limit Comparison Test.*) If both $a_n > 0$ and $b_n > 0$ for all $n \in \mathbb{Z}^+$ and if $\lim_{n\to\infty}\left(\frac{a_n}{b_n}\right) = L$ where $0 < L < \infty$, then either both series $\sum_n a_n$ and $\sum_n b_n$ are convergent or both series are divergent.

(C2) (*The n'th Root Test or Cauchy's Criterion.*)

$$\lim_{n\to\infty} |a_n|^{\frac{1}{n}} \begin{cases} < 1 & \sum_n a_n \quad \text{is convergent} \\ > 1 & \sum_n a_n \quad \text{is divergent} \end{cases}$$

(C3) (*The Ratio Test or D'Alembert's Criterion.*)

$$\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| \begin{cases} < 1 & \sum_n a_n \quad \text{is convergent} \\ > 1 & \sum_n a_n \quad \text{is divergent.} \end{cases}$$

Comment(s). *(On the convergence criteria...)*

1. *The first test uses a comparison against another known convergent series to determine if a series converges, the second and third tests require only knowledge of the series in question.*

2. *We draw attention to the fact that in (C2) and (C3) certain cases are omitted, specifically when* $\lim_{n\to\infty} |a_n|^{\frac{1}{n}} = 1$ *and* $\lim_{n\to\infty} |\frac{a_{n+1}}{a_n}| = 1$*, the root and the ratio tests are both inconclusive.*

3. *The root test (C2) is more powerful than the ratio test (C3), but often (C3) is simpler to work with.*

Proofs of these convergence criteria are not part of this course but they are interesting, so we include a proof of the limit comparison test here in the lecture notes but these are included for fun and will not be examinable!

Proof of (C1): We will make use of the following lemma[2] which is known as the direct comparison test for convergence:

[2] "A subsidiary or intermediate theorem in an argument or proof", from the Greek meaning 'something assumed'. You might like to think of it as meaning a little theorem.

Lemma 6.1. *Suppose that* $0 < a_n \leq b_n$ *for all* $n \in \mathbb{Z}^+$ *then if* $\sum_{n=1}^{\infty} b_n$ *converges then so does* $\sum_{n=1}^{\infty} a_n$.

Proof of lemma: Let the partial sums of the two series be denoted by $S_N \equiv \sum_{n=1}^{N} a_n$ and $T_N \equiv \sum_{n=1}^{N} b_n$ then as $a_n \leq b_n$ for all $n \in \mathbb{Z}$ then $S_N \leq T_N$ for all $N \in \mathbb{Z}$. Consequently (as $0 < a_n \leq b_n$), $\lim_{N\to\infty}(S_N) \leq \lim_{N\to\infty}(T_N) = \sum_n b_n$ as $\sum_n b_n$ is convergent.

Now if $\lim_{n\to\infty}(\frac{a_n}{b_n}) = L$, where $0 < L < \infty$ (as $a_n > 0$ and $b_n > 0$), then (by the definition of the existence of the asymptotic limit) there exists some N such that for all $n > N$

$$-\epsilon < \frac{a_n}{b_n} - L < \epsilon$$

hence,

$$(L-\epsilon)b_n < a_n < (L+\epsilon)b_n.$$

Hence $a_n < (L+\epsilon)b_n$ for all $n > N$ hence if $\sum_n b_n$ is convergent then $(L+\epsilon)\sum_n b_n$ converges and so, by the lemma, $\sum_{n>N} a_n$ is convergent. Now $\sum_{n=1}^{N} a_n$ being a finite sum of real numbers is convergent hence:

$$\sum_n a_n = \sum_{n=1}^{N} a_n + \sum_{n=N}^{\infty} a_n$$

is convergent as it is the sum of two convergent series.

If, alternatively, we know that $\sum_n a_n$ is convergent then by choosing N large enough we can guarantee that $L - \epsilon$ is positive and hence

$$b_n < \frac{a_n}{L-\epsilon}$$

and by the lemma $\sum_n b_n$ is convergent. Hence if either series is convergent, then both are, or, alternatively, if either series is divergent then both are. □

Let us practise using the limit comparison test.

Example 6.1. *Use the limit comparison test (C1) to show that*

$$\sum_{n=1}^{\infty} \frac{1}{n^2}$$

is convergent.

Let $a_n = \frac{1}{n^2}$ *and now we must identify a convergent series* $\sum_n b_n$ *such that* $\lim_{n\to\infty} \frac{a_n}{b_n}$ *exists. We will compare the series with* $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$ *which we have already shown is convergent, so we have* $b_n = \frac{1}{n(n+1)}$ *and* $a_n \geq 0$, $b_n \geq 0$. *Now,*

$$\lim_{n\to\infty} \frac{a_n}{b_n} = \lim_{n\to\infty} \frac{(\frac{1}{n^2})}{(\frac{1}{n(n+1)})} = \lim_{n\to\infty} \left(\frac{n(n+1)}{n^2}\right) = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right) = 1.$$

Hence as the limit exists and $\sum_n b_n$ is convergent then $\sum_n \frac{1}{n^2}$ is convergent.

Example 6.2. *Show that*

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$$

is convergent.

Consider the partial sums:

$$\begin{aligned} S_N &= \sum_{n=1}^{N} \frac{(-1)^{n+1}}{n} \\ &= \sum_{n,\text{odd}}^{N} \frac{1}{n} - \sum_{n,\text{even}}^{N} \frac{1}{n} \\ &= \begin{cases} \sum_{m=1}^{N/2} \frac{1}{2m-1} - \sum_{m=1}^{N/2} \frac{1}{2m} = \sum_{m=1}^{N/2} \frac{1}{2m(2m-1)} & \text{for even } N \\ \sum_{m=1}^{(N+1)/2} \frac{1}{2m-1} - \sum_{m=1}^{(N-1)/2} \frac{1}{2m} = \sum_{m=1}^{(N-1)/2} \frac{1}{2m(2m-1)} + \frac{1}{N} & \text{for odd } N \end{cases} \end{aligned}$$

hence the series will only converge if $\sum_m \frac{1}{2m(2m-1)}$ is convergent. Now we may use the limit comparison test (C1) to compare this series against the convergent series $\sum_n \frac{1}{n^2}$, i.e. taking $a_n = \frac{1}{2n(2n-1)}$ and $b_n = \frac{1}{n^2}$, then

$$\lim_{n\to\infty} \frac{a_n}{b_n} = \lim_{n\to\infty} \frac{\left(\frac{1}{2n(2n-1)}\right)}{\left(\frac{1}{n^2}\right)} = \lim_{n\to\infty} \left(\frac{n^2}{2n(2n-1)}\right) = \lim_{n\to\infty} \left(\frac{n^2}{4n^2-2n}\right) = \frac{1}{4}.$$

Therefore as the limit exists and $\sum_n b_n$ is convergent then $\sum_n \frac{1}{2n(2n-1)}$ is convergent and hence $\sum_n \frac{(-1)^{n+1}}{n}$ is convergent.

6.3 Series as Functions of x.

Definition 6.3.1. *A power series is an expression of the form $S(x) \equiv \sum_{n=n_0}^{\infty} b_n x^n$.*

We may think of a power series as giving a series for any given value of x, that is a power series is a series $\sum_n a_n$ where the terms $a_n = b_n x^n$ each depend on x. We might immediately wonder how $S(x)$ varies as x varies: in particular is $S(x)$ convergent for all or any values of x? Now we may make use of the convergence criteria, in particular Cauchy's $n'th$ root test (C2) and D'Alembert's ratio test (C3) for the series $\sum_n a_n$ by substituting $a_n = b_n x^n$ into the

convergence tests. Doing this we find from (C2) that if

$$\lim_{n\to\infty} |a_n|^{\frac{1}{n}} = \lim_{n\to\infty} |b_n x^n|^{\frac{1}{n}} \begin{cases} < 1 & S(x) \text{ converges} \\ > 1 & S(x) \text{ diverges.} \end{cases}$$

We are interested in finding a condition on x to express the convergence criteria, so it is notationally useful to define $\lim_{n\to\infty} |b_n|^{-\frac{1}{n}} \equiv R_1$ as then the convergence criterion $C2$ is expressed as

$$\lim_{n\to\infty} |b_n x^n|^{\frac{1}{n}} = \lim_{n\to\infty} |b_n|^{\frac{1}{n}} |x^n|^{\frac{1}{n}} = \left(\frac{1}{R_1}\right) \lim_{n\to\infty} |x^n|^{\frac{1}{n}} = \left(\frac{1}{R_1}\right) \lim_{n\to\infty} |x| < 1.$$

Now note that $\lim_{n\to\infty} |x| = |x|$ and therefore the power series $S(x)$ will converge if

$$|x| < R_1$$

and $S(x)$ will diverge if $|x| > R_1$. R_1 is called the radius of convergence as it delimits the points where the power series is well-defined and converges from the points where it diverges. From the convergence criterion (C3) we will find an alternative definition of the radius of convergence. Substituting $a_n = b_n x^n$ into (C3) we find that if

$$\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = \lim_{n\to\infty} \left|\frac{b_{n+1}x^{n+1}}{b_n x^n}\right| = \lim_{n\to\infty} \left|\frac{b_{n+1}x}{b_n}\right| \begin{cases} < 1 & S(x) \text{ converges} \\ > 1 & S(x) \text{ diverges.} \end{cases}$$

Let us define $\lim_{n\to\infty} \left|\frac{b_{n+1}}{b_n}\right|^{-1} \equiv R_2$ then $S(x)$ will converge for all x such that

$$|x| < R_2$$

and will diverge for all x such that $|x| > R_2$. Naturally we would only expect there to be a single value $R = R_1 = R_2$ which is the radius of convergence for a power series. The relation between R_1 and R_2 is not part of this course, but of course there is an interesting interaction between the pair of radii of convergence. At the end of this chapter we will include an appendix for the interested reader who wishes to understand more about the two formulae for the radius of convergence of a power series (which is not an examinable part of the course, but is interesting).

Definition 6.3.2. *The radius of convergence of a power series $\sum_n b_n x^n$ is a non-negative number R such that the series converges for all x such that $|x| < R$ and diverges for all $|x| > R$, where*

$$R = \lim_{n\to\infty} |b_n|^{-\frac{1}{n}}$$

and, where it exists,

$$R = \lim_{n\to\infty} \left|\frac{b_n}{b_{n+1}}\right|.$$

Comment(s). *(On the radius of convergence...)*

1. *For $|x| = R$ there is no general conclusion to be drawn, one must inspect the series being considered using some alternative method.*
2. *If $R = 0$ then the series diverges for all $x \neq 0$.*
3. *If $R = \infty$ the series converges for all $x \in \mathbb{R}$.*
4. *The limit $R = \lim_{n\to\infty} \left|\frac{b_n}{b_{n+1}}\right|$ is often the simplest limit to compute, but it only exists if the power series has consecutive terms which are non-zero. Many power series have only even or odd terms, in which case this limit is not defined. When both definitions of the radius of convergence exist they agree, and in the appendix you can see that the radius coming from the ratio comparison test implies the existence of the other definition (from the root test) of the radius of convergence, and in that case they coincide.*

Example 6.3. *Show that the power series*

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

is convergent for all $x \in \mathbb{R}$.

Here the series has coefficients $b_n = \frac{1}{n!}$, hence

$$R = \lim_{n\to\infty} \left|\frac{b_n}{b_{n+1}}\right| = \lim_{n\to\infty} \left|\frac{(\frac{1}{n!})}{(\frac{1}{(n+1)!})}\right| = \lim_{n\to\infty} \left(\frac{(n+1)!}{n!}\right) = \lim_{n\to\infty} (n+1) = \infty.$$

Hence e^x converges for all $|x| < \infty$, i.e. for all $x \in \mathbb{R}$. At this stage, we may recall (with some satisfaction) all the results that rested upon this observation, including the exponential notation for the trigonometric and hyperbolic functions, Euler's formula, the polar coordinate notation for complex numbers, the roots of unity, trigonometric identities for double and half angle formulae and hence all the integral and derivatives which required these identities.

