Advanced Calculus and Analysis
MA1002

Ian Craw

ii

April 13, 2000, Version 1.3

Additional copies may be obtained from:

Department of Mathematical Sciences
University of Aberdeen
Aberdeen AB9 2TY

DSN: mth200-101982-8

Foreword

These Notes

The notes contain the material that I use when preparing lectures for a course I gave from the mid 1980's until 1994; in that sense they are *my* lecture notes.

> "Lectures were once useful, but now when all can read, and books are so numerous, lectures are unnecessary." *Samuel Johnson, 1799.*

Lecture notes have been around for centuries, either informally, as handwritten notes, or formally as textbooks. Recently improvements in typesetting have made it easier to produce "personalised" printed notes as here, but there has been no fundamental change. Experience shows that very few people are able to use lecture notes as a *substitute* for lectures; if it were otherwise, lecturing, as a profession would have died out by now.

These notes have a long history; a "first course in analysis" rather like this has been given within the Mathematics Department for at least 30 years. During that time many people have taught the course and all have left their mark on it; clarifying points that have proved difficult, selecting the "right" examples and so on. I certainly benefited from the notes that Dr Stuart Dagger had written, when I took over the course from him and this version builds on that foundation, itslef heavily influenced by (Spivak 1967) which was the recommended textbook for most of the time these notes were used.

The notes are written in LATEX which allows a higher level view of the text, and simplifies the preparation of such things as the index on page 101 and numbered equations. You will find that most equations are not numbered, or are numbered symbolically. However sometimes I want to refer back to an equation, and in that case it is numbered within the section. Thus Equation (1.1) refers to the first numbered equation in Chapter 1 and so on.

Acknowledgements

These notes, in their printed form, have been seen by many students in Aberdeen since they were first written. I thank those (now) anonymous students who helped to improve their quality by pointing out stupidities, repetitions misprints and so on.

Since the notes have gone on the web, others, mainly in the USA, have contributed to this gradual improvement by taking the trouble to let me know of difficulties, either in content or presentation. As a way of thanking those who provided such corrections, I endeavour to incorporate the corrections in the text almost immediately. At one point this was no longer possible; the diagrams had been done in a program that had been 'subsequently "upgraded" so much that they were no longer useable. For this reason I had to withdraw the notes. However all the diagrams have now been redrawn in "public

iv

domaian" tools, usually `xfig` and `gnuplot`. I thus expect to be able to maintain them in future, and would again welcome corrections.

Ian Craw
Department of Mathematical Sciences
Room 344, Meston Building
email: `Ian.Craw@maths.abdn.ac.uk`
www: `http://www.maths.abdn.ac.uk/~igc`
April 13, 2000

Contents

List of Figures

Chapter 1

Introduction.

This chapter contains reference material which you should have met before. It is here both to remind you that you have, and to collect it in one place, so you can easily look back and check things when you are in doubt.

You are aware by now of just how sequential a subject mathematics is. If you don't understand something when you first meet it, you usually get a second chance. Indeed you will find there are a number of ideas here which it is essential you now understand, because you will be using them all the time. So another aim of this chapter is to repeat the ideas. It makes for a boring chapter, and perhaps should have been headed "all the things you hoped never to see again". However I am only emphasising things that you will be using in context later on.

If there is material here with which you are not familiar, don't panic; any of the books mentioned in the book list can give you more information, and the first tutorial sheet is designed to give you practice. And ask in tutorial if you don't understand something here.

1.1 The Need for Good Foundations

It is clear that the calculus has many outstanding successes, and there is no real discussion about its viability as a theory. However, despite this, there are problems if the theory is accepted uncritically, because naive arguments can quickly lead to errors. For example the chain rule can be phrased as

$$\frac{df}{dx} = \frac{df}{dy}\frac{dy}{dx},$$

and the "quick" form of the proof of the chain rule — cancel the dy's — seems helpful. However if we consider the following result, in which the pressure P, volume V and temperature T of an enclosed gas are related, we have

$$\frac{\partial P}{\partial V}\frac{\partial V}{\partial T}\frac{\partial T}{\partial P} = -1, \tag{1.1}$$

a result which certainly does not appear "obvious", even though it is in fact true, and we shall prove it towards the end of the course.

Another example comes when we deal with infinite series. We shall see later on that the series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \frac{1}{9} - \frac{1}{10} \cdots$$

adds up to $\log 2$. However, an apparently simple re-arrangement gives

$$\left(1 - \frac{1}{2}\right) - \frac{1}{4} + \left(\frac{1}{3} - \frac{1}{6}\right) - \frac{1}{8} + \left(\frac{1}{5} - \frac{1}{10}\right) \cdots$$

and this clearly adds up to half of the previous sum — or $\log(2)/2$.

It is this need for care, to ensure we can *rely* on calculations we do, that motivates much of this course, illustrates why we emphasise accurate argument as well as getting the "correct" answers, and explains why in the rest of this section we need to revise elementary notions.

1.2   The Real Numbers

We have four infinite sets of familiar objects, in increasing order of complication:

$\mathbb{N}$ — **the Natural numbers** are defined as the set $\{0, 1, 2, \dots, n, \dots\}$. Contrast these with the **positive integers**; the same set without 0.

$\mathbb{Z}$ — **the Integers** are defined as the set $\{0, \pm 1, \pm 2, \dots, \pm n, \dots\}$.

$\mathbb{Q}$ — **the Rational numbers** are defined as the set $\{p/q : p, q \in \mathbb{Z}, q \neq 0\}$.

$\mathbb{R}$ — **the Reals** are defined in a much more complicated way. In this course you will start to see why this complication is necessary, as you use the distinction between $\mathbb{R}$ and $\mathbb{Q}$.

Note: We have a natural inclusion $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$, and each inclusion is proper. The only inclusion in any doubt is the last one; recall that $\sqrt{2} \in \mathbb{R} \setminus \mathbb{Q}$ (i.e. it is a real number that is not rational).

One point of this course is to illustrate the difference between $\mathbb{Q}$ and $\mathbb{R}$. It is subtle: for example when computing, it can be ignored, because a computer always works with a rational approximation to any number, and as such can't distinguish between the two sets. We hope to show that the complication of introducing the "extra" reals such as $\sqrt{2}$ is worthwhile because it gives simpler results.

Properties of $\mathbb{R}$

We summarise the properties of $\mathbb{R}$ that we work with.

Addition: We can add and subtract real numbers exactly as we expect, and the usual rules of arithmetic hold — such results as $x + y = y + x$.

Multiplication: In the same way, multiplication and division behave as we expect, and interact with addition and subtraction in the usual way. So we have rules such as $a(b + c) = ab + ac$. Note that we can divide by any number except 0. We make no attempt to make sense of $a/0$, even in the "funny" case when $a = 0$, so for us $0/0$ is meaningless. Formally these two properties say that (algebraically) $\mathbb{R}$ is a field, although it is not essential at this stage to know the terminology.

Order As well as the algebraic properties, $\mathbb{R}$ has an ordering on it, usually written as "$a > 0$" or "$\geq$". There are three parts to the property:

Trichotomy For any $a \in \mathbb{R}$, exactly one of $a > 0$, $a = 0$ or $a < 0$ holds, where we write $a < 0$ instead of the formally correct $0 > a$; in words, we are simply saying that a number is either positive, negative or zero.

Addition The order behaves as expected with respect to addition: if $a > 0$ and $b > 0$ then $a + b > 0$; i.e. the sum of positives is positive.

Multiplication The order behaves as expected with respect to multiplication: if $a > 0$ and $b > 0$ then $ab > 0$; i.e. the product of positives is positive.

Note that we write $a \geq 0$ if either $a > 0$ or $a = 0$. More generally, we write $a > b$ whenever $a - b > 0$.

Completion The set $\mathbb{R}$ has an additional property, which in contrast is much more mysterious — it is complete. It is this property that distinguishes it from $\mathbb{Q}$. Its effect is that there are always "enough" numbers to do what we want. Thus there are enough to solve any algebraic equation, even those like $x^2 = 2$ which can't be solved in $\mathbb{Q}$. In fact there are (uncountably many) more - all the numbers like π, certainly not rational, but in fact not even an algebraic number, are also in $\mathbb{R}$. We explore this property during the course.

One reason for looking carefully at the properties of $\mathbb{R}$ is to note possible errors in manipulation. One aim of the course is to emphasise accurate explanation. Normal algebraic manipulations can be done without comment, but two cases arise when more care is needed:

Never divide by a number without checking first that it is non-zero.

Of course we know that 2 is non zero, so you don't need to justify dividing by 2, but if you divide by x, you should always say, at least the first time, why $x \neq 0$. If you don't know whether $x = 0$ or not, the rest of your argument may need to be split into the two cases when $x = 0$ and $x \neq 0$.

Never multiply an inequality by a number without checking first that the number is positive.

Here it is even possible to make the mistake with numbers; although it is perfectly sensible to multiply an equality by a constant, the same is not true of an inequality. If $x > y$, then of course $2x > 2y$. However, we have $(-2)x < (-2)y$. If multiplying by an expression, then again it may be necessary to consider different cases separately.

1.1. Example. Show that if $a > 0$ then $-a < 0$; and if $a < 0$ then $-a > 0$.

Solution. This is not very interesting, but is here to show how to use the properties formally.

Assume the result is false; then by trichotomy, $-a = 0$ (which is false because we know $a > 0$), or $(-a) > 0$. If this latter holds, then $a + (-a)$ is the sum of two positives and so is positive. But $a + (-a) = 0$, and by trichotomy $0 > 0$ is false. So the only consistent possibility is that $-a < 0$. The other part is essentially the same argument.

1.2. Example. Show that if $a > b$ and $c < 0$, then $ac < bc$.

Solution. This also isn't very interesting; and is here to remind you that the order in which questions are asked can be helpful. The hard bit about doing this is in Example 1.1. This is an idea you will find a lot in example sheets, where the next question *uses* the result of the previous one. It may dissuade you from dipping into a sheet; try rather to work through systematically.

Applying Example 1.1 in the case $a = -c$, we see that $-c > 0$ and $a - b > 0$. Thus using the multiplication rule, we have $(a - b)(-c) > 0$, and so $bc - ac > 0$ or $bc > ac$ as required.

1.3. Exercise. Show that if $a < 0$ and $b < 0$, then $ab > 0$.

1.3 Inequalities

One aim of this course is to get a *useful* understanding of the behaviour of systems. Think of it as trying to see the wood, when our detailed calculations tell us about individual trees. For example, we may want to know *roughly* how a function behaves; can we perhaps ignore a term because it is small and simplify things? In order to to this we need to estimate — replace the term by something bigger which is easier to handle, and so we have to deal with inequalities. It often turns out that we can "give something away" and still get a useful result, whereas calculating directly can prove either impossible, or at best unhelpful. We have just looked at the rules for manipulating the order relation. This section is probably all revision; it is here to emphasise the need for care.

1.4. Example. Find $\{x \in \mathbb{R} : (x - 2)(x + 3) > 0\}$.

Solution. Suppose $(x - 2)(x + 3) > 0$. Note that if the product of two numbers is positive then either both are positive or both are negative. So *either* $x - 2 > 0$ and $x + 3 > 0$, in which case both $x > 2$ and $x > -3$, so $x > 2$; *or* $x - 2 < 0$ and $x + 3 < 0$, in which case both $x < 2$ and $x < -3$, so $x < -3$. Thus

$$\{x : (x - 2)(x + 3) > 0\} = \{x : x > 2\} \cup \{x : x < -3\}.$$

1.5. Exercise. Find $\{x \in \mathbb{R} : x^2 - x - 2 < 0\}$.

Even at this simple level, we can produce some interesting results.

1.6. Proposition (Arithmetic - Geometric mean inequality). *If $a \geq 0$ and $b \geq 0$ then*

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

Solution. For any value of x, we have $x^2 \geq 0$ (why?), so $(a - b)^2 \geq 0$. Thus

$$a^2 - 2ab + b^2 \geq 0,$$
$$a^2 + 2ab + b^2 \geq 4ab.$$
$$(a + b)^2 \geq 4ab.$$

Since $a \geq 0$ and $b \geq 0$, taking square roots, we have $\dfrac{a + b}{2} \geq \sqrt{ab}$. This is the **arithmetic - geometric mean inequality**. We study further work with inequalities in section 1.7.

1.4 Intervals

We need to be able to talk easily about certain subsets of $\mathbb{R}$. We say that $I \subset \mathbb{R}$ is an **open interval** if

$$I = (a, b) = \{x \in \mathbb{R} : a < x < b\}.$$

Thus an open interval excludes its end points, but contains all the points in between. In contrast a **closed interval** contains both its end points, and is of the form

$$I = [a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}.$$

It is also sometimes useful to have **half - open** intervals like $(a, b]$ and $[a, b)$. It is trivial that $[a, b] = (a, b) \cup \{a\} \cup \{b\}$.

> The two end points a and b are points in $\mathbb{R}$. It is *sometimes* convenient to allow also the possibility $a = -\infty$ and $b = +\infty$; it should be clear from the context whether this is being allowed. If these extensions are being excluded, the interval is sometimes called a *finite* interval, just for emphasis.

Of course we can easily get to more general subsets of $\mathbb{R}$. So $(1, 2) \cup [2, 3] = (1, 3]$ shows that the union of two intervals may be an interval, while the example $(1, 2) \cup (3, 4)$ shows that the union of two intervals *need not* be an interval.

1.7. Exercise. Write down a pair of intervals I_1 and I_2 such that $1 \in I_1$, $2 \in I_2$ and $I_1 \cap I_2 = \emptyset$.

Can you still do this, if you require in addition that I_1 is centred on 1, I_2 is centred on 2 and that I_1 and I_2 have the same (positive) length? What happens if you replace 1 and 2 by any two numbers l and m with $l \neq m$?

1.8. Exercise. Write down an interval I with $2 \in I$ such that $1 \notin I$ and $3 \notin I$. Can you find the largest such interval? Is there a largest such interval if you also require that I is closed?

Given l and m with $l \neq m$, show there is always an interval I with $l \in I$ and $m \notin I$.

1.5 Functions

Recall that $f : D \subset \mathbb{R} \to T$ is a **function** if $f(x)$ is a well defined value in T for each $x \in D$. We say that D is the **domain** of the function, T is the **target space** and $f(D) = \{f(x) : x \in D\}$ is the **range** of f.

Note first that the definition says nothing about a formula; just that the result must be properly defined. And the definition can be complicated; for example

$$f(x) = \begin{cases} 0 & \text{if } x \leq a \text{ or } x \geq b; \\ 1 & \text{if } a < x < b. \end{cases}$$

defines a function on the whole of $\mathbb{R}$, which has the value 1 on the open interval (a, b), and is zero elsewhere [and is usually called the *characteristic* function of the interval (a, b).]

In the simplest examples, like $f(x) = x^2$ the domain of f is the whole of $\mathbb{R}$, but even for relatively simple cases, such as $f(x) = \sqrt{x}$, we need to restrict to a smaller domain, in this case the domain D is $\{x : x \geq 0\}$, since we cannot define the square root of a negative number, at least if we want the function to have real - values, so that $T \subset \mathbb{R}$.

Note that the domain is part of the definition of a function, so changing the domain technically gives a different function. This distinction will start to be important in this course. So $f_1 : \mathbb{R} \to \mathbb{R}$ defined by $f_1(x) = x^2$ and $f_2 : [-2, 2] \to \mathbb{R}$ defined by $f_2(x) = x^2$ are formally *different* functions, even though they both are "x^2" Note also that the range of f_2 is $[0, 4]$. This illustrate our first use of intervals. Given $f : \mathbb{R} \to \mathbb{R}$, we can always restrict the domain of f to an interval I to get a new function. Mostly this is trivial, but sometimes it is useful.

Another natural situation in which we need to be careful of the domain of a function occurs when taking quotients, to avoid dividing by zero. Thus the function

$$f(x) = \frac{1}{x - 3} \quad \text{has domain } \{x \in \mathbb{R} : x \neq 3\}.$$

The point we have excluded, in the above case 3 is sometimes called a **singularity** of f.

1.9. Exercise. Write down the natural domain of definition of each of the functions:

$$f(x) = \frac{x - 2}{x^2 - 5x + 6} \qquad g(x) = \frac{1}{\sin x}.$$

Where do these functions have singularities?

It is often of interest to investigate the behaviour of a function near a singularity. For example if

$$f(x) = \frac{x - a}{x^2 - a^2} = \frac{x - a}{(x - a)(x + a)} \quad \text{for } x \neq a.$$

then since $x \neq a$ we can cancel to get $f(x) = (x + a)^{-1}$. This is of course a different representation of the function, and provides an indication as to how f may be extended through the singularity at a — by giving it the value $(2a)^{-1}$.

1.6   Neighbourhoods

This situation often occurs. We need to be able to talk about a function *near* a point: in the above example, we don't want to worry about the singularity at $x = -a$ when we are discussing the one at $x = a$ (which is actually much better behaved). If we only look at the points distant less than d for a, we are really looking at an interval $(a - d, a + d)$; we call such an interval a **neighbourhood** of a. For traditional reasons, we usually replace the

distance d by its Greek equivalent, and speak of a distance δ. If $\delta > 0$ we call the interval $(a - \delta, a + \delta)$ a neighbourhood (sometimes a δ - neighbourhood) of a. The significance of a neighbourhood is that it is an interval in which we can look at the behaviour of a function without being distracted by other irrelevant behaviours. It usually doesn't matter whether δ is very big or not. To see this, consider:

1.10. Exercise. Show that an open interval contains a neighbourhood of each of its points.

We can rephrase the result of Ex 1.7 in this language; given $l \neq m$ there is some (sufficiently small) δ such that we can find disjoint δ - neighbourhoods of l and m. We use this result in Prop 2.6.

1.7 Absolute Value

Here is an example where it is natural to use a two part definition of a function. We write

$$|x| = \left\{ \begin{array}{ll} x & \text{if } x \geq 0; \\ -x & \text{if } x < 0. \end{array} \right.$$

An equivalent definition is $|x| = \sqrt{x^2}$. This is the **absolute value** or **modulus** of x. It's particular use is in describing distances; we interpret $|x - y|$ as the distance between x and y. Thus

$$(a - \delta, a + \delta) = \{X \in \mathbb{R} : |x - a| < \delta\},$$

so a δ - neighbourhood of a consists of all points which are closer to a than δ.

Note that we can always "expand out" the inequality using this idea. So if $|x - y| < k$, we can rewrite this without a modulus sign as the pair of inequalities $-k < x - y < k$. We sometimes call this "unwrapping" the modulus; conversely, in order to establish an inequality involving the modulus, it is simply necessary to show the corresponding pair of inequalities.

1.11. Proposition (The Triangle Inequality.). *For any $x, y \in \mathbb{R}$,*

$$|x + y| \leq |x| + |y|.$$

Proof. Since $-|x| \leq x \leq |x|$, and the same holds for y, combining these we have

$$-|x| - |y| \leq x + y \leq |x| + |y|$$

and this is the same as the required result. $\qquad \square$

1.12. Exercise. Show that for any $x, y, z \in \mathbb{R}$, $|x - z| \leq |x - y| + |y - z|$.

1.13. Proposition. *For any $x, y \in \mathbb{R}$,*

$$|x - y| \geq \Big| |x| - |y| \Big|.$$

Proof. Using 1.12 we have

$$|x| = |x - y + y| \le |x - y| + |y|$$

and so $|x| - |y| \le |x - y|$. Interchanging the rôles of x and y, and noting that $|x| = |-x|$, gives $|y| - |x| \le |x - y|$. Multiplying this inequality by -1 and combining these we have

$$-|x - y| \le |x| - |y| \le |x - y|$$

and this is the required result. $\square$

1.14. Example. Describe $\{x \in \mathbb{R} : |5x - 3| > 4\}$.

Proof. Unwrapping the modulus, we have either $5x - 3 < -4$, or $5x - 3 > 4$. From one inequality we get $5x < -4 + 3$, or $x < -1/5$; the other inequality gives $5x > 4 + 3$, or $x > 7/5$. Thus

$$\{x \in \mathbb{R} : |5x - 3| > 4\} = (-\infty, -1/5) \cup (7/5, \infty).$$

$\square$

1.15. Exercise. Describe $\{x \in \mathbb{R} : |x + 3| < 1\}$.

1.16. Exercise. Describe the set $\{x \in \mathbb{R} : 1 \le x \le 3\}$ using the absolute value function.

1.8   The Binomial Theorem and other Algebra

At its simplest, the binomial theorem gives an expansion of $(1+x)^n$ for any positive integer n. We have

$$(1 + x)^n = 1 + nx + \frac{n.(n-1)}{1.2}x^2 + \ldots + \frac{n.(n-1).(n-k+1)}{1.2.\ldots.k}x^k + \ldots + x^n.$$

Recall in particular a few simple cases:

$$(1 + x)^3 = 1 + 3x + 3x^2 + x^3,$$
$$(1 + x)^4 = 1 + 4x + 6x^2 + 4x^3 + x^4,$$
$$(1 + x)^5 = 1 + 5x + 10x^2 + 10x^3 + 5x^4 + x^5.$$

There is a more general form:

$$(a + b)^n = a^n + na^{n-1}b + \frac{n.(n-1)}{1.2}a^{n-2}b^2 + \ldots + \frac{n.(n-1).(n-k+1)}{1.2.\ldots.k}a^{n-k}b^k + \ldots + b^n,$$

with corresponding special cases. Formally this result is only valid for any positive integer n; in fact it holds appropriately for more general exponents as we shall see in Chapter 7

Another simple algebraic formula that can be useful concerns powers of differences:

$$a^2 - b^2 = (a - b)(a + b),$$
$$a^3 - b^3 = (a - b)(a^2 + ab + b^2),$$
$$a^4 - b^4 = (a - b)(a^3 + a^2b + ab^2 + b^3)$$

and in general, we have

$$a^n - b^n = (a - b)(a^{n-1} + a^{n-2}b + a^{n-3}b^2 + \ldots + a^b n - 1 + b^{n-1}).$$

Note that we made use of this result when discussing the function after Ex 1.9.

And of course you remember the usual "completing the square" trick:

$$ax^2 + bx + c = a\left(x^2 + \frac{b}{a}x + \frac{b^2}{4a^2}\right) + c - \frac{b^2}{4a}$$

$$= a\left(x + \frac{b}{2a}\right)^2 + \left(c - \frac{b^2}{4a}\right).$$

Chapter 2

Sequences

2.1 Definition and Examples

2.1. Definition. A (real infinite) sequence is a map $a : \mathbb{N} \to \mathbb{R}$

Of course if is more usual to call a function f rather than a; and in fact we usually start labeling a sequence from 1 rather than 0; it doesn't really matter. What the definition is saying is that we can lay out the members of a sequence in a list with a first member, second member and so on. If $a : \mathbb{N} \to \mathbb{R}$, we usually write a_1, a_2 and so on, instead of the more formal $a(1)$, $a(2)$, even though we usually write functions in this way.

2.1.1 Examples of sequences

The most obvious example of a sequence is the sequence of natural numbers. Note that the integers are not a sequence, although we can turn them into a sequence in many ways; for example by enumerating them as $0, 1, -1, 2, -2 \dots$. Here are some more sequences:

Definition	First 4 terms	Limit
$a_n = n - 1$	0, 1, 2, 3	does not exist ($\to \infty$)
$a_n = \dfrac{1}{n}$	$1, \dfrac{1}{2}, \dfrac{1}{3}, \dfrac{1}{4}$	0
$a_n = (-1)^{n+1}$	1, -1, 1, -1	does not exist (the sequence oscillates)
$a_n = (-1)^{n+1}\dfrac{1}{n}$	$1, -\dfrac{1}{2}, \dfrac{1}{3}, -\dfrac{1}{4}$	0
$a_n = \dfrac{n-1}{n}$	$0, \dfrac{1}{2}, \dfrac{2}{3}, \dfrac{3}{4}$	1
$a_n = (-1)^{n+1}\left(\dfrac{n-1}{n}\right)$	$0, -\dfrac{1}{2}, \dfrac{2}{3}, -\dfrac{3}{4}$	does not exist (the sequence oscillates)
$a_n = 3$	3, 3, 3, 3	3

A sequence doesn't have to be defined by a sensible "formula". Here is a sequence you may recognise:-

$$3, \quad 3.1, \quad 3.14, \quad 3.141, \quad 3.1415, \quad 3.14159, \quad 3.141592 \dots$$

11

where the terms are successive truncates of the decimal expansion of π.

Of course we can graph a sequence, and it sometimes helps. In Fig 2.1 we show a sequence of locations of (just the x coordinate) of a car driver's eyes. The interest is whether the sequence oscillates predictably.
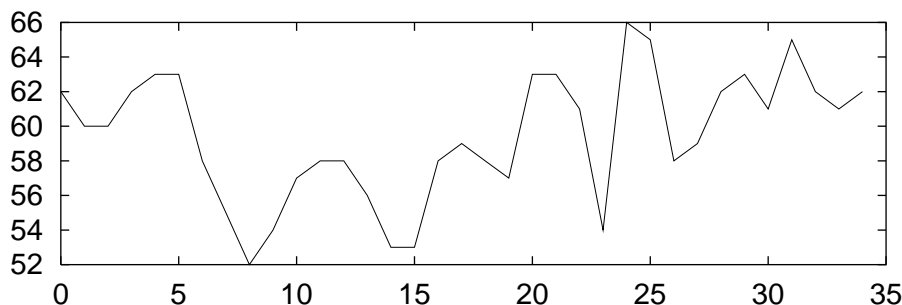
Figure 2.1: A sequence of eye locations.

Usually we are interested in what happens to a sequence "in the long run", or what happens "when it settles down". So we are usually interested in what happens when $n \to \infty$, or in the **limit** of the sequence. In the examples above this was fairly easy to see.

Sequences, and interest in their limits, arise naturally in many situations. One such occurs when trying to solve equations numerically; in Newton's method, we use the standard calculus approximation, that

$$f(a + h) \approx f(a) + h.f'(a).$$

If now we almost have a solution, so $f(a) \approx 0$, we can try to perturb it to $a + h$, which is a true solution, so that $f(a + h) = 0$. In that case, we have

$$0 = f(a + h) = f(a) + h.f'(a) \quad \text{and so} \quad h \approx \frac{f(a)}{f'(a)}.$$

Thus a better approximation than a to the root is $a + h = a - f(a)/f'(a)$.

If we take $f(x) = x^3 - 2$, finding a root of this equation is solving the equation $x^3 = 2$, in other words, finding $\sqrt[3]{2}$ In this case, we get the sequence defined as follows

$$a_1 = 1 \text{while} a_{n+1} = \frac{2}{3}a_n + \frac{2}{3a_n^2} \quad \text{if } n > 1. \tag{2.1}$$

Note that this makes sense: $a_1 = 1$, $a_2 = \frac{2}{3}.1 + \frac{2}{3.1^2}$ etc. Calculating, we get $a_2 = 1.333$, $a_3 = 1.2639$, $a_4 = 1.2599$ and $a_5 = 1.2599$ In fact the sequence does converge to $\sqrt[3]{2}$; by taking enough terms we can get an *approximation* that is as accurate as we need. [You can check that $a_5^3 = 2$ to 6 decimal places.]

Note also that we need to specify the accuracy needed. There is no *single* approximation to $\sqrt[3]{2}$ or π which will always work, whether we are measuring a flower bed or navigating a satellite to a planet. In order to use such a sequence of approximations, it is first necessary to specify an acceptable accuracy. Often we do this by specifying a neighbourhood of the limit, and we then often speak of an ϵ - neighbourhood, where we use ϵ (for error), rather than δ (for distance).

2.2. Definition. Say that a sequence $\{a_n\}$ **converges** to a **limit** l if and only if, given $\epsilon > 0$ there is some N such that

$$|a_n - l| < \epsilon \quad \text{whenever} \quad n \geq N.$$

A sequence which converges to some limit is a **convergent sequence**.

2.3. Definition. A sequence which is not a convergent sequence is **divergent**. We sometimes speak of a sequence **oscillating** or tending to infinity, but properly I am just interested in divergence at present.

2.4. Definition. Say a property $P(n)$ holds **eventually** iff $\exists N$ such that $P(n)$ holds for all $n \geq N$. It holds **frequently** iff given N, there is some $n \geq N$ such that $P(n)$ holds.

We call the n a **witness**; it witnesses the fact that the property is true somewhere at least as far along the sequence as N. Some examples using the language are worthwhile. The sequence $\{-2, -1, 0, 1, 2, \dots\}$ is eventually positive. The sequence $\sin(n!\pi/17)$ is eventually zero; the sequence of natural numbers is frequently prime.

It may help you to understand this language if you think of the sequence of days in the future[1]. You will find, according to the definitions, that it will frequently be Friday, frequently be raining (or sunny), and even frequently February 29. In contrast, eventually it will be 1994, and eventually you will die. A more difficult one is whether Newton's work will eventually be forgotten!

