Engineering mathematics 1 EE106

Introductory analysis and calculus

C. Nash
Mathematical Physics Department
National University of Ireland
Maynooth

cnash@thphys.nuim.ie

Preface

This is a set of notes which supplement my lectures on for course EE106 which is about elementary analyisis and calculus. They are reasonably self contained but should be read *as well as* other written material. There are so many books covering this material that I shall recommend just four. These are:

1. Booth D. J. and Stroud K. A., *Engineering mathematics*, Palgrave MacMillan, (2007).

2. Kreyszig E., *Advanced engineering mathematics*, Wiley, (2010).

3. Hobson M. P. and Riley K. F., *Essential mathematical methods for the physical sciences*, Cambridge University Press, (2011).

4. Spivak M., *Calculus*, Cambridge University Press, (2006).

The first of these four would be an adequate book for this course. The second assumes that differentiation and integration are already known starts with differential equations; it then moves on through many other more advanced topics most of which will be covered in a second year engineering course.

The third book takes a slightly wider viewpoint but covers much the same material as the second. Lastly the fourth book is almost exclusively concerned with calculus rather than its applications. It is very well written and is also rigorous; this latter point means that it is more suitable for a mathematics course than an engineering course, nevertheless it is still worth having a look at in the library. One should always try to read as widely as one can.

Charles Nash

CHAPTER I

Introductory analysis

§ 1. Notation

WE shall use **N** to denote the set of *natural numbers* $1, 2, \ldots$, in other words **N** is the set of *positive integers*. We shall use **Z** to denote *all* the integers $\ldots, -2, -1, 0, 1, 2, \ldots$ Lastly **R** and **C** will denote the real and complex numbers respectively. This same information is often, and more conveniently, displayed by writing

$$\begin{aligned} \mathbf{N} &= \{1, 2, \ldots\} \\ \mathbf{Z} &= \{\ldots, -2, -1, 0, 1, 2, \ldots\} \\ \mathbf{R} &= \{x : x \text{ is a real number}\} \\ \mathbf{C} &= \{z : z \text{ is a complex number}\} \end{aligned} \tag{1.1}$$

Let me just remind the reader that an equation like

$$\mathbf{R} = \{x : x \text{ is a real number}\} \tag{1.2}$$

when read aloud yields

R *is the set of x such that x is a real number*

and similarly for the other examples given above.

§ 2. Sequences

A collection of objects arranged in some particular order is called a *sequence*. Let us look at a few examples of sequences.

Example *The first ten positive integers (arranged in increasing order)*

$$1, 2, 3, 4, 5, 6, 7, 8, 9, 10 \tag{1.3}$$

Notice that these integers are arranged in *increasing order*. This order is a vital property of the sequence: if we *change the order* we create a *new sequence*. For instance we could reverse the order in our sequence $1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ so as to give the new sequence

Example *The first ten positive integers (arranged in decreasing order)*

$$10, 9, 8, 7, 6, 5, 4, 3, 2, 1 \tag{1.4}$$

We see that we could make many more sequences from the first ten positive integers. In fact, we can now see, that there are as many such sequences as there are ways of *ordering* these integers.

If we recall that the number of ways of ordering n things is $n!$ (pronounced n factorial) and

$$n! = n(n-1)(n-2)\cdots 2\cdot 1 \tag{1.5}$$

then we compute that the number of sequences that contain the ten integers $1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ is

$$\begin{aligned} 10! &= 10\cdot 9\cdot 8\cdot 7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1 \\ &= 3628800 \end{aligned} \tag{1.6}$$

which is a pretty large number.

Sequences need not consist of mere numbers but can be made from any sort of object as long as one provides a rule for ordering them. For example one could have a sequence consisting of the top 100 golfers in the world ordered by their earnings in the previous financial year, or the top 10 football teams in a football league ordered by their point totals and so on.

Thus to discuss an arbitrary sequence it is useful to use an algebraic notation and write a sequence of, say, 5 objects or elements, as

$$s_1, s_2, s_3, s_4, s_5 \tag{1.7}$$

Suppose then that we have a sequence consisting of the first seven odd positive integers then we would write this as

$$1, 3, 5, 7, 9, 11, 13 \tag{1.8}$$

But we could also write

$$\begin{gathered} s_1, s_2, s_3, s_4, s_5, s_6, s_7 \\ \text{where } s_1 = 1,\ s_2 = 3,\ s_3 = 5, \ \text{ etc.} \end{gathered} \tag{1.9}$$

More compactly still we could write

$$\begin{gathered} s_1, s_2, s_3, s_4, s_5, s_6, s_7 \\ \text{where } s_n = 2n - 1 \\ \text{and } n = 1, 2, 3, 4, 5, 6, 7 \end{gathered} \tag{1.10}$$

The point being that, as n runs through the 7 values $1, \ldots 7$, the expression $2n + 1$ runs through the 7 odd numbers $1, 3, \ldots 13$ comprising the sequence.

All the preceding sequences contained a *finite* number of elements; this need not be so: many important sequences contain an *infinite* number of elements. We have

Example *An infinite sequence*

$$2, 4, 6, 8, 10, 12, \ldots \tag{1.11}$$

The sequence 1.11 above consists of all the even positive integers and is clearly infinite.

When a sequence is finite it may be quite important just how many sequences with the same number of elements can be constructed.

Example *A National Lottery*

For example consider a national lottery where one chooses 6 numbers from the first 42 positive integers. [1]. We might then construct a sequence of all possible choices of 6 numbers from 42, in this case we are not interested in the order the choices are placed in just the number of elements in the sequence. This number—the number of ways of choosing 6 things from 42—is

$$\begin{aligned} \binom{42}{6} &= \frac{42!}{6!\,36!} \\ &= 5245786 \end{aligned} \tag{1.12}$$

which we see is slightly under 5.25 million.

So the odds against winning the jackpot in this lottery are

$$1 \text{ in } 5245786 \tag{1.13}$$

Let us calculate how expensive it would be to bet on all the possible combinations. If we get 1 bet for €1 euro then 5245786 bets would cost

$$€\,5,245,786 \tag{1.14}$$

This tells us something quite interesting: namely if the jackpot exceeds €5.24 million or so—which it occasionally does—then one can, in theory, emerge a net winner by betting on all the possible 5245786 combinations. There are of course some drawbacks to this: how does one place such a large number of bets in a reasonable time; also one might still lose if someone else also had the winning combination thus causing the jackpot to be split. These drawbacks would be somewhat offset by the large number of winning match 4 and match 5 combinations which would be achieved though.

[1] Nowadays our lottery uses 45 rather than 42, the reader can easily adjust the analysis here to fit that case.

§ 3. Series

If we have a sequence whose elements are real or complex numbers then we can *add together* all its terms obtaining what is called a *series*. So a *sequence* such as

$$1, 4, 7, 10, 13 \tag{1.15}$$

becomes the *series*

$$1 + 4 + 7 + 10 + 13 \tag{1.16}$$

If we denote the series by S then we can write

$$S = 1 + 4 + 7 + 10 + 13 \tag{1.17}$$

and since

$$1 + 4 + 7 + 10 + 13 = 35 \tag{1.18}$$

we describe this fact by saying that "S is a series consisting of 5 terms whose sum is 35". We also labour the obvious somewhat by noting that the symbol S is also equal to the sum of the series so that

$$S = 35 \tag{1.19}$$

§ 4. Arithmetic series

Notice that in the series 1.17 above consecutive terms all differ by the same amount namely 3. Any series which has this property of consecutive terms all differing by the same amount is referred to as an *arithmetic series.* Let us now examine this property in some more detail some algebraic notation will be helpful as we can then be more general.

We only need to specify two things in order to know an arithmetic series completely: these are the *first term* and the *difference between consecutive terms.* Let us denote the first term by

$$a \tag{1.20}$$

and the difference between consecutive terms by

$$d \tag{1.21}$$

Now suppose there are exactly n terms in the arithmetic series then we shall denote a general term in the series by a_i where where each a_i is given by the formula

$$\begin{aligned} a_i &= a + id \\ i &= 0, \ldots (n-1) \end{aligned} \tag{1.22}$$

We can now summarise all this by writing a completely arbitrary *arithmetic series* as the expression

$$\begin{aligned} &a_0 + a_1 + a_2 + \cdots + a_{n-1} \\ \text{where } &a_i = a + id \\ &i = 0, \dots (n-1) \end{aligned} \tag{1.23}$$

To see how this works in practice we return to the series 1.17 which is

$$S = 1 + 4 + 7 + 10 + 13 \tag{1.24}$$

and we see that the first term is 1, so that $a = 1$ and the difference between consecutive terms 3 so $d = 3$. Thus, using $a_i = a + id$ we find that

$$\begin{aligned} a_0 &= a + 0 \cdot d \Rightarrow a_0 = 1 \\ a_1 &= a + d \Rightarrow a_1 = 4 \\ a_2 &= a + 2d \Rightarrow a_2 = 7 \\ &\vdots \\ a_4 &= a + 3d \Rightarrow a_4 = 13 \end{aligned} \tag{1.25}$$

§§ 4.1 The sum of the terms in an arithmetic series

There is quite a simple formula for the sum of the terms in an arithmetic series. It is obtained by writing the series out twice: once in the normal order and the second time in reverse order. Let us see this in action: If we write out a general arithmetic series S with n terms out we get

$$S = a_0 + a_1 + a_2 + \cdots + a_{n-1} \tag{1.26}$$

where we note that the last term is a_{n-1} not a_n because otherwise we would have $n+1$ terms instead of n. Now writing S out in reverse order we get

$$S = a_{n-1} + a_{n-2} + \cdots a_1 + a_0 \tag{1.27}$$

The last step is place both expressions together, one on top of the other, and add them giving

$$\begin{aligned} S &= a_0 + a_1 + a_2 + \cdots + a_{n-1} \\ S &= a_{n-1} + a_{n-2} + \cdots a_1 + a_0 \\ \text{adding} \qquad \Rightarrow 2S &= (a_0 + a_{n-1}) + (a_1 + a_{n-2}) + \cdots (a_{n-1} + a_0) \end{aligned} \tag{1.28}$$

where we draw attention to the fact that we have deliberately bracketed the terms together in pairs. The reason for this is that all the pairs are in fact *equal* to each other—indeed, if we use the property that

$$a_i = a + (i-1)d \tag{1.29}$$

we observe that

$$\begin{aligned}(a_0 + a_{n-1}) &= 2a + (n-1)d \\ (a_1 + a_{n-2}) &= 2a + (n-1)d \\ &\vdots \\ (a_{n-1} + a_0) &= 2a + (n-1)d\end{aligned} \tag{1.30}$$

Hence we conclude that

$$\begin{aligned}2S &= \underbrace{(2a + (n-1)d) + (2a + (n-1)d) + \cdots + (2a + (n-1)d)}_{n \text{ terms}} \\ \Rightarrow 2S &= n(2a + (n-1)d) \\ \Rightarrow S &= \frac{n(2a + (n-1)d)}{2}\end{aligned} \tag{1.31}$$

So the sum of the first n terms of an arithmetic series

$$a + (a + d) + (a + 2d) + \cdots + (a + (n-1)d) \tag{1.32}$$

is S where

$$S = \frac{n(2a + (n-1)d)}{2} \tag{1.33}$$

Example *The sum of the arithmetic series 1.17*

Recall that 1.17 is the series

$$S = 1 + 4 + 7 + 10 + 13 \tag{1.34}$$

with $a = 1$ and $d = 3$ and application of our formula 1.33 gives

$$S = \frac{5(2 \cdot 1 + 4 \cdot 3)}{2} = 35 \tag{1.35}$$

just as it should.