Example 6.4. *Find the radius of convergence for the power series*

$$S(x) = \sum_{n=0}^{\infty} x^n.$$

Here the series has coefficients $b_n = 1$, hence

$$R = \lim_{n\to\infty}\left|\frac{b_n}{b_{n+1}}\right| = \lim_{n\to\infty}\left|\frac{1}{1}\right| = 1.$$

Hence $S(x)$ converges for all $|x| < 1$. From the identity for the geometric sum of a finite number of terms we know that

$$\sum_{n=0}^{N} x^n = \frac{1-x^{N+1}}{1-x} \qquad \forall\, x \neq 1.$$

This is a partial sum for $S(x)$ and as $\lim_{N\to\infty}(\sum_{n=0}^{N} x^n) = S(x)$, and if $|x| < 1$ this sum converges. Hence we have

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \qquad \forall\, |x| < 1$$

where we have observed that $\lim_{N\to\infty}(x^{n+1}) = 0$ for $|x| < 1$.

To gain confidence in the convergence of power series one might compare $\sum_{n=0}^{N} x^n$ for various N with $\frac{1}{1-x}$ by plotting the graphs for $N = 1, 2, 3, \ldots$. These graphs are plotted together with $y = \frac{1}{1-x}$ in the graph in figure 6.1. Notice how the graphs of $y = \sum_{n=0}^{N} x^n$ coincide with $y = \frac{1}{1-x}$ agreement (improving as N increases) in the region where $|x| < 1$. We might phrase this with a different emphasis and say that the function $\frac{1}{1-x}$ has a power series expansion about $x = 0$ for all $|x| < 1$. This foreshadows the idea of the Taylor power series expansion for a function.

Example 6.5. *Find the radius of convergence for* $\sin x$.

The first problem is to write $\sin x$ *as a power series, we have,*

$$\sin(x) = \sum_{m=0}^{\infty}(-1)^m \frac{x^{2m+1}}{(2m+1)!} = \sum_{n=0}^{\infty} b_n x^n.$$

Now we have

$$b_n = \begin{cases} 0 & \text{if } n \text{ is even} \\ \frac{(-1)^m}{(2m+1)!} = \frac{(-1)^{\frac{n-1}{2}}}{n!} & \text{if } n \text{ is odd} \end{cases}$$

N.B. we have taken $n = 2m+1$ to denote odd n and so $m = \frac{n-1}{2}$ has been substituted to find the expressions above. As $|\frac{b_n}{b_{n+1}}|$ is either zero or undefined for any consecutive pair of coefficients, and so the limit as $n \to \infty$ does not exist. Consequently we will use the alternative definition of the radius of convergence

$$R = \lim_{n\to\infty}\left(|b_n|^{-\frac{1}{n}}\right).$$

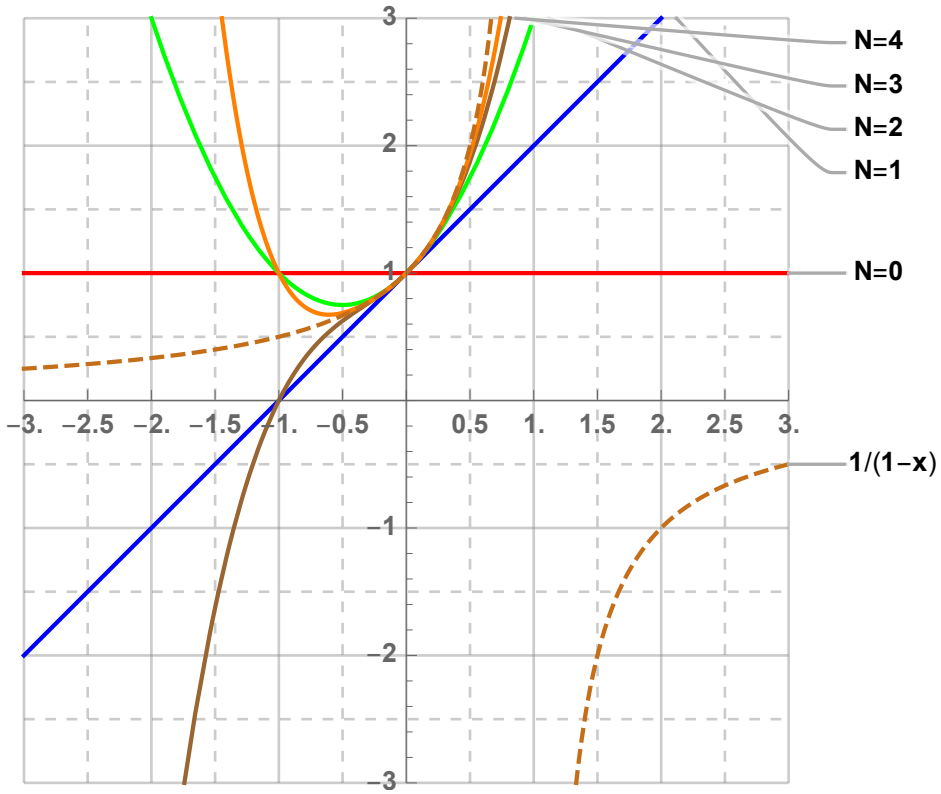


Figure 6.1: The graphs of $y = \sum_{n=0}^{N} x^n$ for $N = 1, 2, 3, 4$ with the various N shown, at the edge of the graph. The dashed line is the graph of $y = \frac{1}{1-x}$.

For odd n this limit becomes

$$R = \lim_{n\to\infty}\left(\left|(-1)^{\frac{n-1}{2}}\frac{1}{n!}\right|^{-\frac{1}{n}}\right) = \lim_{n\to\infty}\left(\left|\frac{1}{n!}\right|^{-\frac{1}{n}}\right) = \lim_{n\to\infty}\left(|n!|^{\frac{1}{n}}\right).$$

To show that this limit exists we can use Stirling's approximation for $n!$ which is given by:

$$n! \approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \qquad \text{for } n >> 1$$

this approximation is valid as $n \to \infty$. Hence we have,

$$R = \lim_{n\to\infty}\left(|n!|^{\frac{1}{n}}\right) \approx \lim_{n\to\infty}\left(|\left(\frac{n}{e}\right)^n\sqrt{2\pi n}|^{\frac{1}{n}}\right) = \lim_{n\to\infty}\left(\frac{n}{e}\right)\lim_{n\to\infty}\left((2\pi n)^{\frac{1}{2n}}\right) = \infty.$$

We are able to split the product of the limits as,

$$\lim_{n\to\infty}(n^{\frac{1}{2n}}) = \lim_{n\to\infty}(e^{\ln(n^{\frac{1}{2n}})}) = \lim_{n\to\infty}(e^{\frac{\ln n}{2n}}) = e^{(\lim_{n\to\infty}(\frac{\ln n}{2n}))} = e^0 = 1.$$

Therefore as the radius of convergence is infinite, the sine function converges for all $x \in \mathbb{R}$. This was a rather difficult conclusion to reach using the root test for convergence, alternatively

we could have argued that $sin(x)$ converges by using the triangle inequality:

$$|\sin(x)| = \left| \sum_{m=0}^{\infty} (-1)^m \frac{x^{2m+1}}{(2m+1)!} \right| \leq \sum_{m=0}^{\infty} \left(\frac{|x|^{2m+1}}{(2m+1)!} \right) \leq \sum_{m=0}^{\infty} \frac{|x|^m}{m!} = e^{|x|}$$

as e^x converges for all $x \in \mathbb{R}$ then $\sin(x)$ converges for all $x \in \mathbb{R}$.

6.4 Taylor's Theorem

Brook Taylor, a contemporary of Isaac Newton, based at Cambridge was one of the finest mathematicians of his generation but he is most frequently remembered as the originator of Taylor series, a way to construct a convergent power series for a function $f(x)$. To express the result efficiently, it is useful to take up some new notation, let

$$f^n(x_0) \equiv \frac{d^n}{dx^n}(f(x))\bigg|_{x=x_0}$$

be a shorthand notation for the n'th derivative of $f(x)$ evaluated at the point $x = x_0$. Using this notation we make the following claim which will form a basis for Taylor's theorem: if a function $f(x)$ can be written as a power series with a radius of convergence $R > 0$ (i.e. for all $x \in \mathbb{R}$ such that $|x| < R$, then $f(x) = \sum_{n=0}^{\infty} b_n x^n$) and if we may exchange the order of the derivative and the summation for an infinite series (as we can for a finite series) then:

$$b_n = \frac{f^n(0)}{n!}$$

if the n'th derivative of $f(x)$ exists at $x = 0$.
Proof of the claim: If

$$S(x) \equiv \sum_{n=0}^{\infty} \frac{f^n(0)}{n!} x^n \qquad \text{for } |x| < R$$

is the same as $f(x)$ then we can confirm this by checking that all of their derivatives at a point $x = 0$ are equal as well as $f(0) = S(0)$. So,

$$\begin{aligned} S^m(x) &= \frac{d^m}{dx^m}\left(\sum_{n=0}^{\infty} \frac{f^n(0)}{n!} x^n \right) \\ &= \sum_{n=0}^{\infty} \frac{f^n(0)}{n!} n(n-1)(n-2)\dots(n-m+1) x^{n-m} \\ &= \sum_{n=m}^{\infty} \frac{f^n(0)}{(n-m)!} x^{n-m}. \end{aligned}$$

Now at $x = 0$ we have

$$S^m(0) = \sum_{n=m}^{\infty} \frac{f^n(0)}{(n-m)!} 0^{n-m} = f^n(0)$$

and $S(0) = f^0(0) = f(0)$. Hence,

$$f(x) = \sum_{n=0}^{\infty} \frac{f^n(0)}{n!} x^n$$

as claimed. □

Now, because we might be (rightly) concerned about moving $\frac{d}{dx}$ (which is a limit) past an infinite sum (which is another limit) we present a proof that sidesteps this issue and uses the fundamental theorem of calculus. Now,

$$\int_0^x f^{n+1}(y)dy = \int_0^x \frac{d}{dy}(f^n(y))dy = f^n(x) - f^n(0)$$

and after rearranging we have,

$$f^n(x) = f^n(0) + \int_0^x f^{n+1}(y)dy.$$

Our intention is to make repeated use of the identity above. By assumption we will presume that f is differentiable (if the function can be written as a convergent series then it will be infinitely differentiable) and we commence with writing the statement above for the case $n = 0$:

$$\begin{aligned} f(x) &= f(0) + \int_0^x f^1(y)dy \\ &= f(0) + x\int_0^1 f^1(xz_1)dz_1 \\ &= f(0) + x\int_0^1 \left(f^1(0) + \int_0^{xz_1} f^2(y_1)dy_1 \right) dz_1 \\ &= f(0) + xf^1(0) + x\int_0^1 \left(\int_0^{xz_1} f^2(y_1)dy_1 \right) dz_1 \end{aligned}$$

where we have made the substitution $y = xz_1$ so that $dy = xdz_1$ and x is a constant. In the final line we have made use of the identity for $f^n(x)$ given above. Our intention is to repeat

this procedure to derive a series expansion for $f(x)$. We have,

$$\begin{aligned}
f(x) &= f(0) + xf^1(0) + x\int_0^1 \left(\int_0^{xz_1} f^2(y_1)dy_1 \right) dz_1 \\
&= f(0) + xf^1(0) + x^2 \int_0^1 z_1 \left(\int_0^1 f^2(xz_1z_2)dz_2 \right) dz_1 \\
&= f(0) + xf^1(0) + x^2 \int_0^1 z_1 \left(\int_0^1 \left(f^2(0) + \int_0^{xz_1z_2} f^3(y_2)dy_2 \right) dz_2 \right) dz_1 \\
&= f(0) + xf^1(0) + x^2 f^2(0) \int_0^1 z_1 dz_1 + x^2 \int_0^1 z_1 \left(\int_0^1 \left(\int_0^{xz_1z_2} f^3(y_2)dy_2 \right) dz_2 \right) dz_1 \\
&= f(0) + xf^1(0) + \frac{f^2(0)}{2}x^2 + x^2 \int_0^1 z_1 \left(\int_0^1 \left(\int_0^{xz_1z_2} f^3(y_2)dy_2 \right) dz_2 \right) dz_1
\end{aligned}$$