Using this language, we can rephrase the definition of convergence as

We say that $a_n \to l$ as $n \to \infty$ iff given any error $\epsilon > 0$ eventually a_n is closer to l then ϵ. Symbolically we have

$$\epsilon > 0 \quad \exists N \quad \text{s.t.} \quad |a_n - l| < \epsilon \quad \text{whenever} \quad n \geq N.$$

Another version may make the content of the definition even clearer; this time we use the idea of neighbourhood:

We say that $a_n \to l$ as $n \to \infty$ iff given any (acceptable) error $\epsilon > 0$ eventually a_n is in the ϵ - neighbourhood of l.

It is important to note that the definition of the limit of a sequence doesn't have a simpler form. If you think you have an easier version, talk it over with a tutor; you may find that it picks up as convergent some sequences you don't want to be convergent. In Fig 2.2, we give a picture that may help. The ϵ- neighbourhood of the (potential) limit l is represented by the shaded strip, while individual members a_n of the sequence are shown as blobs. The definition then says the sequence is convergent to the number we have shown as a potential limit, provided the sequence is eventually in the shaded strip: and this must be true even if we redraw the shaded strip to be narrower, as long as it is still centred on the potential limit.

---

[1] I need to assume the sequence is infinite; you can decide for yourself whether this is a philosophical statement, a statement about the future of the universe, or just plain optimism!
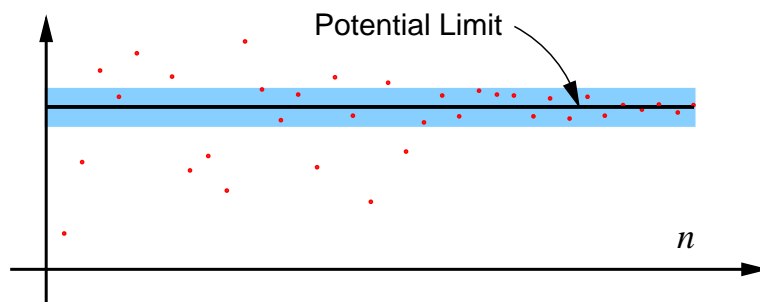
Figure 2.2: A picture of the definition of convergence

2.2   Direct Consequences

With this language we can give some simple examples for which we can use the definition directly.

- If $a_n \to 2$ as $n \to \infty$, then (take $\epsilon = 1$), eventually, a_n is within a distance 1 of 2. One consequence of this is that eventually $a_n > 1$ and another is that eventually $a_n < 3$.

- Let $a_n = 1/n$. Then $a_n \to 0$ as $n \to \infty$. To check this, pick $\epsilon > 0$ and then choose N with $N > 1/\epsilon$. Now suppose that $n \geq N$. We have

$$0 \leq \frac{1}{n} \leq \frac{1}{N} < \epsilon \quad \text{by choice of } N.$$

- The sequence $a_n = n - 1$ is divergent; for if not, then there is some l such that $a_n \to l$ as $n \to \infty$. Taking $\epsilon = 1$, we see that eventually (say after N) , we have $-1 \leq (n-1) - l < 1$, and in particular, that $(n-1) - l < 1$ for all $n \geq N$. thus $n < l + 2$ for all n, which is a contradiction.

2.5. Exercise. Show that the sequence $a_n = (1/\sqrt{n}) \to 0$ as $n \to \infty$.

Although we can work directly from the definition in these simple cases, most of the time it is too hard. So rather than always working directly, we also use the definition to prove some general tools, and then use the tools to tell us about convergence or divergence. Here is a simple tool (or Proposition).

2.6. Proposition. *Let $a_n \to l$ as $n \to \infty$ and assume also that $a_n \to m$ as $n \to \infty$. Then $l = m$. In other words, if a sequence has a limit, it has a unique limit, and we are justified in talking about **the** limit of a sequence.*

Proof. Suppose that $l \neq m$; we argue by contradiction, showing this situation is impossible. Using 1.7, we choose disjoint neighbourhoods of l and m, and note that since the sequence converges, eventually it lies in each of these neighbourhoods; this is the required contradiction. □

We can argue this directly (so this is another version of this proof). Pick $\epsilon = |l - m|/2$. Then eventually $|a_n - l| < \epsilon$, so this holds e.g.. for $n \geq N_1$. Also, eventually $|a_n - m| < \epsilon$,

so this holds eg. for $n \geq N_2$. Now let $N = \max(N_1, N_2)$, and choose $n \geq N$. Then both inequalities hold, and

$$
\begin{aligned}
|l - m| &= |l - a_n + a_n - m| \\
&\leq |l - a_n| + |a_n - m| \quad \text{by the triangle inequality} \\
&< \epsilon + \epsilon = |l - m|
\end{aligned}
$$

2.7. Proposition. *Let $a_n \to l \neq 0$ as $n \to \infty$. Then eventually $a_n \neq 0$.*

Proof. Remember what this means; we are guaranteed that from some point onwards, we *never* have $a_n = 0$. The proof is a variant of "if $a_n \to 2$ as $n \to \infty$ then eventually $a_n > 1$." One way is just to repeat that argument in the two cases where $l > 0$ and then $l < 0$. But we can do it all in one:

Take $\epsilon = |l|/2$, and apply the definition of "$a_n \to l$ as $n \to \infty$". Then there is some N such that

$$
\begin{aligned}
|a_n - l| &< |l|/2 \quad \text{for all } n \geq N \\
\text{Now} \quad l &= l - a_n + a_n. \\
\text{Thus} \quad |l| &\leq |l - a_n| + |a_n|, \quad \text{so} \quad |l| \leq |l|/2 + |a_n|, \\
\text{and} \quad |a_n| &\geq |l|/2 \neq 0.
\end{aligned}
$$

$\square$

2.8. Exercise. Let $a_n \to l \neq 0$ as $n \to \infty$, and assume that $l > 0$. Show that eventually $a_n > 0$. In other words, use the first method suggested above for $l > 0$.

2.3   Sums, Products and Quotients

2.9. Example. Let $a_n = \dfrac{n+2}{n+3}$. Show that $a_n \to 1$ as $n \to \infty$.

Solution. There is an obvious manipulation here:-

$$
a_n = \frac{n+2}{n+3} = \frac{1 + 2/n}{1 + 3/n}.
$$

We hope the numerator converges to $1 + 0$, the denominator to $1 + 0$ and so the quotient to $(1 + 0)/(1 + 0) = 1$. But it is not obvious that our definition does indeed behave as we would wish; we need rules to justify what we have done. Here they are:-

2.10. Theorem. *(New Convergent sequences from old) Let $a_n \to l$ and $b_n \to m$ as $n \to \infty$. Then*

Sums: $a_n + b_n \to l + m$ *as $n \to \infty$;*

Products: $a_n b_n \to lm$ *as $n \to \infty$; and*

Inverses: *provided $m \neq 0$ then $a_n/b_n \to l/m$ as $n \to \infty$.*

Note that part of the point of the theorem is that the new sequences are *convergent.*

Proof. Pick $\epsilon > 0$; we must find N such that $|a_n + b_n - (l + m) < \epsilon$ when $n \geq N$. Now because First pick $\epsilon > 0$. Since $a_n \to l$ as $n \to \infty$, there is some N_1 such that $|a_n - l| < \epsilon/2$ whenever $n > N_1$, and in the same way, there is some N_2 such that $|b_n - m| < \epsilon/2$ whenever $n > N_2$. Then if $N = \max(N_1, N_2)$, and $n > N$, we have

$$|a_n + b_n - (l + m)| < |a_n - l| + |b_n - m| < \epsilon/2 + \epsilon/2 = \epsilon.$$

The other two results are proved in the same way, but are technically more difficult. Proofs can be found in (Spivak 1967). □

2.11. Example. Let $a_n = \dfrac{4 - 7n^2}{n^2 + 3n}$. Show that $a_n \to -7$ as $n \to \infty$.

Solution. A helpful manipulation is easy. We choose to divide both top and bottom by the highest power of n around. This gives:

$$a_n = \frac{4 - 7n^2}{n^2 + 3n} = \frac{\frac{4}{n^2} - 7}{1 + \frac{3}{n}}.$$

We now show each term behaves as we expect. Since $1/n^2 = (1/n).(1/n)$ and $1/n \to 0$ as $n \to \infty$, we see that $1/n^2 \to 0$ as $n \to \infty$, using "product of convergents is convergent". Using the corresponding result for sums shows that $\frac{4}{n^2} - 7 \to 0 - 7$ as $n \to \infty$. In the same way, the denominator $\to 1$ as $n \to \infty$. Thus by the "limit of quotients" result, since the limit of the denominator is $1 \neq 0$, the quotient $\to -7$ as $n \to \infty$.

2.12. Example. In equation 2.1 we derived a sequence (which we claimed converged to $\sqrt[3]{2}$) from Newton's method. We can now show that *provided the limit exists and is non zero,* the limit is indeed $\sqrt[3]{2}$.

Proof. Note first that if $a_n \to l$ as $n \to \infty$, then we also have $a_{n+1} \to l$ as $n \to \infty$. In the equation

$$a_{n+1} = \frac{2}{3}a_n + \frac{2}{3a_n^2}$$

we now let $n \to \infty$ on both sides of the equation. Using Theorem 2.10, we see that the right hand side converges to $\dfrac{2}{3}l + \dfrac{2}{3l^2}$, while the left hand side converges to l. But they are the same sequence, so both limits are the same by Prop 2.6. Thus

$$l = \frac{2}{3}l + \frac{2}{3l^2} \quad \text{and so } l^3 = 2.$$

□

2.13. Exercise. Define the sequence $\{a_n\}$ by $a_1 = 1$, $a_{n+1} = (4a_n + 2)/(a_n + 3)$ for $n \geq 1$. Assuming that $\{a_n\}$ is convergent, find its limit.

2.14. Exercise. Define the sequence $\{a_n\}$ by $a_1 = 1$, $a_{n+1} = (2a_n + 2)$ for $n \geq 1$. Assuming that $\{an\}$ is convergent, find its limit. Is the sequence convergent?

2.15. Example. Let $a_n = \sqrt{n+1} - \sqrt{n}$. Show that $a_n \to 0$ as $n \to \infty$.

Proof. A simple piece of algebra gets us most of the way:

$$
\begin{aligned}
a_n &= \left(\sqrt{n+1} - \sqrt{n}\right) \cdot \frac{\left(\sqrt{n+1} + \sqrt{n}\right)}{\left(\sqrt{n+1} + \sqrt{n}\right)} \\
&= \frac{(n+1) - n}{\left(\sqrt{n+1} + \sqrt{n}\right)} = \frac{1}{\left(\sqrt{n+1} + \sqrt{n}\right)} \to 0 \quad \text{as} \quad n \to \infty.
\end{aligned}
$$

□

2.4 Squeezing

Actually, we can't take the last step yet. It is true and looks sensible, but it is another case where we need more results getting new convergent sequences from old. We really want a good dictionary of convergent sequences.

The next results show that order behaves quite well under taking limits, but also shows why we need the dictionary. The first one is fairly routine to prove, but you may still find these techniques hard; if so, note the result, and come back to the proof later.

2.16. Exercise. Given that $a_n \to l$ and $b_n \to m$ as $n \to \infty$, and that $a_n \le b_n$ for each n, then $l \le m$.

Compare this with the next result, where we can also *deduce* convergence.

2.17. Lemma. *(The Squeezing lemma) Let $a_n \le b_n \le c_n$, and suppose that $a_n \to l$ and $c_n \to l$ as $n \to \infty$. The $\{b_n\}$ is convergent, and $b_n \to l$ as $n \to \infty$.*

Proof. Pick $\epsilon > 0$. Then since $a_n \to l$ as $n \to \infty$, we can find N_1 such that

$$
|a_n - l| < \epsilon \quad \text{for} \quad n \ge N_1
$$

and since $c_n \to l$ as $n \to \infty$, we can find N_2 such that

$$
|c_n - l| < \epsilon \quad \text{for} \quad n \ge N_2.
$$

Now pick $N = \max(N_1, N_2)$, and note that, in particular, we have

$$
-\epsilon < a_n - l \quad \text{and} \quad c_n - l < \epsilon.
$$

Using the given order relation we get

$$
-\epsilon < a_n - l \le b_n - l \le c_n - l < \epsilon,
$$

and using only the middle and outer terms, this gives

$$
-\epsilon < b_n - l < \epsilon \quad \text{or} \quad |b_n - l| < \epsilon \quad \text{as claimed.}
$$

□

Note: Having seen the proof, it is clear we can state an "eventually" form of this result. We don't need the inequalities to hold all the time, only eventually.

2.18. Example. Let $a_n = \dfrac{\sin(n)}{n^2}$. Show that $a_n \to 0$ as $n \to \infty$.

Solution. Note that, whatever the value of $\sin(n)$, we always have $-1 \le \sin(n) \le 1$. We use the squeezing lemma:

$$-\frac{1}{n^2} < a_n < \frac{1}{n^2}. \quad \text{Now} \quad \frac{1}{n^2} \to 0, \quad \text{so} \quad \frac{\sin(n)}{n^2} \to 0 \quad \text{as} \quad n \to \infty.$$

2.19. Exercise. Show that $\sqrt{1 + \dfrac{1}{n}} \to 1$ as $n \to \infty$.

Note: We can now do a bit more with the $\sqrt{n+1} - \sqrt{n}$ example. We have

$$0 \le \frac{1}{\sqrt{n+1} + \sqrt{n}} \le \frac{1}{2\sqrt{n}},$$

so we have our result since we showed in Exercise 2.5 that $(1/\sqrt{n}) \to 0$ as $n \to \infty$.

This illustrates the need for a good bank of convergent sequences. In fact we don't have to use ad-hoc methods here; we can get such results in much more generality. We need the next section to prove it, but here is the results.

2.20. Proposition. *Let f be a continuous function at a, and suppose that $a_n \to a$ as $n \to \infty$. Then $f(a_n) \to f(a)$ as $n \to \infty$.*

Note: This is another example of the "new convergent sequences from old" idea. The application is that $f(x) = x^{1/2}$ is continuous everywhere on its domain which is $\{x \in \mathbb{R} : x \ge 0\}$, so since $n^{-1} \to 0$ as $n \to \infty$, we have $n^{-1/2} \to 0$ as $n \to \infty$; the result we proved in Exercise 2.5.

2.21. Exercise. What do you deduce about the sequence $a_n = \exp(1/n)$ if you apply this result to the continuous function $f(x) = \mathrm{e}^x$?

2.22. Example. Let $a_n = \dfrac{1}{n \log n}$ for $n \ge 2$. Show that $a_n \to 0$ as $n \to \infty$.

Solution. Note that $1 \le \log n \le n$ if $n \ge 3$, because $\log(e) = 1$, log is monotone increasing, and $2 < e < 3$. Thus $n < n \log n < n^2$, when $n \ge 3$ and

$$\frac{1}{n^2} < \frac{1}{n \log n} < \frac{1}{n}. \quad \text{Now} \quad \frac{1}{n} \to 0 \quad \text{and} \quad \frac{1}{n^2} \to 0, \quad \text{so} \quad \frac{1}{n \log n} \to 0 \quad \text{as} \quad n \to \infty.$$

Here we have used the "eventually" form of the squeezing lemma.

2.23. Exercise. Let $a_n = \dfrac{1}{\sqrt{n} \log n}$ for $n \ge 2$. Show that $a_n \to 0$ as $n \to \infty$.

2.5   Bounded sequences

2.24. Definition. Say that $\{a_n\}$ is a **bounded** sequence iff there is some K such that $|a_n| \leq K$ for all n.

This definition is correct, but not particularly useful at present. However, it does provide good practice in working with abstract formal definitions.

2.25. Example. Let $a_n = \dfrac{1}{\sqrt{n} \log n}$ for $n \geq 2$. Show that $\{a_n\}$ is a bounded sequence. [This is the sequence of Exercise 2.23].

2.26. Exercise. Show that the sum of two bounded sequences is bounded.

2.27. Proposition. *An eventually bounded sequence is bounded*

Proof. Let $\{a_n\}$ be an eventually bounded sequence, so there is some N, and a bound K such that $|a_n| \leq K$ for all $n \geq N$. Let $L = \max\{|a_1|, |a_2|, \ldots |a_{N-1}|, K\}$. Then by definition $|a_1| \leq L$, and indeed in the same way, $|a_k| \leq L$ whenever $k < N$. But if $n \geq N$ then $|a_n| \leq K \leq L$, so in fact $|a_n| \leq L$ for all n, and the sequence is actually bounded.   □

2.28. Proposition. *A convergent sequence is bounded*

Proof. Let $\{a_n\}$ be a convergent sequence, with limit l say. Then there is some N such that $|a_n - l| < 1$ whenever $n \geq N$. Here we have used the definition of convergence, taking ϵ, our pre-declared error, to be 1. Then by the triangle inequality,

$$|a_n| \leq |a_n - l| + |l| \leq 1 + |l| \quad \text{for all } n \geq N.$$

Thus the sequence $\{a_n\}$ is eventually bounded, and so is bounded by Prop 2.27.   □

And here is another result on which to practice working from the definition. In order to tackle simple proofs like this, you should start by writing down, using the definitions, the information you are given. Then write out (in symbols) what you wish to prove. Finally see how you can get from the known information to what you need. Remember that if a definition contains a variable (like ϵ in the definition of convergence), then the definition is true *whatever* value you give to it — even if you use $\epsilon/2$ (as we did in 2.10) or ϵ/K, for any constant K. Now try:

2.29. Exercise. Let $a_n \to 0$ as $n \to \infty$ and let $\{b_n\}$ be a bounded sequence. Show that $a_n b_n \to 0$ as $n \to \infty$. [If $a_n \to l \neq 0$ as $n \to \infty$, and $\{b_n\}$ is a bounded sequence, then in general $\{a_n b_n\}$ is *not* convergent. Give an example to show this.]

2.6   Infinite Limits

2.30. Definition. Say that $a_n \to \infty$ as $n \to \infty$ iff given K, there is some N such that $a_n \geq K$ whenever $n \geq N$.

This is just one definition; clearly you can have $a_n \to -\infty$ etc. We show how to use such a definition to get some results rather like those in 2.10. For example, we show

$$a_n = \frac{n^2 + 5n}{3n + 2} \to \infty \quad \text{as} \quad n \to \infty.$$

Pick some K. We have $a_n = n.\left(\dfrac{n+5}{3n+2}\right) = n.b_n$ (say).   Using results from 2.10, we see that $b_n \to 1/3$ as $n \to \infty$, and so, eventually, $b_n > 1/6$ (Just take $\epsilon = 1/6$ to see this). Then for large enough n, we have $a_n > n/6 > K$, providing in addition $n > 6K$. Hence $a_n \to \infty$ as $n \to \infty$.

Note: It is *false* to argue that $a_n = n.(1/3) \to \infty$; you can't let one part of a limit converge without letting the other part converge! And we refuse to do arithmetic with ∞ so can't hope for a theorem directly like 2.10.

Chapter 3

Monotone Convergence

3.1 Three Hard Examples

So far we have thought of convergence in terms of knowing the limit. It is very helpful to be able to deduce convergence even when the limit is itself difficult to find, or when we can only *find* the limit provided it exist. There are better techniques than we have seen so far, dealing with more difficult examples. Consider the following examples:

Sequence definition of e: Let $a_n = \left(1 + \dfrac{1}{n}\right)^n$; then $a_n \to e$ as $n \to \infty$.

Stirling's Formula: Let $a_n = \dfrac{n!}{\sqrt{n}(n/e)^n}$; then $a_n \to \sqrt{2\pi}$ as $n \to \infty$.

Fibonacci Sequence: Define $p_1 = p_2 = 1$, and for $n \geq 1$, let $p_{n+2} = p_{n+1} + p_n$.

Let $a_n = \dfrac{p_{n+1}}{p_n}$; then $a_n \to \dfrac{1 + \sqrt{5}}{2}$ as $n \to \infty$.

We already saw in 2.12 that knowing a sequence **has** a limit can help to **find** the limit. As another example of this, consider the third sequence above. We have

$$\frac{p_{n+2}}{p_{n+1}} = 1 + \frac{p_n}{p_{n+1}} \quad \text{and so} \quad a_{n+1} = 1 + \frac{1}{a_n}. \tag{3.1}$$

Assume now that $a_n \to l$ as $n \to \infty$; then as before in 2.12, we know that $a_{n+1} \to l$ as $n \to \infty$ (check this formally)! Using 2.10 and equation 3.1, we have $l = 1 + 1/l$, or $l^2 - l - 1 = 0$. Solving the quadratic gives

$$a_n = \frac{1 \pm \sqrt{5}}{2},$$

and so these are the only possible limits. Going back to equation 3.1, we see that since $a_1 = 1$, we have $a_n > 1$ when $n > 1$. Thus by 2.16, $l \geq 1$; we can eliminate the negative root, and see that the limit, **if** it exists, is unique. In fact, all of the limits described above are within reach with some more technique, although Stirling's Formula takes quite a lot of calculation.

3.2   Boundedness Again

Of course not all sequences have limits. But there is a useful special case, which will take care of the three sequences described above and many more. To describe the result we need more definitions, which describe extra properties of sequences.

3.1. Definition. A sequence $\{a_n\}$ is **bounded above** if there is some K such that $a_n \leq K$ for all n. We say that K is an **upper bound** for the sequence.

- There is a similar definition of a sequence being **bounded below**.

- The number K may not be the *best* or smallest upper bound. All we know from the definition is that it will be an upper bound.

3.2. Example. The sequence $a_n = 2 + (-1)^n$ is bounded above.

Solution.   Probably the best way to show a sequence is bounded above is to write down an upper bound – i.e. a suitable value for K. In this case, we claim $K = 4$ will do. To check this we must show that $a_n \leq 4$ for all n. But if n is even, then $a_n = 2 + 1 = 3 \leq 4$, while if n is odd, $a_n = 2 - 1 = 1 \leq 4$. So for any n, we have $a_n \leq 4$ and 4 is an upper bound for the sequence $\{a_n\}$. Of course 3 is also an upper bound for this sequence.

3.3. Exercise. Let $a_n = \dfrac{1}{n} + \cos\left(\dfrac{1}{n}\right)$. Show that $\{a_n\}$ is bounded above and is bounded below. [Recall that $|\cos x| \leq 1$ for *all* x.]

3.4. Exercise. Show that a sequence which is bounded above and bounded below is a bounded sequence (as defined in 2.24).

3.2.1   Monotone Convergence

3.5. Definition. A sequence $\{a_n\}$ is an **increasing** sequence if $a_{n+1} \geq a_n$ for every n.

- If we need precision, we can distinguish between a strictly increasing sequence, and a (not necessarily strictly) increasing sequence. This is sometime called a *non-decreasing* sequence.

- There is a similar definition of a *decreasing* sequence.

- What does it mean to say that a sequence is eventually increasing?

- A sequence which is either always increasing or always decreasing is called a *monotone* sequence. Note that an "arbitrary" sequence is not monotone (it will usually sometimes increase, and sometimes decrease).

Nevertheless, monotone sequences *do* happen in real life. For example, the sequence

$$a_1 = 3 \qquad a_2 = 3.1 \qquad a_3 = 3.14 \qquad a_4 = 3.141 \qquad a_5 = 3.1415\ldots$$

is how we often describe the decimal expansion of π. Monotone sequences are important because we can say something useful about them which is not true of more general sequences.

3.6. Example. Show that the sequence $a_n = \dfrac{n}{2n+1}$ is increasing.

Solution. One way to check that a sequence is increasing is to show $a_{n+1} - a_n \geq 0$, a second way is to compute a_{n+1}/a_n, and we will see more ways later. Here,

$$
\begin{aligned}
a_{n+1} - a_n &= \frac{n+1}{2(n+1)+1} - \frac{n}{2n+1} \\
&= \frac{(2n^2 + 3n + 1) - (2n^2 + 3n)}{(2n+3)(2n+1)} \\
&= \frac{1}{(2n+3)(2n+1)} > 0 \quad \text{for all } n
\end{aligned}
$$

and the sequence *is* increasing.

3.7. Exercise. Show that the sequence $a_n = \dfrac{1}{n} - \dfrac{1}{n^2}$ is decreasing when $n > 1$.

If a sequence is increasing, it is an interesting question whether or not it is bounded above. If an upper bound does exist, then it seems as though the sequence can't help converging — there is nowhere else for it to go.
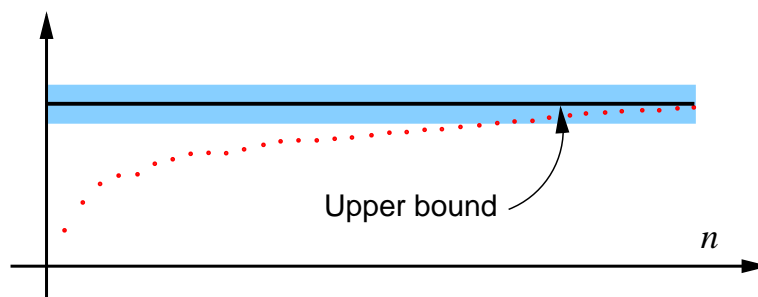
Figure 3.1: A monotone (increasing) sequence which is bounded above seems to converge because it has nowhere else to go!

In contrast, if there is no upper bound, the situation is clear.

3.8. Example. An increasing sequence which is *not* bounded above tends to ∞ (see definition 2.30).

Solution. Let the sequence be $\{a_n\}$, and assume it is not bounded above. Pick K; we show eventually $a_n > K$. Since K is not an upper bound for the sequence, there is some witness to this. In other words, there is some a_N with $a_N > K$. But in that case, since the sequence is increasing monotonely, we have $a_n \geq a_N \geq K$ for all $n \geq N$. Hence $a_n \to \infty$ as $n \to \infty$.

3.9. Theorem (The monotone convergence principle). *Let $\{a_n\}$ be an increasing sequence which is bounded above; then $\{a_n\}$ is a convergent sequence. Let $\{a_n\}$ be a decreasing sequence which is bounded below; then $\{a_n\}$ is a convergent sequence*

Proof. To prove this we need to appeal to the completeness of $\mathbb{R}$, as described in Section 1.2. Details will be given in third year, or you can look in (Spivak 1967) for an accurate deduction from the appropriate axioms for $\mathbb{R}$. $\square$

This is a very important result. It is the first time we have seen a way of deducing the convergence of a sequence without first knowing what the limit is. And we saw in 2.12 that just knowing a limit exists is sometimes enough to be able to find its value. Note that the theorem only deduces an "eventually" property of the sequence; we can change a finite number of terms in a sequence without changing the value of the limit. This means that the result must still be true of we only know the sequence is eventually increasing and bounded above. What happens if we relax the requirement that the sequence is bounded above, to be just *eventually* bounded above? (Compare Proposition 2.27).

3.10. Example. Let a be fixed. Then $a^n \to 0$ as $n \to \infty$ if $|a| < 1$, while if $a > 1$, $a^n \to \infty$ as $n \to \infty$.

Solution. Write $a_n = a^n$; then $a_{n+1} = a.a_n$. If $a > 1$ then $a_{n+1} < a_n$, while if $0 < a < 1$ then $a_{n+1} < a_n$; in either case the sequence is monotone.

Case 1 $0 < a < 1$; the sequence $\{a_n\}$ is bounded below by 0, so by the monotone convergence theorem, $a_n \to l$ as $n \to \infty$. As before note that $a_{n+1} \to l$ as $n \to \infty$. Then applying 2.10 to the equation $a_{n+1} = a.a_n$, we have $l = a.l$, and since $a \neq 1$, we must have $l = 0$.

Case 2 $a > 1$; the sequence $\{a_n\}$ is increasing. Suppose it is bounded above; then as in Case 1, it converges to a limit l satisfying $l = a.l$. This contradiction shows the sequence is *not* bounded above. Since the sequence is monotone, it must tend to infinity (as described in 3.9).

Case 3 $|a| < 1$; then since $-|a^n| \leq a^n \leq |a^n|$, and since $|a^n| = |a|^n \to 0$ as $n \to \infty$, by squeezing, since each outer limit $\to 0$ by case 1, we have $a^n \to 0$ as $n \to \infty$ whenever $|a| < 1$.

3.11. Example. Evaluate $\lim_{n\to\infty} \left(\frac{2}{3}\right)^n$ and $\lim_{n\to\infty}\left(\left(\frac{-2}{3}\right)^n + \left(\frac{4}{5}\right)^n\right)$.

Solution. Using the result that if $|a| < 1$, then $a^n \to 0$ as $n \to \infty$, we deduce that $(-2/3)^n \to 0$ as $n \to \infty$, that $(4/5)^n \to 0$ as $n \to \infty$, and using 2.10, that the second limit is also 0.

3.12. Exercise. Given that $k > 1$ is a fixed constant, evaluate $\lim_{n\to\infty}\left(1 - \frac{1}{k}\right)^n$. How does your result compare with the sequence definition of e given in Sect 3.1.

3.13. Example. Let $a_1 = 1$, and for $n \geq 1$ define $a_{n+1} = \frac{6(1 + a_n)}{7 + a_n}$. Show that $\{a_n\}$ is convergent, and find its limit.