Example *The first 10000 odd numbers*

For this example our series is

$$S = 1 + 3 + 5 + \cdots + 19999 \tag{1.36}$$

which can also be written as

$$\begin{gathered}S = a_0 + a_1 + \cdots + a_{n-1} \\ \text{with} \quad a_i = 2i + 1 \quad \text{and} \quad n = 10000\end{gathered} \tag{1.37}$$

Evidently we have

$$a = 1, \quad d = 2 \tag{1.38}$$

so using 1.33 we compute that

$$\begin{aligned} S &= \frac{n(2 + (n-1)2)}{2} \\ \Rightarrow S &= n^2 \end{aligned} \tag{1.39}$$

which is a useful formulae in its own right. We are dealing with the case where $n = 10000$ though we see that have a completely general formula which gives the sum of the first n odd numbers; in any case for $n = 10000$ we find that

$$S = 10000^2 = 10^8 \tag{1.40}$$

§ 5. Geometric series

In this section we shall deal with what are called *geometric series*. Here is a example.

Example *A geometric series*

$$S = 3 + 6 + 12 + 24 + 48 \tag{1.41}$$

In 1.41 above we see that the first term of the series is 3 and that the next term is got by multiplying 3 by 2 to get 6 and then 6 is multiplied by 2 to get its successor 12 and so on. We can rewrite this to make this pattern more obvious; doing this we obtain

$$\begin{aligned} S &= 3 + 6 + 12 + 24 + 48 \\ &= 3 + 3 \cdot 2 + 3 \cdot 2^2 + 3 \cdot 2^3 + 3 \cdot 2^4 \end{aligned} \tag{1.42}$$

Here is another geometric series

Example *A geometric series where the terms get smaller*

$$S = 5 + \frac{5}{3} + \frac{5}{3^2} + \frac{5}{3^3} + \frac{5}{3^4} + \frac{5}{3^5} \tag{1.43}$$

Now to display an arbitrary geometric series all we have to do is to replace the first term 3 by a and to replace the multiplier 2 by r. An arbitrary geometric series, with n terms, is now given by S where

$$S = a + ar + ar^2 + ar^3 + \cdots + ar^{n-1} \tag{1.44}$$

again note that there are only n terms despite the fact that the last term is ar^{n-1}. A piece of *terminology* to be aware of is that the multiplier r is often referred to as the *ratio* of the series, this is because the ratio of successive terms is equal to r. In algebraic form we are simply observing that, if ar^i is a term in a geometric series then,

$$r = \frac{ar^i}{ar^{i-1}} \tag{1.45}$$

By returning to examples 1.41 and 1.43 we can easily verify that for 1.41 we have

$$a = 3,\ r = 2 \tag{1.46}$$

while for 1.43 we have

$$a = 5,\ r = \frac{1}{5} \tag{1.47}$$

Note that when $r > 1$ the terms in the series get successively *bigger* as the series progresses while for $r < 1$ the terms get *smaller*; it is clear, too, that were we to have $r = 1$ then the terms would stay the same and not change at all, thereby yielding a rather a boring series.

Notice that there is a certain similarity between arithmetic and geometric series namely: in arithmetic series the terms are constructed b repeated *adding* of the constant d, while in geometric series the terms are constructed by repeated *multiplication* by the constant r. Hence the main change in passing from arithmetic to geometric series is to trade addition for multiplication.

§§ 5.1 The sum of the terms in a geometric series

It is now time for us to derive a formula for the sum of the terms in a geometric series.

What we do is to write out S and then subtract from it the quantity rS. This gives us the expressions

$$\begin{aligned} S &= a + ar + ar^2 + ar^3 + \cdots + ar^{n-1} \\ \Rightarrow rS &= ar + ar^2 + ar^3 + \cdots + ar^{n-1} + ar^n \\ \Rightarrow S - rS &= a - ar^n \\ \Rightarrow (1-r)S &= a - ar^n \\ \Rightarrow S &= \frac{(a - ar^n)}{(1-r)} \\ &= \frac{a(1-r^n)}{(1-r)} \end{aligned} \tag{1.48}$$

So to summarise: the sum to n terms of a geometric series with first term a and ratio r is given by

$$S = \frac{a(1-r^n)}{(1-r)} \tag{1.49}$$

We can conclude this section by quickly computing the sums of the terms in our two examples 1.41 and 1.43.

For 1.41 where, $n = 5$, $a = 3$ and $r = 2$ we find that

$$S = \frac{3(1-2^5)}{(1-2)} = 3 \cdot 31 = 93 \tag{1.50}$$

a fact which can be verified by direct calculation.

Lastly for 1.43 we have $n = 6$, $a = 5$ and $r = 1/3$ giving

$$S = \frac{5\left(1-\left(\frac{1}{3}\right)^6\right)}{\left(1-\frac{1}{3}\right)} = \frac{5 \cdot 364}{3^5} = \frac{1820}{243} \tag{1.51}$$

a fact which is again verifiable by direct calculation.

§ 6. Limits and infinite series

A series can have an infinite number of terms: for example consider the series

Example *A nice infinite series*

$$\begin{aligned} S &= 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots \\ &= \sum_{n=1}^{\infty} \frac{1}{n^2} \end{aligned} \tag{1.52}$$

Such a series is called an *infinite series* and its sum S may or *may not* be finite. It turns out that this series 1.52 does have a finite value and in fact it is known that [2]

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \tag{1.53}$$

Unfortunately S can also be infinite rendering the series not so useful as in the following.

Example *A not so nice infinite series*

$$\begin{aligned} S &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots \\ &= \sum_{n=1}^{\infty} \frac{1}{n} \end{aligned} \tag{1.54}$$

[2] This is far from obvious and so we do not worry about how to prove this here; we merely want to use the result. It is easy to convince oneself experimentally of the result by using a calculator or a computer to sum up a finite number of terms.

For this series 1.54 one finds that[3] $S = \infty$. An infinite series can also have a sum S which refuses to settle down to any fixed value as one sums the terms but just *oscillates* as shown next.

Example *An oscillating series*

$$S = 2 - 2 + 2 - 2 + 2 - \cdots \tag{1.56}$$

It is clear that as we add up the terms of 1.56 the answer is 2 or 0 depending on whether we stop after an odd or an even numbers of terms. Thus S refuses to tend to any fixed value. Instead it *oscillates* between the values 0 and 2.

The terminology used to describe the behaviour in these three representative examples of infinite series is that a series such as 1.52 for which S has a *finite* value is called *convergent*; but a series such as 1.54 for which S tends to *infinity* is called *divergent*, the oscillatory series 1.56 is also called *divergent* even though no infinite value of S is encountered. Divergent and convergent series can be treated with more precision by introducing the idea of a limit to which we now turn.

A limit is an extremely widespread and fundamental concept in mathematics and it is time we met it. In the present context it is applied to a sequence but it is applied to functions as well and the same idea underlies both situations. We do not need the notion of the limit for *finite sequences* (we could use it but it we gain nothing) so we shall just explain what happens for *infinite sequences.*

Our definition[4]

Definition (The limit of an infinite sequence) *An infinite sequence*

$$\{s_1, s_2, s_3, \ldots\} \tag{1.57}$$

approaches the limit s as n tends to infinity if we can make s_n as near as we wish to s by requiring that n be large enough. We then write this symbolically as

$$\lim_{n\to\infty} s_n = s \quad \textit{or} \quad s_n \to s, \ \textit{as } n \to \infty \tag{1.58}$$

[3] For those interested in why this series diverges we digress briefly to give a little more information on the matter. If once sums up just N terms one finds that for N large one has

$$1 + \frac{1}{2} + \cdots \frac{1}{N} \to \ln N + \gamma \tag{1.55}$$

where γ is a constant. So as N increases S goes to infinity because $\ln N$ does.

[4] This definition is not as logically tight or rigorous as it would be in an advanced mathematics course but this is deliberate: it consumes more time to be rigorous, and would limit the amount of material we could cover. Rigour does have its place though in an appropriate mathematics course. The same remark will apply to subsequent definitions that we shall give in this course.

Return now to any infinite series

$$S = s_1 + s_2 + s_3 + \cdots = \sum_{i=1}^{\infty} s_i \tag{1.59}$$

and first let us sum up only the first N terms giving what is called the *partial sum* S_N where

$$S_N = s_1 + s_2 + s_3 + \cdots + s_N = \sum_{i=1}^{N} s_i \tag{1.60}$$

By doing this for successively larger and larger values of N we can construct an infinite *sequence* out of these partial sums, namely the sequence

$$\{S_1, S_2, S_3, \ldots\} \tag{1.61}$$

So this is the way we define convergence of the series S:

Definition (Convergence of an infinite series) *The infinite series*

$$s_1 + s_2 + s_3 + \cdots \tag{1.62}$$

converges to the value S *if*

$$\lim_{N\to\infty} S_N = S \quad \textit{or equivalently} \quad S_N \to S, \text{ as } N \to \infty \tag{1.63}$$

In words one says that the series $s_1 + s_2 + s_3 + \cdots$ *converges to* S *if its sequence of partial sums* $\{S_1, S_2, S_3, \ldots\}$ *has the limit* S.

It is important to realise that when an infinite series does *not converge* this can be either because $S_N \to \infty$ as $N \to \infty$ as is the case in 1.54; or because S_N remains finite but *oscillates* as in 1.56. It is time to consider geometric series again.

Example *The convergence of the geometric series* $a + ar + ar^2 + ar^3 + \cdots$ *when* $r < 1$

If we take the infinite geometric series

$$a + ar + ar^2 + ar^3 + \cdots \tag{1.64}$$

then its partial sum S_N is simply what is stated in the formula of 1.49 so we have

$$S_N = \frac{a(1-r^N)}{(1-r)} \tag{1.65}$$

so

$$\lim_{N\to\infty} S_N = \lim_{N\to\infty} \frac{a(1-r^N)}{(1-r)} \tag{1.66}$$

But all we need to know on the RHS of 1.66 is $\lim_{N\to\infty} r^N$ about which we can immediately say that

$$\lim_{N\to\infty} r^N = \begin{cases} \infty, & \text{if } r > 1 \\ 1, & \text{if } r = 1 \\ 0, & \text{if } r < 1 \end{cases} \tag{1.67}$$

But we have [5] the condition $r < 1$ so we have $\lim_{N\to\infty} r^N = 0$ which on insertion into 1.66 gives

$$\lim_{N\to\infty} S_N = \frac{a}{(1-r)} \tag{1.68}$$

So the sum S of the infinite series is given by

$$\begin{aligned} S &= a + ar + ar^2 + ar^3 + \cdots \\ &= \frac{a}{(1-r)}, \text{ provided } r < 1 \end{aligned} \tag{1.69}$$

If $r \geq 1$ it is easy to verify that any geometric series diverges; it is also easy to check from the formula 1.33 that all arithmetic series diverge if we allow the number of terms to become infinite. We shall only consider convergent infinite series. We can do the computation of an infinite sum for a specific example.

Example *The infinite series*

$$S = 5 + \frac{5}{3} + \frac{5}{3^2} + \frac{5}{3^3} + \frac{5}{3^4} + \cdots \tag{1.70}$$

.