where we have substituted $y_1 = xz_1z_2$, where x is a constant and z_1 is treated as constants within the dy_1 integral so that $dy_1 = xz_1dz_2$. Now we may imagine continuing this procedure of substituting $y_n = xz_1z_2 \dots z_n$ at the n'th iteration and using the identity for $f^n(x)$ to expand the integral about $f^n(0)$. Had we carried on we could have confirmed the expansion for the first set of terms, i.e. that

$$f(x) = f(0) + xf^1(0) + \frac{x^2}{2}f^2(0) + \frac{x^3}{3!}f^3(0) + \dots .$$

However if we stopped the procedure after the first N terms of the Taylor series had been generated we would have been left with an integral $R_{N+1}(x)$ which is known as the remainder term:

$$f(x) = \sum_{n=0}^{N} \frac{f^n(0)x^n}{n!} + R_{N+1}(x)$$

where

$$\begin{aligned}
R_1(x) &= x\int_0^1 f^1(xz_1)dz_1 \\
R_2(x) &= x\int_0^1 \left(\int_0^{xz_1} f^2(y_1)dy_1 \right) dz_1 \\
R_3(x) &= x^2 \int_0^1 z_1 \left(\int_0^1 \left(\int_0^{xz_1z_2} f^3(y_2)dy_2 \right) dz_2 \right) dz_1 \\
&\vdots
\end{aligned}$$

The result we have derived above is an example of a Taylor series, a Taylor series $f(x)$ about $x = 0$ is known as the Maclaurin series for $f(x)$. Let us summarise:

Definition 6.4.1. *The Maclaurin series for a function $f(x)$ is a Taylor series expansion about $x = 0$ and to order N is written*

$$f(x) = \sum_{n=0}^{N} \frac{f^n(0)}{n!} x^n + R_{N+1}(x) \qquad \text{where } R_{N+1}(x) = \frac{1}{N!} \int_0^x f^{N+1}(y)(x-y)^N dy.$$

Why can we write the remainder term this way? We can prove this by induction. Let S_N be the statement that

$$R_{N+1}(x) = f(x) - \sum_{n=0}^{N} \frac{f^n(0)}{n!} x^n.$$

We will prove the inductive step first, namely that $S_N \implies S_{N+1}$.

$$\begin{aligned}
R_{N+2}(x) &= \frac{1}{(N+1)!} \int_0^x f^{N+2}(y)(x-y)^{N+1} dy \\
&= \frac{1}{(N+1)!} \int_0^x \frac{d}{dy}(f^{N+1}(y))(x-y)^{N+1} dy \\
&= -\frac{1}{(N+1)!} \int_0^x f^{N+1}(y) \frac{d}{dy}((x-y)^{N+1}) dy + \frac{1}{(N+1)!} \Big[f^{N+1}(y)(x-y)^{N+1} \Big]_{y=0}^{y=x} \\
&= \frac{1}{N!} \int_0^x f^{N+1}(y)(x-y)^N dy - \frac{1}{(N+1)!} f^{N+1}(0)(x)^{N+1} \\
&= R_{N+1}(x) - \frac{1}{(N+1)!} f^{N+1}(0) x^{N+1} \\
&= f(x) - \sum_{n=0}^{N} \frac{f^n(0)}{n!} x^n - \frac{1}{(N+1)!} f^{N+1}(0) x^{N+1} \\
&= f(x) - \sum_{n=0}^{N+1} \frac{f^n(0)}{n!} x^n
\end{aligned}$$

which is the statement S_{N+1}. Now let us check the basis step for $N = 0$, the left-hand-side of the statement S_0 is

$$R_1(x) = \frac{1}{0!} \int_0^x f^1(y)(x-y)^0 dy = \int_0^x f^1(y) dy = f(x) - f(0).$$

While the right-hand-side of S_0 reads

$$f(x) - \sum_{n=0}^{0} \frac{f^n(0)}{n!} \frac{1}{0!} = f(x) - f(0).$$

Therefore S_0 is a true statement and our proof by induction of the structure of the remainder term is complete.

The Maclaurin series give an approximation for $f(x)$ around zero with the discrepancy being given by the remainder term. In general one is interested in expanding about points $x = a$ where a may be non-zero. This more general case is called the Taylor series. The expansion parameter is no longer $x = (x - 0)$ but in the parameter $(x - a)$:

Definition 6.4.2. *The Taylor series to order N around the point $x = a$ of a function $f(x)$ is*

$$f(x) = \sum_{n=0}^{N} \frac{f^n(a)}{n!}(x-a)^n + R_{N+1}(x) \qquad \text{where } R_{N+1}(x) = \frac{1}{N!}\int_a^x f^{N+1}(y)(x-y)^N dy.$$

One can set about proving this theorem in the same manner as we did for the Maclaurin series, but instead of using x one works with the shifted variable $(x - a)$.

Our aim is to get some familiarity with using the Taylor series by computing the Taylor series for some familiar functions.

Example 6.6. *Find the Maclaurin series for the exponential function up to order N and state the remainder term.*

We have $f(x) = e^x$, hence $f^n(x) = \frac{d^n}{dx^n}(e^x) = e^x$. For the Maclaurin series we need to expand about $x = 0$, so we need $f^n(0) = e^0 = 1$ for all n. Hence the Maclaurin series is

$$e^x = \sum_{n=0}^{N} \frac{x^n}{n!} + R_{N+1}(x)$$

where

$$R_{N+1}(x) = \frac{1}{N!}\int_0^x e^y(x-y)^N dy.$$

Although it is not required for the example, it is reassuring to note that the remainder term vanishes as $N \to \infty$ by the following argument. Let $M \in \mathbb{Z}$ such that $M > (x - y) > 0$ (as $y \in [0, x]$ in the integral) then for all $N > M$

$$(x-y)^{N-M} < M^{N-M} < M(M+1)(M+2)\dots(M+(N-M-1)) = \frac{(N-1)!}{(M-1)!}$$

and so,

$$(x-y)^N < (x-y)^M \frac{(N-1)!}{(M-1)!} < M^M \frac{(N-1)!}{(M-1)!}.$$

Therefore

$$0 \leq \lim_{N\to\infty}\left(\frac{(x-y)^N}{N!}\right) \leq \lim_{N\to\infty}\left(M^M \frac{(N-1)!}{(M-1)!N!}\right) = \frac{M^M}{(M-1)!}\lim_{N\to\infty}\left(\frac{1}{N}\right) = 0.$$

Consequently we see that in the limit $N \to \infty$ that the remainder term vanishes,

$$\lim_{N\to\infty}\left(R_{N+1}(x)\right) = \int_0^x e^y \lim_{N\to\infty}\left(\frac{(x-y)^N}{N!}\right)dy = 0.$$

This is consistent with our observation that the infinite power series for the exponential function is convergent for all $x \in \mathbb{R}$.

Example 6.7. *Find the Maclaurin series for the sine function up to order N, where N is an odd, positive integer, and state the remainder term.*

We have $f(x) = \sin x$, hence

$$f^n(x) = \frac{d^n}{dx^n}(\sin x) = \begin{cases} (-1)^{\frac{n}{2}} \sin x & \text{for even } n \\ (-1)^{\frac{n-1}{2}} \cos x & \text{for odd } n. \end{cases}$$

Hence we have

$$f^n(0) = \begin{cases} 0 & \text{for even } n \\ (-1)^{\frac{n-1}{2}} & \text{for odd } n. \end{cases}$$

Hence to find the terms up to order $n = N$ we may write odd $n = 2m+1$ for $m \in \{0, 1, 2, 3, \ldots, \frac{N-1}{2}\}$, (recall that N is odd by the assumption stated in the question) so that

$$\sin x = \sum_{m=0}^{\frac{N-1}{2}} \frac{(-1)^m}{(2m+1)!}x^{2m+1} + R_{N+2}(x)$$

where

$$R_{N+2} = \frac{1}{(N+1)!}\int_0^x f^{N+2}(y)(x-y)^{N+1}dy = \frac{1}{(N+1)!}\int_0^x (-1)^{\frac{N+1}{2}} \cos(y)(x-y)^{N+1}dy.$$

We observe here that as the terms of even order in x are zero, the remainder term after the x^N term (where N is odd) begins with terms proportional to x^{N+2}, hence the remainder term is R_{N+2} rather than R_{N+1}.

Example 6.8. *Find the first N (non-zero) terms, where N is even, in the Taylor series for the cosine function expanded about the point $x = a$.*

We have $f(x) = \cos x$, hence

$$f^n(x) = \frac{d^n}{dx^n}(\cos x) = \begin{cases} (-1)^{\frac{n+1}{2}} \sin x & \text{for odd } n \\ (-1)^{\frac{n}{2}} \cos x & \text{for even } n. \end{cases}$$

Hence we have

$$f^n(a) = \begin{cases} (-1)^{\frac{n+1}{2}} \sin a & \text{for odd } n \\ (-1)^{\frac{n}{2}} \cos a & \text{for even } n. \end{cases}$$

To find the first N terms we do not need to consider the remainder term, and including the first term we need only construct terms up to and including order $N-1$ hence we have

$$\begin{aligned} \cos x &= \cos(a + (x-a)) \\ &= \sum_{n=0}^{N-2} \frac{f^n(a)}{n!}(x-a)^n \\ &= \sum_{n,\text{even}}^{N-2} \frac{(x-a)^n}{n!}(-1)^{\frac{n}{2}} \cos a + \sum_{n,\text{odd}}^{N-1} \frac{(x-a)^n}{n!}(-1)^{\frac{n+1}{2}} \sin a \\ &= \sum_{m=0}^{\frac{N-2}{2}} \frac{(x-a)^{2m}}{(2m)!}(-1)^m \cos a + \sum_{m=0}^{\frac{N-2}{2}} \frac{(x-a)^{2m+1}}{(2m+1)!}(-1)^{m+1} \sin a \end{aligned}$$

Where we have written odd $n = 2m+1$ and even $n = 2m$.