Solution. We can calculate the first few terms of the sequence as follows:

$$a_1 = 1, \qquad a_2 = 1.5 \qquad a_3 = 1.76 \qquad a_4 = 1.89\ldots$$

and it *looks* as though the sequence might be increasing. Let

$$f(x) = \frac{6(1 + x)}{7 + x}, \quad \text{so} \quad f(a_n) = a_{n+1}. \tag{3.2}$$

By investigating f, we hope to deduce useful information about the sequence $\{a_n\}$. We have

$$f'(x) = \frac{(7+x).6 - 6(1+x)}{(7+x)^2} = \frac{36}{((7+x)^2} > 0.$$

Recall from elementary calculus that if $f'(x) > 0$, then f is increasing; in other words, if $b > a$ then $f(b) > f(a)$. We thus see that f is increasing.

Since $a_2 > a_1$, we have $f(a_2) > f(a_1)$; in other words, $a_3 > a_2$. Applying this argument inductively, we see that $a_{n+1} > a_n$ for all n, and the sequence $\{a_n\}$ is increasing.

If x is large, $f(x) \approx 6$, so perhaps $f(x) < 6$ for all x.

$$6 - f(x) = \frac{6(7+x) - 6 - 6x}{7+x} = \frac{36}{7+x} > 0 \quad \text{if} \quad x > -7$$

In particular, we see that $f(x) \leq 6$ whenever $x \geq 0$. Clearly $a_n \geq 0$ for all n, so $f(a_n) = a_{n+1} \leq 6$ for all n, and the sequence $\{a_n\}$ is increasing and bounded above. Hence $\{a_n\}$ is convergent, with limit l (say).

Since also $a_{n+1} \to l$ as $n \to \infty$, applying 2.10 to the defining equation gives $l = \dfrac{6(1+l)}{7+l}$, or $l^2 + 7l = 6 + 6l$. Thus $l^2 + l - 6 = 0$ or $(l+3)(l-2) = 0$. Thus we can only have limits 2 or -3, and since $a_n \geq 0$ for all n, necessarily $l > 0$. Hence $l = 2$.

Warning: There is a difference between showing that f is increasing, and showing that the sequence is increasing. There is of course a relationship between the function f and the sequence a_n; it is precisely that $f(a_n) = a_{n+1}$. What we know is that *if* f is increasing, then the sequence carries on going the way it starts; if it starts by increasing, as in the above example, then it continues to increase. In the same way, if it starts by decreasing, the sequence will continue to decrease. Show this for yourself.

3.14. Exercise. Define the sequence $\{a_n\}$ by $a_1 = 1$, $a_{n+1} = (4a_n + 2)/(a_n + 3)$ for $n \geq 1$. Show that $\{a_n\}$ is convergent, and find it's limit.

3.15. Proposition. *Let $\{a_n\}$ be an increasing sequence which is convergent to l (In other words it is necessarily bounded above). Then l is an upper bound for the sequence $\{a_n\}$.*

Proof. We argue by contradiction. If l is not an upper bound for the sequence, there is some a_N which witnesses this; i.e. $a_N > l$. Since the sequence is increasing, if $n \geq N$, we have $a_n \geq a_N > l$. Now apply 2.16 to deduce that $l \geq a_N > l$ which is a contradiction. $\square$

3.16. Example. Let $a_n = \left(1 + \dfrac{1}{n}\right)^n$; then $\{a_n\}$ is convergent.

Solution. We show we have an increasing sequence which is bounded above. By the binomial theorem,

$$
\begin{aligned}
\left(1 + \frac{1}{n}\right)^n &= 1 + \frac{n}{n} + \frac{n(n-1)}{2!} \cdot \frac{1}{n^2} + \frac{n(n-1)(n-2)}{3!} \cdot \frac{1}{n^3} + \cdots + \frac{1}{n^n} \\
&\leq 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots \\
&\leq 1 + 1 + \frac{1}{2} + \frac{1}{2.2} + \frac{1}{2.2.2} + \cdots \leq 3.
\end{aligned}
$$

Thus $\{a_n\}$ is bounded above. We show it is increasing in the same way.

$$
\begin{aligned}
\left(1 + \frac{1}{n}\right)^n &= 1 + \frac{n}{n} + \frac{n(n-1)}{2!} \cdot \frac{1}{n^2} + \frac{n(n-1)(n-2)}{3!} \cdot \frac{1}{n^3} + \cdots + \frac{1}{n^n} \\
&= 1 + 1 + \frac{1}{2!}\left(1 - \frac{1}{n}\right) + \frac{1}{3!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) + \ldots
\end{aligned}
$$

from which it clear that a_n increases with n. Thus we have an increasing sequence which is bounded above, and so is convergent by the Monotone Convergence Theorem 3.9. Another method, in which we show the limit is actually e is given on tutorial sheet 3.

3.2.2   The Fibonacci Sequence

3.17. Example. Recall the definition of the sequence which is the ratio of adjacent terms of the Fibonacci sequence as follows:- Define $p_1 = p_2 = 1$, and for $n \geq 1$, let $p_{n+2} = p_{n+1} + p_n$. Let $a_n = \dfrac{p_{n+1}}{p_n}$; then $a_n \to \dfrac{1 + \sqrt{5}}{2}$ as $n \to \infty$. Note that we only have to show that this sequence is convergent; in which case we already know it converges to $\dfrac{1 + \sqrt{5}}{2}$.

Solution. We compute the first few terms.

n	1	2	3	4	5	6	7	8	Formula
p_n	1	1	2	3	5	8	13	21	$p_{n+2} = p_n + p_{n+1}$
a_n	1	2	1.5	1.67	1.6	1.625	1.61	1.62	$a_n = p_{n+1}/p_n$

On the basis of this evidence, we make the following *guesses*, and then try to prove them:

- For all n, we have $1 \leq a_n \leq 2$;

- a_{2n+1} is increasing and a_{2n} is decreasing; and

- a_n is convergent.

Note we are really behaving like proper mathematicians here; the aim is simply to use proof to see if earlier guesses were correct. The method we use could be very like that in the previous example; in fact we choose to use induction more.

> Either method can be used on either example, and you should become familiar with both techniques.

Recall that, by definition,

$$
a_{n+1} = \frac{p_{n+1}}{p_n} = \quad = 1 + \frac{1}{a_n}. \tag{3.3}
$$

Since $p_{n+1} \geq p_n$ for all n, we have $a_n \geq 1$. Also, using our guess,

$$
2 - a_{n+1} = 2 - \left(1 + \frac{1}{a_n}\right) = 1 - \frac{1}{a_n} \geq 0, \quad \text{and} \quad a_n \leq 2.
$$

The next stage is to look at the "every other one" subsequences. First we get a relationship like equation 3.3 for these. (We hope these subsequences are going to be better behaved than the sequence itself).

$$a_{n+2} = 1 + \frac{1}{a_{n+1}} = 1 + \frac{1}{1 + \frac{1}{a_n}} = 1 + \frac{a_n}{1 + a_n}. \tag{3.4}$$

We use this to compute how the difference between successive terms in the sequence behave. Remember we are interested in the "every other one" subsequence. Computing,

$$
\begin{aligned}
a_{n+2} - a_n &= \frac{a_n}{1 + a_n} - \frac{a_{n-2}}{1 + a_{n-2}} \\
&= \frac{a_n + a_n a_{n-2} - a_{n-2} - a_n a_{n-2}}{(1 + a_n)(1 + a_{n-2})} \\
&= \frac{a_n - a_{n-2}}{(1 + a_n)(1 + a_{n-2})}
\end{aligned}
$$

In the above, we already know that the denominator is positive (and in fact is at least 4 and at most 9). This means that $a_{n+2} - a_n$ has the same sign as $a_n - a_{n-2}$; we can now use this information on each subsequence. Since $a_4 < a_2 = 2$, we have $a_6 < a_4$ and so on; by induction, a_{2n} forms a decreasing sequence, bounded below by 1, and hence is convergent to some limit α. Similarly, since $a_3 > a_1 = 1$, we have $a_5 > a_3$ and so on; by induction a_{2n} forms an increasing sequence, bounded above by 2, and hence is convergent to some limit β.

Remember that adjacent terms of both of these sequences satisfied equation 3.4, so as usual, the limit satisfies this equation as well. Thus

$$\alpha = 1 + \frac{\alpha}{1 + \alpha}, \qquad \alpha + \alpha^2 = 1 + 2\alpha, \qquad \alpha^2 - \alpha - 1 = 0 \quad \text{and} \quad \alpha = \frac{1 \pm \sqrt{1 + 4}}{2}$$

Since all the terms are positive, we can ignore the negative root, and get $\alpha = (1 + \sqrt{5})/2$. In exactly the same way, we have $\beta = (1 + \sqrt{5})/2$, and both subsequences converge to the same limit. It is now an easy deduction (which is left for those who are interested - ask if you want to see the details) that the Fibonacci sequence itself converges to this limit.

Chapter 4

Limits and Continuity

4.1 Classes of functions

We first met a sequence as a particularly easy sort of function, defined on $\mathbb{N}$, rather than $\mathbb{R}$. We now move to the more general study of functions. However, our earlier work wasn't a diversion; we shall see that sequences prove a useful tool both to investigate functions, and to give an idea of appropriate methods.

Our main reason for being interested in studying functions is as a model of some behaviour in the real world. Typically a function describes the behaviour of an object over time, or space. In selecting a suitable class of functions to study, we need to balance generality, which has chaotic behaviour, with good behaviour which occurs rarely. If a function has lots of good properties, because there are strong restrictions on it, then it can often be quite hard to show that a given example of such a function has the required properties. Conversely, if it is easy to show that the function belongs to a particular class, it may be because the properties of that class are so weak that belonging may have essentially no useful consequences. We summarise this in the table:

Strong restrictions	Weak restrictions
Good behaviour	Bad behaviour
Few examples	Many examples

It turns out that there are a number of "good" classes of functions which are worth studying. In this chapter and the next ones, we study functions which have steadily more and more restrictions on them. Which means the behaviour steadily improves; and at the same time, the number of examples steadily decreases. A perfectly general function has essentially nothing useful that can be said about it; so we start by studying **continuous** functions, the first class that gives us much theory.

In order to discuss functions sensibly, we often insist that we can "get a good look" at the behaviour of the function at a given point, so typically we restrict the domain of the function to be well behaved.

4.1. Definition. A subset U of $\mathbb{R}$ is **open** if given $a \in U$, there is some $\delta > 0$ such that $(a - \delta, a + \delta) \subseteq U$.

In fact this is the same as saying that given $a \in U$, there is some open interval containing a which lies in U. In other words, a set is open if it contains a neighbourhood of each of its

points. We saw in 1.10 that an open interval is an open set. This definition has the effect that if a function is defined on an open set we can look at its behaviour near the point a of interest, from both sides.

4.2   Limits and Continuity

We discuss a number of functions, each of which is worse behaved than the previous one. Our aim is to isolate an imprtant property of a function called **continuity**.

4.2. Example.   1. Let $f(x) = \sin(x)$. This is defined for all $x \in \mathbb{R}$.

   [*Recall we use radians automatically in order to have the derivative of* $\sin x$ *being* $\cos x$.]

   2. Let $f(x) = \log(x)$. This is defined for $x > 0$, and so naturally has a restricted domain. Note also that the domain is an open set.

   3. Let $f(x) = \dfrac{x^2 - a^2}{x - a}$ when $x \neq a$, and suppose $f(a) = 2a$.

   4. Let $f(x) = \begin{cases} \dfrac{\sin x}{x} & \text{when } x \neq 0, \\ = 1 & \text{if } x = 0. \end{cases}$

   5. Let $f(x) = 0$ if $x < 0$, and let $f(x) = 1$ for $x \geq 0$.

   6. Let $f(x) = \sin \dfrac{1}{x}$ when $x \neq 0$ and let $f(0) = 0$.

   In each case we are trying to study the behaviour of the function near a particular point. In example 1, the function is well behaved everywhere, there are no problems, and so there is no need to pick out particular points for special care. In example 2, the function is still well behaved wherever it is defined, but we had to restrict the domain, as promised in Sect. 1.5. In all of what follows, we will assume the domain of all of our functions is suitably restricted.

   We won't spend time in this course discussing standard functions. It is assumed that you know about functions such as $\sin x$, $\cos x$, $\tan x$, $\log x$, $\exp x$, $\tan^{-1} x$ and $\sin^{-1} x$, as well as the "obvious" ones like polynomials and rational functions — those functions of the form $p(x)/q(x)$, where p and q are polynomials. In particular, it is assumed that you know these are differentiable everywhere they are defined. We shall see later that this is quite a strong piece of information. In particular, it means they are examples of continuous functions. Note also that even a function like $f(x) = 1/x$ is continuous, because, wherever it is defined (ie on $\mathbb{R} \setminus \{0\}$), it is continuous.

   In example 3, the function is not defined at a, but rewriting the function

$$\frac{x^2 - a^2}{x - a} = x + a \quad \text{if } x \neq a,$$

we see that as x approaches a, where the function is not defined, the value of the function approaches $2a$. It thus seems very *reasonable* to extend the definition of f by defining $f(a) = 2a$. In fact, what we have observed is that

$$\lim_{x \to a} \frac{x^2 - a^2}{x - a} = \lim_{x \to a} (x + a) = 2a.$$

We illustrate the behaviour of the function for the case when $a = 2$ in Fig 4.1
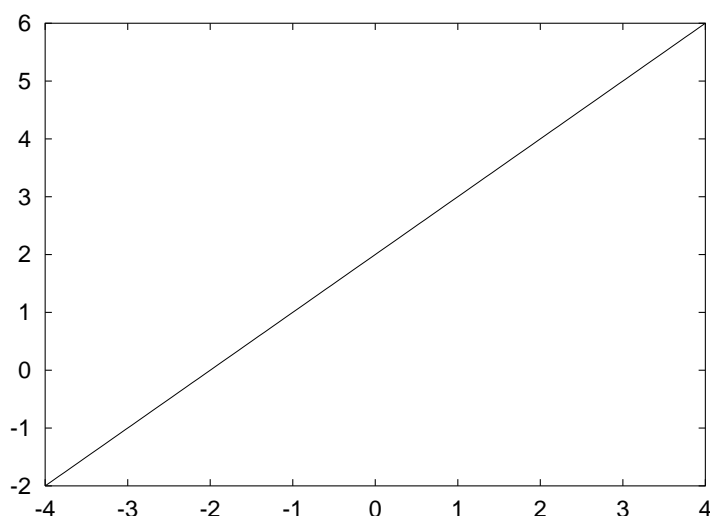

Figure 4.1: Graph of the function $(x^2 - 4)/(x - 2)$ The automatic graphing routine does not even notice the singularity at $x = 2$.

In this example, we can argue that the use of the $(x^2 - a^2)/(x - a)$ was perverse; there was a more natural definition of the function which gave the "right" answer. But in the case of $\sin x/x$, example 4, there was no such definition; we are forced to make the two part definition, in order to define the function "properly" everywhere. So we again have to be careful near a particular point in this case, near $x = 0$. The function is graphed in Fig 4.2, and again we see that the graph shows no evidence of a difficulty at $x = 0$

Considering example 5 shows that these limits need not always exist; we describe this by saying that the limit from the left and from the right both exist, but differ, and the function has a *jump discontinuity* at 0. We sketch the function in Fig 4.3.

In fact this is not the worst that can happen, as can be seen by considering example 6. Sketching the graph, we note that the limit at 0 does not even exists. We prove this in more detail later in 4.23.

The crucial property we have been studying, that of having a definition at a point which is the "right" definition, given how the function behaves near the point, is the property of **continuity**. It is closely connected with the existence of limits, which have an accurate definition, very like the "sequence" ones, and with very similar properties.

4.3. Definition. Say that $f(x)$ **tends to** l as $x \to a$ iff given $\epsilon > 0$, there is some $\delta > 0$ such that whenever $0 < |x - a| < \delta$, then $|f(x) - l| < \epsilon$.

Note that we *exclude* the possibility that $x = a$ when we consider a limit; we are *only* interested in the behaviour of f near a, but not at a. In fact this is very similar to the definition we used for sequences. Our main interest in this definition is that we can now describe continuity accurately.

4.4. Definition. Say that f is **continuous at** a if $\lim_{x \to a} f(x) = f(a)$. Equivalently, f is continuous at a iff given $\epsilon > 0$, there is some $\delta > 0$ such that whenever $|x - a| < \delta$, then $|f(x) - f(a)| < \epsilon$.
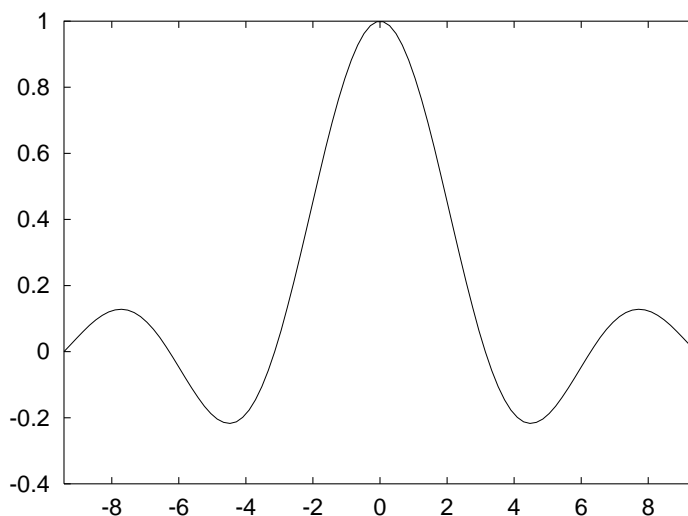
Figure 4.2: Graph of the function $\sin(x)/x$. Again the automatic graphing routine does not even notice the singularity at $x = 0$.
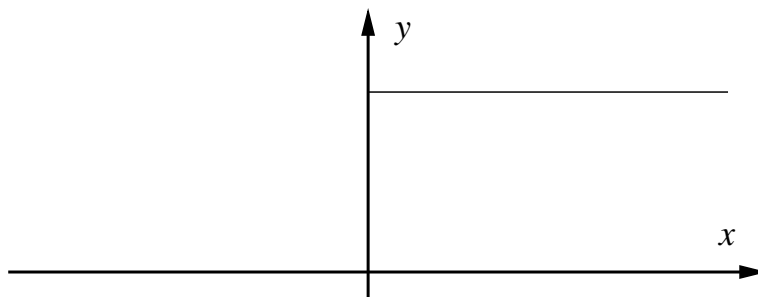
Figure 4.3: The function which is 0 when $x < 0$ and 1 when $x \geq 0$; it has a jump discontinuity at $x = 0$.

Note that in the "epsilon - delta" definition, we no longer need exclude the case when $x = a$. Note also there is a good geometrical meaning to the definition. Given an error ϵ, there is *some* neighbourhood of a such that if we stay in that neighbourhood, then f is trapped within ϵ of its value $f(a)$.

We shall not insist on this definition in the same way that the definition of the convergence of a sequence was emphasised. However, all our work on limts and continuity of functions can be traced back to this definition, just as in our work on sequences, everything could be traced back to the definition of a convergent sequence. Rather than do this, we shall state without proof a number of results which enable continuous functions both to be recognised and manipulated. So you are expected to know the definition, and a few simply $\epsilon - \delta$ proofs, but you can apply (correctly - and always after checking that any needed conditions are satisfied) the standard results we are about to quote in order to do required manipulations.

4.5. Definition. Say that $f : U(open) \to \mathbb{R}$ is **continuous** if it is continuous at each point
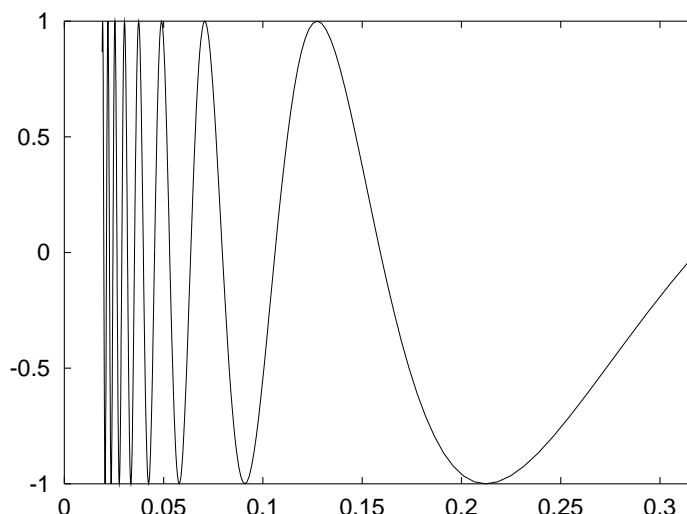
Figure 4.4: Graph of the function $\sin(1/x)$. Here it is easy to see the problem at $x = 0$; the plotting routine gives up near this singularity.

$a \in U$.

Note: This is important. The function $f(x) = 1/x$ is defined on $\{x : x \neq 0\}$, and is a continuous function. We cannot usefully define it on a larger domain, and so, by the definition, it is continuous. This is an example where the naive "can draw it without taking the pencil from the paper" definition of continuity is not helpful.

4.6. Example. Let $f(x) = \dfrac{x^3 - 8}{x - 2}$ for $x \neq 2$. Show how to define $f(2)$ in order to make f a continuous function at 2.

Solution. We have

$$\frac{x^3 - 8}{x - 2} = \frac{(x - 2)(x^2 + 2x + 4)}{(x - 2)} = (x^2 + 2x + 4)$$

Thus $f(x) \to (2^2 + 2.2 + 4) = 12$ as $x \to 2$. So defining $f(2) = 12$ makes f continuous at 2, (and hence for all values of x).

[Can you work out why this has something to do with the derivative of $f(x) = x^3$ at the point $x = 2$?]

4.7. Exercise. Let $f(x) = \dfrac{\sqrt{x} - 2}{x - 4}$ for $x \neq 4$. Show how to define $f(4)$ in order to make f a continuous function at 4.

Sometimes, we can work out whether a function is continuous, directly from the definition. In the next example, it looks as though it is going to be hard, but turns out to be quite possible.

4.8. Example. Let $f(x) = \begin{cases} x \sin\left(\dfrac{1}{x}\right) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$ Then f is continuous at 0.

Solution. We prove this directy from the definition, using that fact that, for all x, we have $|\sin(x) \leq 1|$. Pick $\epsilon > 0$ and choose $\delta = \epsilon$ [We know the answer, but $\delta = \epsilon/2$, or any value of δ with $0 < \delta \leq \epsilon$ will do]. Then if $|x| < \delta$,

$$\left| x \sin \frac{1}{x} - 0 \right| = \left| x \sin \frac{1}{x} \right| = |x|.\left| \sin \frac{1}{x} \right| \leq |x| < \delta \leq \epsilon$$

as required. Note that this is an example where the product of a continuous and a discontinuous function is continuous. The graph of this function is shown in Fig 4.5.
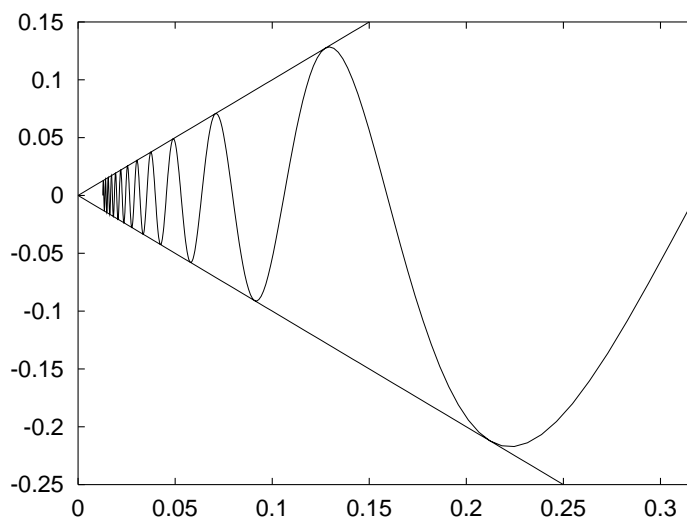
Figure 4.5: Graph of the function $x.\sin(1/x)$. You can probably see how the discontinuity of $sin(1/x)$ gets absorbed. The lines $y = x$ and $y = -x$ are also plotted.

4.3   One sided limits

Although sometimes we get results directly, it is usually helpful to have a larger range of techniques. We give one here and more in section 4.4.

4.9. Definition. Say that $\lim_{x \to a-} f(x) = l$, or that f has a **limit from the left** iff given $\epsilon > 0$, there is some $\delta > 0$ such that whenever $a - \delta < x < a$, then $|f(x) - f(a)| < \epsilon$.

There is a similar definition of "limit from the right", written as $\lim_{x \to a+} f(x) = l$

4.10. Example. Define $f(x)$ as follows:-

$$f(x) = \begin{cases} 3 - x & \text{if } x < 2; \\ 2 & \text{if } x = 2; \\ x/2 & \text{if } x > 2. \end{cases}$$

Calculate the left and right hand limits of $f(x)$ at 2.

Solution. As $x \to 2-$, $f(x) = 3 - x \to 1+$, so the left hand limit is 1. As $x \to 2+$, $f(x) = x/2 \to 1+$, so the right hand limit is 1. Thus the left and right hand limits agree (and disagree with $f(2)$, so f is not continuous at 2).

Note our convention: if $f(x) \to 1$ and always $f(x) \geq 1$ as $x \to 2-$, we say that $f(x)$ tends to 1 **from above**, and write $f(x) \to 1+$ etc.

4.11. Proposition. *If $\lim_{x \to a} f(x)$ exists, then both one sided limts exist and are equal. Conversely, if both one sided limits exits and are equal, then $\lim_{x \to a} f(x)$ exists.*

This splits the problem of finding whether a limit exists into two parts; we can look on either side, and check first that we have a limit, and second, that we get the same answer. For example, in 4.2, example 5, both 1-sided limits exist, but are not equal. There is now an obvious way of checking continuity.

4.12. Proposition. (Continuity Test) *The function f is continuous at a iff both one sided limits exits and are equal to $f(a)$.*

4.13. Example. Let $f(x) = \begin{cases} x^2 & \text{for } x \leq 1, \\ x & \text{for } x \geq 1. \end{cases}$ Show that f is continuous at 1. [In fact f is continuous everywhere].

Solution. We use the above criterion. Note that $f(1) = 1$. Also

$$\lim_{x \to 1-} f(x) = \lim_{x \to 1-} x^2 = 1 \quad \text{while} \quad \lim_{x \to 1+} f(x) = \lim_{x \to 1+} x = 1 = f(1).$$

so f is continuous at 1.

4.14. Exercise. Let $f(x) = \begin{cases} \dfrac{\sin x}{x} & \text{for } x < 0, \\ \cos x & \text{for } x \geq 0. \end{cases}$ Show that f is continuous at 0. [In fact f is continuous everywhere]. [Recall the result of 4.2, example 4]

4.15. Example. Let $f(x) = |x|$. Then f is continuous in $\mathbb{R}$.

Solution. Note that if $x < 0$ then $|x| = -x$ and so is continuous, while if $x > 0$, then $|x| = x$ and so also is continuous. It remains to examine the function at 0. From these identifications, we see that $\lim_{x \to 0-} |x| = 0+$, while $\lim_{x \to 0+} |x| = 0+$. Since $0+ = 0- = 0 = |0|$, by the 4.12, $|x|$ is continuous at 0

4.4 Results giving Coninuity

Just as for sequences, building continuity directly by calculating limits soon becomes hard work. Instead we give a result that enables us to build new continuous functions from old ones, just as we did for sequences. Note that if f and g are functions and k is a constant, then $k.f$, $f + g$, fg and (often) f/g are also functions.

4.16. Proposition. *Let f and g be continuous at a, and let k be a constant. Then $k.f$, $f + g$ and fg are continuous at f. Also, if $g(a) \neq 0$, then f/g is continuous at a.*

Proof. We show that $f + g$ is continuous at a. Since, by definition, we have $(f + g)(a) = f(a) + g(a)$, it is enough to show that

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

Pick $\epsilon > 0$; then there is some δ_1 such that if $|x - a| < \delta_1$, then $|f(x) - f(a)| < \epsilon/2$. Similarly there is some δ_2 such that if $|x - a| < \delta_2$, then $|g(x) - g(a)| < \epsilon/2$. Let $\delta = \min(\delta_1, \delta_2)$, and pick x with $|x - a| < \delta$. Then

$$|f(x) + g(x) - (f(a) + g(a))| \leq |f(x) - f(a)| + |g(x) - g(a)| < \epsilon/2 + \epsilon/2 = \epsilon.$$

This gives the result. The other results are similar, but rather harder; see (Spivak 1967) for proofs. $\qquad\square$

Note: Just as when dealing with sequences, we need to know that f/g is defined in some neighbourhood of a. This can be shown using a very similar proof to the corresponding result for sequences.