Here we can observe straightaway that

$$a = 5, \quad \text{and} \quad r = \frac{1}{3} \tag{1.71}$$

[5] Actually we also allow r to be negative as well as positive and to take account of that the condition $r < 1$ must be replaced by $-1 < r < 1$, i.e. $|r| < 1$. Still more: r could be complex but then the condition $|r| < 1$ is still sufficient though $|r|$ is now extended to complex numbers r cf. §6.

so 1.69 gives us at once the result that

$$S = \frac{5}{(1 - 1/3)} = \frac{15}{2} \tag{1.72}$$

Before leaving this subject we shall consider a slightly more exotic application of summing infinite geometric series.

Example *A mathematical snowflake*

First we must make the snowflake by using triangles as the building blocks. We start with the triangle

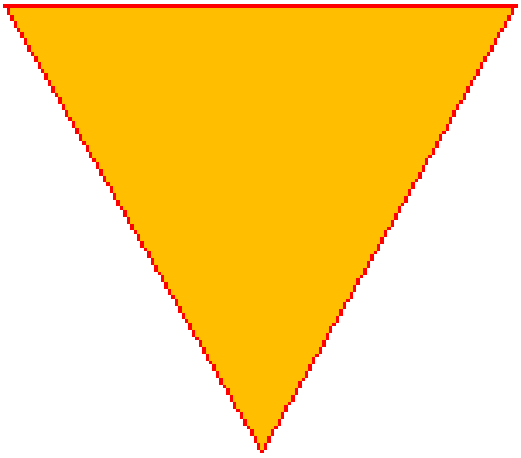


Fig. 1: The basic triangle for the snowflake.

Now we add 3 smaller triangles on to the sides of this basic triangle yielding what is shown in fig. 2

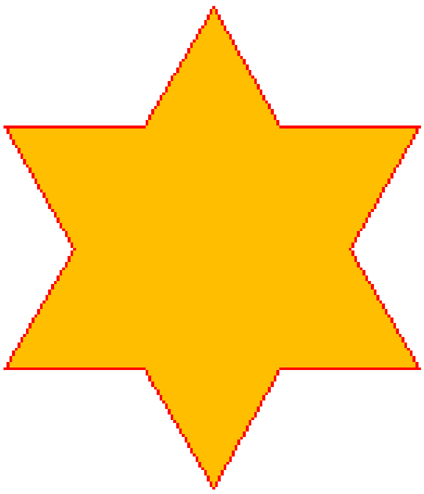


Fig. 2: The second figure for the snowflake.

We can do this yet again to get fig. 3

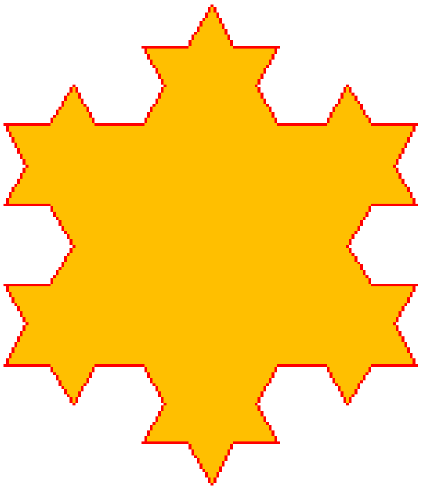


Fig. 3: The third figure for the snowflake.

and doing it once more gives

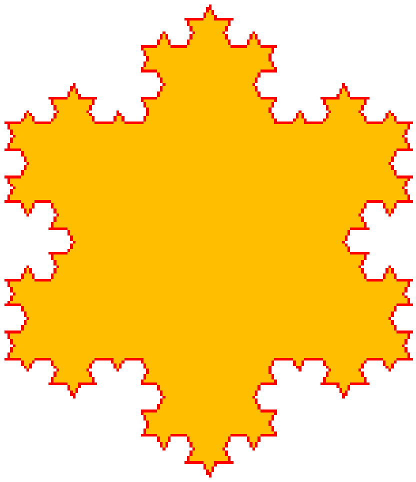


Fig. 4: The fourth figure for the snowflake.

Now we have a figure which looks a bit like a snowflake and the idea is to repeat this process an infinite number of times. The figure that results from this process is then our snowflake[6]

Having made our snowflake the calculational task we set ourselves is to find the area contained inside it. This turns out to be given by an infinite geometric series which we now

[6] This object is sometimes called the *Koch snowflake* after a mathematician called Koch. Its perimeter is actually infinitely long and this jagged perimeter is an example of what is called a *fractal*. Fractals turns out to have applications in many areas including telecommunications and image analysis and compression.

obtain and sum.

Example *The area of the snowflake*

Let the area of the large equilateral triangle shown in fig. 1 be

$$a \tag{1.73}$$

then the smaller triangles added on in fig. 2 are also equilateral and have a side which is 1/3 of the length of that of the large triangle. Thus they have an area of $a/3^2$ or

$$\frac{a}{9} \tag{1.74}$$

Since three triangles were added on to get fig. 2 the area of the object depicted in fig. 2 is

$$a + \frac{3a}{9} \tag{1.75}$$

Next time, to get fig. 3, we add even smaller triangles whose area is 1/9 that of the previous ones, that is their area is

$$\frac{a}{9^2} \tag{1.76}$$

and since a count shows that we add 2 of these for each vertex of fig. 2 then we add $6 \cdot 2 = 3 \cdot 4$ of these triangles giving the area of the "flake" of fig. 3 to be

$$a + \frac{3a}{9} + \frac{3 \cdot 4a}{9^2} \tag{1.77}$$

It is then easy to check, but we shall just accept, that the area, A say, after doing this infinitely many times is given by

$$\begin{aligned} A &= a + \frac{3a}{9} + \frac{3 \cdot 4a}{9^2} + \frac{3 \cdot 4^2 a}{9^3} + \cdots \\ &= a + \frac{3}{9}\left(a + \frac{4a}{9} + \frac{4^2 a}{9^2} + \cdots\right) \end{aligned} \tag{1.78}$$

But we can easily see that the expression

$$a + \frac{4a}{9} + \frac{4^2 a}{9^2} + \cdots \tag{1.79}$$

is an infinite geometric series. hence we know that

$$\begin{aligned} &a + \frac{4a}{9} + \frac{4^2 a}{9^2} + \cdots \\ &= \frac{a}{(1 - 4/9)} = \frac{9a}{5} \end{aligned} \tag{1.80}$$

and if we substitute this result into 1.78 we find that

$$\begin{aligned} A &= a + \frac{3}{9}\frac{9a}{5} \\ &= \frac{8a}{5} \end{aligned} \tag{1.81}$$

So we have learned the interesting fact that the snowflake has an area given by

$$A = \frac{8a}{5} \tag{1.82}$$

where a is the area of the basic triangle of fig. 1.

§§ 6.1 Convergence of infinite series in general: the comparison and ratio tests

We have seen that an (infinite) geometric series converges when $r < 1$; for other infinite series there are several well known tests which are used to tell whether an infinite series converges. We shall just look at two of these: they are called *the comparison test* and *the ratio test*. Both are widely used.

§§ 6.2 The comparison test

Take two infinite series

$$\sum_{n=1}^{\infty} a_n \text{ and } \sum_{n=1}^{\infty} b_n \tag{1.83}$$

and suppose that

$$0 \leq a_n \leq b_n \quad \text{for all } n \tag{1.84}$$

Then

$$\sum_{n=1}^{\infty} b_n \text{ convergent} \Rightarrow \sum_{n=1}^{\infty} a_n \text{ convergent} \tag{1.85}$$

So the convergence of one series is decided by its comparison with another. Here is a sample of the the comparison test in action.

Example *Convergence of the two series*

$$\sum_{n=1}^{\infty} \frac{1}{n^3} \; \textit{and} \; \sum_{n=1}^{\infty} \frac{1}{n^2} \tag{1.86}$$

For this example we set

$$a_n = \frac{1}{n^3} \text{ and } b_n = \frac{1}{n^2} \tag{1.87}$$

and we observe that

$$\frac{1}{n^3} \leq \frac{1}{n^2}, \text{ for } n = 1, 2, 3, 4, \ldots \tag{1.88}$$

But we know already that $\sum_{n=1}^{\infty} (1/n^2)$ converges because we learn from consulting 1.53 that

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \tag{1.89}$$

Hence, applying the comparison test 1.85, we conclude that

$$\sum_{n=1}^{\infty} \frac{1}{n^3} \quad \text{converges} \tag{1.90}$$

Incidentally the same argument would work with the series

$$\sum_{n=1}^{\infty} \frac{1}{n^p}, \qquad \text{for } p > 2 \tag{1.91}$$

So the comparison test also gives the result that [7]

$$\sum_{n=1}^{\infty} \frac{1}{n^p} \quad \text{converges for } p > 2 \tag{1.92}$$

§§ **6.3 The ratio test**

Let

$$\begin{aligned} &s_1 + s_2 + s_3 + \cdots \\ &= \sum_{n=1}^{\infty} s_n \end{aligned} \tag{1.93}$$

be an infinite series. Form the *ratio* s_{n+1}/s_n; now if we define r by writing

$$r = \lim_{n \to \infty} \left| \frac{s_{n+1}}{s_n} \right| \tag{1.94}$$

then the series $\sum_{n=1}^{\infty} s_n$ *converges* if

$$r < 1 \tag{1.95}$$

We can see how the ratio test at work in the following example.

[7] Actually, just for information, we add that still more is known: this is that $\sum_{n=1}^{\infty}(1/n^p)$ converges precisely when $p > 1$.

Example *Convergence of the series*

$$\sum_{n=1}^{\infty} \frac{1}{n!} \tag{1.96}$$

Forming the desired ratio we see that we have to evaluate the limit

$$\begin{aligned} r &= \lim_{n\to\infty} \left| \frac{\frac{1}{(n+1)!}}{\frac{1}{n!}} \right| \\ &= \lim_{n\to\infty} \frac{1}{(n+1)} \\ &= 0 \end{aligned} \tag{1.97}$$

So we have found that $r < 1$ and so

$$\sum_{n=1}^{\infty} \frac{1}{n!} \tag{1.98}$$

converges. In fact most of you probably know already that if $e = 2.7182818284\ldots$ is the base of natural logarithms then

$$e = \sum_{n=0}^{\infty} \frac{1}{n!} \tag{1.99}$$

and

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \tag{1.100}$$

so for e to be a finite number at all the series 1.98 has to converge.

§ 7. Limits, functions and continuity

So far we have considered limits that arose in the study of sequences and series; however limits also arise all the time when working with functions. Still more important, as we shall see later, is the fact that the whole of differential and integral calculus is based on appropriate limits of functions.

We begin with a definition similar to that used for sequences in 1.57 above.

Definition (The limit of a function) *A function $f(x)$ approaches the limit L near $x = a$ if we can make $f(x)$ as close as we like to L by requiring x to be close enough to a. Symbolically we then write*

$$\lim_{x\to a} f(x) = L \quad \text{or} \quad f(x) \to L, \text{ as } x \to a \tag{1.101}$$

The first point to notice is that if function $f(x)$ has a graph which passes through a point a and has a value L there then its limit at a is indeed L. For example consider the perfectly ordinary graph shown in fig. 5 below. This is the graph of the function $f(x) = x^2 - 10x + 2$ and when $x = 12$ this function $f(x)$ has the value 26.