Example 6.9. *Find the first 4 (non-zero) terms, in the Maclaurin series for the function*

$$f(x) = \ln(1-x).$$

We have $f(x) = \ln(1-x)$, hence we compute and so we find,

n	$f^n(x) = \frac{d^n}{dx^n}(\ln(1-x))$	$f^n(0)$
0	$\ln(1-x)$	0
1	$\frac{-1}{(1-x)}$	-1
2	$\frac{-1}{(1-x)^2}$	-1
3	$\frac{-2}{(1-x)^3}$	-2
4	$\frac{-6}{(1-x)^4}$	-6

$$\begin{aligned} \ln(1-x) &= \sum_{n=0}^{4} \frac{f^n(0)}{n!}x^n + R_5(x) \\ &= -x - \frac{1}{2}x^2 - \frac{1}{3!}(2)x^3 - \frac{1}{4!}(6)x^4 + R_5(x) \\ &= -x - \frac{1}{2}x^2 - \frac{1}{3}x^3 - \frac{1}{4}x^4 + R_5(x). \end{aligned}$$

If we had continued the expansion to higher order terms the apparent pattern would continue and we would find

$$\ln(1-x) = -\sum_{n=1}^{\infty} \frac{x^n}{n}$$

so long as the general remainder term $R_{N+1}(x)$ vanishes as $N \to \infty$. So let us check when this is the case:

$$\begin{aligned}
\lim_{N\to\infty}(R_{N+1}(x)) &= \lim_{N\to\infty}\left(\frac{1}{N!}\int_0^x f^{N+1}(y)(x-y)^N dy\right) \\
&= -\lim_{N\to\infty}\left(\frac{1}{N!}\int_0^x N!\frac{(x-y)^N}{(1-y)^{N+1}} dy\right) \\
&= -\lim_{N\to\infty}\left(\int_0^x \frac{(x-y)^N}{(1-y)^{N+1}} dy\right) \\
&= -\int_0^x \lim_{N\to\infty}\left(\frac{(x-y)^N}{(1-y)^N}\right)\frac{1}{1-y} dy
\end{aligned}$$

Now

$$\lim_{N\to\infty}\left(\frac{(x-y)^N}{(1-y)^N}\right)$$

converges to zero if $|\frac{x-y}{1-y}| < 1$, i.e. if $|x| < 1$, but if $x > 1$ the remainder term diverges. This is a sign that the power series

$$\ln(1-x) = -\sum_{n=1}^{\infty} \frac{x^n}{n}$$

converges only if $|x| < 1$, we can show this is the case by identifying the radius of convergence. We have $b_n = -\frac{1}{n}$ and so,

$$R = \lim_{n\to\infty}\left|\frac{b_{n+1}}{b_n}\right| = \lim_{n\to\infty}\left|\frac{(\frac{1}{n+1})}{(\frac{1}{n})}\right| = \lim_{n\to\infty}\left|\frac{n}{n+1}\right| = \lim_{n\to\infty}\left|\frac{1}{1+\frac{1}{n}}\right| = 1.$$

Therefore, the series expansion,

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \dots$$

converges for $|x| < 1$.

We can now re-state the definition of the analytic function

Definition 6.4.3. *A function which can be locally written as a convergent power series is called an analytic function.*

That is if the Taylor series expansion for a function about a point $x = a$ is convergent for some non-zero radius R about the point $x = a$, i.e. for all x $|x - a| < R$, and if this is true for all $a \in \mathbb{R}$, then the function is analytic. For example, as the exponential function has a convergent power series for all x it is an analytic function.

Now we can make a few simple observations about such analytic functions, as we can always find a convergent power series expansion, then the function is differentiable (as we have no trouble in taking the derivative of a power series or infinite polynomial function) and so it is also continuous. Hence the analytic functions are the purist's idea of a well-defined function of one variable. This observation presents the end of the long-term quest for this course.

6.4.1 Short-cuts for finding a Taylor Expansion

There are many ways to find Taylor series which build upon well-known Taylor series without needing to compute the series term by term. Here we list a few neat observations and short-cuts to Taylor and Maclaurin series.

Now

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

is an analytic function (it converges for all $x \in \mathbb{R}$) and the map $x \to -x$ is a symmetry of $\mathbb{R}$ so e^{-x} is also convergent, i.e. we have,

$$e^x = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \dots .$$

We may make a similar observation for

$$\ln(1 - x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \dots$$

which is convergent for $x \in (-1, 1)$, hence

$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$$

is convergent for $x \in (-1, 1)$. We may use (well-defined) operations on analytic functions to find other series expansions, e.g.

$$\frac{1}{1+x} = \frac{d}{dx}(\ln(1 + x)) = \frac{d}{dx}(x - \frac{x^2}{2} + \frac{x^3}{3} - \dots) = 1 - x + x^2 - x^3 + \dots$$

which we might have deduced by summing the geometric series on the right when $|x| < 1$. A

final, more involved example, of manipulating convergent series is

$$\begin{aligned}\frac{1}{\cos(x)} &= \frac{1}{1-\frac{x^2}{2!}+\frac{x^4}{4!}-\ldots} \\ &\equiv \frac{1}{1+Y} \\ &= 1-Y+Y^2-Y^3+\ldots \\ &= 1-(-\frac{x^2}{2!}+\frac{x^4}{4!}-\ldots)+(-\frac{x^2}{2!}+\frac{x^4}{4!}-\ldots)^2+\ldots \\ &= 1+\frac{x^2}{2!}-\frac{x^4}{4!}+\frac{x^4}{4}+\ldots \\ &= 1+\frac{x^2}{2!}+\frac{5x^4}{4!}+\ldots\end{aligned}$$

where we have limited $|Y|<1$, notice that in general $Y=\cos(x)-1\in[-2,0]$, so this restriction on Y corresponds to constraining $0\leq\cos(x)\leq 1$. By attempting other manipulations of functions where they are know to be analytic one can come up with other wonderful examples of Taylor series.

We conclude the course with a wonderful observation which is built upon the Taylor series and is fantastically useful for computing limits.

6.5 l'Hôpital's Rule

The rule is named after the 17^{th} century French mathematician Guillaume de l'Hôpital (1661-1704) and l'Hôpital's rule provides a method for evaluating limits which appear to involve indefinite forms (e.g the wretched $\frac{0}{0}$ or $\frac{\infty}{\infty}$). Should we wish to find the limit $\lim_{x\to 0}\left(\frac{f(x)}{g(x)}\right)$ for which $f(0)=0$ and $g(0)=0$ and we are able to find the functions $f(x)$ and $g(x)$ as power series expanded about $x=0$ then we have:

$$\begin{aligned}\lim_{x\to 0}\left(\frac{f(x)}{g(x)}\right) &= \lim_{x\to 0}\left(\frac{f(0)+xf^1(0)+\frac{1}{2}x^2f^2(0)+\ldots}{g(0)+xg^1(0)+\frac{1}{2}x^2g^2(0)+\ldots}\right) \\ &= \lim_{x\to 0}\left(\frac{xf^1(0)+\frac{1}{2}x^2f^2(0)+\ldots}{xg^1(0)+\frac{1}{2}x^2g^2(0)+\ldots}\right) \\ &= \lim_{x\to 0}\left(\frac{f^1(0)+\frac{1}{2}xf^2(0)+\ldots}{g^1(0)+\frac{1}{2}xg^2(0)+\ldots}\right) \\ &= \frac{f^1(0)}{g^1(0)}\end{aligned}$$

where we have substituted $f(0) = 0$ and $g(0) = 0$ before taking the limit. So long as $g^1(0) \neq 0$ this result gives a well-defined limit. Let us summarise the generalisation of this result which is known as l'Hôpital's Rule:

Theorem 6.1. *(l'Hôpital's Rule) Let $f(x)$ and $g(x)$ be two functions which are differentiable on the interval I (except possibly at the point $x = x_0$) such that $\lim_{x\to x_0}(f(x)) = 0$, $\lim_{x\to x_0}(g(x)) = 0$ and $g'(x) \neq 0$ for all $x \in I \setminus x_0$ then*

$$\lim_{x\to x_0}\left(\frac{f(x)}{g(x)}\right) = \lim_{x\to x_0}\left(\frac{f'(x)}{g'(x)}\right)$$

where $x_0 \in I$.

It is possible to further generalise this statement of the theorem. To see this we will consider again the example where $x_0 = 0$ but now consider functions such that $\lim_{x\to 0}(f(x)) = \pm\infty$ and $\lim_{x\to 0}(g(x)) = \pm\infty$ then we have,

$$\lim_{x\to 0}\left(\frac{f(x)}{g(x)}\right) = \lim_{x\to 0}\left(\frac{1/g(x)}{1/f(x)}\right) = \lim_{x\to 0}\left(-\frac{g'(x)/(g(x))^2}{-f'(x)/(f(x))^2}\right)$$

where we have been able to invoke l'Hôpital's rule as we have $\lim_{x\to 0}(1/f(x)) = 0$ and $\lim_{x\to 0}(1/g(x)) = 0$. Now if the limit $\lim_{x\to 0}\left(\frac{(g(x))^2}{(f(x))^2}\right)$ exists we can multiply the above by it to find

$$\lim_{x\to 0}\left(\frac{g(x)}{f(x)}\right) = \lim_{x\to 0}\left(\frac{g'(x)}{f'(x)}\right).$$

Example 6.10. *Use l'Hôpital's rule to determine the following limit:*

$$\lim_{x\to 0}\left(\frac{a^x - 1}{x}\right).$$

We note that $\lim_{x\to 0}(a^x - 1) = 0$ and $\lim_{x\to 0}(x) = 0$, so we have,

$$\lim_{x\to 0}\left(\frac{a^x - 1}{x}\right) = \lim_{x\to 0}\left(\frac{a^x \ln(a)}{1}\right) = \ln(a).$$

Where to take the derivative we note that if $y = a^x$ then $\ln(y) = x\ln(a)$ and on taking the derivative we have

$$\frac{1}{y}\frac{dy}{dx} = \ln(a)$$

and hence

$$\frac{dy}{dx} = a^x \ln(a).$$

Example 6.11. *Use l'Hôpital's rule to determine the following limit:*

$$\lim_{x\to 0}\left(\frac{\sin(x)}{x}\right).$$

We note that $\lim_{x\to 0}(\sin(x)) = 0$ *and* $\lim_{x\to 0}(x) = 0$*, so we have,*

$$\lim_{x\to 0}\left(\frac{\sin(x)}{x}\right) = \lim_{x\to 0}\left(\frac{\cos(x)}{1}\right) = 1.$$

Example 6.12. *Use l'Hôpital's rule to determine the following limit:*

$$\lim_{x\to 0}\left(\frac{2\sin(x)-\sin(2x)}{x-\sin(x)}\right).$$

We note that $\lim_{x\to 0}(2\sin(x)-\sin(2x)) = 0$ *and* $\lim_{x\to 0}(x-\sin(x)) = 0$*, so we have,*

$$\begin{aligned}\lim_{x\to 0}\left(\frac{2\sin(x)-\sin(2x)}{x-\sin(x)}\right) &= \lim_{x\to 0}\left(\frac{2\cos(x)-2\cos(2x)}{1-\cos(x)}\right)\\ &= \lim_{x\to 0}\left(\frac{-2\sin(x)+4\sin(2x)}{\sin(x)}\right)\\ &= \lim_{x\to 0}\left(\frac{-2\cos(x)+8\cos(2x)}{\cos(x)}\right)\\ &= 6.\end{aligned}$$

Example 6.13. *Use l'Hôpital's rule to determine the following limit:*

$$\lim_{x\to 0^+}\left(x\ln(x)\right).$$

We note that

$$x\ln(x) = \frac{\ln(x)}{1/x}$$

and as $\lim_{x\to 0^+}(\ln(x)) = \infty$ *and* $\lim_{x\to 0^+}(1/x) = \infty$*, so we have,*

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

---

l'Hôpital's rule is constructed upon the assumptions that

- $f(x)$ and $g(x)$ have well-defined Taylor expansions around the limit point, and

- the limit $\lim_{x\to x_0}\left(\frac{f'(x)}{g'(x)}\right)$ exists while $f(x_0)=0$ and $g(x_0)=0$ or $f(x_0)=\pm\infty$ and $g(x_0)=\pm\infty$.

Although l'Hôpital's rule seems like a panacea for all limits, the assumptions around its construction means one must take care when using it. For example consider the limit

$$\lim_{x\to 2}\left(\frac{2x^2-3x+1}{5x+4}\right)=\frac{3}{14}$$

as the quotient is well-defined at the limit point, if we had mistakenly applied l'Hôpital's rule we would have evaluated:

$$\lim_{x\to 2}\left(\frac{\frac{d}{dx}(2x^2-3x+1)}{\frac{d}{dx}(5x+4)}\right)=\lim_{x\to 2}\left(\frac{4x-3}{5}\right)=1$$

which is not the correct answer. Another example where the rule is not valid is the limit

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

If we had used l'Hôpital's rule we would have evaluated

$$\lim_{x\to\infty}\left(\frac{\frac{d}{dx}(x+\sin(x))}{\frac{d}{dx}(x)}\right)=\lim_{x\to\infty}\left(\frac{1+\cos(x)}{1}\right)$$

and this limit does not exist, so l'Hôpital's rule was not of use here and some other method is needed[3].