4.17. Proposition. *Let f be continuous at a, and let g be continuous at $f(a)$. Then $g \circ f$ is continuous at a*

Proof. Pick $\epsilon > 0$. We must find $\delta > 0$ such that if $|x - a| < \delta$, then $g(f(x)) - g(f(a))| < \epsilon$. We find δ using the given properties of f and g. Since g is continuous at $f(a)$, there is some $\delta_1 > 0$ such that if $|y - f(a)| < \delta_1$, then $|g(y) - g(f(a))| < \epsilon$. Now use the fact that f is continuous at a, so there is some $\delta > 0$ such that if $|x - a| < \delta$, then $|f(x) - f(a)| < \delta_1$. Combining these results gives the required inequality. $\qquad\square$

4.18. Example. The function in Example 4.8 is continuous everywhere. When we first studied it, we showed it was continuous at the "difficult" point $x = 0$. Now we can deduce that it is continuous everywhere else.

4.19. Example. The function $f : x \longmapsto \sin^3 x$ is continuous.

Solution. Write $g(x) = \sin(x)$ and $h(x) = x^3$. Note that each of g and h are continuous, and that $f = g \circ h$. Thus f is continuous.

4.20. Example. Let $f(x) = \tan\left(\dfrac{x^2 - a^2}{x^2 + a^2}\right)$. Show that f is continuous at every point of its domain.

Solution. Let $g(x) = \dfrac{x^2 - a^2}{x^2 + a^2}$. Since $-1 < g(x) < 1$, the function is properly defined for all values of x (whilst $\tan x$ is undefined when $x = (2k + 1)\pi/2$), and the quotient is continuous, since each term is, and since $x^2 + a^2 \neq 0$ for any x. Thus f is continuous, since $f = \tan \circ g$.

4.21. Exercise. Let $f(x) = \exp\left(\dfrac{1 + x^2}{1 - x^2}\right)$. Write down the domain of f, and show that f is continuous at every point of its domain.

As another example of the use of the definitions, we can give a proof of Proposition 2.20

4.22. Proposition. *Let f be a continuous function at a, and let $a_n \to a$ as $n \to \infty$. Then $f(a_n) \to f(a)$ as $n \to \infty$.*

Proof. Pick $\epsilon > 0$ we must find N such that $|a_n - f(a)| < \epsilon$ whenever $n \geq N$. Now since f is continuous at a, we can find δ such that if $|x - a| < \delta$, then $|f(x) - f(a)| < \epsilon$. Also, since $a_n \to a$ as $n \to \infty$ there is some N (taking δ for our epsilon — but anything smaller, like $\delta = \epsilon/2$ etc would work) such that $|a_n - a| < \delta$ whenever $n \geq N$. Combining these we see that if $n \geq N$ then $|a_n - f(a)| < \epsilon$, as required. □

We can consider the example $f(x) = \sin(1/x)$ with this tool.

4.23. Example. Suppose that $\lim_{x \to 0} \sin(1/x) = l$; in other words, assume, to get a contradiction, that the limit exists. Let $x_n = 1/(\pi n)$; then $x_n \to 0$ as $n \to \infty$, and so by assumption, $\sin(1/x_n) = \sin(n\pi) = 0 \to l$ as $n \to \infty$. Thus, just by looking at a single sequence, we see that the limit (if it exists) can only be l. But instead, consider the sequence $x_n = 2/(4n+1)\pi$, so again $x_n \to 0$ as $n \to \infty$. In this case, $\sin(1/x_n) = \sin((4n+1)\pi/2) = 1$, and we must also have $l = 1$. Thus l does not exist.

Note: Sequences often provide a quick way of demonstrating that a function is *not* continuous, while, if f is well behaved on *each* sequence which converges to a, then in fact f is continuous at a. The proof is a little harder than the one we have just given, and is left until next year.

4.24. Example. We know from Prop 4.22 together with Example 4.8 that if $a_n \to 0$ as $n \to \infty$, then

$$a_n \sin\left(\frac{1}{a_n}\right) \to 0 \quad \text{as} \quad n \to \infty.$$

Prove this directly using squeezing.

Solution. The proof is essentially the same as the proof of Example 4.8. We have

$$0 \leq \left| a_n \sin\frac{1}{a_n} \right| = |a_n|.\left| \sin\frac{1}{a_n} \right| \leq |a_n| \to 0 \quad \text{as} \quad n \to \infty.$$

4.5 Infinite limits

There are many more definitions and results about limits. First one that is close to the sequence definition:

4.25. Definition. Say that $\lim_{x \to \infty} f(x) = l$ iff given $\epsilon > 0$, there is some K such that whenever $x > K$, then $|f(x) - l| < \epsilon$.

4.26. Example. Evaluate $\lim_{x \to \infty} \dfrac{x^2 + 3}{3x^2 + 2x + 1}$.

Solution. The idea here should be quite familiar from our sequence work. We use the fact that $1/x \to 0$ as $x \to \infty$. Thus

$$\frac{x^2 + 3}{3x^2 + 2x + 1} = \frac{1 + 3/x^2}{3 + 2/x + 1/x^2} \to \frac{1}{3} \quad \text{as} \quad x \to \infty.$$

4.27. Definition. Say that $\lim_{x\to\infty} f(x) = \infty$ iff given $L > 0$, there is some K such that whenever $x > K$, then $f(x) > L$.

The reason for working on proofs from the definition is to be able to check what results of this type are trivially true without having to find it in a book. For example

4.28. Proposition. *Let $g(x) = 1/f(x)$. Then $g(x) \to 0+$ as $x \to \infty$ iff $f(x) \to \infty$ as $x \to \infty$. Let $y = 1/x$. Then $y \to 0+$ as $x \to \infty$; conversely, $y \to \infty$ as $x \to 0+$*

Proof. Pick $\epsilon > 0$. We show there is some K such that if $x > K$, then $0 < y < \epsilon$; indeed, simply take $K = 1/\epsilon$. The converse is equally trivial. $\qquad\square$

4.6   Continuity on a Closed Interval

So far our results have followed because f is continuous at a particular point. In fact we get the best results from assuming rather more. Indeed the results in this section are precisely why we are interested in discussing continuity in the first place. Although some of the results are "obvious", they only follow from the continuity property, and indeed we present counterexamples whenever that fails. So in order to be able to use the following helpful results, we must first be able to check our functions satisfy the hypothesis of continuity. That task is of course what we have just been concentrating on.

4.29. Definition. Say that f is continuous on $[a, b]$ iff f is continuous on (a, b), and if, in addition, $\lim_{x\to a+} f(x) = f(a)$, and $\lim_{x\to b-} f(x) = f(b)$.

We sometimes refer to f as being continuous on a compact interval. Such an f has some very nice properties.

4.30. Theorem (Intermediate Value Theorem). *Let f be continuous on the compact interval $[a, b]$, and assume that $f(a) < 0$ and $f(b) > 0$. Then there is some point c with $a < c < b$ such hat $f(c) = 0$.*

Proof. We make no attempt to prove this formally, but sketch a proof with a pair of sequences and a repeated bisection argument. It is also noted that each hypothesis is necessary. $\qquad\square$

4.31. Example. Show there is at least one root of the equation $x - \cos x = 0$ in the interval $[0, 1]$.

Proof. Apply the Intermediate Value Theorem to f on the closed interval $[0, 1]$. The function is continuous on that interval, and $f(0) = -1$, while $f(1) = 1 - \cos(1) > 0$. Thus there is some point $c \in (0, 1)$ such that $f(c) = 0$ as required. $\qquad\square$

4.32. Exercise. Show there is at least one root of the equation $x - \mathrm{e}^{-x} = 0$ in the interval $(0, 1)$.

4.33. Corollary. *Let f be continuous on the compact interval $[a, b]$, and assume there is some constant h such that $f(a) < h$ and $f(b) > h$. Then there is a point c with $a < c < b$ such that $f(c) = h$.*

Proof. Apply the Intermediate Value Theorem to $f - h$ on the closed interval $[0, 1]$. Note that by considering $-f + h$ we can cope with the case when $f(a) > h$ and $f(b) < h$. $\square$

Note: This theorem is the reason why continuity is often loosely described as a function you can draw without taking your pen from the paper. As we have seen with $y = 1/x$, this can give an inaccurate idea; it is in fact more akin to connectedness.

4.34. Theorem. (Boundedness) *Let f be continuous on the compact interval $[a, b]$. Then there are constants M and m such that $m < f(x) < M$ for all $x \in [a, b]$. In other words, we are guaranteed that the graph of f is bounded both above and below.*

Proof. This again uses the completeness of $\mathbb{R}$, and again no proof is offered. Note also that the hypotheses are all needed. $\square$

4.35. Theorem. (Boundedness) *Let f be continuous on the compact interval $[a, b]$. Then there are points x_0 and x_1 such that $f(x_0) < f(x) < f(x_1)$ for all $x \in [a, b]$. In other words, we are guaranteed that the graph of f is bounded both above and below, and that these global extrema are actually attained.*

Proof. This uses the completeness of $\mathbb{R}$, and follows in part from the previous result. If M is the least upper bound given by theorem 4.34. Consider the function $g(x) = (M - f(x))^{-1}$. This is clearly continuous. If there is some point at which $f(y) = M$, there is nothing to prove; otherwise g(x) is defined everywhere, and is continuous, and hence bounded. This contradicts the fact that M was a least upper bound for f.

Note also that the hypotheses are all needed. $\square$

Chapter 5

Differentiability

5.1 Definition and Basic Properties

In This section we continue our study of classes of functions which are suitably restricted. Again we are passing from the general to the particular. The next most particular class of function we study after the class of continuous functions is the class of differentiable functions. We discuss the definition, show how to get "new functions from old" in what by now is a fairly routine way, and prove that this is a smaller class: that every differentiable function is continuous, but that there are continuous functions that are not differentiable. Informally, the difference is that the graph of a differentiable function may not have any sharp corners in it.

As with continuous functions, our aim is to show that there are many attractive properties which hold for differentiable functions that don't hold in general for continuous functions. One we discuss in some detail is the ease with which certain limits can be evaluated, by introducing l'Hôpital's rule. Although we don't prove this, the corresponding results are false for continuous functions.

We take the view that much of this material has already been discussed last year, so we move fairly quickly over the basics.

5.1. Definition. Let U be an open subset of $\mathbb{R}$, and let $f : U \to \mathbb{R}$. We say that f is differentiable at $a \in U$ iff

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a} \quad \text{or equivalently,} \quad \lim_{h \to 0} \frac{f(a + h) - f(a)}{h} \quad \text{exists.}$$

The limit, if it exists, is written as $f'(a)$.

We say that f is differentiable in U if and only if it is differentiable at each a in U.

Note that the **Newton quotient** is not defined when $x = a$, nor need it be for the definition to make sense. But the Newton quotient, if it exists, can be extended to be a continuous function at a by defining its value at that point to be $f'(a)$. Note also the emphasis on the existence of the limit. Differentiation is as much about showing the *existence* of the derivative, as *calculating* the value of the derivative.

5.2. Example. Let $f(x) = x^3$. Show, directly from the definition, that $f'(a) = 3a^2$. Compare this result with 4.6.

Solution.   This is just another way of asking about particular limits, like 4.2; we must compute

$$\lim_{x \to a} \frac{x^3 - a^3}{x - a} = \lim_{x \to a} \frac{(x - a)(x^2 + xa + a^2)}{x - a} = \lim_{x \to a}(x^2 + 2xa + a^2) = 3a^2.$$

5.3. Exercise. Let $f(x) = \sqrt{x}$. Show, directly from the definition, that $f'(a) = 1/2\sqrt{a}$ when $a \neq 0$. What function do you have to consider in the particular case when $a = 4$?

Just as with continuity, it is impractical to use this definition every time to compute derivatives; we need results showing how to differentiate the usual class of functions, and we assume these are known from last year. Thus we assume the rules for differentiation of sums products and compositions of functions, together with the known derivatives of elementary functions such as sin, cos and tan; their reciprocals sec, cosec and cot; and exp and log.

5.4. Proposition. *Let f and g be differentiable at a, and let k be a constant. Then $k.f$, $f + g$ and fg are differentiable at f. Also, if $g(a) \neq 0$, then f/g is differentiable at a. Let f be differentiable at a, and let g be differentiable at $f(a)$. Then $g \circ f$ is differentiable at a. In addition, the usual rules for calculating these derivatives apply.*

5.5. Example. Let $f(x) = \tan\left(\dfrac{x^2 - a^2}{x^2 + a^2}\right)$ for $a \neq 0$. Show that f is differentiable at every point of its domain, and calculate the derivative at each such point.

Solution.   This is the same example we considered in 4.20. There we showed the domain was the whole of $\mathbb{R}$, and that the function was continuous everywhere. Let $g(x) = \dfrac{x^2 - a^2}{x^2 + a^2}$. Then g is properly defined for all values of x, and the quotient is differentiable, since each term is, and since $x^2 + a^2 \neq 0$ for any x since $a \neq 0$. Thus f is differentiable using the chain rule since $f = \tan \circ g$, and we are assuming known that the elementary functions like tan are differentiable.

Finally to actually calculate the derivative, we have:-

$$\begin{aligned} f'(x) &= \sec^2\left(\frac{x^2 - a^2}{x^2 + a^2}\right) . \frac{(x^2 + a^2).2x - ((x^2 - a^2).2x)}{(x^2 + a^2)^2} \\ &= \frac{4a^2 x}{(x^2 + a^2)^2} . \sec^2\left(\frac{x^2 - a^2}{x^2 + a^2}\right). \end{aligned}$$

5.6. Exercise. Let $f(x) = \exp\left(\dfrac{1 + x^2}{1 - x^2}\right)$. Show that f is differentiable at every point of its domain, and calculate the derivative at each such point.

The first point in our study of differentiable functions is that it is *more* restrictive for a function to be differentiable, than simply to be continuous.

5.7. Proposition. *Let f be differentiable at a. Then f is continuous at a.*

Proof. To establish continuity, we must prove that $\lim_{x \to a} f(x) = f(a)$. Since the Newton quotient is known to converge, we have for $x \neq a$,

$$f(x) - f(a) = \frac{f(x) - f(a)}{x - a}.(x - a) \to f'(a).0 \quad \text{as } x \to a.$$

Hence f is continuous at a. □

5.8. Example. Let $f(x) = |x|$; then f is continuous everywhere, but *not* differentiable at 0.

Solution. We already know from Example 4.15 that $|x|$ is continuous. We compute the Newton quotient directly; recall that $|x| = x$ if $x \geq 0$, while $|x| = -x$ if $x < 0$. Thus

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

Thus both of the one-sided limits exist, but are unequal, so the limit of the Newton quotient does not exist.

5.2 Simple Limits

Our calculus of differentiable functions can be used to compute limits which otherwise prove troublesome.

5.9. Proposition (l'Hôpital's rule: simple form). *Let f and g be functions such that $f(a) = g(a) = 0$ while $f'(a)$ and $g'(a)$ both exits and $g'(a) \neq 0$. Then*

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

Proof. Since $f(a) = g(a) = 0$, provided $x \neq a$, we have

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f(x) - f(a)}{x - a} \frac{x - a}{g(x) - g(a)} \to \frac{f'(a)}{g'(a)} \quad \text{as} \quad x \to a,$$

where the last limit exists, since $g'(a) \neq 0$. □

5.10. Remark. If $f'(a)$ and $g'(a)$ exist, computing $\lim_{x \to a} \dfrac{f(x)}{g(x)}$ is easy by 4.16, since f and g must be continuous at a by Proposition 5.7, unless we get an indeterminate form $0/0$ or ∞/∞ for the formal quotient. In fact l'Hôpitals rule helps in both cases, although we need to develop stronger forms.

5.11. Example. Show that $\lim_{x \to 0} \dfrac{3x - \sin x}{x} = 2$.

Solution. Note first that we cannot get the result trivially from 4.16, since since $g(a) = 0$ and so we get the indeterminate form $0/0$. However, we are in a position to apply the simple form of l'Hôpital, since $x' = 1 \neq 0$. Applying the rule gives

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

5.12. Example. Show that $\lim\limits_{x \to 0} \dfrac{\sqrt{1+x}-1}{x} = 1/2$.

Solution.   Note again that we cannot get the result trivially from 4.16, since this gives the indeterminate 0/0 form, because $g(a) = 0$. However, we are in a position to apply the simple form of l'Hôpital, since $x' = 1 \neq 0$. Applying the rule gives

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

5.13. Exercise. Evaluate $\lim\limits_{x \to 0} \dfrac{\log(1+x)}{\sin x}$.

5.14. Example. (Spurious, but helps to remember!) Show that $\lim\limits_{x \to 0} \dfrac{\sin x}{x} = 1$.

Solution.   This is spurious because we need the limit to calculate the derivative in the first place, but applying l'Hôpital certainly gives the result.

5.3   Rolle and the Mean Value Theorem

We can combine our definition of derivative with the Intermediate Value Theorem to give a useful result which is in fact the basis of most elementary applications of the differential calculus. Like the results on continuous functions, it is a global result, and so needs continuity and differentiability on a whole interval.

5.15. Theorem (Rolle's Theorem). *Let f be continuous on $[a, b]$, and differentiable on (a, b), and suppose that $f(a) = f(b)$. Then there is some c with $a < c < b$ such that $f'(c) = 0$.*

   Note: The theorem guarantees that the point c exists somewhere. It gives *no* indication of how to find c. Here is the diagram to make the point geometrically:
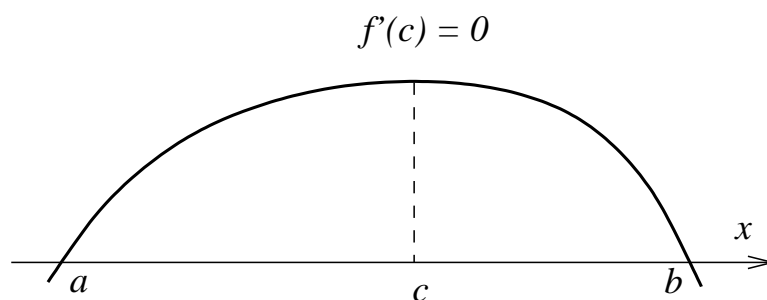
Figure 5.1: If f crosses the axis twice, somewhere between the two crossings, the function is flat. The accurate statement of this "obvious" observation is Rolle's Theorem.

Proof. Since f is continuous on the compact interval $[a, b]$, it has both a global maximum and a global minimum. Assume first that the global maximum occurs at an interior point $c \in (a, b)$. In what follows, we pick h small enough so that $c + h$ always lies in (a, b). Then

If $h > 0$, $\dfrac{f(c + h) - f(c)}{h} \leq 0$, and so $\lim\limits_{h \to 0+} \dfrac{f(c + h) - f(c)}{h} \leq 0$, since we know the limit exists.

Similarly, if $h < 0$, $\dfrac{f(c + h) - f(c)}{h} \geq 0$, and so $\lim\limits_{h \to 0+} \dfrac{f(c + h) - f(c)}{h} \geq 0$, since we know the limit exists. Combining these, we see that $f'(c) = 0$, and we have the result in this case.

A similar argument applies if, instead, the global minimum occurs at the interior point c. The remaining situation occurs if both the global maximum and global minimum occur at end points; since $f(a) = f(b)$, it follows that f is constant, and any $c \in (a, b)$ will do. $\qquad\square$

5.16. Example. Investigate the number of roots of each of the polynomials

$$p(x) = x^3 + 3x + 1 \quad \text{and} \quad q(x) = x^3 - 3x + 1.$$

Solution. Since $p'(x) = 3(x^2 + 1) > 0$ for all $x \in \mathbb{R}$, we see that p has at most one root; for if it had two (or more) roots there would be a root of $p'(x) = 0$ between them by Rolle. Since $p(0) = 1$, while $p(-1) = -3$, there is at least one root by the Intermediate Value Theorem. Hence p has exactly one root.

We have $q'(x) = 3(x^2 - 1) = 0$ when $x = \pm 1$. Since $q(-1) = 3$ and $q(1) = -1$, there is a root of q between -1 and 1 by the Intermediate Value Theorem. Looking as $x \to \infty$ and as $x \to -\infty$ shows here are three roots of q.

5.17. Exercise. Show that the equation $x - e^{-x} = 0$ has exactly one root in the interval $(0, 1)$.

Our version of Rolle's theorem is valuable as far as it goes, but the requirement that $f(a) = f(b)$ is sufficiently strong that it can be quite hard to apply sometimes. Fortunately the geometrical description of the result — that somewhere the tangent is parallel to the axis, does have a more general restatement.

5.18. Theorem (The Mean Value Theorem). *Let f be continuous on $[a, b]$, and differentiable on (a, b). Then there is some c with $a < c < b$ such that*

$$\frac{f(b) - f(a)}{b - a} = f'(c) \quad \text{or equivalently} \quad f(b) = f(a) + (b - a)f'(c).$$

Proof. We apply Rolle to a suitable function; let

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

Then h is continuous on the interval $[a, b]$, since f is, and in the same way, it is differentiable on the open interval (a, b). Also, $f(b) = 0$ and $f(a) = 0$. We can thus apply Rolle's theorem to h to deduce there is some point c with $a < c < b$ such that $h'(c) = 0$. Thus we have

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

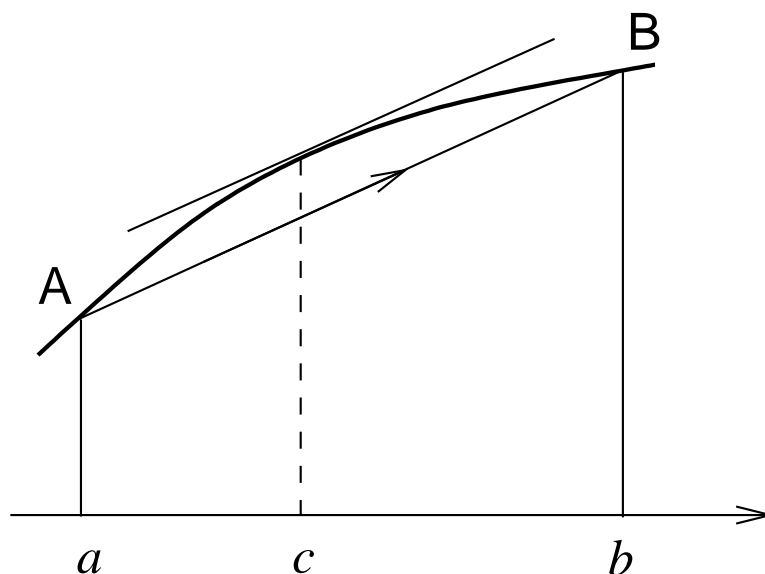which is the required result. $\qquad\square$

Figure 5.2: Somewhere inside a chord, the tangent to f will be parallel to the chord. The accurate statement of this common-sense observation is the Mean Value Theorem.

5.19. Example. The function f satisfies $f'(x) = \dfrac{1}{5 - x^2}$ and $f(0) = 2$. Use the Mean Value theorem to estimate $f(1)$.

Solution. We first estimate the given derivative for values of x satisfying $0 < x < 1$. Since for such x, we have $0 < x^2 < 1$, and so $4 < 5 - x^2 < 5$. Inverting we see that

$$\frac{1}{5} < f'(x) < \frac{1}{4} \quad \text{when} \quad 0 < x < 1.$$

Now apply the Mean Value theorem to f on the interval $[0, 1]$ to obtain some c with $0 < c < 1$ such that $f(1) - f(0) = f'(c)$. From the given value of $f(0)$, we see that $2.2 < f(1) < 2.25$

5.20. Exercise. The function f satisfies $f'(x) = \dfrac{1}{5 + \sin x}$ and $f(0) = 0$. Use the Mean Value theorem to estimate $f(\pi/2)$.

Note the "common sense" description of what we have done. If the derivative doesn't change much, the function will behave linearly. Note also that this gives meaning to the approximation

$$f(a + h) \approx f(a) + hf'(a).$$

We now see that the accurate version of this replaces $f'(a)$ by $f'(c)$ for some c between a and $a + h$.

5.21. Theorem. (The Cauchy Mean Value Theorem) *Let f and g be both continuous on $[a, b]$ and differentiable on (a, b). Then there is some point c with $a < c < b$ such that*

$$g'(c)\Big(f(b) - f(a)\Big) = f'(c)\Big(g(b) - g(a)\Big).$$

In particular, whenever the quotients make sense, we have

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

Proof. Let $h(x) = f(x)\Big(g(b) - g(a)\Big) - g(x)\Big(f(b) - f(a)\Big)$, and apply Rolle's theorem exactly as we did for the Mean Value Theorem. Note first that since both f and g are continuous on $[a, b]$, and differentiable on (a, b), it follows that h has these properties. Also $h(a) = f(a)g(b) - g(a)f(b) = h(b)$. Thus we may apply Rolle to h, to deduce there is some point c with $a < c < b$ such that $h'(c) = 0$. But

$$h'(c) = f'(c)\Big(g(b) - g(a)\Big) - g'(c)\Big(f(b) - f(a)\Big)$$

Thus

$$f'(c)\Big(g(b) - g(a)\Big) = g'(c)\Big(f(b) - f(a)\Big)$$

This is one form of the Cauchy Mean Value Theorem for f and g. If $g'(c) \neq 0$ for any possible c, then the Mean Value theorem shows that $g(b) - g(a) \neq 0$, and so we can divide the above result to get

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

giving a second form of the result.   □

Note: Taking $g(x) = x$ recovers the Mean Value Theorem.

5.4   l'Hôpital revisited

We can get a much more useful form of l'Hôpital's rule using the Cauchy Mean Value Theorem, rather than working, as we did in 5.9, directly from the definition of the derivative.

5.22. Proposition (l'Hôpital's rule: general form). . *Let f and g be functions such that $f(a) = g(a) = 0$, and suppose that f and g are differentiable on an open interval I containing a, and that $g'(x) \neq 0$, except perhaps at a. Then*

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

provided the second limit exists.

Proof. Pick $x > a$ and apply the Cauchy Mean Value Theorem to the interval $[a, x]$, to find c with $a < c < x$ such that

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

Then $\lim\limits_{x \to a+} \dfrac{f(x)}{g(x)} = \lim\limits_{c \to a+} \dfrac{f'(c)}{g'(c)} = \lim\limits_{x \to a+} \dfrac{f'(x)}{g'(x)}$, since we know the actual limit (not just the one sided limit) exists. Now repeat with $x < a$ to get the result.   □

5.23. Example. Evaluate $\lim\limits_{x\to 0} \dfrac{1-\cos x}{x^2} = \dfrac{1}{2}$.

Solution.   We have

$$\lim_{x\to 0} \frac{1-\cos x}{x^2} = \lim_{x\to 0} \frac{\sin x}{2x} = \frac{1}{2},$$

where the use of l'Hôpital is justified since the second limit exists. Note that you **can't** differentiate top and bottom again, and still expect to get the correct answer; one of the hypotheses of l'Hôpital is that the first quotient is of the 0/0 form.

5.24. Example. Use l'Hôpital to establish the following:

$$\lim_{x\to 0} \frac{\sqrt{1+x}-1-x/2}{x^2} = -\frac{1}{8}.$$

Solution.   We have

$$\lim_{x\to 0} \frac{\sqrt{1+x}-1-x/2}{x^2} = \lim_{x\to 0} \frac{(1/2)(1+x)^{-1/2}-1/2}{2x} = \lim_{x\to 0} \frac{(1/2)(-1/2)(1+x)^{-3/2}}{2} = \frac{-1}{8},$$

The use of l'Hôpital is justified the second time, since the third limit exists; since we now know the second limit exists, the use of l'Hôpital is justified the first time.

5.25. Exercise. Evaluate $\lim\limits_{x\to 0}\left(\dfrac{1}{\sin x} - \dfrac{1}{x}\right) = \lim\limits_{x\to 0}\dfrac{x-\sin x}{x\sin x}$.

5.5   Infinite limits

We can use Proposition 4.28 to get results about infinite limits.

5.26. Example. Evaluate $\lim\limits_{x\to\infty} x\log\left(1+\dfrac{1}{x}\right)$.

Solution.

$$\begin{aligned}
\lim_{x\to\infty} x\log\left(1+\frac{1}{x}\right) &= \lim_{x\to\infty} \frac{\log(1+y)}{y} \quad \text{writing } y = 1/x, \\
&= \lim_{y\to 0+} \frac{\log(1+y)}{y} \\
&= \lim_{y\to 0} \frac{\log(1+y)}{y} = 1.
\end{aligned}$$

The last step is valid, since the final limit exists by l'Hôpital; note also that this gives another way of finding $a_n = (1+1/n)^n$.