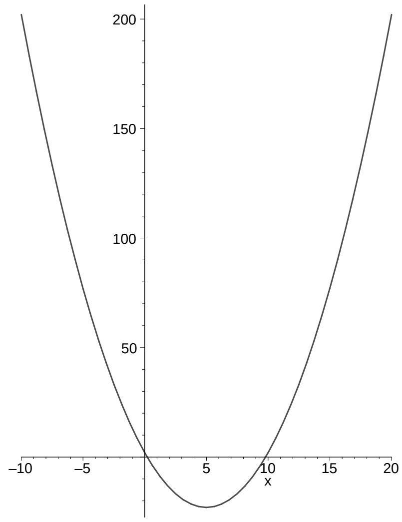


Fig. 5: The function $x^2 - 10x + 2$.

So, for this function $x^2 - 10x + 2$, we can say

$$\lim_{x\to 12} x^2 - 10x + 2 = 26 \tag{1.102}$$

The second point to notice about limits of functions (and this one is not so trivial as the first one) is that a function may have a limit L but my never attain the value L but just get arbitrarily close to it. This phenomenon is shown in fig. 6.

Fig. 6: The functions e^{-x} and $1/x$.

Figure 6 shows the graph of e^{-x} on the left and the graph of $1/x$ on the right.

For the case of e^{-x} we see that, as $x \to \infty$, then $e^{-x} \to 0$; but e^{-x} never actually *equals* zero (unless $x = \infty$ which is not allowed) it just gets closer and closer to zero. Hence we can say that

$$\lim_{x\to\infty} e^{-x} = 0 \tag{1.103}$$

While for the graph of $1/x$ we observe that, if we set x equal to a *positive value*, say $x = 1$, and then let $x \to 0$, then $1/x$ tends to $+\infty$ and at $x = 0$ itself $1/x$ is infinite and so undefined.

Alternatively we also can see that if we set x equal to a *negative value*, say $x = -1$ then $1/x$ tends to $-\infty$ and is still undefined at $x = 0$. Hence $1/x$ is trying to have *two* illegal values at 0: The first is seen when we approach 0 through positive values from the right we write this one as

$$\lim_{x\to 0^+} \frac{1}{x} = \infty \tag{1.104}$$

where the notation $x \to 0^+$ means that 0 is approached from the *right* by going through *positive* values. The second is seen by approaching 0 through negative values and is written symbolically as

$$\lim_{x\to 0^-} \frac{1}{x} = -\infty \tag{1.105}$$

where the notation $x \to 0^-$ this time means that 0 is approached from the *left* through *negative* values.

We summarise matters by noting what was in common about the limits of the two functions e^{-x} and $1/x$ just discussed: this was that for each function the limit evaluated was never properly attained by the function.

In the first case this was because x would have to have an illegal value (namely ∞) to allow the function to reach its limiting value; while in the second case x takes on a perfectly legal limiting value but the function would have to take on illegal values (namely $\mp\infty$) for the limit to be attained. The upshot is the same in both cases and can be summarised by writing[8]

$$\lim_{x\to a} f(x) = L \quad \textbf{BUT} \quad \begin{cases} f(a) \neq L \\ \text{because either } a \text{ or } L \text{ is an illegal value} \end{cases} \tag{1.106}$$

This matter of whether $f(a) = L$ or not, when $\lim_{x\to a} f(x) = L$, is really what underlies the notion of *continuity*—a notion we are now ready to examine.

§§ 7.1 Continuity

Informally we can describe a continuous function as being one whose graph, once begun, can be completed without lifting the writing implement from the paper. With this description we

[8] For the reader who needs reminding and doesn't want to turn the pages back we gives the values of a and L for both examples. For $f(x) = e^{-x}$ we had $a = \infty$ and $L = 0$ and for $f(x) = 1/x$ we had $x = 0$ and $L = \mp\infty$.

can immediately display an example of a continuous and a discontinuous function cf. fig. 7 below.

Fig. 7: The functions $x^4 - 2x^2$ and $f(x) = \begin{cases} x, & \text{if } x > 1 \\ x - 1, & \text{if } x \leq 1 \end{cases}$

In fig. 7 the function on the *left* is continuous while the function $f(x)$ on the *right* is discontinuous since it 'jumps' at the value $x = 1$. The value $x = 1$ is then called *a point of discontinuity* of $f(x)$. We can easily use limits to get to the bottom of what happens at the point of discontinuity of f.

We simply investigate the value $x = 1$ by evaluating the pair of limits

$$\lim_{x \to 1^-} f(x) \quad \text{and} \quad \lim_{x \to 1^+} f(x) \tag{1.107}$$

What we find from fig. 17 is that the limit from the *left* $\lim_{x \to 1^-} f(x)$ is given by

$$\lim_{x \to 1^-} f(x) = 0 \tag{1.108}$$

while the limit from the *right* $\lim_{x \to 1^+} f(x)$ is given by

$$\lim_{x \to 1^+} f(x) = 1 \tag{1.109}$$

and we note that these disagree. This then the is 'signature' of the discontinuity namely the fact that

$$\lim_{x \to 1^-} f(x) \neq \lim_{x \to 1^+} f(x) \tag{1.110}$$

The more precise definition of continuity can now be unveiled.

Definition (Continuity) *A function $f(x)$ is continuous at $x = a$ if* [9]

$$\lim_{x \to a} f(x) = f(a) \tag{1.111}$$

It is clear too for $\lim_{x\to a} f(x) = f(a)$ then both the left and right limits must both *exist and agree* and they cannot do this when $\lim_{x\to a^-} f(x) \neq \lim_{x\to a^+} f(x)$.

For convenience of examination we display together our two examples of discontinuous functions in fig. 8

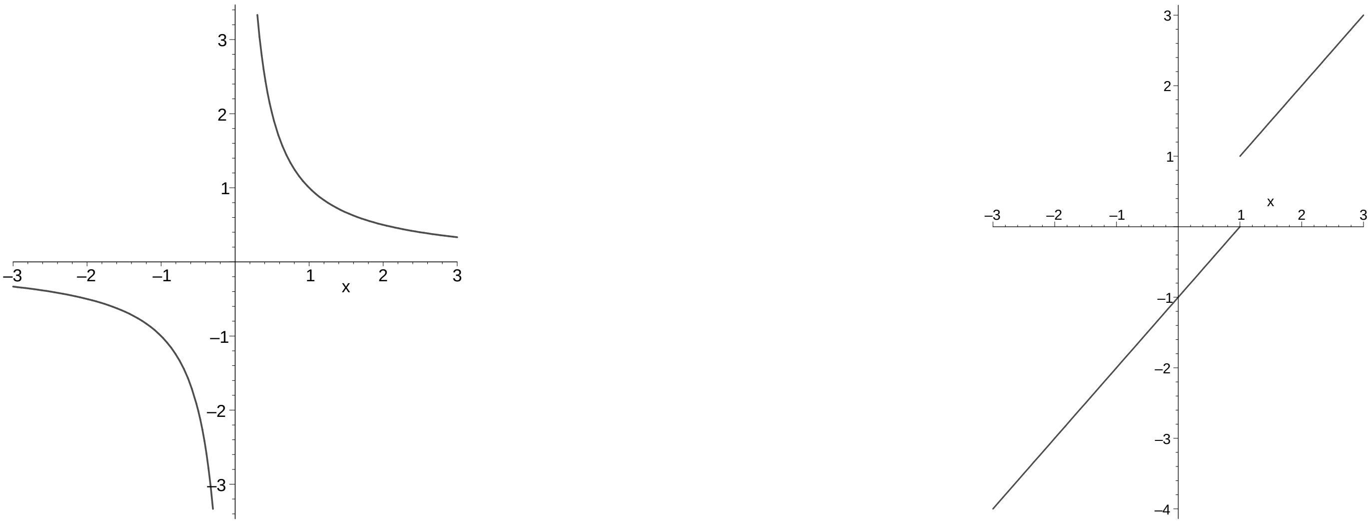


Fig. 8: The function $1/x$ and the function $f(x) = \begin{cases} x, & \text{if } x > 1 \\ x - 1, & \text{if } x \le 1 \end{cases}$

$$\frac{1}{x} \text{ has } \begin{cases} \lim\limits_{x\to 0^-} \dfrac{1}{x} = -\infty \\ \lim\limits_{x\to 0^+} \dfrac{1}{x} = \infty \end{cases} \qquad \text{and} \quad f(x) \text{ has } \begin{cases} \lim\limits_{x\to 1^-} f(x) = 0 \\ \lim\limits_{x\to 1^+} f(x) = 1 \end{cases} \tag{1.112}$$

It is now easy to see take in the visual signs of the discontinuities from the graphs and also the more precise information given in the caption 1.112.

Discontinuous functions are sometimes thought to be abnormal or in some way undesirable. This is a mistake. In fact discontinuous functions arise quite naturally in switching electrical systems on and off. Also the current practice of *digitising* many forms of transmitted data such as telephone conversations, modem data, and radio and television signals shows that functions which take *integer* values as opposed to any *real* value are very common.

[9] It is important to realise that the point $x = a$ at which a function may or may not be continuous is of necessity *finite.* one does not evaluate functions at infinite points. This means that the example of 1.103 where we pointed out that $\lim_{x\to\infty} e^{-x} = 0$ is not an example of a discontinuity, it is simply an example of a function which does not attain one of its limiting values. In fact the function e^{-x} is continuous *everywhere*—that is it is continuous for all finite x.

Example *A simple switch*

If we take just a electrical switch which is simply turned on and off repeatedly generating a constant voltage V when on and 0 when off the we get the simple square wave graph below which is manifestly discontinuous.

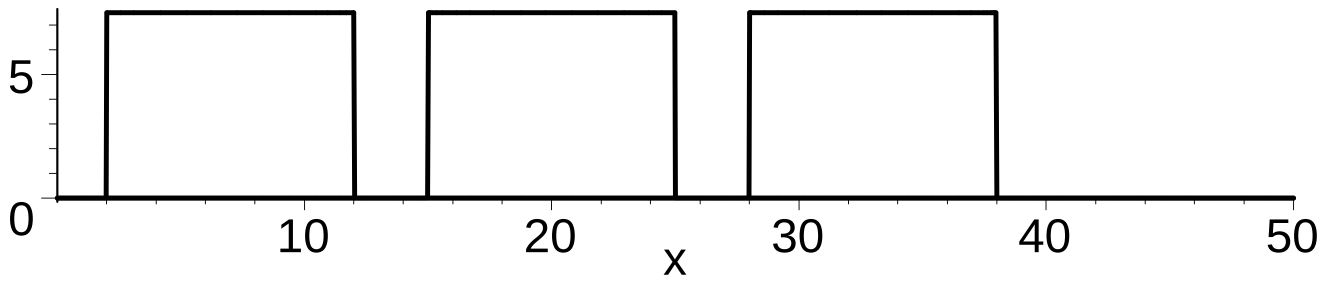


Fig. 9: A square wave

Example *A digital thermometer*

Another discontinuous function is obtained if we take the output in ^0C, say, for a series of measurements made with a thermometer placed inside a car engine; the output of this device being digital. One can imagine that this output being fed into the energy management computer—a feature of all present day cars—so that it may make adjustments to other engine parameters such as fuel supply rate, air mixture rate, spark plug and valve timings and so on. In any case the graph is just a somewhat periodic looking collection of dots and is depicted in fig. 10.