[3]We made use of $\lim_{x\to 0}(\frac{sin(1/x)}{1/x})=0$ to deduce the limit by using a change of the limit variable $y=\frac{1}{x}$.

Appendix

6.A The Ratio Test Implies the Root Test.

[4] Given a power series

$$S(x) \equiv \sum_{n=1}^{\infty} a_n x^n$$

we have two formulae for its radius of convergence[5]:

$$R_1 = \lim_{n\to\infty} (|a_n|^{-\frac{1}{n}}) \tag{6.1}$$

$$R_2 = \lim_{n\to\infty} \left|\frac{a_n}{a_{n+1}}\right|. \tag{6.2}$$

For different series it is useful to prefer one of these definitions for a radius of convergence over another. In this appendix we will show that if R_2 exists, then R_1 also exists and in that case $R_1 = R_2$ giving a single radius of convergence from the pair of definitions.

Let us commence by recalling the $\epsilon - \delta$ definition for the existence of a limit to equation 6.1 before putting the resulting expression in a helpful format.

If the limit defining R_2 exists then

$$|\left|\frac{a_n}{a_{n+1}}\right| - R_2| < \epsilon_2 \qquad \forall\, \epsilon_2 > 0, \quad n > N$$

where N is some sufficiently large value of $n \in \mathbb{Z}$ chosen such that the inequality is satisfied. Now it will be useful to consider the natural logarithm of R_2 in order to later compare it with R_1. We have,

$$\ln(R_2) \equiv \ln(\lim_{n\to\infty} \left|\frac{a_n}{a_{n+1}}\right|) = \lim_{n\to\infty} \Big(\ln|a_n| - \ln|a_{n+1}| \Big).$$

Now the definition of the existence of the limit give us:

$$|\ln|a_n| - \ln|a_{n+1}| - \ln(R_2)| < \epsilon \qquad \forall\, \epsilon > 0, \quad n > N. \tag{6.3}$$

Now to compare this with the limit R_1 we will also take the natural logarithm of R_1, so,

$$\ln(R_1) \equiv \ln(\lim_{n\to\infty} |a_n|^{-\frac{1}{n}}) = \lim_{n\to\infty} (\ln(|a_n|^{-\frac{1}{n}})) = \lim_{n\to\infty} (-\frac{1}{n} \ln|a_n|) \tag{6.4}$$

[4] The material in this appendix is provided for interest and is not an examinable part of the course.

[5] Recall that $|x| < R$ defines the values of x for which the power series is convergent and well-defined and R is the radius of convergence.

and by the definition of the existence of the limit we have,

$$|-\frac{1}{n}\ln|a_n| - \ln(R_1)| < \epsilon_1 \qquad \forall\,\epsilon_1 > 0, \quad n > N. \tag{6.5}$$

Our aim now is to show that equation (6.3) implies equation (6.5) and furthermore that $R_1 = R_2$.

Given equation (6.3) we may say that

$$\sum_{n=N}^{M} |\ln|a_n| - \ln|a_{n+1}| - \ln(R_2)| < \sum_{n=N}^{M} \epsilon = (M-N+1)\epsilon \tag{6.6}$$

where $M \in \mathbb{Z}$ and $M \geq N$. We may develop the left-hand-side of the equation above by using the triangle inequality (i.e. $|A+B| \leq |A| + |B|$), so that,

$$\begin{aligned}
\sum_{n=N}^{M} |\ln|a_n| - \ln|a_{n+1}| - \ln(R_2)| &\geq \left| \sum_{n=N}^{M} (\ln|a_n| - \ln|a_{n+1}| - \ln(R_2)) \right. \\
&= \Big| \ln|a_N| - \ln|a_{N+1}| + \ln|a_{N+1}| - \ln|a_{N+2}| + \ldots \\
&\qquad + \ln|a_M| - \ln|a_{M+1}| - (M-N+1)\ln(R_2) \Big| \\
&= \left| \ln|a_N| - \ln|a_{M+1}| - (M-N+1)\ln(R_2) \right|.
\end{aligned}$$

Hence from equation (6.6) we have,

$$\left| \ln|a_N| - \ln|a_{M+1}| - (M-N+1)\ln(R_2) \right| < (M-N+1)\epsilon.$$

Now we remark that as $M \geq N$ then $M - N \geq 0$ and $M - N + 1 > 0$. Also we note that $M - N + 1 = M - (N-1) \leq M$. It is also worth remarking that N is a fixed number, fixed by the definition in equation (6.3). Following these observations we see that if we divide the expression above by $M+1$ and take the limit as $M \to \infty$ we have:

$$\lim_{M\to\infty} \left| \frac{1}{M+1}\ln|a_N| - \frac{1}{M+1}\ln|a_{M+1}| - \frac{M-N+1}{M+1}\ln(R_2) \right| < \lim_{M\to\infty} \left(\frac{M-N+1}{M+1}\epsilon \right).$$

Since N is fixed, when we take the limit we find the expression above simplifies to

$$\lim_{M\to\infty} \left| -\frac{1}{M+1}\ln|a_{M+1}| \right| - \ln(R_2) < \epsilon.$$

In finding the above we have used $\lim_{M\to\infty} \left| \frac{1}{M+1}\ln|a_N| \right| = 0$ (N is fixed and so a_N is constant as $M \to \infty$) and $\lim_{M\to\infty} \left(\frac{M-N+1}{M+1} \right) = \lim_{M\to\infty} \left(1 - \frac{N}{M+1} \right) = 1$. Now from equation (6.4) we

have that,

$$\lim_{M\to\infty}\left|-\frac{1}{M+1}\ln|a_{M+1}|\right| = \ln(R_1),$$

hence,

$$\ln(R_1)-\ln(R_2) = \ln\left(\frac{R_1}{R_2}\right) < \epsilon \qquad \forall\,\epsilon.$$

This is true for all $\epsilon > 0$ which implies that $R_1 = R_2$. Implicitly our work is now done, as if R_2 exists then we have shown that it is equal to R_1 and hence R_1 exists. However we set out to show that equation (6.3) implies equation (6.5). This will be achieved in a simple step from the equation we have derived above:

$$\lim_{M\to\infty}\left|-\frac{1}{M+1}\ln|a_{M+1}|\right| - \ln(R_2) = \lim_{M\to\infty}\left|-\frac{1}{M+1}\ln|a_{M+1}| - \ln(R_2)\right| < \epsilon.$$

Therefore there exists $M_0 \in \mathbb{Z}^+$ such that for all $M > M_0$ then

$$\left|-\frac{1}{M+1}\ln|a_{M+1}| - \ln(R_2)\right| < \epsilon$$

which is equation (6.5) where $M+1 = n$, $M_0+1 = N$, $\epsilon = \epsilon_1$ and $R_2 = R_1$.

In this appendix we have shown that the ratio test for the convergence of a power series implies the root test for convergence.

7. Tutorial Exercises

CALCULUS 1 TUTORIAL EXERCISES I

In this tutorial you will practise working with injective, surjective and bijective functions. Questions marked with $\star$ should be submitted to your tutor in your tutorial class for marking and feedback.

Tutor's Example[1]

Let $f : A \to B$ be given by

$$f(x) = \frac{\sin x}{x}.$$

Identify the largest domain $A \in \mathbb{R}$ for which f is well-defined. Sketch the graph of f. If $B = (-\infty, n] \in \mathbb{R}$ identify the smallest value of n for which f is well-defined.

---

1. State the maximum domains and minimum ranges in $\mathbb{R}$ for which the following are well-defined functions:

   (a) $f(x) = \frac{(x+2)^2}{(x-2)^2}$

   (b) $g(x) = \frac{1}{x}$

   (c) $h(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & x = 0 \\ 1 & x > 0. \end{cases}$

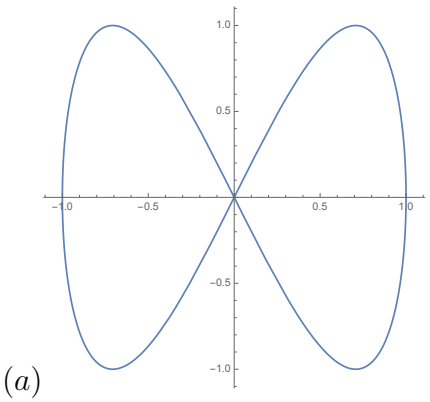
   The function $h(x)$ is a common definition of the Heaviside step function, named after Oliver Heaviside.

2. For each of the three functions in question 1, sketch the graph and qualitatively describe the slope of each graph - in particular state whether the slope is positive, negative or does not exist.

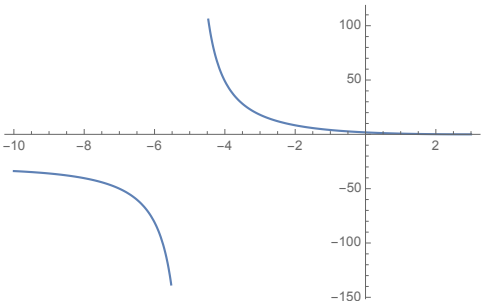
$\star$ 3. Which of the following graphs could correspond to well-defined functions? State whether the functions could be bijections. You may assume that the graphs are plotted over the full domain and range in each case (i.e. this means you can assume the range is such that

[1] To be shown by the tutor on the board at the start of the tutorial

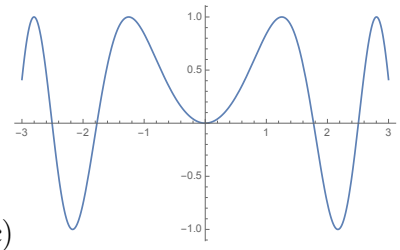
each function is surjective).



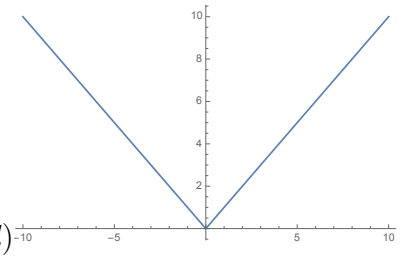
(a)



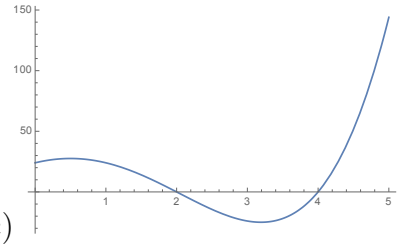
(b)



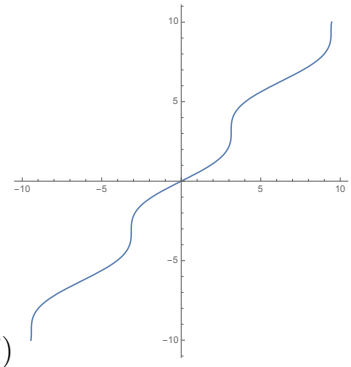
(c)



(d)



(e)



(f)

⋆ 4. Sketch the following functions and for each case state whether the function is injective, surjective, neither or if the function is not well-defined?

(a) $y : [3, \infty) \to \mathbb{R}$ given by $y(x) = x^2 - 6x + 9$.

(b) $y : \mathbb{R} \to \mathbb{R}$ given by $y(x) = (x - 2)^3$.

(c) $y : \mathbb{R}^+ \to \mathbb{R}^+$ given by $y(x) = (x - 2)(x + 1)(x - 3)$.

(d) $y : \mathbb{R} \to \mathbb{R}^+$ given by $y(x) = \sqrt{x^2 + 4x + 4}$.

(e) $y : \mathbb{R} \to \mathbb{R}$ given by $y(x) = 2^x - 3^x$.

(f) $y : \mathbb{R} \to \mathbb{R}$ given by $y(x) = x^4$.

(g) $y : \mathbb{R}^+ \to [1, \infty)$ given by $y(x) = x + 2^x$.