5.27. Exercise. Evaluate $\lim\limits_{x\to\infty} x\sin\left(\dfrac{1}{x}\right)$.

5.28. Proposition (l'Hôpital's rule: infinite limits). *Let f and g be functions such that $\lim_{x\to\infty} f(x) = \lim_{x\to\infty} g(x) = \infty$, and suppose that $\lim\limits_{x\to\infty} \dfrac{f'(x)}{g'(x)}$ exists. Then*

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

Proof. (Sketch for interest — not part of the course). Pick $\epsilon > 0$ and choose a such that

$$\lim_{x \to \infty} \left| \frac{f'(x)}{g'(x)} - l \right| < \epsilon \quad \text{for all } x > a.$$

Then pick K such that if $x > K$, then $g(x) - g(a) \neq 0$. By Cauchy,

$$\frac{f'(c)}{g'(c)} = \frac{f(x) - f(a)}{g(x) - g(a)} \quad \text{for all } x > K.$$

Note that although c depends on x, we always have $c > a$. Then

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a)}{g(x) - g(a)} \cdot \frac{f(x)}{f(x) - f(a)} \cdot \frac{g(x) - g(a)}{g(x)},$$
$$\to l.1.1 \quad \text{as } x \to \infty.$$

$\square$

5.5.1 (Rates of growth)

One interest in these results is to see how fast functions grow as $x \to \infty$. This is explored further in the exercises. But important results are:

- The function e^x increases faster than any power of x.

- x^α increases faster than any power of $\log x$ if $\alpha > 0$.

5.6 Taylor's Theorem

We have so far explored the Mean Value theorem, which can be rewritten as

$$f(a + h) = f(a) + hf'(c)$$

where c is some point between a and $a + h$. [By writing the definition of c in this way, we have a statement that works whether $h > 0$ or $h < 0$.] We have already met the approximation

$$f(a + h) \sim f(a) + hf'(a)$$

when we studied the Newton - Raphson method for solving an equation, and have already observed that the Mean Value Theorem provides a more accurate version of this. Now consider what happens when f is a polynomial of degree n

$$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_{n-1} x^{n-1} + a_n x^n.$$

Note that $f(0) = a_0$. Differentiating gives

$$f'(x) = a_1 + 2a_2 x + 3a_3 x^2 \ldots + (n-1)a_{n-1} x^{n-2} + n a_n x^{n-1},$$

and so $f'(0) = a_1$. Again, we have

$$f''(x) = 2a_2 + 3.2a_3 x \ldots + (n-1)(n-2)a_{n-1} x^{n-3} + n(n-1)a_n x^{n-2},$$

and $f''(0) = 2a_2$. After the next differentiation, we get $f'''(0) = 3!a_3$, while after k differentiations, we get, $f^{(k)}(0) = k!a_k$, provided $k \leq n$. Thus we can rewrite the polynomial, using its value, and the value of its derivatives at 0, as

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \ldots + \frac{f^{(n-1)}(0)}{(n-1)!}x^{n-1} + \frac{f^{(n)}(0)}{n!}x^n.$$

This opens up the possibility of representing more general functions than polynomials in this way, and so getting a generalisation of the Mean Value Theorem.

5.29. Theorem (Taylors Theorem - Lagrange form of Remainder). *Let f be continuous on $[a, x]$, and assume that each of f', f'', $\ldots$, $f^{(n+1)}$ is defined on $[a, x]$. Then we can write*

$$f(x) = P_n(x) + R_n(x),$$

where $P_n(x)$, the Taylor polynomial of degree n about a, and $R_n(x)$, the corresponding remainder, are given by

$$\begin{aligned} P_n(x) &= f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(n)}(a)}{n!}(x-a)^n, \\ R_n(x) &= \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}, \end{aligned}$$

where c is some point between a and x.

We make no attempt to prove this, although the proof can be done with the tools we have at our disposal. Some quick comments:

- the theorem is also true for $x < a$; just restate it for the interval $[x, a]$ etc;

- if $n = 0$, we have $f(x) = f(a) + (x-a)f'(c)$ for some c between a and x; this is a restatement of the Mean Value Theorem;

- if $n = 1$, we have

$$f(x) = f(a) + (x-a)f'(x) + \frac{f''(c)}{2!}(x-a)^2$$

for some c between a and x; this often called the **Second Mean Value Theorem**;

- in general we can restate **Taylor's Theorem** as

$$f(x) = f(a) + (x-a)f'(x) + \ldots + \frac{f^{(n)}(a)}{2!}(x-a)^n + \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1},$$

for some c between a and x;

- the special case in which $a = 0$ has a special name; it is called **Maclaurin's Theorem**;

- just as with Rolle, or the Mean Value Theorem, there is **no** useful information about the point c.

We now explore the meaning and content of the theorem with a number of examples.

5.30. Example. Find the Taylor polynomial of order n about 0 for $f(x) = e^x$, and write down the corresponding remainder term.

Solution.   There is no difficulty here in calculating derivatives — clearly $f^{(k)}(x) = e^x$ for all k, and so $f^{(k)}(0) = 1$. Thus by Taylor's theorem,

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^2}{2!} + \dots \frac{x^n}{n!} + \frac{x^{n+1}}{(n+1)!} e^c$$

for some point c between 0 and x. In particular,

$$P_n(x) = 1 + x + \frac{x^2}{2!} + \frac{x^2}{2!} + \dots \frac{x^n}{n!} \quad \text{and} \quad R_n(x) = \frac{x^{n+1}}{(n+1)!} e^c.$$

We can actually say a little more about this example if we recall that x is *fixed*. We have

$$e^x = P_n(x) + R_n(x) = P_n(x) + \frac{x^{n+1}}{(n+1)!} e^c$$

We show that $R_n(x) \to 0$ as $n \to \infty$, so that (again for fixed x), the sequence $P_n(x) \to e^x$ as $n \to \infty$. If $x < 0$, $e^c < 1$, while if $x \geq 1$, then since $c < x$, we have $e^c < e^x$. thus

$$|R_n(x)| = \left| \frac{x^{n+1}}{(n+1)!} e^c \right| \leq \frac{|x|^{n+1}}{(n+1)!} \max(e^x, 1) \to 0 \quad \text{as } n \to \infty.$$

We think of the limit of the polynomial as forming a series, the **Taylor series** for e^x. We study series (and then Taylor series) in Section 7.

5.31. Example. Find the Taylor polynomial of order 1 about a for $f(x) = e^x$, and write down the corresponding remainder term.

Solution. Using the derivatives computed above, by Taylor's theorem,

$$e^x = e^a + (x - a) e^a + \frac{(x - a)^2}{2!} e^c$$

for some point c between a and x. In particular,

$$P_1(x) = e^a + (x - a) e^a \quad \text{and} \quad R_1(x) = \frac{(x - a)^2}{2!} e^c.$$

5.32. Example. Find the Maclaurin polynomial of order $n > 3$ about 0 for $f(x) = (1 + x)^3$, and write down the corresponding remainder term.

Solution. We have

$$\begin{aligned}
f(x) &= (1 + x)^3 \\
f'(x) &= 3(1 + x)^2 \\
f''(x) &= 6(1 + x) \\
f'''(x) &= 6 \\
f^{(n)}(x) &= 0 \quad \text{if } n > 3.
\end{aligned} \qquad
\begin{aligned}
f(0) &= 1 \\
f'(x) &= 3 \\
f''(x) &= 6 \\
f'''(x) &= 6
\end{aligned}$$

and so, by Taylor's theorem

$$(1+x)^3 = 1 + 3x + \frac{6}{2!}x^2 + \frac{6}{3!}x^3,$$

a result we could have got directly, but which is at least reassuring.

5.33. Example. Find the Taylor polynomial of order n about 0 for $f(x) = \sin x$, and write down the corresponding remainder term.

Solution.   There is no difficulty here in calculating derivatives — we have

$$\begin{aligned}
f(x) &= \sin x \\
f'(x) &= \cos x \\
f''(x) &= -\sin x \\
f'''(x) &= -\cos x \\
f^{(4)}(x) &= \sin x \quad \text{and so on.}
\end{aligned} \qquad \begin{aligned}
f(0) &= 0 \\
f'(x) &= 1 \\
f''(x) &= 0 \\
f'''(x) &= -1.
\end{aligned}$$

Thus by Taylor's theorem,

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots + (-1)^{n+1}\frac{x^{2n+1}}{(2n+1)!} + \ldots$$

Writing down the remainder term isn't particularly useful, but the important point is that

$$|R_{2n+1}(x)| \leq \left|\frac{x^{2n+3}}{(2n+3)!}\right| \to 0 \quad \text{as } n \to \infty.$$

5.34. Exercise. Recall that $\cosh x = \dfrac{e^x + e^{-x}}{2}$, and that $\sinh x = \dfrac{e^x - e^{-x}}{2}$. Now check the shape of the following Taylor polynomials:

$$\begin{aligned}
\cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots + (-1)^n\frac{x^{2n}}{2n!} + \ldots \\
\sinh x &= x + \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots + \frac{x^{2n+1}}{(2n+1)!} + \ldots \\
\cosh x &= 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots + \frac{x^{2n}}{2n!} + \ldots
\end{aligned}$$

5.35. Example. Find the maximum error in the approximation

$$\sin(x) \sim x - \frac{x^3}{3!}$$

given that $|x| < 1/2$.

Solution.   We use the Taylor polynomial for $\sin x$ of order 4 about 0, together with the corresponding remainder. Thus

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!}\cos c$$

for some c with $0 < c < 1/2$ or $-1/2 < c < 0$. In any case, since $|x| < 1/2$,

$$\left| \frac{x^5}{5!} \cos c \right| \leq \left| \frac{x^5}{5!} \right| \leq \left| \frac{1}{2^5.5!} \right| \leq \frac{1}{120.32}.$$

Warning: The Taylor polynomial always exists, providing f is suitably differentiable. But it need not be useful. Consider the example

$$f(x) = \begin{cases} \exp(-1/x^2) & \text{if } x > 0; \\ 0 & \text{if } x \leq 0. \end{cases}$$

The interest in f is at 0; it is well behaved everywhere else. It turns out that

$$f(0) = f'(0) = f''(0) = \ldots = f^{(n)}(0) = \ldots = 0.$$

So the Taylor polynomial of degree n for f about 0 is $P_n(x) = 0 + 0x + 0x^2 + \ldots + 0.x^n = 0$, and so for every n, $R_n(x) = f(x)$. Clearly in this case, P_n tells us nothing useful about the function.

5.36. Example. Find the Taylor polynomial of order n about 0 for $f(x) = (1+x)^\alpha$, and note that this gives a derivation of the binomial theorem. In fact, the remainder $|R_n(x)| \to 0$ as $n \to \infty$, provided $|x| < 1$.

Solution. There is again no difficulty here in calculating derivatives — we have

$$\begin{array}{ll}
f(x) = (1+x)^\alpha & f(0) = 1 \\
f'(x) = \alpha(1+x)^{\alpha-1} & f'(x) = \alpha \\
f''(x) = \alpha(\alpha-1)(1+x)^{\alpha-2} & f''(x) = \alpha(\alpha-1) \\
f'''(x) = \alpha(\alpha-1)(\alpha-2)(1+x)^{\alpha-3} & f'''(x) = \alpha(\alpha-1)(\alpha-2) \\
\text{and in general} & \\
f^{(n)}(x) = \alpha(\alpha-1)\ldots(\alpha-n+1)(1+x)^{\alpha-n} & f^{(n)} = \alpha(\alpha-1)\ldots(\alpha-n+1).
\end{array}$$

Thus by Taylor's theorem,

$$(1+x)^\alpha = 1 + \alpha x - \frac{\alpha(\alpha-1)}{2!}x^2 + \frac{\alpha(\alpha-1)(\alpha-2)}{3!}x^3 +$$
$$\ldots + \frac{\alpha(\alpha-1)\ldots(\alpha-n+1)}{n!}x^n + \ldots.$$

The remainder is not hard to deal with, but we omit the proof; in fact $|R_n(x)| \to 0$ when $n \to \infty$.

Note also that if $\alpha > 0$ is an *integer*, say $\alpha = n$ then $|R_n(x)| = 0$ and $f(x) = P_n(x)$. This is another way to get the Binomial theorem described in Section 1.8.

Chapter 6

Infinite Series

In this section we return to study a particular kind of sequence, those built by adding up more and more from a given collection of terms. One motivation comes from section 5.6, in which we obtained a sequence of approximating polynomials $\{P_n\}$ to a given function. It is more natural to think of adding additional terms to the polynomial, and as such we are studying **series**. However, there is a closely related sequence; the sequence of partial sums.

6.1 Arithmetic and Geometric Series

Consider the sum

$$a + (a + r) + (a + 2r) + \cdots + (a + nr)$$

We are trying to add up the terms in an **arithmetic progression**. A small amount of notation makes the addition easy.

$$
\begin{aligned}
\text{Let } S_n &= a + (a + r) + (a + 2r) + \cdots + (a + nr), \\
S_n &= (a + nr) + (a + (n - 1)r) + \cdots + (a + r) + a, \\
\text{so } 2S_n &= (2a + nr) + (2a + nr) + \cdots + (2a + nr), \text{ and} \\
S_n &= (n + 1).\frac{(a + (a + nr))}{2}.
\end{aligned}
$$

Note that if $r > 0$ then $S_n \to \infty$ as $n \to \infty$, while if $r < 0$ then $S_n \to -\infty$ as $n \to \infty$.

We next consider a **geometric progression (or series)**:

$$
\begin{aligned}
\text{Let } S_n &= a + ar + ar^2 + \cdots + ar^n, \\
rS_n &= ar + ar^2 + ar^3 + \cdots + ar^{n+1}, \\
\text{so } (1 - r)S_n &= a(1 - r^{n+1}), \text{ and} \\
S_n &= \frac{a(1 - r^{n+1})}{1 - r} \quad \text{if} \quad r \neq 1.
\end{aligned}
$$

Note that if $|r| < 1$, then $S_n \to \dfrac{a}{1 - r}$ as $n \to \infty$.

If $r > 1$, say $r = 1 + k$,

$$
\begin{aligned}
S_n &= a\left(\frac{(1+k)^{n+1} - 1}{k}\right), \\
&> \frac{a}{k}\left(1 + (n+1)k - 1\right) > a(n+1) \to \infty \quad \text{if} \quad a > 0.
\end{aligned}
$$

6.1. Example. Find $\displaystyle\sum_{n=1}^{\infty}\left(\frac{1}{2^n} + \frac{1}{3^n}\right)$.

Solution.

$$
\sum\left(\frac{1}{2^n} + \frac{1}{3^n}\right) = \sum\frac{1}{2^n} + \sum\frac{1}{3^n} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots + \frac{1}{3} + \frac{1}{9} + \cdots
$$

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

6.2. Exercise. Find $\displaystyle\sum_{n=1}^{\infty}\left(7\frac{1}{3^n} - 4\frac{1}{2^n}\right)$.

6.2   Convergent Series

6.3. Definition. Let $\{a_n\}$ be a sequence, and let

$$
S_n = a_1 + a_2 + \cdots + a_n, \quad \text{the } n^{th} \text{ partial sum.}
$$

If $\displaystyle\lim_{n\to\infty} S_n$ exists, we say that $\sum a_n$ is a **convergent series**, and write $\displaystyle\lim_{n\to\infty} S_n = \sum a_n$.

Thus a series is convergent if and only if it's sequence of partial sums is convergent. The limit of the sequence of partial sums is the **sum of the series**. A series which is *not* convergent, is a **divergent series**.

6.4. Example. The series $\sum r^n$ is convergent with sum $1/(1 - r)$, provided that $|r| < 1$. For other values of r, the series is divergent; in particular, the series $\sum(-1)^n$ is divergent.

Solution. We noted above that when $|r| < 1$, $S_n \to a/(1 - r)$ as $n \to \infty$; note particular cases;

$$
\sum_{n=1}^{\infty} \frac{1}{2^n} = 1 \quad \text{or equivalently,} \quad \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = 1.
$$

6.5. Example. The sum $\displaystyle\sum \frac{1}{n(n+1)}$ is convergent with sum 1.

Solution. We can compute the partial sums explicitly:

$$
S_n = \sum_{k=1}^{n} \frac{1}{k(k+1)} = \sum_{k=1}^{n}\left(\frac{1}{k} - \frac{1}{k+1}\right) = 1 - \frac{1}{n+1} \to 1 \quad \text{as} \quad n \to \infty.
$$

6.6. Example. The sum $\sum \dfrac{1}{n}$ is divergent.

Solution. We estimate the partial sums:

$$
\begin{aligned}
S_n &= \frac{1}{1} + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \cdots + \frac{1}{8}\right) + \left(\frac{1}{9} + \cdots + \frac{1}{15}\right) + \cdots + \frac{1}{n} \\
&> 1 + \frac{1}{2} + \frac{2}{4} + \frac{4}{8} > 2\frac{1}{2} \quad \text{if } n \ge 15 \\
&> 1 + \frac{1}{2} + \frac{2}{4} + \frac{4}{8} + \frac{8}{16} > 3 \quad \text{if } n \ge 31 \\
&\to \quad \infty \quad \text{as} \quad n \to \infty.
\end{aligned}
$$

6.7. Example. The sum $\sum \dfrac{1}{n^2}$ is convergent. [Actually the sum is $\pi^2/6$, but this is much harder.]
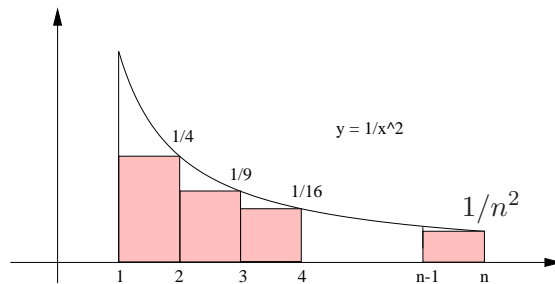

Figure 6.1: Comparing the area under the curve $y = 1/x^2$ with the area of the rectangles below the curve

Solution. We estimate the partial sums. Since $1/n^2 > 0$, clearly $\{S_n\}$ is an increasing sequence. We show it is bounded above, whence by the Monotone Convergence Theorem (3.9), it is convergent. From the diagram,

$$
\frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} \quad < \quad \int_1^n \frac{d\,x}{x^2}, \quad \text{and so}
$$

$$
S_n \quad < \quad 1 + \left[-\frac{1}{x}\right]_1^n \le 2 - \frac{1}{n}.
$$

Thus $S_n < 2$ for all n, the sequence of partial sums is bounded above, and the series is convergent.

6.8. Proposition. *Let $\sum a_n$ be convergent. Then $a_n \to 0$ as $n \to \infty$.*

Proof. Write $l = \lim_{n\to\infty} S_n$, and recall from our work on limits of sequences that $S_{n-1} \to l$ as $n \to \infty$. Then

$$
a_n = (a_1 + a_2 + \ldots a_n) - (a_1 + a_2 + \ldots a_{n-1}) = S_n - S_{n-1} \to l - l \quad \text{as} \quad n \to \infty.
$$

$\square$

6.9. Remark. This gives a *necessary* condition for the convergence of a series; it is *not* sufficient. For example we have seen that $\sum 1/n$ is divergent, even though $1/n \to 0$ as $n \to \infty$.

6.10. Example. The sum $\sum \dfrac{1}{n}$ is divergent (Graphical method).

Solution. We estimate the partial sums. Since $1/n > 0$, clearly $\{S_n\}$ is an increasing sequence. We show it is not bounded above, whence by the note after 3.9, the sequence of partial sums $\to \infty$ as $n \to \infty$.
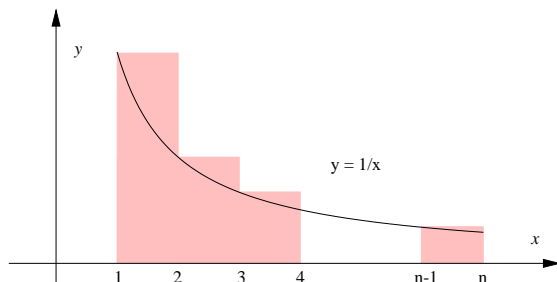

Figure 6.2: Comparing the area under the curve $y = 1/x$ with the area of the rectangles above the curve

From the diagram,

$$1 + \frac{1}{2} + \cdots + \frac{1}{n} > \int_1^n \frac{d\,x}{x} > \frac{1}{2} + \cdots + \frac{1}{n}.$$

Writing $S_n = 1 + \dfrac{1}{2} + \cdots + \dfrac{1}{n}$, we have $S_n > \log n > S_n - 1$, or equivalently $1 + \log n > S_n > \log n$ for all n. Thus $S_n \to \infty$ and the series is divergent. [There is a much better estimate; the difference $S_n - \log n \to \gamma$ as $n \to \infty$, where γ is Euler's constant.]

6.11. Proposition. *Let $\sum a_n$ and $\sum b_n$ be convergent. Then $\sum(a_n + b_n)$ and $\sum c.a_n$ are convergent.*

Proof. This can be checked easily directly from the definition; it is in effect the same proof that the sum of two convergent sequences is convergent etc. □

6.3   The Comparison Test

We have already used the Monotone Convergence Theorem in studying simple series. In fact it is a lot more useful. When we know the behaviour of some simple series, we can deduce many more results by comparison as follows.

6.12. Theorem. (The Comparison Test) *Assume that $0 \le a_n \le b_n$ for all n, and suppose that $\sum b_n$ is convergent. Then $\sum a_n$ is convergent.*

Proof. Define

$$
\begin{aligned}
S_n &= a_1 + a_2 + \cdots + a_n, \\
T_n &= b_1 + b_2 + \cdots b_n.
\end{aligned}
$$

Then by hypothesis, $0 \leq S_n \leq T_n$. Since $\{T_n\}$ is a convergent sequence, it is a bounded sequence by Prop 2.28. In particular, it is bounded above, so there is some K such that $T_n \leq K$ for all n. Thus $S_n \leq K$ for all n, so $\{S_n\}$ is a sequence that is bounded above; since $a_n \geq 0$, $S_{n+1} = S_n + a_n \geq S_n$ and $\{S_n\}$ is an increasing sequence. Thus by the Monotone Convergence Theorem, it is a convergent sequence $\qquad\square$

.

6.13. Example. Let $a_n = \dfrac{2n}{3n^3 - 1}$ and let $b_n = \dfrac{1}{n^2}$. Then $\sum a_n$ is convergent

Solution. For $n \geq 1$, $n^3 \geq 1$, so $3n^3 - 1 \geq 2n^3$. Thus

$$a_n = \frac{2n}{3n^3 - 1} \leq \frac{2n}{2n^3} = b_n.$$

Since we know that $\sum b_n$ is convergent, so is $\sum a_n$.

6.14. Remark. The conclusions of Theorem 6.12 remain true even if we only have $a_n \leq b_n$ eventually; for if it holds for $n \geq N$, we replace the inequality by

$$S_n \leq T_n + a_1 + a_2 + \cdots + a_N$$

and this then holds for all n.

6.15. Example. Let $a_n = \dfrac{\log n}{n}$, and compare with $b_n = \dfrac{1}{n}$.

Solution. Note that if $n \geq 3$, then $\log n > 1$. We can thus use the "eventually" form of the comparison test; we have

$$a_n = \frac{\log n}{n} > \frac{1}{n} = b_n,$$

We deduce divergence, for if $\sum a_n$ were convergent, it follows that that $\sum b_n$ was convergent, which it isn't!

6.16. Corollary (Limiting form of the Comparison Test). *Suppose that $a_n > 0$ and $b_n > 0$, and that there is some constant k such that $\lim\limits_{n \to \infty} \dfrac{a_n}{b_n} = k > 0$. Then $\sum a_n$ is convergent iff $\sum b_n$ is convergent.*

Proof. Assume first that $\sum b_n$ is convergent. Since $a_n/b_n \to k$ as $n \to \infty$, eventually (take $\epsilon = k > 0$), we have $a_n \leq 2kb_n$, which is convergent by 6.11. Hence $\sum a_n$ is convergent by 6.12. To get the converse, note that $b_n/a_n \to 1/k$ as $n \to \infty$, so we can use the same argument with a_n and b_n interchanged. $\qquad\square$

6.17. Example. Let $a_n = \dfrac{n}{n^2 + 1}$ and let $b_n = \dfrac{1}{n}$. Then $\sum a_n$ is divergent by the limiting form of the comparison test.

Solution. Note that the terms are all positive, so we try to apply the limiting form of the comparison test directly.

$$\frac{a_n}{b_n} = \frac{n}{n^2 + 1} \cdot \frac{n}{1} = \frac{n^2}{n^2 + 1} \to 1 \quad \text{as } n \to \infty.$$

Since the limit is non-zero, the use of the limiting form of the comparison test is valid, and we see that $\sum a_n$ is divergent.

Note also we need our work on sequences in Section 2 to evaluate the required limit.

This is all very well, but as with the "new sequences from old" programme, we need a few reference sequences before we can get further. One set is the geometric series, which we have already met.

6.18. Exercise. Let $a_n = \dfrac{2^n + 7}{3^n - 1}$ and let $b_n = \left(\dfrac{2}{3}\right)^n$. Use the limiting form of the comparison test to show that $\sum a_n$ is convergent.

We also know about $\sum 1/n^\alpha$, at least when $\alpha \geq 2$ when it converges, by comparison with $\sum 1/n^2$, and when $\alpha \leq 1$ when it diverges, by comparison with $\sum 1/n$.

6.19. Proposition. *The sum* $\sum \dfrac{1}{n^\alpha}$ *is convergent when* $\alpha > 1$.

Solution. Assume $\alpha > 1$; we estimate the partial sums. Since $1/n^\alpha > 0$, clearly $\{S_n\}$ is an increasing sequence. Let $S_n = 1 + \dfrac{1}{2^\alpha} + \cdots + \dfrac{1}{n^\alpha}$, and consider the graph of $y = 1/x^\alpha$, noting that y is a decreasing function of x (which is where we use that fact that $\alpha > 1$). From a diagram which is essentially the same as that of Fig 6.1,

$$\text{shaded area} \quad < \int_1^n \frac{d\,x}{x^\alpha}$$

$$\frac{1}{2^\alpha} + \cdots + \frac{1}{n^\alpha} < \left[\frac{1}{1-\alpha}\frac{1}{x^{\alpha-1}}\right]_1^n$$

$$S_n - 1 < \frac{1}{\alpha - 1}\left(1 - \frac{1}{n^{\alpha-1}}\right)$$

$$S_n < \frac{1}{\alpha - 1} + 1$$

Thus the sequence of partial sums is bounded above, and the series converges.

6.20. Exercise. Let $a_n = \dfrac{n}{\sqrt{n^5 + n + 1}}$ and let $b_n = \dfrac{1}{n^{3/2}}$. Use the limiting form of the comparison test to show that $\sum a_n$ is convergent.

We can consider the method of comparing with integrals as an "integral test" for the convergence of a series; rather than state it formally, note the method we have used.

6.21. Theorem (The Ratio Test). *Let* $\sum a_n$ *be a series, and assume that* $\lim \dfrac{|a_{n+1}|}{|a_n|} \to r$ *as* $n \to \infty$. *Then if* $r < 1$, *the series is convergent, if* $r > 1$, *the series is divergent, while if* $r = 1$, *the test gives no information.*

Proof. A proof follows by comparing with the corresponding geometric series with ratio r. Details will be given in full in the third year course. $\square$

6.22. Example. Let $a_n = \dfrac{2^n (n!)^2}{(2n)!}$. Then $\sum a_n$ is convergent.

Solution. We look the ratio of adjacent terms in the series (of positive terms).

$$\frac{(n+1)^{\text{st}} \text{ term}}{n^{\text{th}} \text{ term}} = \frac{a_{n+1}}{a_n} = \frac{2^{n+1}(n+1)!(n+1)!}{(2n+2)!} \frac{(2n)!}{2^n n! n!}$$

$$= \frac{2(n+1)^2}{(2n+2)(2n+1)} = \frac{n+1}{2n+1} \to \frac{1}{2} \quad \text{as } n \to \infty.$$

Since the ratio of adjacent terms in the series tends to a limit which is < 1, the series converges by the ratio test.

6.4 Absolute and Conditional Convergence

So far most of our work has been with series all of whose terms are positive. There is a good reason for this; there is very little we can say about series with mixed signs. Indeed there is just one useful result at this level, which is the topic of this section.

The easiest case occurs when the series really can be thought of as a series of positive terms.

6.23. Definition. The series $\sum_{n=1}^{\infty} a_n$ is **absolutely convergent** iff the series $\sum_{n=1}^{\infty} |a_n|$ is convergent.

6.24. Definition. The series $\sum_{n=1}^{\infty} a_n$ is **conditionally convergent** if and only if the series $\sum_{n=1}^{\infty} a_n$ is convergent but not absolutely convergent.

6.25. Example. Show that the series $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^2}$ is absolutely convergent.

Solution. We have $\sum_{n=1}^{\infty} \left| \frac{(-1)^{n+1}}{n^2} \right| = \sum_{n=1}^{\infty} \frac{1}{n^2}$ and this is convergent by 6.19. So $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^2}$ is absolutely convergent.

Note: We choose to work with the sign $(-1)^{n+1}$ rather than $(-1)^n$ simply for tidiness; it is usual to start a series with a positive term, so the coefficient of a_1 is chosen to be $+$. Thus that of a_2 must be $-$ etc if the series is to alternate.

6.26. Exercise. Show that the series $\sum_{n=2}^{\infty} \frac{(-1)^n}{n^2 \log n}$ is absolutely convergent.

6.27. Exercise. Show that the series $\sum_{n=1}^{\infty} \frac{\sin n}{n^2}$ is absolutely convergent. [Hint: note that $|\sin n| \leq 1$, and use the comparison test.]