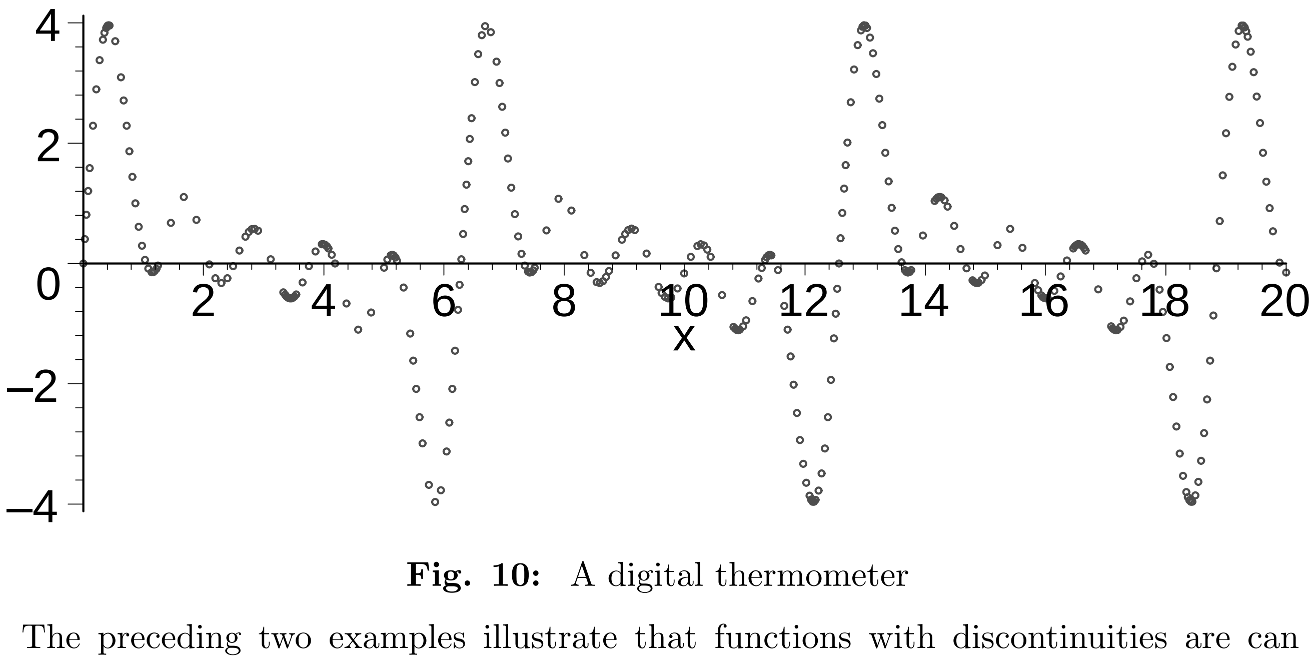


Fig. 10: A digital thermometer

The preceding two examples illustrate that functions with discontinuities are can be perfectly practical and real functions not just esoteric pathological mathematical examples. Note too that a graph such as that of the digital thermometer, which just consists of dots, is discontinuous *everywhere* not just at one point as in the second graph of fig. 8, or at a series of isolated points as in fig. 9.

Example *A capacitor being charged*

As our final example we take a continuous function. The graph of 28 shows the function $Q(t$ as a function of time t where $Q(t)$ is the charge on a capacitor as it is being charged up. Incidentally the function $Q(t)$ being plotted is given by

$$Q(t) = 1 - \exp(-4.5t) \tag{1.113}$$

so that the capacitor will have a charge $Q = 1$ only if t becomes infinite. [10] Hence in theory, the capacitor takes an infinite time to charge fully; of course in practice it reaches 99% of its final charge in a finite time.

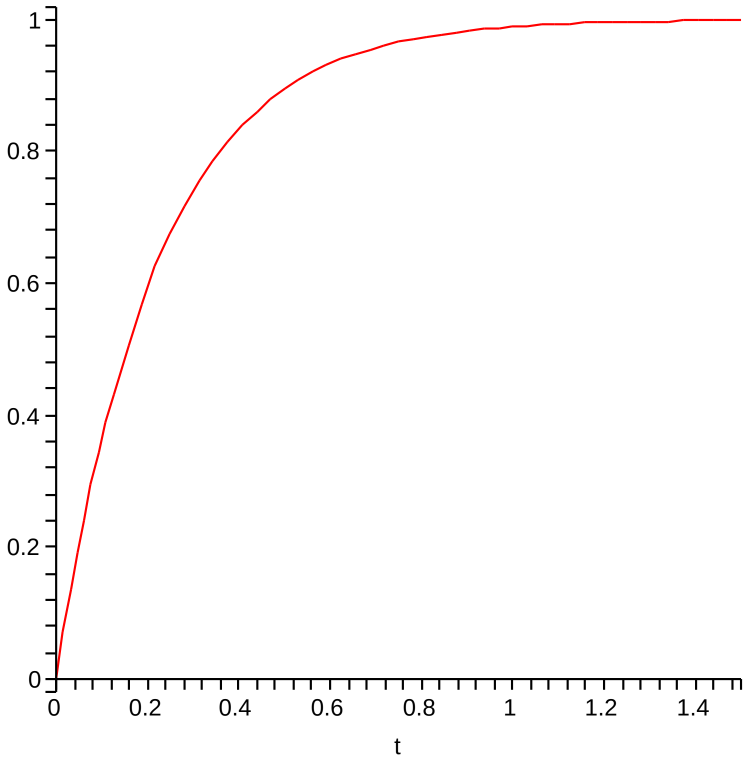


Fig. 11: A charging capacitor

Having been introduced to the notions of limits and continuous functions we are ready to begin our work on calculus; this is the topic of the next chapter.

[10] In the language of limits this is just the statement that $\lim_{t\to\infty} Q(t) = 1$.

CHAPTER II
Differential calculus

§ 1. Derivatives

HISTORICALLY derivatives grew out of a desire to have a general way to draw or construct a tangent to any curve. If the curve was a circle, an ellipse, or a parabola tangents could easily be constructed using geometrical rules but none of these rules gave a hint to a general method. We shall see now that the key technical step to drawing a general tangent is to think of it as a *limit*.

A tangent to a curve at a point p is a straight line which just touches the curve at p cf. fig 12.

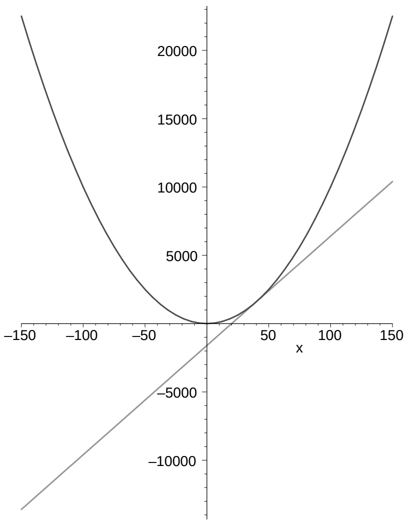


Fig. 12: The curve $f(x) = x^2 + 3$ and one of its tangents

Unfortunately we cannot *define* a tangent to a curve as a straight line which just touches it in one point. To see why this is so examine the two graphs shown in fig. 13. The first graph shows a straight line which is a tangent to a curve but the line touches the curve in two points only one of which—the rightmost one—is a tangential point. The second graph shows a parabola with two lines—one vertical and one slanted—passing through the same point on the parabola. However note that, though both these lines touch the parabola at just one point, only one of these lines—the slanted one—is a tangent.

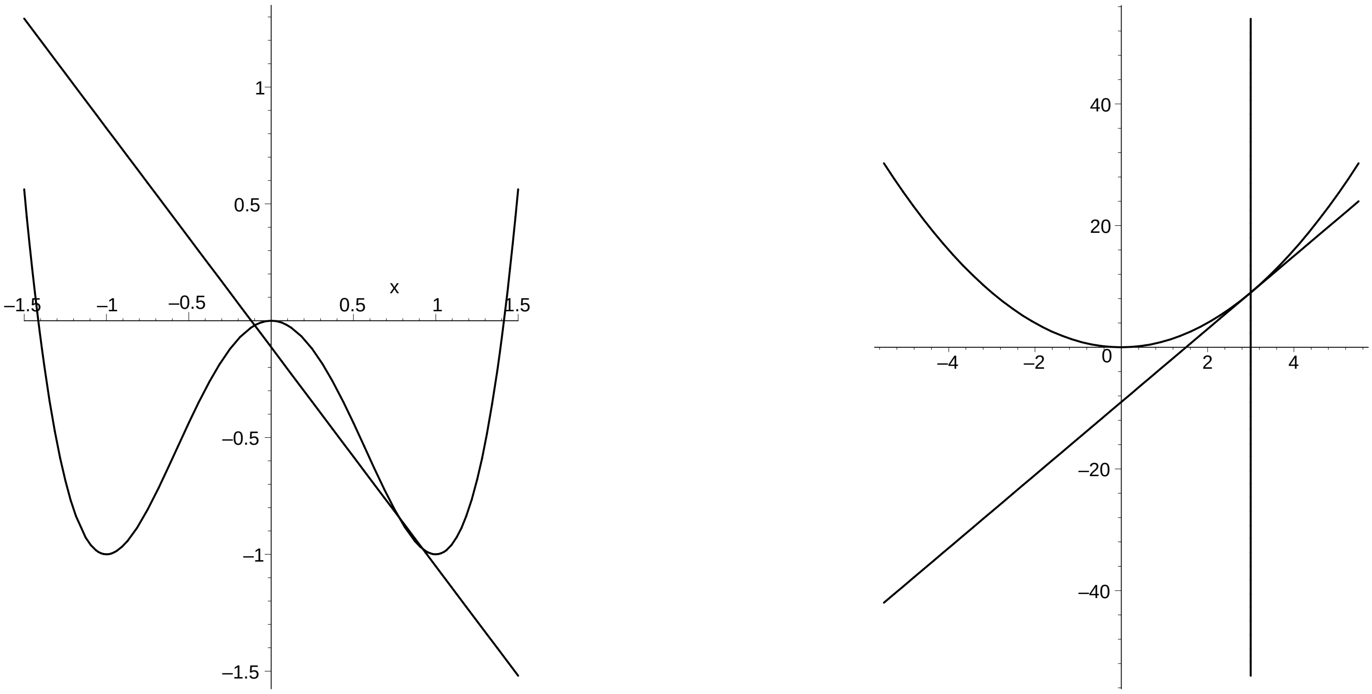


Fig. 13: The curves $f(x) = x^4 - 2x^2$ and $f(x) = x^2$ teaching a lesson about tangents

We now turn to the method of tangent construction that actually works; we shall see that this method will require us to use a limit.

Suppose then that we want to construct a tangent to the function $f(x)$ at the point x and then calculate its slope. We accomplish this task by doing just two things: First we draw a straight line through the two points with coordinates $(x, f(x))$ and $(x+h, f(x+h))$ as shown in fig. 14.