(h) $y : [3, \infty) \to \mathbb{R}$ given by $y(x) = x^4 + 2x^3 - 13x^2 - 14x + 24$.

5. *(For discussion)* A set S is countably infinite if it can be put in one-to-one correspondence with the natural numbers $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$, i.e. if a bijection $f : S \to \mathbb{N}$ can be constructed. Discuss whether the following are countably infinite sets:

   (a) The set of positive even integers.

   (b) The set of all integers $\mathbb{Z}$.

   (c) The set of ordered pairs of the natural numbers, given by $\{(n, m) \,|\, \forall\, n, m \in \mathbb{N}\}$ N.B. the order of the pair of natural numbers matters, so that $(1, 2) \neq (2, 1)$.

   (d) The set of rational numbers $\mathbb{Q} = \{\, p/q \,|\, \forall\, p, q \in \mathbb{Z} \text{ and } q \neq 0\}$.

   (e) The real numbers $\mathbb{R}$.

6. *Mathematica*[2] Graphs are plotted on Mathematica using the command `Plot[]`. For example if you open Mathematica, type

$$\texttt{Plot[x\^{}2, \{x, -5, 5\}]}$$

   then push the shift and the enter key simultaneously, you will get a plot of $f(x) = x^2$ for $x \in [-5, 5]$ within the real numbers. There are many more options which can be added to the command so that:

$$\texttt{Plot[x\^{}2, \{x, -5, 5\}, GridLines-> Automatic, PlotRange-> \{\{-10, 10\}, \{-50, 50\}\}]}$$

   draws the parabola $f(x) = x^2$ for the values of x in the domain $[-5, 5]$, on the axes with grid lines added in automatically and with horizontal axes running from -10 to 10 and vertical axes from -50 to 50. Practise using Mathematica to confirm your graph sketches of the functions in problem 1 (some research will be required to plot $h(x)$ in part (c)).

7. *A challenging problem*[3]. Consider the set of all lists of positive integers, L, i.e. the elements of L are lists of positive integers $\{n_1, n_2, \ldots, n_N\}$ and each $n_i \in \mathbb{Z}^+$, e.g. $\{1, 4, 6\}$, $\{4, 1, 6, 6, 6, 4\}$ and $\{7, 52, 23456, 3, 504, 10023\}$ are elements of L. Argue that L is countably infinite and construct a bijection from L to $\mathbb{N}$.

[2]Occasionally there will be a challenge to use Mathematica - these challenges are for you to try outside the tutorial class using the college's computer rooms.

[3]All challenging problems in the tutorial problem sets are beyond the scope of the course, and are not examinable, but they are, hopefully interesting and inspiring.

CALCULUS 1 TUTORIAL EXERCISES II

In this tutorial you will work with inverse functions, exponents, logarithms, polynomial functions and trigonometric functions. Questions marked with $\star$ should be submitted to your tutor in your tutorial class for marking and feedback.

Tutor's Example[4]

Consider the function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = (x+1)(x-2)(x+3).$$

Prove that the function is NOT a bijection. Identify domains (which together cover $\mathbb{R}$) over which $f(x)$ becomes invertible.

$\star$ 8. Identify domains (which together cover $\mathbb{R}$) over which $f : \mathbb{R} \to \mathbb{R}$ becomes invertible, where

(a) $f(x) = -x^3 + x^2 + 9x - 9$,

(b) $f(x) = \tan(x^2)$,

(c) $f(x) = 2^x + 2^{-x}$.

$\star$ 9. Find the value of x for which

$$4^{\frac{x}{y}+\frac{y}{x}} = 32$$

and

$$\log_3(x-y) + \log_3(x+y) = 1.$$

[Hat-tip: this problem was set at math10.com[5].]

$\star$ 10. For what value of n does $\log_2(3)\log_3(4)\log_4(5)\ldots\log_n(n+1) = 10$? [Hat-tip: this problem was set at math10.com[6].]

11. Use the definitions of the sine, cosine and tangent functions, to prove the following pair of trigonometric identities:

(i) $1 + \tan^2 x = \sec^2 x$ and

[4]To be shown by the tutor on the board at the start of the tutorial

[5]See `https://www.math10.com/problems/logarithmic-equation-problems/difficult/`

[6]See `https://www.math10.com/problems/logarithmic-equation-problems/difficult/`

(ii) $1 + \cot^2 x = \operatorname{cosec}^2 x$.

Both of these identities arise from using Pythagoras' theorem for right-angled triangles. Identify the right-angled triangles for which Pythagoras' theorem gives the identities above.

12. We have expressed e^x as the infinite sum $\sum_{k=0}^{\infty} \frac{x^n}{n!}$, but written it notationally as a power of Euler's number e. In this question show that the infinite sum expression of e^x obeys the following properties of exponents:

(i) $e^x e^y = e^{x+y}$

(ii) $e^{-x} = \frac{1}{e^x}$

You may wish to use the binomial expansion $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$.

13. *Mathematica.* Use Mathematica to plot $f(x) = \sum_{k=0}^{N} \frac{1}{k!} x^k$ for $N = 2$, 5 and 10 and compare your plot to the exponential function denoted by `Exp[x]` in Mathematica. Plot all four graphs on the same set of axes for the domain $x \in [-5, 5]$. N.B. The Mathematica command for a summation is `Sum[]` which takes two arguments: the expression to be summed and the range of the summation variable. For example the sum of the first 10 square numbers is found using `Sum[k^2, {k, 1, 10}]`.

14. *A challenging problem.* In the lecture notes we considered differential equations of the form

$$\frac{d^n}{dx^n}(f(x)) = Af(x).$$

When $A = 1$ and $n = 1$ the solution satisfying $f(0) = 1$ is the exponential function; when $A = 1$ and $n = 2$ the solutions are the hyperbolic functions ($f(x) = \cosh(x)$ when $f(0) = 1$ and $f'(0) = 0$ and $f(x) = \sinh(x)$ when $f(0) = 0$ and $f'(0) = 1$); when $A = -1$ and $n = 2$ the solutions are the trigonometric functions ($f(x) = \cos(x)$ when $f(0) = 1$ and $f'(0) = 0$ and $f(x) = \sin(x)$ when $f(0) = 0$ and $f'(0) = 1$). What are the solutions to the equation when

(i) $A = 1$, $n = 3$ for $f(0) = 1$, $f'(0) = 0$ and $f''(0) = 1$? And

(ii) $A = -1$, $n = 3$ for $f(0) = 1$, $f'(0) = 0$ and $f''(0) = 1$?

Give power series (infinite polynomial functions) for each solution.

CALCULUS 1 TUTORIAL EXERCISES III

In this tutorial you will practise working with the inverse trigonometric functions, the hyperbolic functions, trigonometric functions with complex arguments, double-angle formulae and composition of functions. Questions marked with ⋆ should be submitted to your tutor in your tutorial class for marking and feedback.

Tutor's Example[7]

Prove that

$$a\cos\theta + b\sin\theta = c\sin(\theta + \alpha)$$

for some $c, \alpha \in \mathbb{R}$ and $c \geq 0$.

⋆ 15. Express the following combinations of functions in the form $c\sin(\theta + \alpha)$, where c and α are real constants which you have to find (assume $|\beta| < \pi/2$):

(a) $\sin(\theta) + \cos(\theta)$ (b) $\sqrt{3}\sin(\theta) - \cos(\theta)$ (c) $\sin(\theta) - \tan(\beta)\cos(\theta)$

⋆ 16. Use the explicit formula for the function $\operatorname{arcsinh} : \mathbb{R} \to \mathbb{R}$, namely $\operatorname{arcsinh}(y) = \ln(|y + \sqrt{y^2+1}|)$, to show that

$$\operatorname{arcsinh}(\sinh(x)) = x \quad \text{for all} \quad x \in \mathbb{R}$$
$$\sinh(\operatorname{arcsinh}(y)) = y \quad \text{for all} \quad y \in \mathbb{R}$$

17. Use the exponential forms for the sinh and cosh functions to derive an analytic expression for arctanh .

⋆ 18. Show that $\cos(\pi/12) = \frac{1}{2}\sqrt{2+\sqrt{3}}$.
(Hint: use the formula for $\cos(2\theta)$ in terms of $\cos(\theta)$)

19. Find formulae that express the hyperbolic functions $\sinh(x)$, $\cosh(x)$ and $\tanh(x)$ as ratios of polynomials of $t = \tanh(x/2)$.

[7] To be shown by the tutor on the board at the start of the tutorial

⋆ 20. Find all solutions $\theta \in \mathbb{R}$ (if any) of the equation $\cot(\theta) + \tan(\theta) = \alpha$, for the following three cases:

(i) $\alpha = 1$ (ii) $\alpha = 2$ (iii) $\alpha = 4$

21. Let $\alpha, \beta \in \mathbb{R}$. Rewrite each of the following expressions as a constant times a product of trigonometric functions:
(i) $\sin(\alpha - \beta) + \sin(\alpha + \beta)$ (ii) $\cos(\alpha - \beta) - \cos(\alpha + \beta)$

22. Find the values of

(a) $\arcsin(\sin(5\pi/6))$ (b) $\arcsin(\sin(7\pi/6))$ (c) $\arccos(\cos(7\pi/6))$
(d) $\arccos(\cos(11\pi/6))$

23. Let $f : \mathbb{R} \to \mathbb{R}$ be such that $f(x) = x^3$, and $g : \mathbb{R} \to \mathbb{R}$ be such that $g(x) = x + 2$.

(a) What is the formula for $f \circ g$?

(b) What is the formula for $g \circ f$?

(c) For which values of $x \in \mathbb{R}$ are the formulae equal? (Hint: you will need to solve a quadratic.)

(d) Are the formulae equal for all $x \in \mathbb{R}$?

(e) Prove that function composition is <u>not commutative</u> in general, i.e. $f \circ g \neq g \circ f$ (Hint: find a counter-example f and g such that there is some x such that $(f \circ g)(x) \neq (g \circ f)(x)$).

24. *A challenging problem.* Assuming that

$$2x(y^2 - 1) + 2y(x^2 - 1) = (1 + x^2)(1 + y^2)$$

and

$$4z(1 - y^2) + 4y(1 - z^2) = (1 + z^2)(1 + y^2)$$

where $x, y, z \in \mathbb{R}$, find the value of the following expression:

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

Hint:Use the substitutions $x = \tan\alpha$, $y = \tan\beta$ and $z = \tan\gamma$ to rewrite the expressions.

CALCULUS 1 TUTORIAL EXERCISES IV

In this tutorial you will work with the limit, continuous functions, the intermediate value theorem, asymptotic limits and the algebra of limits. Questions marked with $\star$ should be submitted to your tutor in your tutorial class for marking and feedback.

Tutor's Example[8]

Find the value of k so that

$$f(x) = \begin{cases} \frac{x^2-4}{x-2} & x < 2 \\ (x+1)^2 + k & x \geq 2 \end{cases}$$

is a continuous function.

$\star$ 25. For each function below identify the value of k that makes the function continuous:

(a) $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = \begin{cases} \frac{x^2+2x-15}{x-3} & x < 3 \\ x^2 - kx + 1 & x \geq 3. \end{cases}$

(b) $f : [-3, 3] \to \mathbb{R}$ given by $f(x) = \begin{cases} \frac{3x-6}{k(x^3+2x^2-11x+6)} & x < 2 \\ \frac{7x}{12} - 1 & x \geq 2. \end{cases}$

(c) $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = \begin{cases} 2e^{3x} & x \leq 0 \\ \frac{\sin(kx)}{x} & x > 0. \end{cases}$

$\star$ 26. If they exist, evaluate the following asymptotic limits:

(a) $\lim_{x\to\infty} \left[\frac{27x^{26}+24x^{17}-5}{9x^{26}-27x^{17}-5}\right]$

(b) $\lim_{x\to-\infty} \left[- x \sin\left(\frac{1}{5x}\right)\right]$.