Our interest in absolutely convergent series starts by observing that they are in fact all convergent. Indeed this is the easiest way to show a series is convergent if the terms are not all positive.

6.28. Proposition. *An absolutely convergent series is convergent.*

Proof. Assume that $\sum_{n=1}^{\infty} a_n$ is absolutely convergent, and define

$$a_n^+ = \begin{cases} a_n & \text{if } a_n > 0 \\ 0 & \text{if } a_n \le 0 \end{cases} \quad \text{and} \quad a_n^- = \begin{cases} |a_n| & \text{if } a_n < 0 \\ 0 & \text{if } a_n \ge 0 \end{cases}$$

The point of this definition is that

$$0 \le a_n^+ \le |a_n| \quad \text{and} \quad 0 \le a_n^- \le |a_n| \quad \text{for all n,} \tag{6.1}$$

so we have two new series of positive terms, while

$$|a_n| = a_n^+ + a_n^- \quad \text{and} \quad a_n = a_n^+ - a_n^-.$$

Using equation 6.1 to compare with the convergent series $\sum_{n=1}^{\infty} |a_n|$, we see that each of

$$\sum_{n=1}^{\infty} a_n^+ \quad \text{and} \quad \sum_{n=1}^{\infty} a_n^-$$

is a convergent series of positive terms. Thus

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} a_n^+ - a_n^- = \sum_{n=1}^{\infty} a_n^+ + \sum_{n=1}^{\infty} a_n^-$$

is also convergent using 6.11.   □

This gives one way of proving that a series is convergent even if the terms are not all positive, and so we can't use the comparison test directly. There is essentially only one other way, which is a very special, but useful case known as Leibniz theorem, or the theorem on alternating signs, or the **alternating series test**. We give the proof because the argument is so like the proof of the convergence of the ratio of adjacent terms in the Fibonacci series 3.1.

Warning: Note how we usually talk about the "Fibonacci series", even though it is a *sequence* rather than a series. Try not to be confused by this popular but inaccurate usage.

6.29. Theorem. *Leibniz Theorem* Let $\{a_n\}$ be a **decreasing** sequence of **positive** terms such that $a_n \to 0$ as $n \to \infty$. Then the series

$$\sum_{n=1}^{\infty} (-1)^{n+1} a_n \quad \text{is convergent.}$$

Proof. Write S_n for the nth partial sum of the series $\sum_{n=1}^{\infty} (-1)^{n+1} a_n$. We show this sequence has the same type of oscillating behaviour that the corresponding sequence of partial sums in the Fibonacci example. By definition, we have

$$s_{2n+1} = a_1 - a_2 + a_3 - \ldots + a_{2n-1} - a_{2n} + a_{2n+1}$$
$$s_{2n-1} = a_1 - a_2 + a_3 - \ldots + a_{2n-1}$$

and so, subtracting, we have

$$s_{2n+1} = s_{2n-1} - a_{2n} + a_{2n+1}.$$

Since $\{a_n\}$ is a decreasing sequence, $a_{2n} > a_{2n+1}$ and so $s_{2n+1} < s_{2n-1}$. Thus we have a decreasing sequence

$$s_1 > s_3 > s_5 > \ldots > s_{2n-1} > s_{2n+1} > \ldots .$$

Similarly $s_{2n} > s_{2n-2}$ and we have an increasing sequence

$$s_2 < s_4 < s_6 \ldots < s_{2n-2} < s_{2n} < \ldots .$$

Also

$$s_{2n+1} = s_{2n} + a_{2n+1} > s_{2n}$$

Thus

$$s_2 < s_4 < s_6 \ldots < s_{2n-2} < s_{2n} < s_{2n+1} < s_{2n-1} < \ldots s_5 < s_3 < s_1$$

and the sequence s_1, s_3, s_5, ... is a decreasing sequence which is bounded below (by s_2), and so by 3.9 is convergent to α (say). Similarly s_2, s_4, s_6, ... is an increasing sequence which is bounded above (by s_1), and so by 3.9 is convergent to β (say) also

$$s_{2n+1} - s_{2n} = a_{2n+1}$$

and so letting $n \to \infty$

$$\alpha - \beta = 0$$

So $\alpha = \beta$, and all the partial sums are tending to α, so the series converges. □

6.30. Example. Show that the series $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$ is conditionally convergent.

Solution. We have $\displaystyle\sum_{n=1}^{\infty} \left| \frac{(-1)^{n+1}}{n} \right| = \sum_{n=1}^{\infty} \frac{1}{n}$ and this is divergent by 6.19; thus the series is *not* absolutely convergent. We show using 6.29 that this series is still convergent, and so is conditionally convergent.

Write $a_n = 1/n$, so $a_n > 0$, $a_{n+1} < a_n$ and $a_n \to 0$ as $n \to \infty$. Thus all the conditions of Leibniz's theorem are satisfied, and so the series $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$ is convergent.

6.31. Proposition (Re-arranging an Absolutely convergent Series). *Let $\sum_{n=1}^{\infty} a_n$ be an absolutely convergent series and suppose that $\{b_n\}$ is a re-arrangement of $\{a_n\}$. Then $\sum_{n=1}^{\infty} b_n$ is convergent, and*

$$\sum_{n=1}^{\infty} b_n = \sum_{n=1}^{\infty} a_n.$$

Proof. See next year, or (Spivak 1967); the point here is that we need *absolute convergence* before series behave in a reasonable way.   □

Warning: It is not useful to re-arrange conditionally convergent series (remember the rearrangement I did in section 1.1). There is a result which is an extreme form of this:

> Pick $x \in \mathbb{R}$, and let $\sum_{n=1}^{\infty} a_n$ be a conditionally convergent series. then there is a re-arrangement $\{b_n\}$ of $\{a_n\}$ such that $\sum_{n=1}^{\infty} b_n = x$!

In other words, we can re-arrange to get any answer we want!

6.5   An Estimation Problem

This section shows how we can use a lot of the earlier ideas to produce what is often wanted in practice — results which are an approximation, together with an indication of how good the approximation is.

> Find how accurate the approximation obtained by just using the first ten terms is, to $\sum_{n=1}^{\infty} \dfrac{1}{n^2}$.

Again we are going to use geometrical methods for this. Our geometric statement follows from the diagram, and is the assertion that the area of the rectangles below the curve is less than the area under the curve, which is less than the area of the rectangles which contain the curve.
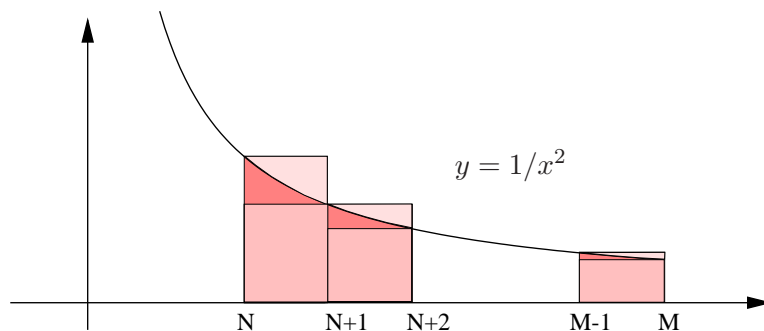

Figure 6.3: An upper and lower approximation to the area under the curve

Writing this out geometrically gives the statement:

$$\sum_{n=N+1}^{M} \frac{1}{n^2} \leq \int_{N}^{M} \frac{dx}{x^2} \leq \sum_{n=N}^{M-1} \frac{1}{n^2}$$

We can evaluate the middle integral:

$$\int_{N}^{M} \frac{dx}{x^2} = \left[-\frac{1}{x} \right]_{N}^{M} = \frac{1}{N} - \frac{1}{M}.$$

For convenience, we define

$$S = \sum_{n=1}^{\infty} \frac{1}{n^2} \quad \text{and} \quad S_N = \sum_{n=1}^{N} \frac{1}{n^2}.$$

We can now express our inequality in these terms:

$$S_M - S_N \le \frac{1}{N} - \frac{1}{M} \le S_{M-1} - S_{N-1}$$

Next, let $M \to \infty$, so $S_M \to S$, and $1/M \to 0$. We have

$$S - S_N \le \frac{1}{N} \le S - S_{N-1}$$

Replacing N by $N + 1$, gives another inequality, which also holds, namely

$$S - S_{N+1} \le \frac{1}{N+1} \le S - S_N,$$

and combining these two, we have

$$S - S_{N+1} \le \frac{1}{N+1} \le S - S_N \le \frac{1}{N} \le S - S_{N-1}.$$

In particular, we have both upper and lower bounds for $S - S_N$, as

$$\frac{1}{N+1} \le S - S_N \le \frac{1}{N}.$$

To make the point that this is a useful statement, we now specialise to the case when $N = 10$. Then

$$\frac{1}{11} \le S - S_{10} \le \frac{1}{10} \quad \text{or} \quad 0 \le S - \left(S_{10} + \frac{1}{11} \right) \le \frac{1}{10} - \frac{1}{11} = \frac{1}{110}.$$

Our conclusion is that although S_{10} is not a very good approximation, we can describe the error well enough to get a much better approximation.

Chapter 7

Power Series

7.1 Power Series and the Radius of Convergence

In Section 5.6, we met the idea of writing $f(x) = P_n(x) + R_n(x)$, to express a function in terms of its Taylor polynomial, together with a remainder. We even saw in 5.30 that, for some functions, the remainder $R_n(x) \to 0$ as $n \to \infty$ for each fixed x. We now recognise this as showing that certain series converge.

We have more effective ways of showing that such a series converges — we can use that ratio test. But note that such a test will only show that a series converges, not that it converges to the function used to generate it in the first place. We saw an example of such a problem in the Warning before Example 5.36.

To summarise the results we had in Section 5.6,

7.1. Proposition. *The following series converge for all values of x to the functions shown:*

$$\mathrm{e}^x = 1 + x + \frac{x^2}{2!} + \frac{x^2}{2!} + \ldots \frac{x^n}{n!} + \ldots$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots + (-1)^{n+1}\frac{x^{2n+1}}{(2n+1)!} + \ldots$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots + (-1)^n\frac{x^{2n}}{2n!} + \ldots$$

$$\sinh x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots + \frac{x^{2n+1}}{(2n+1)!} + \ldots$$

$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots + \frac{x^{2n}}{2n!} + \ldots$$

These are all examples of the subject of this section; they are real power series, which we can use to define functions. The corresponding functions are the best behaved of all the classes of functions we meet in this course; indeed are as well behaved as could possibly be expected. We shall see in this section that this class of functions are really just "grown up polynomials", and that almost any manipulation valid for polynomials remains valid for this larger class of function.

7.2. Definition. A **real power series** is a series of the form $\sum a_n x^n$, where the a_n are real numbers, and x is a real variable.

We are thus dealing with a whole collection of series, one for each different value of x. Our hope is that there is some coherence; that the behaviour of series for different values of x are related in some sensible way.

7.3. Example. The geometric series $\sum_{n=0}^{\infty} x^n$ is another example of a power series we have already met. We saw this series is convergent for all x with $|x| < 1$.

It turns out that a power series is usually best investigated using the ratio test, Theorem 6.21. And the behaviour of power series is in fact very coherent.

7.4. Theorem (Radius of Convergence). *Suppose $\sum a_n x^n$ is a power series. Then one of the following happens:*

- *$\sum a_n x^n$ converges only when $x = 0$; or*

- *$\sum a_n x^n$ converges absolutely for all x; or*

- *there is some number $R > 0$ such that $\sum a_n x^n$ converges absolutely for all x with $|x| < R$, and diverges for all x with $|x| > R$.*

No statement is made in the third case about what happens when $x = R$.

7.5. Definition. The number R described above is called the **radius of convergence** of the power series. By allowing $R = 0$ and $R = \infty$, we can consider *every* power series to have a radius of convergence.

Thus every power series has a radius of convergence. We sometimes call the interval $(-R, R)$, where the power series is guaranteed to converge, the **interval of convergence**. It is characterised by the fact that the series converges (absolutely) inside this interval and diverges outside the interval.

- The word "radius" is used, because in fact the same result is true for complex series, and then we have a genuine circle of convergence, with convergence for all (complex) z with $|z| < R$, and guaranteed divergence whenever $|z| > R$.

- Note the power of the result; we are guaranteed that when $|x| > R$, the series diverges; it can't even converge "accidentally" for a few $x's$ with $|x| > R$. Only on the circle of convergence is there ambiguity.

This regularity of behaviour makes it easy to investigate the radius of convergence of a power series using the ratio test.

7.6. Example. Find the radius of convergence of the series $\sum \dfrac{nx^n}{2^{n+1}}$.

Solution. Recall that the ratio test only applies to series of positive terms, so we look at the ratio of the moduli.

$$
\begin{aligned}
\frac{|(n+1)^{\text{st}} \text{ term}|}{|n^{\text{th}} \text{ term}|} &= \frac{(n+1)|x|^{n+1}}{2^{n+2}} \frac{2^{n+1}}{n|x|^n} \\
&= \frac{(n+1)|x|}{n.2} \to \frac{|x|}{2} \quad \text{as } n \to \infty.
\end{aligned}
$$

Thus the given series diverges if $|x| > 2$ and converges absolutely (and so of course converges) if $|x| < 2$. Hence it has radius of convergence 2.

7.7. Example. Find the radius of convergence of the series $\sum \dfrac{(-1)^n n! x^n}{n^n}$.

Solution. This one is a little more subtle than it looks, although we have met the limit before. Again we look at the ratio of the moduli of adjacent terms.

$$
\begin{aligned}
\frac{|(n+1)^{\text{st}} \text{ term}|}{|n^{\text{th}} \text{ term}|} &= \frac{(n+1)! |x|^{n+1}}{(n+1)^{n+1}} \frac{n^n}{n! |x|^n} = (n+1).|x|.\frac{n^n}{(n+1)^{(n+1)}} \\
&= \frac{n^n}{(n+1)^n} |x| = \left(\frac{n}{n+1}\right)^n |x| \\
&= \frac{|x|}{\left(\dfrac{n+1}{n}\right)^n} \to \frac{|x|}{\mathrm{e}} \quad \text{as } n \to \infty.
\end{aligned}
$$

Here we have of course used the result about e given in Section 3.1 to note that

$$
\left(1 + \frac{1}{n}\right)^n \to \mathrm{e} \quad \text{as } n \to \infty.
$$

Thus the given series diverges if $|x| > \mathrm{e}$ and converges absolutely (and so of course converges) if $|x| < \mathrm{e}$. Hence it has radius of convergence e.

7.8. Exercise. Find the radius of convergence of the series $\displaystyle\sum_{n=0}^{\infty} \frac{x^n}{n^2 + 1}$

We noted that the theorem gives no information about what happens when $x = R$, i.e. on the circle of convergence. There is a good reason for this — it is quite hard to predict what happens. Consider the following power series, all of which have radius of convergence 2.

$$
\sum_{n=1}^{\infty} \frac{x^n}{2^n} \qquad \sum_{n=1}^{\infty} \frac{x^n}{n2^n} \qquad \sum_{n=1}^{\infty} \frac{x^n}{n^2 2^n}.
$$

The first is divergent when $x = 2$ and when $x = -2$, the second converges when $x = -2$, and diverges when $x = 2$, while the third converges both when $x = 2$ and when $x = -2$. These results are all easy to check by direct substitution, and using Theorem 6.29.

7.2 Representing Functions by Power Series

Once we know that a power series has a radius of convergence, we can use it to define a function. Suppose the power series $\sum a_n x^n$ has radius of convergence $R > 0$, and let $I = (-R, R)$. We now define a function f on this open interval I as follows:

$$
f(x) = \sum_{n=1}^{\infty} a_n x^n \qquad \text{for } x \in I.
$$

It turns out that this is the last, and best behaved of the classes of functions we study in this course.

In fact all of what we say below remains true when $R = \infty$, provided we interpret the open interval I as $\mathbb{R}$.

7.9. Theorem. *Let $\sum a_n x^n$ be a power series with radius of convergence $R > 0$. Let I be the open interval $(-R, R)$, and define $f(x) = \sum a_n x^n$ for $x \in I$. Then $\sum n a_n x^{n-1}$ has radius of convergence R, f is differentiable on I, and*

$$f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1} \quad \text{for } x \in I.$$

We summarise this result by saying that we can differentiate a power series term - by - term everywhere inside the circle of convergence. If $R = \infty$, then this can be done for all x.

Proof. Quite a lot harder than it looks; we need to be able to re-arrange power series, and then use the Mean Value Theorem to estimate differences, and show that even when we add an infinite number of errors, they don't add up to too much. It can be found e.g. in (Spivak 1967). $\square$

7.10. Corollary. *Let f and I be defined as 7.9. Then f has an indefinite integral defined on I, given by*

$$G(x) = \sum_{n=0}^{\infty} \frac{a_n}{(n+1)} x^{n+1} \quad \text{for } x \in I.$$

Proof. Apply 7.9 to G to see that $G'(x) = f(x)$, which is the required result. $\square$

7.3   Other Power Series

We now derive some further power series to add to the collection described in 7.1. Starting with the geometric series

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots + x^n + \ldots \qquad \text{for } |x| < 1, \tag{7.1}$$

we replace x by $-x$ to get

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \ldots + (-1)^n x^n + \ldots \qquad \text{for } |x| < 1.$$

Integrating both sides then gives

$$\log(1+x) = K + x - \frac{x^2}{2} + \frac{x^3}{3} + \ldots + (-1)^n \frac{x^{n+1}}{n+1} + \ldots \quad \text{for } |x| < 1,$$

where K is a constant of integration. Evaluating both sides when $x = 0$ shows that $K = 0$, and so we get the series

$$\log(1+x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{n+1}}{n+1} \quad \text{valid for } |x| < 1. \tag{7.2}$$

Note: It is easy to get this result directly from the Taylor Series. The next one is not quite so easy.

We return to equation 7.1, and replace x by $-x^2$ to get

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + \ldots + (-1)^n x^{2n} + \ldots \quad \text{for } |x| < 1.$$

Again integrating both sides, we have

$$\arctan(x) = K + x - \frac{x^3}{3} + \frac{x^5}{5} + \ldots + (-1)^n \frac{x^{2n+1}}{2n+1} + \ldots \quad \text{for } |x| < 1,$$

where K is a constant of integration. Again putting $x = 0$ shows that $K = 0$, and so

$$\arctan(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{2n+1} \quad \text{valid for } |x| < 1. \tag{7.3}$$

7.11. Example. Find the radius of convergence R of the power series

$$\frac{x^2}{2} - \frac{x^3}{3.2} + \frac{x^4}{4.3} - \frac{x^5}{5.4} \ldots + \frac{(-1)^n x^n}{n(n-1)} + \ldots$$

By differentiation or otherwise, find the sum of the series for $|x| < R$.

[You may assume, without proof, that

$$\int \log(1+x)\, dx = K + x \log(1+x) - x + \log(1+x),$$

for some constant of integration K.]

Solution. Apply the ratio test to the given power series. Then

$$\left| \frac{a_{n+1}}{a_n} \right| = \left| \frac{x^{n+1} n(n-1)}{x^n n(n+1)} \right| \to |x| \quad \text{as} \quad n \to \infty.$$

Thus the new series has radius of convergence 1. Denote its sum by $f(x)$, defined for $|x| < 1$. Inside the circle of convergence, it is permissible to differentiate term - by - term, and thus $f'(x) = \log(1+x)$ for $|x| < 1$, since they have the same power series. Hence

$$f(x) \quad = \quad \int \log(1+x)\, dx = K + x \log(1+x) - \int \frac{x+1-1}{1+x}\, dx \tag{7.4}$$

$$= \quad K + x \log(1+x) - x + \log(1+x). \tag{7.5}$$

Putting $x = 0$ shows that $K = 0$ and so $f(x) = (1+x)\log(1+x) - x$.

We have now been able to derive a power series representation for a function without working directly from the Taylor series, and doing the differentiations — which can often prove very awkward. Nevertheless, we have still found the Taylor coefficients.

7.12. Proposition. *Let $\sum a_n x^n$ be a power series, with radius of convergence $R > 0$, and define*

$$f(x) = \sum_{n=0}^{\infty} a_n x^n \quad \text{for } |x| < R.$$

Then $a_n = \dfrac{f^{(n)}(0)}{n!}$, so the given series is the Taylor (or Maclaurin) series for f

Proof. We can differentiate n times by 7.9 and we still get a series with the same radius of convergence. Also, calculating exactly as in the start of Section 5.6, we see that the derivatives satisfy $f^{(n)}(0) = n!a_n$, giving the uniqueness result. $\square$

7.13. Example. Let $f(x) = \dfrac{1}{1 - x^3}$. Calculate $f^{(n)}(0)$.

Solution. We use the binomial theorem to get a power series expansion about 0,

$$\frac{1}{1 - x^3} = 1 + x^3 + x^6 + x^9 + \ldots + x^{3n} + \ldots \quad \text{valid for } |x| < 1.$$

We now read off the various derivatives. Clearly $f^{(n)}(0) = 0$ unless n is a multiple of 3, while $f^{(3k)}(0) = (3k)!$ by 7.12.

7.4   Power Series or Function

We have now seen that when a power series is used to define a function, then that function is *very* well behaved, and we can manipulate it by manipulating, in the obvious way, the corresponding power series. However there are snags. A function has *one* definition which works everywhere it makes sense (at least for simple functions), whereas the power series corresponding to a function depends also on the point about which the expansion is happening. An example will probably make this clearer than further discussion.

7.14. Example. Give the power series expansions for the function $f(x) = \dfrac{1}{1 - x}$.

Solution. We can already do this about 0 by the Binomial Theorem; we have:

$$\frac{1}{1 - x} = 1 + x + x^2 + x^3 + \ldots + x^n + \ldots \quad \text{for } |x| < 1.$$

To expand about a different point, e.g. about 3, write $y = x - 3$. Then

$$\frac{1}{1 - x} = \frac{1}{1 - (y + 3)} = \frac{1}{-2 - y} = -\frac{1}{2}.\frac{1}{1 - y/2},$$

and again using the Binomial Theorem on the last representation, we have

$$\frac{1}{1 - x} = -\frac{1}{2}\left(1 + y/2 + y^2/4 + y^3/9 + \ldots + y^n/2^n + \ldots\right) \quad \text{for } |y/2| < 1.$$

Writing this in terms of x gives

$$\frac{1}{1 - x} = -\frac{1}{2}\left(1 + \frac{x - 3}{2} + \frac{(x - 3)^2}{4} + \ldots + \frac{(x - 3)^n}{2^n} + \ldots\right) \quad \text{for } |x - 3| < 2.$$

It should be no surprise that this is the Taylor series for the same function about the point 3. And it is in fact not an accident that the radius of convergence of the new series is 2. More investigation (quite a lot more - mainly for complex functions) shows the radius of convergence is always that of the *largest* circle that can be fitted into the domain of definition of the function. And that is why it is of interest to sometimes consider power series as complex power series. The power series expansion for $(1 + x^2)^{-1}$ has radius of convergence 1. This seems implausible viewed with real spectacles, but totally explicable when it is realised that the two points i and $-i$ are stopping the expansion from being valid in a larger circle.

7.5   Applications*

This section will not be formally examined, but is here to show how we can get more interesting results from power series.

7.5.1   The function e^x grows faster than any power of x

Specifically, we claim that for any $\alpha \in \mathbb{R}$, $\lim_{x \to \infty} x^{-\alpha} e^x = \infty$.

We have

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \quad \text{and so} \quad x^{-\alpha} e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Given α, we can always find some N such that $N - \alpha \geq 1$. Next note that provided $x > 0$, each term in the series for $x^{-\alpha} e^x$ is positive, and hence the sum is greater than any given term. So in particular,

$$x^{-\alpha} e^x > \frac{x^{N-\alpha}}{N!} \geq \frac{x}{N!} \quad \text{if } x > 1.$$

In particular, since N is fixed, $\dfrac{x}{N!} \to \infty$ as $x \to \infty$, giving the result claimed.

7.5.2   The function $\log x$ grows more slowly than any power of x

Specifically, we claim that for any $\beta > 0$, $\lim_{x \to \infty} x^{-\beta} \log x = 0$.

We are interested in what happens when $x \to \infty$, so we will restrict to the situation when $x > 0$. Put $y = \beta \log x$, noting this is possible since $x > 0$. Thus $y/\beta = \log x$, or equivalently, $x = e^{y/b}$. Thus we have $x^\beta = e^y$, and so

$$x^{-\beta} \log x = \frac{y}{\beta} . e^{-y} \quad \text{when } y = \beta \log x. \tag{7.6}$$

Since $\beta > 0$, as $x \to \infty$, so $y \to \infty$. But from our previous result, as $y \to \infty$, so $\beta y^{-1} e^Y \to \infty$, which is the same as saying that $\dfrac{y}{\beta} . e^{-y} \to 0$ as $y \to \infty$. this is the required result by equation 7.6.

7.5.3   The probability integral $\int_0^\alpha e^{-x^2} \, dx$

The normal distribution is a very common model throughout the whole of science for the situation when things occur "at random". In particular, probability theory attempts to predict what will happen "on average", perhaps for computing risks and premiums on life insurance; in so doing one is often led to consider an integral of the form

$$I = \int_0^\alpha e^{-x^2} \, dx.$$

It turns out that this integral cannot be evaluated using the usual tricks — substitution, integration by parts etc. But a power series representation and Corollary 7.10 *can* help.

Thus

$$\mathrm{e}^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

$$\mathrm{e}^{-x^2} = \sum_{n=0}^{\infty} (-1)^n \frac{(-x^2)^n}{n!} = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{n!}$$

and we can integrate this term-by term, by Corollary 7.10, to get

$$\int \mathrm{e}^{-x^2} \, dx = K + \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)n!}.$$

Performing a definite integral removes the constant of integration, to give

$$\int_0^{\alpha} \mathrm{e}^{-} x^2 \, dx = \sum_{n=0}^{\infty} \frac{(-1)^n \alpha^{2n+1}}{(2n+1)n!}.$$

The partial sums of the power series on the right can be computed, and converge quite quickly, so we have a practical method of evaluating the integral, even thought we can't "do" the integral.

7.5.4   The number e is irrational

We use a power series argument, together with a proof by contradiction, as was done to show that $\sqrt{2}$ was irrational. So assume, to get our contradiction, that e is rational, so it can be written in the form $\mathrm{e} = a/b$, where a and b are integers. Note that this means $\mathrm{e}^{-1} = b/a$. From the exponential series,

$$\frac{b}{a} = \mathrm{e}^{-1} = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!}$$

and so, multiplying through by $a!$,

$$\frac{b}{a}.a! = \mathrm{e}^{-1} = \sum_{n=0}^{\infty} \frac{(-1)^n.a!}{n!}.$$

Thus, we have

$$b(a-1)! = a! - \frac{a!}{1!} + \frac{a!}{2!} - \frac{a!}{3!} + \dots$$

$$= a! - \frac{a!}{1!} + \frac{a!}{2!} - \frac{a!}{3!} + \dots + \frac{(-1)^a a!}{a!}$$

$$+ (-1)^{a+1} \left\{ \frac{1}{a+1} - \frac{1}{(a+1)(a+2)} + \frac{1}{(a+1)(a+2)(a+3)} - \dots \right\}.$$

The left hand side is an integer, as is each of the terms in the sum

$$a! - \frac{a!}{1!} + \frac{a!}{2!}\frac{a!}{3!} + \dots + \frac{(-1)^a a!}{a!};$$

it follows that the rest of that equation is an integer, so

$$(-1)^{a+1} \left\{ \frac{1}{a+1} - \frac{1}{(a+1)(a+2)} + \frac{1}{(a+1)(a+2)(a+3)} - \cdots \right\}$$

is an integer. But this is an alternating sequence of positive terms, which decreases to 0, and so by 6.29, is convergent to a sum which lies between the first two partial sums. The first partial sum is $\frac{1}{a+1}$ while the second is $\frac{1}{a+2}$ (check this). So there must be an integer between $\frac{1}{a+1}$ and $\frac{1}{a+2}$. Since there is not, this contradiction demonstrates our initial assertion.

Chapter 8

Differentiation of Functions of Several Variables

We conclude with two chapters which are really left over from last year's calculus course, and which should help to remind you of the techniques you met then. We shall mainly be concerned with differentiation and integration of functions of more than one variable. We describe

- *how* each process can be done;

- *why* it is interesting, in terms of applications; and

- how to interpret the process geometrically.

In this chapter we concentrate on differentiation, and in the last one, move on to integration.

8.1 Functions of Several Variables

Last year, you did a significant amount of work studying functions, typically written as $y = f(x)$, which represented the variation that occurred in some (dependent) variable y, as another (independent) variable, x, changed. For example you might have been interested in the height y after a given time x, or the area y, enclosed by a rectangle with sides x and $10 - x$. Once the function was known, the usual rules of calculus could be applied, and results such as the time when the particle hits the ground, or the maximum possible area of the rectangle, could be calculated. In the earlier part of the course, we have extended this work by taking a more rigorous approach to a lot of the same ideas.