This line is not yet a tangent to the curve, but in any case, its slope is

$$\frac{f(x+h) - f(x)}{h} \tag{2.1}$$

Secondly we send $h \to 0$ which makes our line become the tangent to $f(x)$ at x; its slope is then given by the value of the expression 2.1 as $h \to 0$ i.e. by the limit

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \tag{2.2}$$

It is this quantity which is called the **derivative** of $f(x)$ at x—we record this symbolically by writing

$$\frac{df}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \tag{2.3}$$

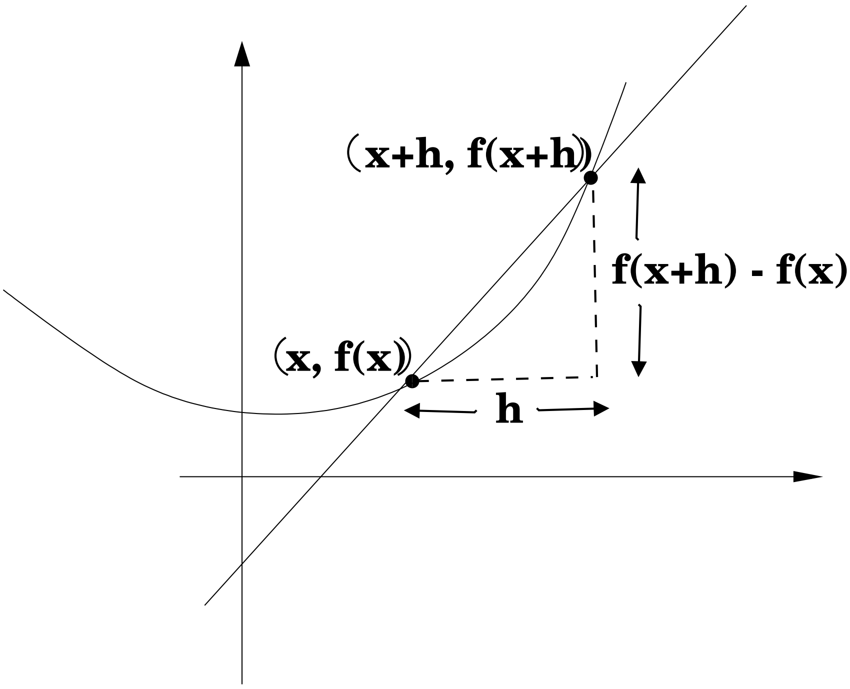


Fig. 14: The curve $(f(x)$ and the line through the points $(x, f(x))$ and $(x+h, f(x+h))$

The derivative is also often written as f' or $f'(x)$. To summarise the notation we have

$$f' = \frac{df}{dx} \tag{2.4}$$

We proceed at once to our first calculation of a derivative.

Example *The derivative of x^n for $n \in \mathbf{N}$*

So our task is to calculate[1]

$$\frac{df}{dx}, \qquad \text{where} \qquad f(x) = x^n,\ n \in \mathbf{N} \tag{2.5}$$

We have

$$\begin{aligned} \frac{f(x+h)-f(x)}{h} &= \frac{(x+h)^n - x^n}{h} \\ &= \frac{\left(x^n + \binom{n}{1}x^{n-1}h + \binom{n}{2}x^{n-2}h^2 + \cdots + \binom{n}{n-1}xh^{n-1} + h^n\right) - x^n}{h} \\ &= \binom{n}{1}x^{n-1} + \binom{n}{2}x^{n-2}h + \cdots + \binom{n}{n-1}xh^{n-2} + h^{n-1} \end{aligned} \tag{2.6}$$

[1] We shall need the binomial theorem for this calculation so we remind the forgetful that it says that

$$(x+y)^m = x^m + \binom{m}{1}x^{m-1}y + \binom{m}{2}x^{m-2}y^2 + \cdots + \binom{m}{m-1}xy^{m-1} + y^m$$

But now the limit is easy and we find that

$$\begin{aligned}\lim_{h\to 0}\frac{f(x+h)-f(x)}{h} &= \lim_{h\to 0}\left\{\binom{n}{1}x^{n-1}+\binom{n}{2}x^{n-2}h+\cdots+\binom{n}{n-1}xh^{n-2}+h^{n-1}\right\}\\ &= \binom{n}{1}x^{n-1}\\ &= nx^{n-1},\quad \text{since } \binom{n}{1}=n\end{aligned} \tag{2.7}$$

Hence our result has the following simple form

$$\frac{dx^n}{dx} = nx^{n-1},\ n\in\mathbf{N} \tag{2.8}$$

We can even allow the n in x^n to have the value 0 because it is a simple matter to check that [2]

$$\frac{dx^0}{dx}\equiv\frac{d1}{dx}=0 \tag{2.10}$$

Still more is true: the n in x^n can have *any* real value: positive, negative, rational or irrational [3]; so we have

$$\frac{dx^n}{dx} = nx^{n-1},\ n\in\mathbf{R} \tag{2.11}$$

It is easy to check that derivatives have the following useful properties

$$\begin{aligned}\frac{d}{dx}\left(f(x)+g(x)\right) &= \frac{df(x)}{dx}+\frac{dg(x)}{dx}, &&\text{for any two functions } f(x) \text{ and } g(x)\\ \frac{d}{dx}\left(Cf(x)\right) &= C\frac{df(x)}{dx}, &&\text{for any constant } C \text{ and function } f(x)\end{aligned} \tag{2.12}$$

These properties 2.12 make it instantly possible to compute many more derivatives. For instance using the first equality in 2.12 we can immediately conclude that

$$\begin{aligned}\frac{d}{dx}(x^3+x^7) &= \frac{dx^3}{dx}+\frac{dx^7}{dx}\\ &= 3x^2+7x^6\end{aligned} \tag{2.13}$$

[2] The argument goes like this

$$\begin{aligned}\frac{d1}{dx} &= \lim_{h\to 0}\frac{(1-1)}{h}\\ &= 0\end{aligned} \tag{2.9}$$

[3] The proof of this statement for irrational n is not quite as trivial as might be assumed; it requires a proof of the binomial theorem for irrational n and we do not give the details here.

while using 2.12 part two we find that

$$\frac{d}{dx}(100x^{10}) = 1000x^9 \tag{2.14}$$

Finally applying both properties together we have the useful result that

$$\frac{d}{dx}\left(100x^{10} + x^3 + x^7\right) = 1000x^9 + 3x^2 + 7x^6 \tag{2.15}$$

All this means that we can now differentiate any *polynomial* in x—i.e. any expression of the form

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \tag{2.16}$$

where the quantities $a_0, a_1, \ldots a_n$ are constants. The derivative of this polynomial is given by

$$\frac{d}{dx}\left(a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0\right) = na_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + \cdots + 2a_2 x + a_1 \tag{2.17}$$

§§ 1.1 The derivatives of the expressions $1/f(x)$ and $f(x)g(x)$

We must forge ahead in learning more about how to differentiate. The first task we can polish off quickly is the derivative of

$$\frac{1}{f(x)} \tag{2.18}$$

This is how it goes

$$\begin{aligned}
\frac{d}{dx}\left(\frac{1}{f(x)}\right) &= \lim_{h\to 0}\left\{\frac{1}{f(x+h)} - \frac{1}{f(x)}\right\}\frac{1}{h} \\
&= \lim_{h\to 0}\left\{\frac{f(x) - f(x+h)}{f(x+h)f(x)}\right\}\frac{1}{h} \\
&= -\lim_{h\to 0}\left\{\frac{f(x+h) - f(x)}{h}\frac{1}{f(x+h)f(x)}\right\} \\
&= -\lim_{h\to 0}\left\{\frac{f(x+h) - f(x)}{h}\right\}\lim_{h\to 0}\left\{\frac{1}{f(x+h)f(x)}\right\} \\
&= -\left(\frac{1}{f(x)}\right)^2 \frac{df}{dx}
\end{aligned} \tag{2.19}$$

So we have established that

$$\begin{aligned}
\frac{d}{dx}\left(\frac{1}{f(x)}\right) &= -\left(\frac{1}{f(x)}\right)^2 \frac{df}{dx} \\
&= -\frac{f'}{f^2}
\end{aligned} \tag{2.20}$$

We can now display our latest new skill in the following small calculation

Example *The derivative of*

$$\frac{1}{(a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0)} \tag{2.21}$$

Using 2.20 above we easily calculate that

$$\frac{d}{dx}\left\{\frac{1}{(a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0)}\right\} = -\frac{(na_n x^{n-1} + \cdots + 2a_2 x + a_1)}{(a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0)^2} \tag{2.22}$$

Turning to our next job which concerns the product $f(x)g(x)$, we use the simple trick of adding and subtracting the quantity $f(x+h)g(x)$ and calculate that

$$\begin{aligned}
\frac{d}{dx}(f(x)g(x)) &= \lim_{h\to 0}\left\{\frac{f(x+h)g(x+h) - f(x)g(x)}{h}\right\} \\
&= \lim_{h\to 0}\left\{\frac{(f(x+h)g(x+h) - f(x+h)g(x)) + (f(x+h)g(x) - f(x)g(x))}{h}\right\} \\
&= \lim_{h\to 0} f(x+h)\left\{\frac{g(x+h) - g(x)}{h}\right\} + \lim_{h\to 0}\left\{\frac{f(x+h) - f(x)}{h}\right\} g(x) \\
&= \left\{\lim_{h\to 0} f(x+h)\right\} \lim_{h\to 0}\left\{\frac{g(x+h) - g(x)}{h}\right\} + \lim_{h\to 0}\left\{\frac{f(x+h) - f(x)}{h}\right\} g(x) \\
&= f(x)\frac{dg(x)}{dx} + \frac{df(x)}{dx}g(x)
\end{aligned} \tag{2.23}$$

This formula is sometimes called the *Leibnitz product formula* and, to summarise, we have just proved that

$$\frac{d}{dx}(f(x)g(x)) = f(x)\frac{dg(x)}{dx} + \frac{df(x)}{dx}g(x) \tag{2.24}$$

The product formula can be put to use at once; let us have a look at one instance of it in action.

Example *The derivative of*

$$(1 + 3x^2 + 6x^4 + 10x^8)(1 + x + x^2 + x^3) \tag{2.25}$$

Using 2.24 we obtain the result

$$\begin{aligned}
\frac{d}{dx}\left((1 + 3x^2 + 6x^4 + 10x^8)(1 + x + x^2 + x^3)\right) &= (1 + 3x^2 + 6x^4 + 10x^8)(1 + 2x + 3x^2) + \\
&\quad (6x + 24x^3 + 80x^7)(1 + x + x^2 + x^3)
\end{aligned} \tag{2.26}$$

where we note the convenience of not having to multiply out all the brackets.

§§ 1.2 The derivative of f/g

An immediate application of the two preceding results 2.20 and 2.24 is to the computation of the derivative of the *quotient* of two functions, that is the quantity

$$\frac{d}{dx}\left\{\frac{f(x)}{g(x)}\right\} \tag{2.27}$$

So, using 2.24 we obtain

$$\begin{aligned}\frac{d}{dx}\left\{\frac{f(x)}{g(x)}\right\} &= f(x)\frac{d}{dx}\left\{\frac{1}{g(x)}\right\} + \frac{df(x)}{dx}\frac{1}{g(x)}\\ &= -f(x)\frac{1}{g^2(x)}\frac{dg(x)}{dx} + \frac{df(x)}{dx}\frac{1}{g(x)}, \quad \text{using 2.20}\\ &= \frac{g(x)\frac{df(x)}{dx} - f(x)\frac{dg(x)}{dx}}{g^2(x)}\\ &= \frac{gf' - fg'}{g^2}\end{aligned} \tag{2.28}$$

and this is sometimes called the *quotient rule.* As a token of what we are now able to differentiate we just quote some expressions—we shall leave the actual computation of their derivatives to the reader.

Example *Some expressions we can now differentiate*

$$\begin{gathered}\frac{(1-22x^2-77x^4+10x^8)}{(1-x+x^2-x^3+x^4)}, \ \frac{(x+3x^3-12x^4+x^6)(1+x^2+x^6+x^{24})}{(1-x+x^3-x^5-x^7)}\\ \frac{(1+x^2+x^4+x^6+x^8))(x^5-5x^7+x^{11})}{(1+x+3x^2-5x^7+6x^9)(x-x^2+x^3+x^6)}\end{gathered} \tag{2.29}$$

It is now time for the derivative of just one more variety of function.