(c) $\lim_{y\to\infty} \left[e^{-2yx} \tanh(4y)\right]$ for $x > 0$.

27. Let $n \in \mathbb{Z}^+$, and calculate the limit $\lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n$.
Hint: substitute $n^{-1} = x$ and try to use limits calculated in the lectures.

[8] To be shown by the tutor on the board at the start of the tutorial

28. Calculate the following limits, if they exist, without using power series:

(a) $\lim_{x\to 0} \frac{1-\cos(x)}{x^2}$ (b) $\lim_{x\to 0} \dfrac{\sin(7x)}{\tan(2x)}$

(c) $\lim_{x\to 0} \frac{\sin(x^2)}{x}$ (d) $\lim_{x\to 0} \cosh(x)$

29. Calculate the following limits, if they exist, using the power series representations of the hyperbolic functions:

(a) $\lim_{x\to 0} \frac{\sinh(x)}{x}$ (b) $\lim_{x\to 0} \dfrac{1-\cosh(x)}{x}$

(c) $\lim_{x\to 0} \frac{1-\cosh(x)}{x^2}$ (d) $\lim_{x\to 0} \dfrac{\tanh(x)}{x}$

30. Calculate the following limits by making clever substitutions:

(a) $\lim_{x\to \pi} \frac{\sin(x)}{\pi-x}$ (b) $\lim_{x\to 0} \dfrac{\arcsin(x)}{x}$

(c) $\lim_{x\to 0} \frac{\arcsin(3x)}{\tan(5x)}$ (d) $\lim_{x\to 0} \dfrac{\operatorname{arctanh}(x)}{x}$

31. *A challenging problem*[9]. Consider a fixed circle C_1 defined by the equation

$$(x-1)^2 + y^2 = 1$$

and a shrinking circle C_2 with radius r and center the origin. P is the point $(0, r)$, Q is the upper point of intersection of the two circles, and R is the point of intersection of the line PQ and the x-axis. What happens to R as C_2 shrinks, that is, as $r \to 0^+$?

[9]This problem is perhaps not so challenging but the result is interesting - I first saw this question on the blog `https://mrchasemath.wordpress.com/2010/02/23/really-fun-limit-problem/` where you can find access to a Java applet which will help you understand the solution.

CALCULUS 1 TUTORIAL EXERCISES V

In this tutorial you will do some more work with limits, deriving fundamental limits using the sandwich theorem and simplifying complicated limit problems. Questions marked with $\star$ should be submitted to your tutor in your tutorial class for marking and feedback.

Tutor's Example[10]

Use the sandwich theorem to calculate the limit

$$\lim_{x\to 0}\left(x\sin\left(\frac{1}{x}\right)\right).$$

32. Use the sandwich theorem to show that $\lim_{x\to 0}(x^4\cos(2/x)) = 0$.

33. Evaluate the following limits

    (a) $\lim_{x\to 0}\left(\frac{\sin(5x)}{x}\right)$

    (b) $\lim_{x\to 0}\left(\frac{\sin(3x)}{\sin(8x)}\right)$

    (c) $\lim_{x\to 0}\left(\frac{\cot(5x)}{\operatorname{cosec}(2x)}\right)$

    (d) $\lim_{x\to 0}\left(\frac{\tan(x)}{x}\right)$

    (e) $\lim_{x\to \pi/4}\left(\frac{\sin(x)-\cos(x)}{\cos(2x)}\right)$

34. Let $f:\mathbb{R}\to\mathbb{R}$ be a function satisfying the inequality

$$4x-9 \le f(x) \le x^2-4x+7$$

    for all $x \ge 0$. Compute $\lim_{x\to 4}(f(x))$.

35. Let $f:\mathbb{R}\to\mathbb{R}$ be a function satisfying the inequality

$$\frac{\sin((27-9\sqrt{5})(x-\sqrt{5}))}{3(x-\sqrt{5})} \le f(x) \le \frac{x^4-3x^3-x^2+15x-20}{(x-\sqrt{5})(x+\sqrt{5})}$$

    for all $x \in \mathbb{R}$. Compute $\lim_{x\to\sqrt{5}}(f(x))$.

$\star$ 36. Calculate the following limits, if they exist, using any suitable method:

[10] To be shown by the tutor on the board at the start of the tutorial

(a) $\lim_{x\to\infty} \frac{\sin(x)}{x}$

(b) $\lim_{x\to\infty} \frac{\cos(x)}{x^2}$

(c) $\lim_{x\to\infty} x(e^{1/x}-1)$

(d) $\lim_{x\to 0^+} x^x$

(e) $\lim_{x\to\infty} \frac{\ln(x)}{x^2+1}$

(f) $\lim_{x\to 0^+} \frac{x^x-1}{x\ln(x)}$

(g) $\lim_{x\to 0^+} e^{1/x}$

(h) $\lim_{x\to 0^+} x^{\sin(x)}$

(i) $\lim_{x\to 0} \frac{a^x-b^x}{x}$ where $a>b>0$

(j) $\lim_{x\to 0^+} \frac{(x+1)\ln(x)}{\sin(x)}$

(k) $\lim_{x\to 0^-} e^{1/x}$

(l) $\lim_{x\to\infty} xe^{1/x}$

37. *Mathematica* Use Mathematica to check your answers to problem 33. The Mathematica command for finding a limit is `Limit[]` and it takes at least two arguments, the first being the expression to be evaluated under the limit and the second entry showing the value the variable approaches in the limit. For example to compute $\lim_{x\to 0}(\frac{\sin(x)}{x})$ one types

$$\texttt{Limit[Sin[x]/x, x-> 0]}$$

before pushing shift and enter simultaneously to find the result.

38. *A challenging problem.* Compute the limit

$$\lim_{x\to 0}\left(\frac{1}{\sin^2(x)}+\frac{1}{\tan^2(x)}-\frac{2}{x^2}\right).$$

CALCULUS 1 TUTORIAL EXERCISES VI

In this tutorial you will revise and work with the derivative. Questions marked with $\star$ should be submitted to your tutor in your tutorial class for marking and feedback.

Tutor's Example[11]

Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = \sqrt{1 + \sin(x)})$. Compute $\frac{df}{dx}$ and sketch the graph of $f(x)$ for $x \in [-\pi, \pi]$. Is the function differentiable for $x \in [-\pi, \pi]$?

$\star$ 39. Calculate the following derivatives:

(a) $\frac{d}{dx}\ e^{-x^2}$

(b) $\frac{d}{dx}\ \ln(\tan(x))$

(c) $\frac{d}{dx}\ (x\ln(x) - x)$

(d) $\frac{d}{dx}\ \arcsin(x^2)$

(e) $\frac{d}{dx}\arctan(e^x)$

(f) $\frac{d}{dx}\ e^{\sin(x)}$

(g) $\frac{d}{dx}\ xe^{\arctan(x)}$

(h) $\frac{d}{dx}\ \ln|\arcsin(x)|$

$\star$ 40. Calculate the following derivatives:

(a) $\frac{d}{dx}\ \arcsin\left(\frac{1-x}{1+x}\right)$

(b) $\frac{d}{dx}\ 3^{\sin(x)}$

(c) $\frac{d}{dx}\ \ln|x + \sqrt{x^2 - a^2}|$

(d) $\frac{d}{dx}\ \frac{\sqrt{1+x}}{x}$

(e) $\frac{d}{dx}\ \frac{1}{x}\arccos\left(\sqrt{1 - x^2}\right)$

(f) $\frac{d}{dx}\ x^{x\sin(x)}$

(g) $\frac{d}{dx}\ \ln\left(\cosh(x) + \sinh(x)\right)$

(h) $\frac{d}{dx}\ \ln\left(\sqrt{(1 + x^2)/(1 - x^2)}\right)$

41. Let $f^n(x)$ denote the $n'th$ derivative with respect to x of the function $f : \mathbb{R} \to \mathbb{R}$, where $n \in \mathbb{Z}^+$. Calculate $f^n(x)$ at $x = 0$ when

(a) $f(x) = e^x$.

(b) $f(x) = \sin(x)$.

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

42. Differentiate $\sin(e^{\arcsin(x)})$ with respect to x.

43. Use the result $\frac{d}{dx}(\ln(x)) = \frac{1}{x}$ to find the derivative of x^x with respect to x for $x > 0$.

[11] To be shown by the tutor on the board at the start of the tutorial

44. Consider the hyperbola defined as the set of points satisfying $x^2 - y^2 = 16$. Let $A = (-5, y_A)$, $B = (5, y_B)$ and C the point where the tangent to the hyperbola at A meets the tangent to the hyperbola at B. Compute the area of the triangle ABC.

45. *Mathematica.* Computation of derivatives in Mathematica uses the operation `D[]`. For example, the operation `D[x^2+3x+2,x]` computes the derivative of x^2+3x+2 with respect to x. To compute higher derivatives Mathematica uses the formulation `D[f,{x,n}]` to give $\frac{d^n f}{dx^n}$. Use Mathematica to check your answers to questions 40, 41 part (c) and 42.

46. *A challenging problem.*[12] Show that the function

$$f(x) = \begin{cases} x^4(2 + \sin\left(\frac{1}{x}\right)) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

has an extreme value at a point where the sign of the derivative does not make a simple change.

[12]This problem is, perhaps, less challenging than it is interesting. The function is taken from the marvellous book "Counterexamples in Analysis" by Bernard R. Gelbaum and John M. H. Olmstead.

CALCULUS 1 TUTORIAL EXERCISES VII

In this tutorial you will work with derivatives of implicit functions, inverse functions, parametric functions, and the mean value theorem. Questions marked with ⋆ should be submitted to your tutor in your tutorial class for marking and feedback.

Tutor's Example[13]

The curve defined as the set of points (x, y) satisfying the equation

$$y^2 = x^3 + x^2$$

is an example of an *elliptic curve*. Find an expression for $\frac{dy}{dx}$ and use the result to sketch the curve.

⋆ 47. Each of the following is an equation that determines y as an implicit function of x. Find in all cases an expression for dy/dx.

(a) $x^2 + y^2 = 1$ (b) $y^3 + x^3 = 1$

(c) $\sinh(x) + \cosh(y) = 1$ (d) $xy - e^{x+y} = 2$

(e) $\sin(y) + y = x^3$ (f) $y^2 + x(x-1)(x+1) = 0$

⋆ 48. Each of the following are a pair of equations which determine x and y in terms of a parameter $t \in \mathbb{R}$, which defines a function $y(x)$ implicitly. Find in all cases an expression for dy/dx.

(a) $x = \cos(t),\ y = \sin(t)$ (b) $x = \cosh(t),\ y = \sinh(t)$

(c) $x = t + \sin(t),\ y = \cos(t)$ (d) $x = t^3,\ y = t^2$

(e) $x = e^{2t},\ y = \tanh(t)$ (f) $x = e^t \cos(t),\ y = e^t \sin(t)$

49. Consider the function $y(x)$ defined parametrically through

$$x(t) = e^{t+1} + 1, \qquad y(t) = e^{t^2}$$

where $t \in \mathbb{R}$ is the curve parameter.

[13] To be shown by the tutor on the board at the start of the tutorial

(a) Find an expression for $\frac{dy}{dx}$ in terms of t from this parametric form.

(b) Find an explicit expression for y in terms of x that no longer involves t.

(c) Use the result of (b) to find $\frac{dy}{dx}$ in terms of x, and verify that this agrees with what you got in (a).

50. Find all values c for which the mean value theorem is satisfied for the functions:

(a) $f(x) = 2x^3 + 5x^2 - 23x + 10$ on $[1, 2]$

(b) $f(x) = \sin(x)$ on $[\pi/4, 3\pi/4]$.

(c) $f(x) = \tanh(x)$ on $[0, 1]$ (give your answer as a function of e).

51. Use both the intermediate value theorem and the mean value theorem to prove that $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^3 - 2x^2 + 4x - 8$ has exactly one real root.