We are going to do the same thing now for functions of several variables. For example the height y of a particle may depend on the position x and the time t, so we have $y = f(x, t)$; the volume V of a cylinder depends on the radius r of the base and its height h, and indeed, as you know, $V = \pi r^2 h$; or the pressure of a gas may depend on its volume V and temperature T, so $P = P(V, T)$. Note the trick I have just used; it is often convenient to use P both for the (defendant) variable, and for the function itself: we don't always need separate symbols as in the $y = f(x)$ example.

When studying the real world, it is unusual to have functions which depend solely on a single variable. Of course the single variable situation is a little simpler to study, which is
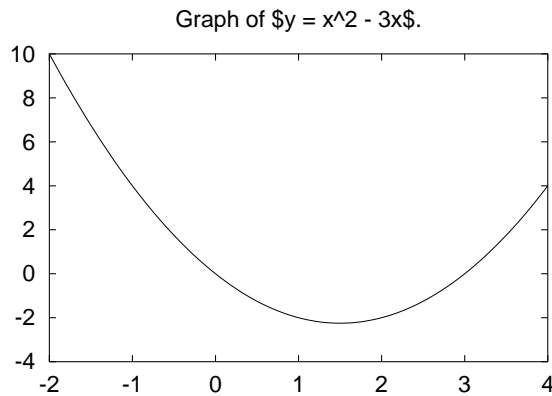
Graph of $y = x^2 - 3x$.

Figure 8.1: Graph of a simple function of one variable

why we started with it last year. And just as last year, we shall usually have a "standard" function name; instead of $y = f(x)$, we often work with $z = f(x, y)$, since most of the extra complications occur when we have two, rather than one (independent) variable, and we don't need to consider more general cases like $w = f(x, y, z)$, or even $y = f(x_1, x_2, \ldots, x_n)$.

Graphing functions of Several Variables

One way we tried to understand the function $y = f(x)$ was by drawing its graph, as shown in Fig 8.1. We then used such a graph to pick out points such as the local minimum at $x = 3/2$, and to see how we could get the same result using calculus.

Working with two or more independent variables is more complicated, but the ideas are familiar. To plot $z = f(x, y)$ we think of z as the **height** of the function f at the point (x, y), and then try to sketch the resulting surface in three dimensions. So we represent a function as a **surface** rather than a curve.

8.1. Example. Sketch the surface given by $z = 2 - x/2 - 2y/3$

Solution. We know the surface will be a plane, because z is a linear function of x and y. Thus it is enough to plot three points that the plane passes through. This gives Fig. 8.2.
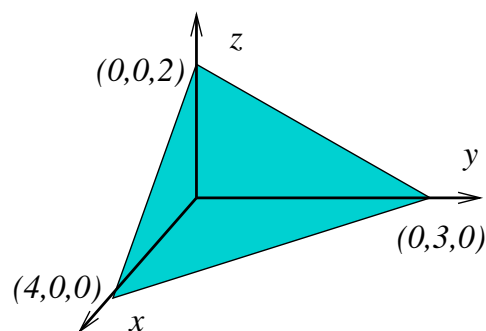
Figure 8.2: Sketching a function of two variables

Of course it is easy to sketch something as simple as a plane. There are graphical

difficulties when dealing with more complicated functions, which make sketching and visu-alisation rather harder than for functions of one variable. And if there are three or more independent variables, there is really no good way of visualising the behaviour of the func-tion directly. But for just two independent variables, there are some tricks.

8.2. Example. Sketch the surface given by $z = x^2 - y^2$.
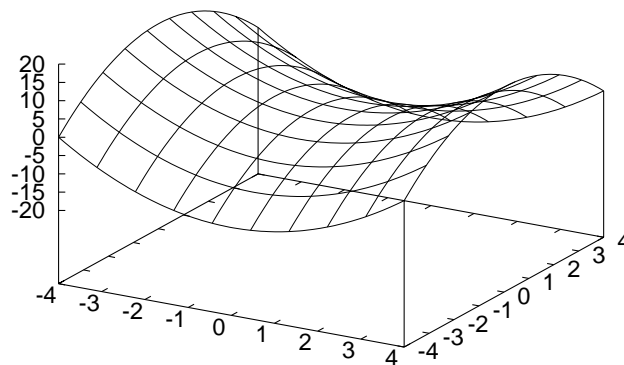
Solution. We can represent the surface directly by drawing it as shown in Fig. 8.3.

Figure 8.3: Surface plot of $z = x^2 - y^2$.

Such a representation is easy to create using suitable software and Fig. 8.3 shows the resulting surface. We now describe how to looking at similar examples without such a program. One approach is to draw a contour map of the surface, and then use the usual tricks to visualise the surface.

For the surface $z = x^2 - y^2$, the points where $z = 0$ lie on $x^2 = y^2$, so form the lines $y = x$ and $y = -x$. We can continue in this way, and look at the points where $z = 1$; so $x^2 - y^2 = 1$. This is one of the hyperbolae shown Fig. 8.4; indeed, fixing z at different values shows the contours (lines of constant height or z value) are all the same shape, but with different constants. We this get the alternative representation as a contour map shown in Fig. 8.4.

A final way to confirm that you have the right view of the surface is to section it in different planes. So far we have looked at the intersection of the planes $z = k$ with the surface $z = x^2 - y^2$ for different values of the constant k. If instead we fix x, at the value a, then $z = a^2 - y^2$. Each of these curves is a parabola with its vertex upwards, at the point $y = 0$, $z = a^2$.

8.3. Exercise. By looking at the curves where z is constant, or otherwise, sketch the surface given by $z = \sqrt{x^2 + y^2}$.
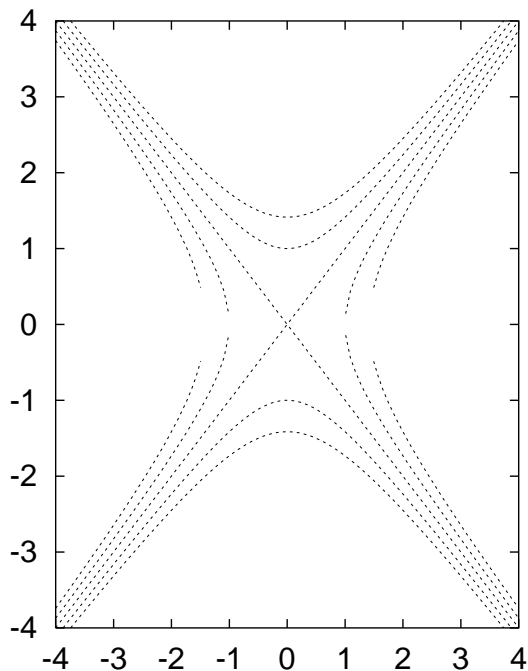
Figure 8.4: Contour plot of the surface $z = x^2 - y^2$. The missing points near the x - axis are an artifact of the plotting program.

Continuity

As you might expect, we say that a function f of two variables is **continuous** at (x_0, y_0) if

$$\lim_{x \to x_0, y \to y_0} f(x, y) = f(x_0, y_0).$$

The only complication comes when we realise that there are many different ways if which $x \to x_0$ and $y \to y_0$. We illustrate with a simple example.

8.4. Example. Investigate the continuity of $f(x, y) = \dfrac{2xy}{x^2 + y^2}$ at the point $(0, 0)$.

Solution. Consider first the case when $x \to 0$ along the x-axis, so that throughout the process, $y = 0$. We have

$$f(x, 0) = \frac{2x.0}{x^2 + 0} = 0 \to 0 \quad \text{as } x \to 0.$$

Next consider the case when $x \to 0$ and $y \to 0$ on the line $y = x$, so we are looking at the special case when $x = y$. We have

$$f(x, x) = \frac{2x^2}{x^2 + x^2} = 1 \to 1 \quad \text{as } x \to 0.$$

Of course f is only continuous if it has the same limit however $x \to 0$ and $y \to 0$, and we have now seen that it doesn't; so f is not continuous at $(0, 0)$.

Although we won't go into it, the usual "putting together" theorems show that f is continuous everywhere else.

8.2   Partial Differentiation

The usual rules for differentiation apply when dealing with several variables, but we now require to treat the variables one at a time, keeping the others constant. It is for this reason that a new symbol for differentiation is introduced. Consider the function

$$f(x, y) = \frac{2y}{y + \cos x}$$

We can consider y fixed, and so treat it as a constant, to get a partial derivative

$$\frac{\partial f}{\partial x} = \frac{2y \sin x}{(2y + \cos x)^2}$$

where we have differentiated with respect to x as usual. Or we can treat x as a constant, and differentiate with respect to y, to get

$$\frac{\partial f}{\partial y} = \frac{(2y + \cos x).2 - 2y.2}{(2y + \cos x)^2} = \frac{2 \cos x}{(2y + \cos x)^2}.$$

Although a partial derivative is itself a function of several variables, we often want to evaluate it at some fixed point, such as (x_0, y_0). We thus often write the partial derivative as

$$\frac{\partial f}{\partial x}(x_0, y_0).$$

There are a number of different notations in use to try to help understanding in different situations. All of the following mean the same thing:-

$$\frac{\partial f}{\partial x}(x_0, y_0), \qquad f_1(x_0, y_0), \qquad f_x(x_0, y_0) \quad \text{and} \quad D_1 F(x_0, y_0).$$

Note also that there is a simple definition of the derivative in terms of a Newton quotient:-

$$\frac{\partial f}{\partial x} = \lim_{\delta x \to 0} \frac{f(x_0 + \delta x, y_0) - f(x_0, y_0)}{\delta x}$$

provided of course that the limit exists.

8.5. *Example.* Let $z = \sin(x/y)$. Compute $x \dfrac{\partial z}{\partial x} + y \dfrac{\partial z}{\partial y}$.

Solution. Treating first y and then x as constants, we have

$$\frac{\partial z}{\partial x} = \frac{1}{y} \cos\left(\frac{x}{y}\right) \quad \text{and} \quad \frac{\partial z}{\partial y} = \frac{-x}{y^2} \cos\left(\frac{x}{y}\right).$$

thus

$$x \frac{\partial z}{\partial x} + y \frac{\partial z}{\partial y} = \frac{x}{y} \cos\left(\frac{x}{y}\right) - \frac{x}{y} \cos\left(\frac{x}{y}\right) = 0.$$

Note: This equation is an equation satisfied by the function we started with, which involves both the function, and its partial derivatives. We shall meet a number of examples of such a **partial differential equation** later.

8.6. Exercise. Let $z = \log(x/y)$. Show that $x\dfrac{\partial z}{\partial x} + y\dfrac{\partial z}{\partial y} = 0$. The fact that the last two function satisfy the same differential equation is not a co-incidence. With our next result, we can see that for *any* suitably differentiable function f, the function $z(x, y) = f(x/y)$ satisfies this partial differential equation.

8.7. Exercise. Let $z = f(x/y)$, where f is suitably differentiable. Show that $x\dfrac{\partial z}{\partial x} + y\dfrac{\partial z}{\partial y} = 0$.

Because the definitions are really just version of the 1-variable result, these examples are quite typical; most of the usual rules for differentiation apply in the obvious way to partial derivatives exactly as you would expect. But there are variants. Here is how we differentiate compositions.

8.8. Theorem. *Assume that f and all its partial derivatives f_x and f_y are continuous, and that $x = x(t)$ and $y = y(t)$ are themselves differentiable functions of t. Let*

$$F(t) = f(x(t), y(t)).$$

Then F is differentiable and

$$\frac{dF}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

Proof. Write $x = x(t)$, $x_0 = x(t_0)$ etc. Then we calculate the Newton quotient for F.

$$
\begin{aligned}
F(t) - F(t_0) &= f(x, y) - f(x_0, y_0) \\
&= f(x, y) - f(x_0, y) + f(x_0, y) - f(x_0, y_0) \\
&= \frac{\partial f}{\partial x}(\xi, y)(x - x_0) + \frac{\partial f}{\partial y}(x_0, \eta)(y - y_0)
\end{aligned}
$$

Here we have used the Mean Value Theorem (5.18) to write

$$f(x, y) - f(x_0, y) = \frac{\partial f}{\partial x}(\xi, y)(x - x_0)$$

for some point ξ between x and x_0, and have argued similarly for the other part. Note that ξ, pronounced "Xi" is the Greek letter "x"; in the same way η, pronounced "Eta" is the Greek letter "y". Thus

$$\frac{F(t) - F(t_0)}{t - t_0} = \frac{\partial f}{\partial x}(\xi, y)\frac{(x - x_0)}{t - t_0} + \frac{\partial f}{\partial y}(x_0, \eta)\frac{(y - y_0)}{t - t_0}$$

Now let $t \to t_0$, and note that in this case, $x \to x_0$ and $y \to y_0$; and since ξ and η are trapped between x and x_0, and y and y_0 respectively, then also $\xi \to x_0$ and $\eta \to y_0$. The result then follows from the continuity of the partial derivatives. $\square$

8.9. Example. Let $f(x, y) = xy$, and let $x = \cos t$, $y = \sin t$. Compute $\dfrac{df}{dt}$ when $t = \pi/2$.

Solution. From the chain rule,

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = -y(t)\sin t + x(t)\cos t = -1.\sin(\pi/2) = -1.$$

The chain rule easily extends to the several variable case; only the notation is complicated. We state a typical example

8.10. Proposition. *Let* $x = x(u,v)$, $y = y(u,v)$ *and* $z = z(u,v)$, *and let* f *be a function defined on a subset* $U \in \mathbb{R}^3$, *and suppose that all the partial derivatives of* f *are continuous. Write*

$$F(u,v) = f(x(u,v), y(u,v), z(u,v)).$$

Then

$$\frac{\partial F}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial u} \quad and \quad \frac{\partial F}{\partial v} == \frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial v}.$$

The introduction of the domain of f above, simply to show it was a function of three variables is clumsy. We often do it more quickly by saying

Let $f(x,y,z)$ have continuous partial derivatives

This has the advantage that you are reminded of the names of the variables on which f acts, although strictly speaking, these names are not bound to the corresponding places. This is an example where we adopt the notation which is very common in engineering maths. But note the confusion if you ever want to talk about the value $f(y,z,x)$, perhaps to define a new function $g(x,y,z)$.

8.11. Example. Assume that $f(u,v,w)$ has continuous partial derivatives, and that

$$u = x - y, \qquad v = y - z \qquad w = z - x.$$

Let

$$F(x,y,z) = f(u(x,y,z), v(x,y,z), w(x,y,z)).$$

Show that

$$\frac{\partial F}{\partial x} + \frac{\partial F}{\partial y} + \frac{\partial F}{\partial z} = 0.$$

Solution. We apply the chain rule, noting first that from the change of variable formulae, we have

$$\frac{\partial u}{\partial x} = 1, \qquad \frac{\partial v}{\partial x} = 0, \qquad \frac{\partial w}{\partial x} = -1,$$

$$\frac{\partial u}{\partial y} = -1, \qquad \frac{\partial v}{\partial y} = 1, \qquad \frac{\partial w}{\partial y} = 0,$$

$$\frac{\partial u}{\partial z} = 0, \qquad \frac{\partial v}{\partial z} = -1, \qquad \frac{\partial w}{\partial z} = 1.$$

Then

$$\frac{\partial F}{\partial x} = \frac{\partial f}{\partial u}.1 + \frac{\partial f}{\partial v}.0 + \frac{\partial f}{\partial w}.-1 = \frac{\partial f}{\partial u} - \frac{\partial f}{\partial w}$$

$$\frac{\partial F}{\partial y} = \frac{\partial f}{\partial u}.-1 + \frac{\partial f}{\partial v}.1 + \frac{\partial f}{\partial w}.0 = -\frac{\partial f}{\partial u} + \frac{\partial f}{\partial v}$$

$$\frac{\partial F}{\partial z} = \frac{\partial f}{\partial u}.0 + \frac{\partial f}{\partial v}.-1 + \frac{\partial f}{\partial w}.1 = -\frac{\partial f}{\partial v} + \frac{\partial f}{\partial w}$$

Adding then gives the result claimed.

8.3   Higher Derivatives

Note that a partial derivative is itself a function of two variables, and so further partial derivatives can be calculated. We write

$$\frac{\partial}{\partial x}\frac{\partial f}{\partial x} = \frac{\partial^2 f}{\partial x^2}, \qquad \frac{\partial}{\partial x}\frac{\partial f}{\partial y} = \frac{\partial^2 f}{\partial x \partial y}, \qquad \frac{\partial}{\partial y}\frac{\partial f}{\partial x} = \frac{\partial^2 f}{\partial y \partial x}, \qquad \frac{\partial}{\partial y}\frac{\partial f}{\partial y} = \frac{\partial^2 f}{\partial y^2}.$$

This notation generalises to more than two variables, and to more than two derivatives in the way you would expect. There is a complication that does not occur when dealing with functions of a single variable; there are *four* derivatives of second order, as follows:

$$\frac{\partial^2 f}{\partial x^2}, \qquad \frac{\partial^2 f}{\partial x \partial y}, \qquad \frac{\partial^2 f}{\partial y \partial x} \quad \text{and} \quad \frac{\partial^2 f}{\partial y^2}.$$

Fortunately, when f has mild restrictions, the order in which the differentiation is done doesn't matter.

8.12. Proposition. *Assume that all second order derivatives of f exist and are continuous. Then the mixed second order partial derivatives of f are equal. i.e.*

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}.$$

8.13. Example. Suppose that $f(x, y)$ is written in terms of u and v where $x = u + \log v$ and $y = u - \log v$. Show that, with the usual convention,

$$\frac{\partial^2 f}{\partial u^2} = \frac{\partial^2 f}{\partial x^2} + 2\frac{\partial^2 f}{\partial x \partial y} + \frac{\partial^2 f}{\partial y^2}$$

and

$$v^2 \frac{\partial^2 f}{\partial v^2} = \frac{\partial f}{\partial y} - \frac{\partial f}{\partial x} + \frac{\partial^2 f}{\partial x^2} + 2\frac{\partial^2 f}{\partial x \partial y} + \frac{\partial^2 f}{\partial y^2}$$

You may assume that all second order derivatives of f exist and are continuous.

Solution. Using the chain rule, we have

$$\frac{\partial f}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \quad \text{and} \quad \frac{\partial f}{\partial v} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v} = \frac{1}{v}\frac{\partial f}{\partial x} - \frac{1}{v}\frac{\partial f}{\partial y}.$$

Thus using both these and their operator form, we have

$$\frac{\partial^2 f}{\partial u^2} = \frac{\partial}{\partial u}\left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial x} + \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial x \partial y} + \frac{\partial^2 f}{\partial y \partial x} + \frac{\partial^2 f}{\partial y^2},$$

while differentiating with respect to v, we have

$$\begin{aligned}
\frac{\partial^2 f}{\partial v^2} &= \frac{\partial}{\partial v}\left(\frac{1}{v}\frac{\partial f}{\partial x} - \frac{1}{v}\frac{\partial f}{\partial y}\right) = -\frac{1}{v^2}\frac{\partial f}{\partial x} + \frac{1}{v}\frac{\partial}{\partial v}\left(\frac{\partial f}{\partial x}\right) + \frac{1}{v^2}\frac{\partial f}{\partial y} - \frac{1}{v}\frac{\partial}{\partial v}\left(\frac{\partial f}{\partial y}\right) \\
&= -\frac{1}{v^2}\frac{\partial f}{\partial x} + \frac{1}{v}\left(\frac{1}{v}\frac{\partial^2 f}{\partial x^2} - \frac{1}{v}\frac{\partial^2 f}{\partial y \partial x}\right) + \frac{1}{v^2}\frac{\partial f}{\partial y} - \frac{1}{v}\left(\frac{1}{v}\frac{\partial^2 f}{\partial x \partial y} - \frac{1}{v}\frac{\partial^2 f}{\partial y^2}\right) \\
&= \frac{1}{v^2}\left(\frac{\partial f}{\partial y} - \frac{\partial f}{\partial x} + \frac{\partial^2 f}{\partial x^2} - 2\frac{\partial^2 f}{\partial x \partial y} + \frac{\partial^2 f}{\partial y^2}\right).
\end{aligned}$$

8.4   Solving equations by Substitution

One of the main interests in partial differentiation is because it enables us to write down how we expect the natural world to behave. We move away from 1-variable results as soon as we have properties which depend on e.g. at least one space variable, together with time. We illustrate with just one example, designed to whet the appetite for the whole subject of mathematical physics.

Assume the displacement of a length of string at time t from its rest position is described by the function $f(x, t)$. This is illustrated in Fig 8.4. The laws of physics describe how the string behaves when released in this position, and allowed to move under the elastic forces of the string; the function f satisfies the *wave equation*

$$\frac{\partial^2 f}{\partial t^2} = c^2 \frac{\partial^2 f}{\partial x^2}.$$
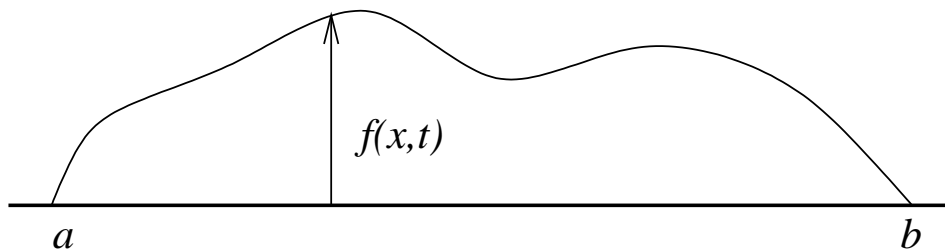
Figure 8.5: A string displaced from the equilibrium position

8.14. Example. Solve the equation

$$\frac{\partial^2 F}{\partial u \partial v} = 0.$$

Solution.   Such a function is easy to integrate, because the two variables appear independently. So $\dfrac{\partial F}{\partial v} = g_1(v)$, where g_1 is an arbitrary (differentiable) function. since when differentiated with respect to u we are given that $\dfrac{\partial^2 F}{\partial u \partial v} = 0$. Thus we can integrate with respect to v to get

$$F(u, v) = \int g_1(v) dv + h(u) = g(v) + h(u),$$

where h is also an arbitrary (differentiable) function.

8.15. Example. Rewrite the wave equation using co-ordinates $u = x - ct$ and $v = x + ct$.

Solution.   Write $f(x, t) = F(u, v)$ and now in principle confuse F with f, so we can tell them apart only by the names of their arguments. In practice we use different symbols to help the learning process; but note that in a practical case, all the F's that appear below, would normally be written as f's By the chain rule

$$\frac{\partial}{\partial x} = \left(\frac{\partial}{\partial u}\right).1 + \left(\frac{\partial}{\partial v}\right).1 \quad \text{and} \quad \frac{\partial}{\partial t} = \left(\frac{\partial}{\partial u}\right).(-c) + \left(\frac{\partial}{\partial v}\right).c.$$

differentiating again, and using the operator form of the chain rule as well,

$$
\begin{aligned}
\frac{\partial^2 f}{\partial t^2} &= \left(c\frac{\partial}{\partial v} - c\frac{\partial}{\partial u}\right)\left(c\frac{\partial F}{\partial v} - c\frac{\partial F}{\partial u}\right) \\
&= c^2\frac{\partial^2 F}{\partial v^2} - c^2\frac{\partial^2 F}{\partial u\partial v} - c^2\frac{\partial^2 F}{\partial v\partial u} + c^2\frac{\partial^2 F}{\partial u^2} \\
&= c^2\left(\frac{\partial^2 F}{\partial v^2} + \frac{\partial^2 F}{\partial u^2}\right) - 2c^2\frac{\partial^2 F}{\partial u\partial v}
\end{aligned}
$$

and similarly

$$
\frac{\partial^2 f}{\partial x^2} = \left(\frac{\partial^2 F}{\partial v^2} + \frac{\partial^2 F}{\partial u^2}\right) + 2\frac{\partial^2 F}{\partial u\partial v}.
$$

Substituting in the wave equation, we thus get

$$
4c^2\frac{\partial^2 F}{\partial u\partial v} = 0,
$$

an equation which we have already solved. Thus solutions to the wave equation are of the form $f(u) + g(v)$ for any (suitably differentiable) functions f and g. For example we may have $\sin(x - ct)$. Note that this is not just any function; it must be constant when $x = ct$.

8.16. Exercise. Let $F(x,t) = \log(2x + 2ct)$ for $x > -ct$, where c is a fixed constant. Show that

$$
\frac{\partial^2 F}{\partial t^2} - c^2\frac{\partial^2 F}{\partial x^2} = 0.
$$

Note that this is simply checking a particular case of the result we have just proved.

8.5   Maxima and Minima

As in one variable calculations, one use for derivatives in several variables is in calculating maxima and minima. Again as for one variable, we shall rely on the theorem that if f is continuous on a closed bounded subset of $\mathbb{R}^2$, then it has a global maximum and a global minimum. And again as before, we note that these must occur *either* at a local maximum or minimum, or else on the boundary of the region. Of course in $\mathbb{R}$, the boundary of the region usually consisted of a pair of end points, while in $\mathbb{R}^2$, the situation is more complicated. However, the principle remains the same. And we can test for local maxima and minima in the same way as for one variable.

8.17. Definition. Say that $f(x, y)$ has a **critical point** at (a, b) if and only if

$$
\frac{\partial f}{\partial x}(a, b) = \frac{\partial f}{\partial y}(a, b) = 0.
$$

It is clear by comparison with the single variable result, that a necessary condition that f have a local extremum at (a, b) is that it have a critical point there, although that is not a sufficient condition. We refer to this as the **first derivative test**.

We can get more information by looking at the second derivative. Recall that we gave a number of different notations for partial derivatives, and in what follows we use f_x rather than the more cumbersome $\dfrac{\partial f}{\partial x}$ etc. This idea extends to higher derivatives; we shall use

$$f_{xx} \quad \text{instead of} \quad \frac{\partial^2 f}{\partial x^2}, \quad \text{and} \quad f_{xy} \quad \text{instead of} \quad \frac{\partial^2 f}{\partial x \partial y} \quad \text{etc.}$$

8.18. Theorem (Second Derivative Test). *Assume that* (a, b) *is a critical point for* f. *Then*

- *If, at* (a, b), *we have* $f_{xx} < 0$ *and* $f_{xx}f_{yy} - f_{xy}^2 > 0$, *then* f *has a **local maximum** at* (a, b).

- *If, at* (a, b), *we have* $f_{xx} > 0$ *and* $f_{xx}f_{yy} - f_{xy}^2 > 0$, *then* f *has a **local minimum** at* (a, b).

- *If, at* (a, b), *we have* $f_{xx}f_{yy} - f_{xy}^2 < 0$, *then* f *has a **saddle point** at* (a, b).

*The test is **inconclusive** at* (a, b) *if* $f_{xx}f_{yy} - f_{xy}^2 = 0$, *and the investigation has to be continued some other way.*

Note that the discriminant is easily remembered as

$$\Delta = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix} = f_{xx}f_{yy} - f_{xy}^2$$

A number of very simple examples can help to remember this. After all, the result of the test should work on things where we can do the calculation anyway!

8.19. Example. Show that $f(x, y) = x^2 + y^2$ has a minimum at $(0, 0)$.

Of course we know it has a global minimum there, but here goes with the test:
Solution. We have $f_x = 2x$; $f_y = 2y$, so $f_x = f_y$ precisely when $x = y = 0$, and this is the only critical point. We have $f_{xx} = f_{yy} = 2$; $f_{xy} = 0$, so $\Delta = f_{xx}f_{yy} - f_{xy}^2 = 4 > 0$ and there is a local minimum at $(0, 0)$.

8.20. Exercise. Let $f(x, y) = xy$. Show there is a unique critical point, which is a saddle point

Proof. We give an indication of how the theorem can be derived — or if necessary how it can be remembered. We start with the two dimensional version of Taylor's theorem, see section 5.6. We have

$$f(a + h, b + k) \sim f(a, b) + h\frac{\partial f}{\partial x}(a, b) + k\frac{\partial f}{\partial y}(a, b) + \frac{1}{2}\left(h^2\frac{\partial^2 f}{\partial x^2} + 2kh\frac{\partial^2 f}{\partial x \partial y} + k^2\frac{\partial^2 f}{\partial y^2}\right)$$

where we have actually taken an expansion to second order and assumed the corresponding remainder is small.