§§ 1.3 The chain rule: The derivative of $f(g(x))$

The final formula we routinely need for working with derivatives is called the *chain rule.* This is the formula that tells us how to differentiate a function which is presented as *a function of a function* so that it looks like

$$f(g(x)) \tag{2.30}$$

For example if one had $f(x) = \sin(x)$ and $g(x) = x^4$ then we would have

$$f(g(x)) = \sin(x^4) \tag{2.31}$$

Steaming ahead we compute that

$$\begin{aligned}
\frac{df(g(x))}{dx} &= \lim_{h\to 0}\left\{\frac{f(g(x+h))-f(g(x))}{h}\right\} \\
&= \lim_{h\to 0}\left\{\frac{f(g(x+h))-f(g(x))}{g(x+h)-g(x)}\right\}\left\{\frac{g(x+h)-g(x)}{h}\right\} \\
&= \lim_{h\to 0}\left\{\frac{f(g(x+h))-f(g(x))}{g(x+h)-g(x)}\right\}\lim_{h\to 0}\left\{\frac{g(x+h)-g(x)}{h}\right\} \\
&= \lim_{h\to 0}\left\{\frac{f(g(x+h))-f(g(x))}{g(x+h)-g(x)}\right\}\frac{dg}{dx}
\end{aligned} \tag{2.32}$$

Now if we just use the elementary fact that

$$\begin{aligned}
g(x+h) &= g(x) + g(x+h) - g(x) \\
&= g(x) + k, \quad \text{where } k = g(x+h) - g(x) \\
\Rightarrow f(g(x+h)) &= f(g(x)+k)
\end{aligned} \tag{2.33}$$

and substitute $f(g(x+h)) = f(g(x)+k)$ in the last line of 2.32 we find that

$$\begin{aligned}
\frac{df(g(x))}{dx} &= \lim_{h\to 0}\left\{\frac{f(g(x)+k)-f(g(x))}{k}\right\}\frac{dg}{dx} \\
&= \frac{df(g)}{dg}\frac{dg}{dx}
\end{aligned} \tag{2.34}$$

and our limits have been evaluated. Hence our result is that

$$\frac{df(g(x))}{dx} = \frac{df(g)}{dg}\frac{dg}{dx} \tag{2.35}$$

and this formula 2.35 is called the *chain rule* or sometimes the derivative of *a function of a function*. Bringing these four useful formulae 2.20, 2.24, 2.28 and 2.35 together, and using the f' style of notation instead of df/dx in the first three, we get the rather easier to remember compact forms

$$\begin{aligned}
\left(\frac{1}{f}\right)' &= -\frac{f'}{f^2} \\
(fg)' &= fg' + f'g \\
\left(\frac{f}{g}\right)' &= \frac{gf' - fg'}{g^2} \\
\frac{df(g(x))}{dx} &= \frac{df(g)}{dg}\frac{dg}{dx}
\end{aligned} \tag{2.36}$$

§ 2. Derivatives of trigonometric and exponential functions

The trigonometric functions—all of which we would like to be able to differentiate—are all built out of $\sin(x)$ and $\cos(x)$ so we shall now compute the derivatives of $\sin(x)$ and $\cos(x)$. Actually since

$$\cos(x) = \sqrt{1 - \sin^2(x)} \tag{2.37}$$

we only have to compute the derivative of $\sin(x)$ and the derivatives of every other trigonometric function will then follow from this one computation.

§§ 2.1 The derivatives of $\sin(x)$ and $\cos(x)$

Proceeding forward, if we bear in mind the formula

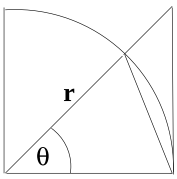
$$\sin(a) - \sin(b)) = 2\cos\left(\frac{a+b}{2}\right)\sin\left(\frac{a-b}{2}\right) \tag{2.38}$$

we have

$$\begin{aligned} \frac{d}{dx}(\sin(x)) &= \lim_{h\to 0}\frac{\sin(x+h)-\sin(x)}{h} \\ &= \lim_{h\to 0}\frac{2\cos\left(x+\frac{h}{2}\right)\sin\left(\frac{h}{2}\right)}{h} \\ &= \cos(x)\lim_{h\to 0}\frac{\sin(\frac{h}{2})}{\frac{h}{2}} \\ &= \cos(x),\ \text{since } \lim_{h\to 0}\frac{\sin(\theta)}{\theta} = 1 \end{aligned} \tag{2.39}$$

So we have found that [4]

[4] We just quoted the result $\lim_{h\to 0}\frac{\sin(\theta)}{\theta} = 1$. For the reader who is interested in its proof we sketch its main idea which is geometric: Consider the sector of the circle and the two triangles shown in the figure.



Since the large triangle contains the sector, which in turn contains the smaller triangle, then computation of their three areas gives the inequality

$$\begin{aligned} &\frac{1}{2}r^2\sin(\theta) < \frac{r^2\theta}{2} < \frac{1}{2}r^2\tan(\theta) \\ &\Rightarrow 1 < \frac{\theta}{\sin(\theta)} < \frac{1}{\cos(\theta)} \\ &\Rightarrow \cos(\theta) < \frac{\sin(\theta)}{\theta} < 1, \quad \text{if we invert everything} \end{aligned}$$

and the limit then easily follows on sending $\theta \to 0$.

$$\frac{d}{dx}(\sin(x)) = \cos(x), \qquad \text{or } \sin'(x) = \cos(x) \tag{2.40}$$

We can now move on to $\cos(x)$. We shall use the the chain rule as follows

$$\begin{aligned} \cos(x) &= \sqrt{1-\sin^2(x)} \\ \Rightarrow \frac{d}{dx}(\cos(x)) &= \left(\frac{1}{2}\right)\left(1-\sin^2(x)\right)^{-1/2}\frac{d}{dx}(1-\sin^2(x)) \\ &= \left(\frac{1}{2}\right)\frac{(-2\sin(x)\sin'(x))}{\sqrt{1-\sin^2(x)}} \\ &= -\frac{\sin(x)\cos(x)}{\cos(x)} \\ &= -\sin(x) \end{aligned} \tag{2.41}$$

so the result is

$$\frac{d}{dx}(\cos(x)) = -\sin(x), \qquad \text{or } \cos'(x) = -\sin(x) \tag{2.42}$$

§§ 2.2 The derivatives of the other trigonometric functions

It is now perfectly straightforward to calculate the derivatives of the remaining standard trigonometric functions. These functions, together with their definitions for the forgetful, are

$$\tan(x),\ \sec(x),\ \cot(x) \text{ and } \operatorname{cosec}(x) \tag{2.43}$$

and have the definitions

$$\tan(x) = \frac{\sin(x)}{\cos(x)},\ \sec(x) = \frac{1}{\cos(x)},\ \cot(x) = \frac{1}{\tan(x)} \text{ and } \operatorname{cosec}(x) = \frac{1}{\sin(x)} \tag{2.44}$$

We leave it to the reader to verify the formulae for their derivatives given below

$$\begin{array}{cc} \tan'(x) = \sec^2(x), & \sec'(x) = \sec(x)\tan(x) \\ \cot'(x) = -\operatorname{cosec}^2(x), & \operatorname{cosec}'(x) = -\operatorname{cosec}(x)\cot(x) \end{array} \tag{2.45}$$

§§ 2.3 The derivatives of exponential functions

An *exponential function* is a function $f(x)$ of the form [5]

$$f(x) = a^x, \qquad \text{where } a \text{ is a constant} \tag{2.46}$$

[5] Actually one allows such a function to be multiplied by a constant C so

$$f(x) = Ca^x$$

is also an exponential function

it acquires its name from the fact that the power x in a^x is also sometimes called an *exponent*. The most famous exponential function is the one obtained by setting

$$\begin{gathered} a = e, \\ \text{where } e = 2.71828\ldots \text{ is the base of natural logarithms} \end{gathered} \tag{2.47}$$

This function

$$e^x \tag{2.48}$$

is often called **the exponential function** and is often denoted by

$$\exp(x) \tag{2.49}$$

so remember that e^x and $\exp(x)$ stand for the *same* function, i.e.

$$e^x = \exp(x) \tag{2.50}$$

The exponential function e^x has the following infinite series expansion

$$\begin{aligned} e^x &= \sum_{n=0}^{\infty} \frac{x^n}{n!} \\ &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \end{aligned} \tag{2.51}$$

The derivative of e^x turns out to be equal to e^x—in other words e^x is its own derivative[6] This is easy to prove if we differentiate both sides of 2.51: Doing this gives

$$\begin{aligned} \frac{d}{dx}\left(\sum_{n=0}^{\infty} \frac{x^n}{n!}\right) &= \frac{d}{dx}\left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots\right) \\ &= \left(0 + 1 + \frac{2x}{2!} + \frac{3x^2}{3!} + \cdots\right) \\ &= \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots\right) \\ &= \sum_{n=0}^{\infty} \frac{x^n}{n!} \\ &= e^x \end{aligned} \tag{2.52}$$

[6] In fact, instead of defining e^x by the infinite series 2.51, we could search for all functions $f(x)$ which are their own derivative—that is $f' = f$—and we would then find that f has to be given by the series

$$f(x) = C \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

So setting $C = 1$ gives $f = e^x$.

So
$$\frac{d}{dx}e^x = e^x \tag{2.53}$$

Before leaving e^x we shall also find the derivative of the logarithm function [7] $\ln(x)$ We remind the reader that $\ln(x)$ is the *inverse function* to e^x and vice-versa—i.e. we have

$$\ln(e^x) = x \quad \textbf{and} \quad e^{\ln(x)} = x \tag{2.54}$$

In any case we can find the derivative of $\ln(x)$ by differentiating the equation $e^{\ln(x)} = x$: if we do this—and remember to use the chain rule 2.35—we get

$$\begin{aligned}
\frac{d}{dx}e^{\ln(x)} &= \frac{d}{dx}x \\
\Rightarrow e^{\ln(x)}\frac{d}{dx}\ln(x) &= 1 \\
\Rightarrow \frac{d}{dx}\ln(x) &= \frac{1}{e^{\ln(x)}} \\
\Rightarrow \frac{d}{dx}\ln(x) &= \frac{1}{x}, \quad \text{since } e^{\ln(x)} = x
\end{aligned} \tag{2.55}$$

So we now know that
$$\frac{d}{dx}\ln(x) = \frac{1}{x} \tag{2.56}$$

The other exponential function we want to differentiate is simply $f = a^x$ for any constant a. If we use the log function $\ln(x)$ then we have

$$f = a^x = e^{x\ln(a)} \tag{2.57}$$

Hence we can do the differentiation immediately and the result is that

$$\begin{aligned}
\frac{d}{dx}a^x &= \frac{d}{dx}e^{x\ln(a)} \\
&= e^{x\ln(a)}\frac{d}{dx}(x\ln(a)) \\
&= e^{x\ln(a)}\ln(a) \\
&= a^x\ln(a)
\end{aligned} \tag{2.58}$$

Recapitulating, we have found that

$$\frac{d}{dx}e^x = e^x,\ \frac{d}{dx}a^x = a^x\ln(a),\ \frac{d}{dx}\ln(x) = \frac{1}{x} \tag{2.59}$$

[7] The notation $\ln(x)$ denotes the logarithm of x to the base e also called the *natural logarithm* of x

§§ 2.4 The hyperbolic functions $\sinh(x)$, $\cosh(x)$ and $\tanh(x)$

Closely related to the trigonometric functions $\sin(x)$, $\cos(x)$ and $\tan(x)$ are the functions $\sinh(x)$, $\cosh(x)$ and $\tanh(x)$ which are officially called *hyperbolic sine*, *hyperbolic cosine* and *hyperbolic tan* respectively. However sinh, cosh and tanh are usually pronounced as "sinsh", "cosh" and "tansh"respectively. These functions are made of various combinations of $e^{\mp x}$; their definitions are

$$\begin{aligned} \sinh(x) &= \frac{e^x - e^{-x}}{2} \\ \cosh(x) &= \frac{e^x + e^{-x}}{2} \\ \tanh(x) &= \frac{\sinh(x)}{\cosh(x)} \end{aligned} \tag{2.60}$$

They have quite a few analogous properties to their trigonometric cousins sin, cos and tan but, unlike sin, cos and tan, they are *not* periodic functions. Among their properties we quote the following, which the reader should find easy to verify from their definitions

(i) $\cosh^2(x) - \sinh^2(x) = 1$
(ii) $\sinh(x + y) = \sinh(x)\cosh(y) + \cosh(x)\sinh(y)$
(iii) $\sinh'(x) = \cosh(x)$
(iv) $\cosh'(x) = \sinh(x)$
(v) $\tanh'(x) = \text{sech}^2(x)$, where $\text{sech}(x) = 1/\cosh(x)$.