52. *Mathematica* The plotting of graphs of parametric functions can be achieved in Mathematica by using `ParametricPlot[]`. The operation takes a pair of arguments, the first being the set of coordinates $(x(t), y(t))$ for the function and the second being the range of the parametric variable t. For example `ParametricPlot[{Cos[t],Sin[t]},{t,0,Pi}` plots the semicircle given by the coordinates $x = \cos(t)$, $y = \sin(t)$ for for $t \in [0, \pi]$. Practise using this command by plotting the parametric curves in question 48, you will wish to choose a range of t that gives an informative plot, rather than using $t \in \mathbb{R}$.

53. *A challenging problem...* Show that

$$\frac{d^n}{dx^n}(uv) = \sum_{k=0}^{n} \binom{n}{k} \frac{d^n u}{dx^n} \frac{d^{n-k} v}{dx^{n-k}}$$

and use this formula to find expressions constraining $f^n(0)$ where $f(x) = \tan(x)$ (and $f^n(x) = \frac{d^n f}{dx^n}$). As far as possible, use these relations to find a closed expression for $f^n(0)$.

CALCULUS 1 TUTORIAL EXERCISES VIII

In this tutorial you will work with the integral, in particular the Riemann integral and the fundamental theorems of calculus. Questions marked with ⋆ should be submitted to your tutor in your tutorial class for marking and feedback.

Tutor's Example[14]

Evaluate the Riemann integral of the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2 - 4x + 4$ on the interval $[-1, 4]$ from first principles by constructing appropriate staircase functions and applying the sandwich theorem.

54. Construct staircase functions and then evaluate from first principles the Riemann integral of $f : \mathbb{R} \to \mathbb{R}$ given by

    (a) $f(x) = \exp(x)$ on the interval $[0, 1]$.

    (b) $f(x) = x^3 - 6x^2 + 11x - 6$ on the interval $[1, 3]$.

    (c) $f(x) = \cosh(x)$ on $[a, b]$ where $a, b \in \mathbb{R}$ with $a < b$.

⋆ 55. Find a 'primitive' for each of the following functions, i.e. a function $F(x)$ such that $F'(x) = f(x)$. Prove your claims.

(a)	$f(x) = 1/(1+x)$	(b)	$f(x) = 1/(1-x)^2$
(c)	$f(x) = 1/(2x+1)^3$	(d)	$f(x) = 1/(1+x^2)$
(e)	$f(x) = 1/(1-x^2)$	(f)	$f(x) = 1/\sqrt{1-x^2}$
(g)	$f(x) = 1/\sqrt{1+x^2}$	(h)	$f(x) = xe^{-\frac{1}{2}x^2}$

56. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a continuous function and that $g : \mathbb{R} \to \mathbb{R}$ is a differentiable function. Let $F(x) = \int_0^{g(x)} f(y)dy$. Calculate $F'(x)$.

57. Let $F(x) = \int_1^{x^2} y^{-1}dy$. What is $F'(x)$?

[14] To be shown by the tutor on the board at the start of the tutorial

⋆ 58. Let $f : \mathbb{R} \to \mathbb{R}$ be given by

$$f(x) = \begin{cases} x^2 & 0 \leq x \leq 1 \\ 1 & 1 < x < 5 \\ (x+6)^2 & 5 \leq x \leq 6 \end{cases}$$

(a) Sketch the graph of $f(x)$ for $x \in [0, 6]$.

(b) Evaluate $\int_0^6 f(x)dx$.

(c) Let $F(x) = \int_0^x f(y)dy$. Find an explicit expression for $F(x)$ for $x \in [0, 6]$ and sketch the graph of F.

(d) Find $F'(x)$ for $x \in [0, 6]$.

59. Use the fundamental theorem of calculus to show that, upon substitution of $x = x(t)$,

$$\int_{x_1}^{x_2} f(x)dx = \int_{t_1}^{t_2} f(x(t))\frac{dx}{dt}dt.$$

60. *Mathematica.* Mathematica evaluates standard integrals using the operation `Integrate[]`. For example, to find the primitive of x^2 one evaluates `Integrate[x^2,x]`, the first argument gives the function to be integrated and the second argument gives the variable to integrate against. For definite integrals one must also specify a range for the variable, so that to find $\int_2^3 x^2 dx$ one evaluates `Integrate[x^2,{x,2,3}]`. Use Mathematica to check your answers to problem 55.

61. *A challenging problem.*[15] Calculate the following indefinite integral:

$$\int \frac{1}{1+x^4}dx.$$

[15]This problem was suggested by Mr. Joshua Brazier.

CALCULUS 1 TUTORIAL EXERCISES IX

In this tutorial you will revise further techniques to integrate functions and some applications of the integral. Questions marked with ⋆ should be submitted to your tutor in your tutorial class for marking and feedback.

Tutor's Example[16]

Use integration by parts (twice) to calculate the indefinite integral

$$\int x^2 \cos(2x)dx.$$

⋆ 62 Calculate the following indefinite integrals using the method of substitution:

(a) $\displaystyle\int \frac{e^x dx}{1+e^x}$ put $x = \ln(y)$

(b) $\displaystyle\int \sqrt{1-x^2}dx$ put $x = \sin(y)$

(c) $\displaystyle\int \frac{xdx}{\sqrt{1-x^2}}$ put $x^2 = 1-y$

(d) $\displaystyle\int \frac{dx}{\sqrt{a^2+x^2}}$ put $x = at$

(e) $\displaystyle\int \frac{2x^3 dx}{1+x}$ put $x = t-1$

(f) $\displaystyle\int \frac{xdx}{1+\sqrt{x}}$ put $x = t^2$

63 Calculate the following indefinite integrals using the method of substitution:

(a) $\int \cos(\sin(x))\cos(x)dx$ (b) $\displaystyle\int \frac{2xdx}{1+x^4}$

(c) $\int \frac{e^x dx}{e^{2x}-1}$ (d) $\displaystyle\int \frac{dx}{2\sqrt{x}\sqrt{1-x}}$

(e) $\int \frac{xdx}{\sqrt{1-x^2}}$ (f) $\displaystyle\int \frac{x^2 dx}{\sqrt{1-x^2}}$

(g) $\int \frac{dx}{(1-x^2)^{3/2}}$ (h) $\displaystyle\int \frac{xdx}{(1-x^2)^{3/2}}$

(i) $\int \frac{x^2 dx}{(1-x^2)^{3/2}}$ (j) $\displaystyle\int \frac{dx}{\cos(x)}$

[16]To be shown by the tutor on the board at the start of the tutorial

⋆ 64. Calculate the following indefinite integrals using integration by parts:

(a) $\int x\sinh(x)dx$ (b) $\int (1+x^2)e^x dx$

(c) $\int \arcsin(x)dx$ (d) $\int x^3\sin(x)dx$

(e) $\int x^2\Big(\ln(x)\Big)^2 dx$ (f) $\int x\arctan(x)dx$

⋆ 65. Define the indefinite integrals $I_n(x)$ for $n \in \mathbb{Z}^+$ as follows:

$$I_n(x) = \int \frac{dx}{(1+x^2)^n}$$

Use integration by parts to derive the following recursion formula:

$$I_n(x) = \frac{x}{2(n-1)(1+x^2)^{n-1}} + \frac{2n-3}{2n-2}\, I_{n-1}(x)$$

Use the recursion formula to find the indefinite integral $\int (1+x^2)^{-2}dx$.

66. Define the indefinite integrals $I_n(x)$ for $n \in \mathbb{Z}$ as $I_n(x) = \int \cos^n(x)dx$. Use integration by parts to derive the following recursion formula:

$$I_n(x) = \frac{1}{n}\sin(x)\cos^{n-1}(x) + \frac{n-1}{n}\, I_{n-2}(x)$$

Use the recursion formula to find the indefinite integral $\int \cos^4(x)dx$.

67. Use the method of partial fractions to calculate the following integrals:

(a) $\int \frac{dx}{x^2+x-6}$ (b) $\int \frac{xdx}{(1-x)^2(1+x)}$

(c) $\int \frac{dx}{2x^2+x-1}$ (d) $\int \frac{xdx}{12x^2-7x-12}$

68. *A challenging problem.* (Taken from the Indian Institute of Technology Joint Entrance Exam in 2006.)
Evaluate the following

$$5050\frac{\int_0^1 (1-x^{20})^{100}dx}{\int_0^1 (1-x^{20})^{101}dx}.$$

[Hint: commence by considering $I_n = \int_0^1 (1-x^m)^n dx$.]

CALCULUS 1 TUTORIAL EXERCISES X

In this tutorial you will revise the radius of convergence and Taylor (and Maclaurin) series. Questions marked with ⋆ should be submitted to your tutor in your tutorial class for marking and feedback.

Tutor's Example[17]

Find the radius of convergence for the power series

$$\sum_n \frac{x^n}{3^n}.$$

69. Calculate the length L of the spiral in the plane described by the parametric equations $y(t) = vt\cos(\omega t)$ and $x(t) = vt\sin(\omega t)$ from $t = 0$ to $t = \tau$, where v, ω are constants.

⋆ 70. Calculate the length L of the curve in the plane described by the function $y = \ln(\cos(x))$, between the points $x = -\pi/4$ and $x = \pi/4$.

71. Find out for the following series whether they are convergent or divergent, using the various results stated and/or derived in the lectures or otherwise:

(a) $\sum_n n^{-3}$ (b) $\sum_n (-1)^n n^{-2}$

(c) $\sum_n 1/\sqrt{n}$ (d) $\sum_n n^\alpha e^{-n}$ $(\alpha > 0)$

⋆ 72. Find the radii of convergence for the following power series:

(a) $\sum_n x^n/n^3$ (b) $\sum_n (-1)^n x^n/n^2$

(c) $\sum_n x^n/\sqrt{n}$ (d) $\sum_n x^n n^\alpha e^{-n}$ $(\alpha > 0)$

[17] To be shown by the tutor on the board at the start of the tutorial

73. Derive the Taylor expansions for the following functions, up to order $N = 3$ about the point $x = 0$ for the first three functions and up to order N for the last one, and give in each case an exact expression for the remainder term:

(a) $f(x) = \tan(x)$ (b) $f(x) = \tanh(x)$

(c) $f(x) = x^3$ (d) $f(x) = \sqrt{1+x}\, e^{\sin(x)}$

⋆ 74. Write the following two functions as power series of the form $\sum_{n=0}^{\infty} b_n x^n$, and determine for each the associated radius of convergence:

(a) $f(x) = 1/(1-x)^2$ (b) $f(x) = \arctan(x)$

Substitute $x = 1$ into your result under (b) (why is it not obvious that this is allowed?) and derive an expression for the number π as a series (the so-called Leibniz series).

75. Find the first three terms (i.e. up to order x^3) of the Taylor expansions about $x = 0$ for the following two functions, by combining and/or manipulation the Taylor expansions of other functions that you know:

(a) $f(x) = \tanh(x)$ (b) $f(x) = \ln\left(\dfrac{1+2x}{1-2x}\right)$

76. Show that for $x \in \mathbb{R}$ the series representation $e^x = \sum_{n=0}^{\infty} x^n/n!$ of the exponential function has the following properties (you may assume that you can interchange differentiation and summation):

$$\frac{d}{dx}e^{ax} = ae^{ax} \qquad e^0 = 1$$

77. *A challenging problem...* Show that the function defined by

$$f(x) = \begin{cases} 0 & x \leq 0 \\ e^{-\frac{1}{x}} & x > 0 \end{cases}$$

is a continuous and a differentiable function, whose first and second derivatives at $x = 0$ are both zero. You will need to use differentiation from first principles to find this result. By considering the form of the higher derivatives (but without computing them explicitly) argue that all the derivatives of this function at $x = 0$ are zero. With the Maclaurin series for $f(x)$ in mind, comment on whether or not you find these results surprising.