We are looking at a critical point, so for any pair (h, k), we have $h\dfrac{\partial f}{\partial x}(a, b) + k\dfrac{\partial f}{\partial y}(a, b) = 0$ and everything hinges on the behaviour of the second order terms. It is thus enough to study the behaviour of the quadratic $Ah^2 + 2Bhk + Ck^2$, where we have written

$$A = \frac{\partial^2 f}{\partial x^2}, \qquad B = \frac{\partial^2 f}{\partial x \partial y}, \quad \text{and} \quad C = \frac{\partial^2 f}{\partial y^2}.$$

Assuming that $A \neq 0$ we can write

$$Ah^2 + 2Bhk + Ck^2 = A\left(h + \frac{Bk}{A}\right)^2 + \left(C - \frac{B^2}{A}\right)k^2$$

$$= A\left(h + \frac{Bk}{A}\right)^2 + \left(\frac{\Delta}{A}\right)k^2$$

where we write $\Delta = CA - B^2$ for the discriminant. We have thus expressed the quadratic as the sum of two squares. It is thus clear that

- if $A < 0$ and $\Delta > 0$ we have a local maximum;

- if $A > 0$ and $\Delta > 0$ we have a local minimum; and

- if $\Delta < 0$ then the coefficients of the two squared terms have opposite signs, so by going out in two different directions, the quadratic may be made either to increase or to decrease.

Note also that we could have completed the square in the same way, but starting from the k term, rather than the h term; so the result could just as easily be stated in terms of C instead of A ☐

8.21. Example. Let $f(x, y) = 2x^3 - 6x^2 - 3y^2 - 6xy$. Find and classify the critical points of f. By considering $f(x, 0)$, or otherwise, show that f does not achieve a global maximum.

Solution. We have $f_x = 6x^2 - 12x - 6y$ and $f_y = -6y - 6x$. Thus critical points occur when $y = -x$ and $x^2 - x = 0$, and so at $(0, 0)$ and $(1, -1)$. Differentiating again, $f_{xx} = 12x - 12$, $f_{yy} = -6$ and $f_{xy} = -6$. Thus the discriminant is $\Delta = -6.(12x - 12) - 36$. When $x = 0$, $\Delta = 36 > 0$ and since $f_{xx} = -12$, we have a local maximum at $(0, 0)$. When $x = 1$, $\Delta = -36 < 0$, so there is a saddle at $(1, -1)$.

To see there is no global maximum, note that $f(x, 0) = 2x^3(1 - 3/x) \to \infty$ as $x \to \infty$, since $x^3 \to \infty$ as $x \to \infty$.

8.22. Exercise. Find the extrema of $f(x, y) = xy - x^2 - y^2 - 2x - 2y + 4$.

8.23. Example. An open-topped rectangular tank is to be constructed so that the sum of the height and the perimeter of the base is 30 metres. Find the dimensions which maximise the surface area of the tank. What is the maximum value of the surface area? [You may assume that the maximum exists, and that the corresponding dimensions of the tank are strictly positive.]

Solution. Let the dimensions of the box be as shown.

Let the area of the surface of the material be S. Then

$$S = 2xh + 2yh + xy,$$

and since, from our restriction on the base and height,

$$30 = 2(x + y) + h, \quad \text{we have} \quad h = 30 - 2(x + y).$$

Substituting, we have

$$S = 2(x + y)\Big(30 - 2(x + y)\Big) + xy = 60(x + y) - 4(x + y)^2 + xy,$$
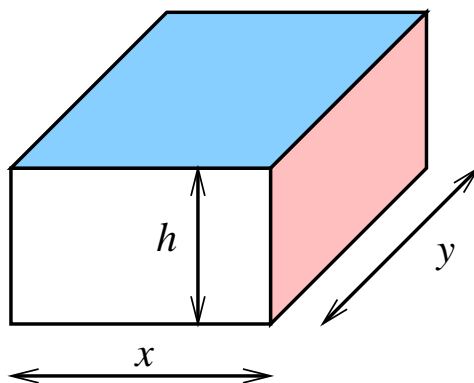
Figure 8.6: A dimensioned box

and for physical reasons, S is defined for $x \geq 0$, $y \geq 0$ and $x + y \leq 15$.

A global maximum (which we are given exists) can only occur on the boundary of the domain of definition of S, or at a critical point, when $\dfrac{\partial S}{\partial x} = \dfrac{\partial S}{\partial y} = 0$. On the boundary of the domain of definition of S, we have $x = 0$ or $y = 0$ or $x + y = 15$, in which case $h = 0$. We are given that we may ignore these cases. Now

$$
\begin{aligned}
S &= -4x^2 - 4y^2 - 7xy + 60x + 60y, \quad \text{so} \\
\frac{\partial S}{\partial x} &= -8x - 7y + 60 = 0, \\
\frac{\partial S}{\partial y} &= -8y - 7x + 60 = 0.
\end{aligned}
$$

Subtracting gives $x = y$ and so $15x = 60$, or $x = y = 4$. Thus $h = 14$ and the surface area is $S = 16(-4 - 4 - 7 + 15 + 15) = 240$ square metres. Since we are given that a maximum exists, this must be it. [If both sides of the surface are counted, the area is doubled, but the critical proportions are still the same.]

Sometimes a function necessarily has an absolute maximum and absolute minimum — in the following case because we have a continuous function defined on a closed bounded subset of $\mathbb{R}^2$, and so the analogue of 4.35 holds. In this case exactly as in the one variable case, we need only search the boundary (using ad - hoc methods, which in fact reduce to 1-variable methods) and the critical points in the interior, using our ability to find local maxima.

8.24. Example. Find the absolute maximum and minimum values of

$$
f(x, y) = 2 + 2x + 2y - x^2 - y^2
$$

on the triangular plate in the first quadrant bounded by the lines $x = 0$, $y = 0$ and $y = 9 - x$

Solution. We know there is a global maximum, because the function is continuous on a closed bounded subset of $\mathbb{R}^2$. Thus the absolute max will occur either in the interior, at a critical point, or on the boundary. If $y = 0$, investigate $f(x, 0) = 2 + 2x - x^2$, while if $x = 0$, investigate $f(0, y) = 2 + 2y - y^2$. If $y = 9 - x$, investigate

$$
f(x, 9 - x) = 2 + 2x + 2(9 - x) - x^2 - (9 - x)^2
$$

for an absolute maximum. In fact extreme may occur when $(x, y) = (0, 1)$ or $(1, 0)$ or $(0, 0)$ or $(9, 0)$ or $(0, 9)$, or $(9/2, 9/2)$. At these points, f takes the values $-41/2, 2, 3, -61$.

Next we seek critical points in the interior of the plate,

$$f_x = 2 - 2x = 0 \quad \text{and} \quad f_y = 2 - 2y = 0.$$

so $(x, y) = (1, 1)$ and $f(1, 1) = 4$, so this must be the global maximum. Can check also using the second derivative test, that it is a *local* maximum.

8.6   Tangent Planes

Consider the surface $F(x, y, z) = c$, perhaps as $z = f(x, y)$, and suppose that f and F have continuous partial derivatives. Suppose now we have a smooth curve on the surface, say $\phi(t) = (x(t), y(t), z(t))$. Then since the curve lies in the surface, we have

$$F(x(t), y(t), z(t)) = c,$$

and so, applying the chain rule, we have

$$\frac{dF}{dt} = \frac{\partial F}{\partial x}\frac{dx}{dt} + \frac{\partial F}{\partial y}\frac{dy}{dt} + \frac{\partial F}{\partial z}\frac{dz}{dt} = 0.$$

or, writing this in terms of vectors, we have

$$\nabla F.\mathbf{v}(t) = \left(\frac{\partial F}{\partial x}, \frac{\partial F}{\partial z}, \frac{\partial F}{\partial z}\right) \cdot \left(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right) = 0.$$

Since the RH vector is the velocity of a point on the curve, which lies on the surface, we see that the left hand vector must be the normal to the curve.

Note that we have defined the **gradient** vector ∇F associated with the function F by

$$\nabla F = \left(\frac{\partial F}{\partial x}, \frac{\partial F}{\partial z}, \frac{\partial F}{\partial z}\right).$$

8.25. Theorem. *The tangent to the surface $F(x, y, z) = c$ at the point (x_0, y_0, z_0) is given by*

$$\frac{\partial F}{\partial x}(x - x_0) + \frac{\partial F}{\partial y}(y - y_0) + \frac{\partial F}{\partial z}(z - z_0) = 0.$$

Proof. This is a simple example of the use of vector geometry. Given that (x_0, y_0, z_0) lies on the surface, and so in the tangent, then for any other point (x, y, z) in the tangent plane, the vector $(x - x_0, y - y_0, z - z_0)$ must lie in the tangent plane, and so must be normal to the normal to the curve (i.e. to ∇F). Thus $(x - x_0, y - y_0, z - z_0)$ and ∇F are perpendicular, and that requirement is the equation which gives the tangent plane. □

8.26. Example. Find the equation of the tangent plane to the surface

$$F(x, y, z) = x^2 + y^2 + z - 9 = 0$$

at the point $P = (1, 2, 4)$.

Solution. We have $\nabla F|_{(1,2,4)} = (2, 4, 1)$, and the equation of the tangent plane is

$$2(x - 1) + 4(y - 2) + (z - 4) = 0.$$

8.27. Exercise. Show that the tangent plane to the surface $z = 3xy - x^3 - y^3$ is horizontal only at $(0, 0, 0)$ and $(1, 1, 1)$.

8.7 Linearisation and Differentials

We obtained a geometrical view of the function $f(x, y)$ by considering the surface $z = f(x, y)$, or $F(x, y, z) = z - f(x, y) = 0$. Note that the tangent plane to this surface at the point $(x_0, y_o, f(x_0, y_0))$ lies close to the surface itself. Just as in one variable, we used the tangent line to approximate the graph of a function, so we shall use the tangent plane to approximate the surface defined by a function of two variables.

The equation of our tangent plane is

$$\frac{\partial F}{\partial x}(x - x_0) + \frac{\partial F}{\partial y}(y - y_0) + \frac{\partial F}{\partial z}(z - f(x_0, y_0)) = 0,$$

and writing the derivatives in terms of f, we have

$$\frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0) + (-1)(z - f(x_0, y_0)) = 0,$$

or, writing in terms of z, the height of the tangent plane above the ground plane,

$$z = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0).$$

Our assumption that the tangent plane lies close to the surface is that $z \approx f(x, y)$, or that

$$f(x, y) \approx f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0).$$

We call the right hand side the **linear approximation** to f at (x_0, y_0).

We can rewrite this with $h = x - x_0$ and $k = y - y_0$, to get

$$f(x, y) - f(x_0, y_0) \approx h\frac{\partial f}{\partial x}(x_0, y_0) + k\frac{\partial f}{\partial y}(x_0, y_0), \quad \text{or} \quad df \approx h\left.\frac{\partial f}{\partial x}\right|_{(x_0, y_0)} + k\left.\frac{\partial f}{\partial y}\right|_{(x_0, y_0)}$$

This has applications; we can use it to see how a function changes when its independent variables are subjected to small changes.

8.28. Example. A cylindrical oil tank is 25 m high and has a radius of 5 m. How sensitive is the volume of the tank to small variations in the radius and height.

Solution. Let V be the volume of the cylindrical tank of height h and radius r. Then $V = \pi r^2 H$, and so

$$\begin{aligned} dV &= \left.\frac{\partial V}{\partial r}\right|_{(r_0, h_0)} dr + \left.\frac{\partial V}{\partial h}\right|_{(r_0, h_0)} dh, \\ &= 250\pi dr + 25\pi dh. \end{aligned}$$

Thus the volume is 10 times as sensitive to errors in measuring r as it is to measuring h.

Try this with a short fat tank!

8.29. Example. A cone is measured. The radius has a measurement error of 3%, and the height an error of 2%. What is the error in measuring the volume?

Solution.   The volume V of a cone is given by $V = \pi r^2 h / 3$, where r is the radius of the cone, and h is the height. Thus

$$
\begin{aligned}
dV &= \frac{2}{3}\pi r h\, dr + \frac{1}{3}\pi r^2\, dh \\
&= 2.(0.03).\frac{\pi r^2 h}{3} + \frac{\pi r^2 h}{3}.(0.02) = V(0.06 + 0.02)
\end{aligned}
$$

Thus there is an 8% error in measuring the volume.

8.30. Exercise. The volume of a cylindrical oil tank is to be calculated from measured values of r and h. What is the percentage error in the volume, if r is measured with an accuracy of 2%, and h measured with an accuracy of 0.5%.

8.8   Implicit Functions of Three Variables

Finally in this section we discuss another application of the chain rule. Assume we have variables x, y and z, related by the equation $F(x, y, z) = 0$. Then assuming that $F_z \neq 0$, there is a version of the implicit function theorem which means we can *in principle*, write $z = z(x, y)$, so we can "solve for z". Doing this gives

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

Now differentiate both sides partially with respect to x. We get

$$
0 = \frac{\partial}{\partial x} F(x, y, z(x, y)) = \frac{\partial F}{\partial x}\frac{\partial x}{\partial x} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial x} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial x},
$$

and so since $\dfrac{\partial y}{\partial x} \neq 0$,

$$
0 = \frac{\partial F}{\partial x} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial x} \quad\text{and}\quad \frac{\partial z}{\partial x} = -\frac{\dfrac{\partial F}{\partial x}}{\dfrac{\partial F}{\partial z}}
$$

Now assume in the same way that $F_x \neq 0$ and $F_y \neq 0$, so we can get two more relations like this, with $x = x(y, z)$ and $y = y(z, x)$. Then form three such equations,

$$
\frac{\partial z}{\partial x}.\frac{\partial x}{\partial y}.\frac{\partial y}{\partial z} = -\left(\frac{\dfrac{\partial F}{\partial x}}{\dfrac{\partial F}{\partial z}}\right)\left(\frac{\dfrac{\partial F}{\partial y}}{\dfrac{\partial F}{\partial x}}\right)\left(\frac{\dfrac{\partial F}{\partial z}}{\dfrac{\partial F}{\partial y}}\right) = -1!
$$

We met this result as Equation 1.1 in Chapter 1, when it seemed totally counter-intuitive!

Chapter 9

Multiple Integrals

9.1 Integrating functions of several variables

Recall that we think of the integral in two different ways. In one way we interpret it as the area under the graph $y = f(x)$, while the fundamental theorem of the calculus enables us to compute this using the process of "anti-differentiation" — undoing the differentiation process.

We think of the area as

$$\sum f(x_i)\, dx_i = \int f(x)\, dx,$$

where the first sum is thought of as a limiting case, adding up the areas of a number of rectangles each of height $f(x_i)$, and width dx_i. This leads to the natural generalisation to several variables: we think of the function $z = f(x, y)$ as representing the height of f at the point (x, y) in the plane, and interpret the integral as the sum of the volumes of a number of small boxes of height $z = f(x, y)$ and area $dx_i dy_j$. Thus the volume of the solid of height $z = f(x, y)$ lying above a certain region R in the plane leads to integrals of the form

$$\int\int_R = \sum_{i=1}^{n}\sum_{j=1}^{m} f(x_i, y_j) dx_i dy_j = \lim S_{mn}.$$

We write such a **double integral** as $\displaystyle\int\int_R f(x, y)\, dA.$

9.2 Repeated Integrals and Fubini's Theorem

As might be expected from the form, in which we can sum over the elementary rectangles $dx\, dy$ in any order, the order does not matter when calculating the answer. There are two important orders — where we first keep x constant and vary y, and then vary x; and the opposite way round. This gives rise to the concept of the repeated integral, which we write as

$$\int\left(\int_R f(x, y)\, dx\right) dy \quad \text{or} \quad \int\left(\int_R f(x, y)\, dy\right) dx.$$

Our result that the order in which we add up the volume of the small boxes doesn't matter is the following, which also formally shows that we evaluate a double integral as any of the possible repeated integrals.

9.1. Theorem (Fubini's theorem for Rectangles). *Let $f(x, y)$ be continuous on the rectangular region $R : a \leq x \leq b; c \leq y \leq d$. Then*

$$\int\int_R f(x, y)\, dA = \int_c^d \left(\int_a^b f(x, y)\, dx \right) dy = \int_a^b \left(\int_c^d f(x, y)\, dy \right) dx$$

Note that this is something like an inverse of partial differentiation. In doing the first inner (or repeated) integral, we keep y constant, and integrate with respect to x. Then we integrate with respect to y. Of course if f is a particularly simply function, say $f(x, y) = g(x)h(y)$, then it doesn't matter which order we do the integration, since

$$\int\int_R f(x, y)\, dA = \int_a^b g(x)\, dx \int_c^d h(y)\, dy.$$

We use the Fubini theorem to actually evaluate integrals, since we have no direct way of calculating a double (as opposed to a repeated) integral.

9.2. Example. Integrate $z = 4 - x - y$ over the region $0 \leq x \leq 2$ and $0 \leq y \leq 1$. Hence calculate the volume under the plane $z = 4 - x - y$ above the given region.

Solution. We calculate the integral as a **repeated integral**, using Fubini's theorem.

$$V = \int_{x=0}^2 \int_{y=0}^1 (4 - x - y)\, dy\, dx = \int_{x=0}^2 \left[(4y - xy - y^2/2) \right]_0^1 dx = \int_{x=0}^2 (4 - x - 1/2)\, dx \text{ etc.}$$

From our interpretation of the integral as a volume, we recognise V as volume under the plane $z = 4 - x - y$ which lies above $\{(x, y) \mid 0 \leq x \leq 2; 0 \leq y \leq 1\}$.

9.3. Exercise. Evaluate $\int_0^3 \int_0^2 (4 - y^2)\, dy\, dx$, and sketch the region of integration.

In fact Fubini's theorem is valid for more general regions than rectangles. Here is a pair of statements which extend its validity.

9.4. Theorem (Fubini's theorem — Stronger Form). *Let $f(x, y)$ be continuous on a region R*

- *if R is defined as $a \leq x \leq b; g_1(x) \leq y \leq g_2(x)$. Then*

$$\int\int_R f(x, y)\, dA = \int_a^b \left(\int_{g_1(x)}^{g_2(x)} f(x, y)\, dy \right) dx$$

- *if R is defined as $c \leq y \leq d; h_1(y) \leq x \leq h_2(y)$. Then*

$$\int\int_R f(x, y)\, dA = \int_c^d \left(\int_{h_1(y)}^{h_2(y)} f(x, y)\, dx \right) dy$$
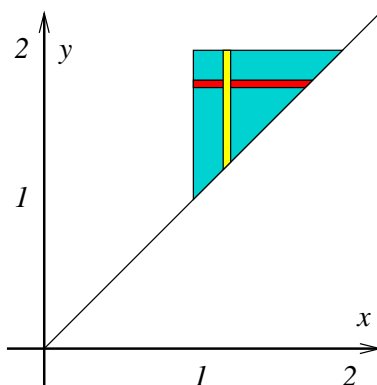
Figure 9.1: Area of integration.

Proof. We give no proof, but the reduction to the earlier case is in principle simple; we just extend the function to be defined on a rectangle by making it zero on the extra bits. The problem with this as it stands is that the extended function is not continuous. However, the difficulty can be fixed. □

This last form enables us to evaluate double integrals over more complicated regions by passing to one of the repeated integrals.

9.5. Example. Evaluate the integral

$$\int_1^2 \left(\int_x^2 \frac{y^2}{x^2} \, dy \right) dx$$

as it stands, and sketch the region of integration.

Reverse the order of integration, and verify that the same answer is obtained.

Solution. The diagram in Fig.9.1 shows the area of integration.

We first integrate in the given order.

$$\int_1^2 \left(\int_x^2 \frac{y^2}{x^2} \, dy \right) dx = \int_1^2 \left[\frac{y^3}{3x^2} \right]_x^2 dx = \int_1^2 \left(\frac{8}{3x^2} - \frac{x}{3} \right) dx$$

$$= \left[-\frac{8}{3x} - \frac{x^2}{6} \right]_1^2 = \left(-\frac{4}{3} - \frac{2}{3} \right) - \left(-\frac{8}{3} - \frac{1}{6} \right) = \frac{5}{6}.$$

Reversing the order, using the diagram, gives

$$\int_1^2 \left(\int_1^y \frac{y^2}{x^2} \, dx \right) dy = \int_1^2 \left[-\frac{y^2}{x} \right]_1^y dy = \int_1^2 \left(-y + y^2 \right) dy$$

$$= \left[-\frac{y^2}{2} - \frac{y^3}{3} \right]_1^2 = \left(-2 + \frac{8}{3} \right) - \left(-\frac{1}{2} + \frac{1}{3} \right) = \frac{5}{6}.$$

Thus the two orders of integration give the same answer.

Another use for the ideas of double integration just automates a procedure you would have used anyway, simply from your knowledge of 1-variable results.
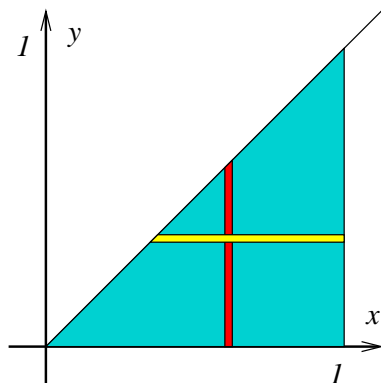
Figure 9.2: Area of integration.

9.6. Example. Find the area of the region bounded by the curve $x^2 + y^2 = 1$, and above the line $x + y = 1$.

Solution. We recognise an area as numerically equal to the volume of a solid of height 1, so if R is the region described, the area is

$$\int\int_R 1\, dx\, dy = \int_0^1 \left(\int_{y=1-x}^{y=\sqrt{1-x^2}} dy \right) dx = \int_0^1 \sqrt{1-x^2} - (1-x)\, dx = \dots .$$

And we also find that Fubini provides a method for actually calculating integrals; sometimes one way of doing a repeated integral is much easier than the other.

9.7. Example. Sketch the region of integration for

$$\int_0^1 \left(\int_y^1 x^2\, e^{xy}\, dx \right) dy.$$

Evaluate the integral by reversing the order of integration.

Solution. The diagram in Fig.9.2 shows the area of integration.

Interchanging the given order of integration, we have

$$\int_0^1 \left(\int_y^1 x^2\, e^{xy}\, dx \right) dy = \int_0^1 \left(\int_0^x x^2\, e^{xy}\, dy \right) dx$$

$$= \int_0^1 \left[\frac{x^2\, e^{xy}}{x} \right]_0^x dx$$

$$= \int_0^1 x\, e^{x^2}\, dx - \int_0^1 x\, dx$$

$$= \left[\frac{1}{2} e^{x^2} - \frac{x^2}{2} \right]_0^1 = \frac{1}{2} (e - 2) .$$

9.8. Exercise. Evaluate the integral

$$\int\int_R (\sqrt{x} - y^2)\, dy\, dx,$$

where R is the region bounded by the curves $y = x^2$ and $x = y^4$.

9.3 Change of Variable — the Jacobian

Another technique that can sometimes be useful when trying to evaluate a double (or triple etc) integral generalise the familiar method of integration by substitution.

Assume we have a change of variable $x = x(u, v)$ and $y = y(u, v)$. Suppose that the region S' in the uv - plane is transformed to a region S in the xy - plane under this transformation. Define the **Jacobian** of the transformation as

$$J(u, v) = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial(x, y)}{\partial(u, v)}.$$

It turns out that this correctly describes the relationship between the element of area $dx\, dy$ and the corresponding area element $du\, dv$.

With this definition, the change of variable formula becomes:

$$\int \int_S f(x, y)\, dx\, dy = \int \int_S' f(x(u, v), y(u, v)) \, |J(u, v)| \ du\, dv.$$

Note that the formula involves the *modulus* of the Jacobian.

9.9. Example. Find the area of a circle of radius R.

Solution. Let A be the disc centred at 0 and radius R. The area of A is thus $\displaystyle \int \int_A dx\, dy$. We evaluate the integral by changing to polar coordinates, so consider the usual transformation $x = r\cos\theta$, $y = r\sin\theta$ between Cartesian and polar co-ordinates. We first compute the Jacobian;

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

Thus

$$J(r, \theta) = \begin{vmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} \\ \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r(\cos^2\theta + \sin^2\theta) = r.$$

We often write this result as

$$dA = dx\, dy = r\, dr\, d\theta$$

Using the change of variable formula, we have

$$\int \int_A dx\, dy = \int \int_A |J(r, \theta)|\, dr\, d\theta = \int_0^R \int_0^{2\pi} r\, dr\, d\theta = 2\pi \frac{R^2}{2}.$$

We thus recover the usual area of a circle.

Note that the Jacobian $J(r, \theta) = r > 0$, so we did indeed take the modulus of the Jacobian above.

9.10. Example. Find the volume of a ball of radius 1.

Solution. Let V be the required volume. The ball is the set $\{(x, y, z) \mid x^2 + y^2 + z^2 \leq 1\}$. It can be thought of as twice the volume enclosed by a hemisphere of radius 1 in the upper half plane, and so

$$V = 2 \int_D \sqrt{1 - x^2 - y^2} \, dx \, dy$$

where the region of integration D consists of the unit disc $\{(x, y) \mid x^2 + y^2 \leq 1\}$. Although we can try to do this integration directly, the natural co-ordinates to use are plane polars, and so we instead do a change of variable first. As in 9.9, if we write $x = r \cos \theta$, $y = r \sin \theta$, we have $dx \, dy = r \, dr \, d\theta$. Thus

$$\begin{aligned}
V = 2 \int_D \sqrt{1 - x^2 - y^2} \, dx \, dy &= 2 \int_D (\sqrt{1 - r^2}) \, r \, dr \, d\theta \\
&= 2 \int_0^{2\pi} d\theta \int_0^1 \sqrt{1 - r^2} \, dr \\
&= 4\pi \left[-\frac{(1 - r^2)^{3/2}}{3} \right]_0^1 \\
&= \frac{4\pi}{3}.
\end{aligned}$$

Note that after the change of variables, the integrand is a product, so we are able to do the dr and $d\theta$ parts of the integral at the same time.

And finally, we show that the same ideas work in 3 dimensions. There are (at least) two co-ordinate systems in $\mathbb{R}^3$ which are useful when cylindrical or spherical symmetry arises. One of these, **cylindrical polars** is given by the transformation

$$x = r \cos \theta, \qquad y = r \sin \theta, \qquad z = z,$$

and the Jacobian is easily calculated as

$$\frac{\partial(x, y, z)}{\partial(r, \theta, z)} = r \quad \text{so} \quad dV = dx \, dy \, dz = r \, dr \, d\theta \, dz.$$

The second useful co-ordinate system is **spherical polars** with transformation

$$x = r \sin \phi \cos \theta, \qquad y = r \sin \phi \sin \theta, \qquad z = r \cos \phi.$$

The transformation is illustrated in Fig 9.3.

It is easy to check that Jacobian of this transformation is given by

$$dV = r^2 \sin \phi \, dr \, d\phi \, d\theta = dx \, dy \, dz.$$

9.11. Example. The moment of inertia of a solid occupying the region R, when rotated about the z - axis is given by the formula

$$I = \int \int \int_R (x^2 + y^2) \rho \, dV.$$

Calculate the moment of inertia about the z-axis of the solid of unit density which lies outside the cylinder of radius a, inside the sphere of radius $2a$, and above the $x - y$ plane.
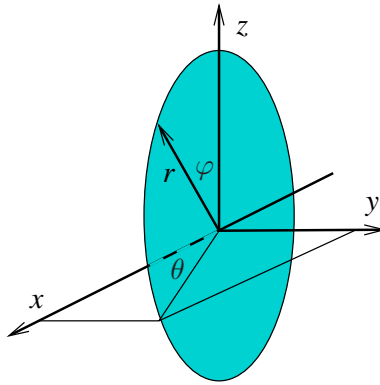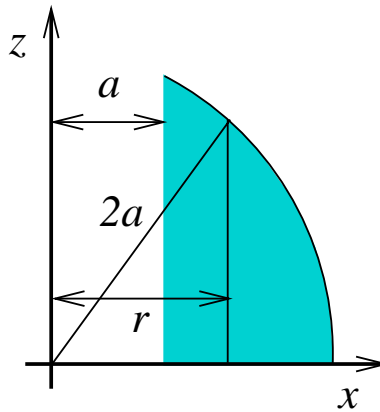
Figure 9.3: The transformation from Cartesian to spherical polar co-ordinates.

Figure 9.4: Cross section of the right hand half of the solid outside a cylinder of radius a and inside the sphere of radius $2a$

Solution. Let I be the moment of inertia of the given solid about the z-axis. A diagram of a cross section of the solid is shown in Fig 9.4.

We use cylindrical polar co-ordinates (r, θ, z); the Jacobian gives $dx\,dy\,dz = r\,dr\,d\theta\,dz$, so

$$\begin{aligned} I & = \int_0^{2\pi} d\theta \int_a^{2a} r\,dr \int_0^{\sqrt{4a^2-r^2}} r^2\,dz \\ & = 2\pi \int_a^{2a} r^3 \sqrt{4a^2 - r^2}\,dr. \end{aligned}$$

We thus have a single integral. Using the substitution $u = 4a^2 - r^2$, you can check that the integral evaluates to $22\sqrt{3}\pi a^5/5$.

9.12. Exercise. Show that

$$\int \int \int z^2\,dxdydz = \frac{4}{15}\pi$$

(where the integral is over the unit ball $x^2 + y^2 + z^2 \le 1$) first by using spherical polars, and then by doing the z integration first and using plane polars.

Bibliography

Spivak, M. (1967), *Calculus*, W. A. Benjamin.

Index Entries