§§ 2.5 The inverse trigonometric functions $\arcsin(x)$, $\arccos(x)$ and $\arctan(x)$

The trigonometric functions all have inverses and we would like to be able to compute their derivatives as they occur in calculations from time to time. The inverse function to $\sin(x)$ is denoted by $\arcsin(x)$ and, because it is an *inverse function* it satisfies the pair of equations[8]

$$\arcsin(\sin(x)) = x, \ \sin(\arcsin(x)) = x \tag{2.61}$$

It is very useful to remember that $\arcsin(x)$ is simply *an angle*—hence if one writes

$$f = \arcsin(x) \tag{2.62}$$

it may help to read this aloud—or silently—as the phrase

"f is the angle whose sine is x"

The other inverse trigonometric functions are defined in the same way giving us

$$\begin{aligned} \arccos(\cos(x)) = x, \ \cos(\arccos(x)) = x \\ \arctan(\tan(x)) = x, \ \tan(\arctan(x)) = x \end{aligned} \tag{2.63}$$

[8] Sometimes, instead of $\arcsin(x)$, the notation $\sin^{-1}(x)$ is used but this is dangerous because it can be confused with $1/sin(x)$ to which it is definitely *not* equal.

Finally to compute their derivatives one uses the same method as was used to obtain the derivative of $\ln(x)$ in 2.55. We start, therefore, with

$$\begin{aligned} f &= \arcsin(x) \\ \Rightarrow \sin(f(x)) &= x \\ \Rightarrow \frac{d}{dx}\{\sin(f(x))\} &= 1 \\ \Rightarrow \frac{d\sin(f)}{df}\frac{df}{dx} &= 1 \\ \Rightarrow \cos(f)\frac{df}{dx} &= 1 \\ \Rightarrow \frac{df}{dx} &= \frac{1}{\cos(f)} \\ &= \frac{1}{\sqrt{1-\sin^2(f)}}, \text{ since } \cos(f) = \sqrt{1-\sin^2(f)} \\ &= \frac{1}{\sqrt{1-x^2}}, \text{ since } \sin(f) = x \end{aligned} \tag{2.64}$$

So we have shown that

$$\arcsin'(x) = \frac{1}{\sqrt{1-x^2}} \tag{2.65}$$

The same technique can be used by the reader to calculate $\arccos'(x)$ and $\arctan'(x)$. These calculations will show that

$$\arccos'(x) = -\frac{1}{\sqrt{1-x^2}} \quad \text{and} \quad \arctan'(x) = \frac{1}{1+x^2} \tag{2.66}$$

§ 3. The significance of derivatives

Derivatives have many, many uses and we are now going to examine some of them.

We shall see that a relatively *small* knowledge of f' can reveal a *large* amount about a function f; this knowledge is usually acquired by studying the *maxima and minima* of f.

It is also important to realise that f', as well as being the slope of tangent, has an interpretation as a *rate of change.*

Thus, if f is a *distance* f' is a *velocity*, if f is a *velocity* f' is an *acceleration*, if f is the *amount of charge* in a conductor f' is the *current*, if f is momentum (in a given direction) f' is the force in that direction, if f is the energy of a *motor* f' is the *power*, if f is *data* inside a communications channel f' is the rate of data transmission in the channel, and so on.

§§ 3.1 Critical points, maxima and minima

We shall begin with a look at maxima and minima. The first fact that needs to be well noted is the significance of the *sign* of the derivative of any function $f(x)$. The *sign* of $f'(x)$—at a given point x—tells one whether $f(x)$ is *increasing* or *decreasing* at x. To see how this arises just remember that the $f'(x)$ is the tan of an angle and recall that $\tan(\theta)$ has the property that

$$\begin{aligned} \theta < \frac{\pi}{2} &\Rightarrow \tan(\theta) > 0 \\ \theta > \frac{\pi}{2} &\Rightarrow \tan(\theta) < 0 \end{aligned} \tag{2.67}$$

cf. fig. 15

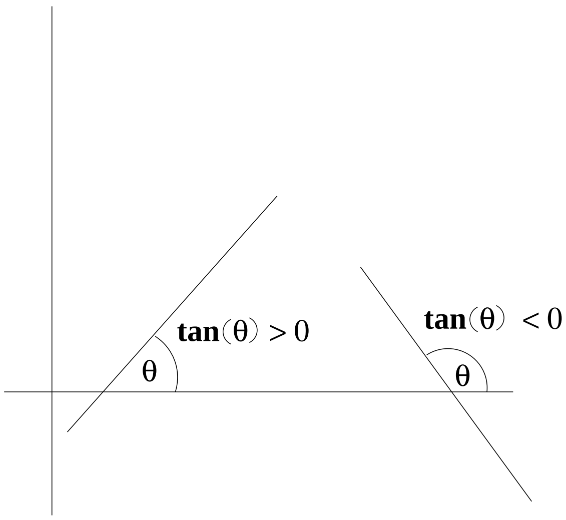


Fig. 15: The sign of $\tan(\theta)$

Now if we examine the graph of $f(x)$ in fig. 16 we see that at a point where $f(x)$ is *increasing* the slope of the tangent is that of an angle *less* than $\pi/2$; while, if $f(x)$ is *decreasing* the slope of the tangent is that of an angle *larger* than $\pi/2$.

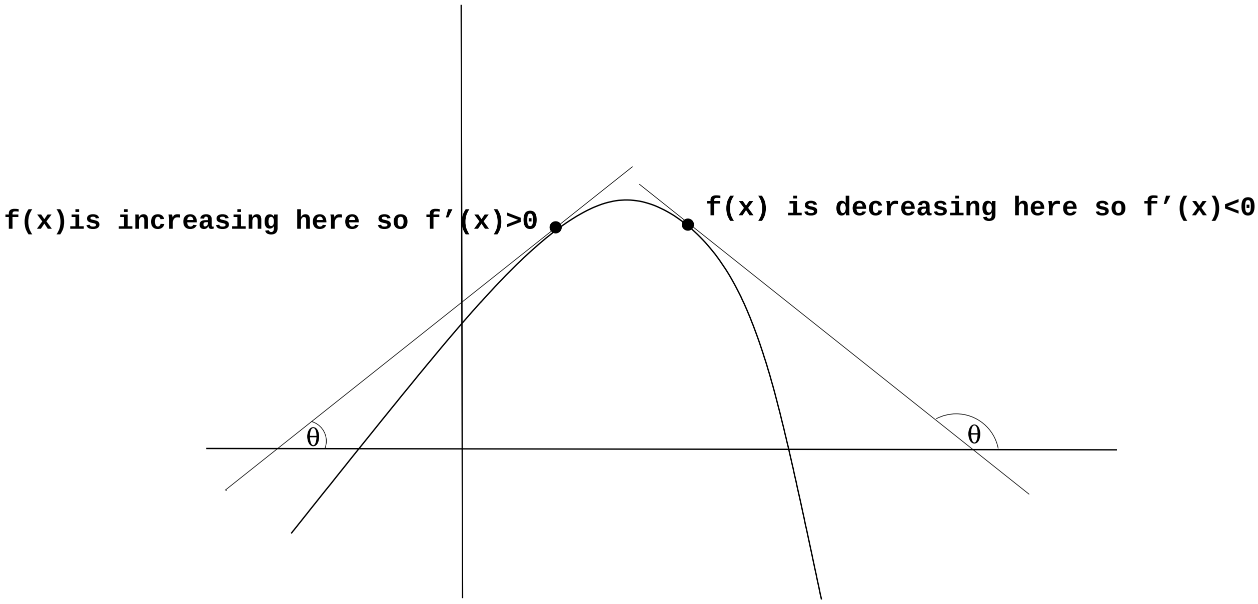


Fig. 16: The meaning of the sign of $f'(x)$

Hence we have the very useful statement that

$$\begin{aligned} f(x) \text{ increasing at the point } x &\Rightarrow f'(x) > 0 \\ f(x) \text{ decreasing at the point } x &\Rightarrow f'(x) < 0 \end{aligned} \tag{2.68}$$

However it is possible for an f, which is increasing and so has $f' > 0$, to *change over* and start decreasing rendering $f' < 0$. Thus the point a, say, at which that happens will be a *maximum* and so f' must pass through zero there; i.e. we must have

$$f'(a) = 0, \quad \text{when } a \text{ is a maximum} \tag{2.69}$$

In an exactly similar way a point a at which changes from being decreasing to increasing is a minimum and so f' must also pass through zero there; i.e. we must again have

$$f'(a) = 0, \quad \text{when } a \text{ is a minimum} \tag{2.70}$$

In general, then, when $f'(x)$ is *zero* at a we *usually*, but *not always*, find that a is a point where $f(x)$ has a *maximum* or a *minimum*. In any case a piece of terminology must now be defined: that of a *critical point* [9]

Definition (A critical point) *A critical point of a function f is a number x such that*

$$f'(x) = 0 \tag{2.71}$$

[9] Sometimes the term *stationary point* is used to mean a point where $f'(x) = 0$ but we shall use the more standard term *critical point*.

We move on at once to show how examples of maxima and minima and to give an example of a critical point which is neither a maximum nor a minimum.

Example *A critical point which is a minimum*

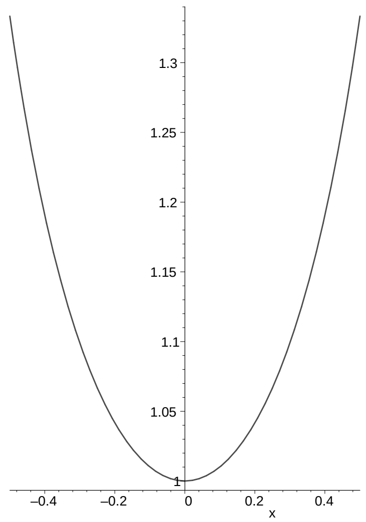


Fig. 17: The function $f = 1/(1-x^2)$. It has $f'(0) = 0$ and has a *minimum* at $x = 0$.

Example *A critical point which is a maximum*

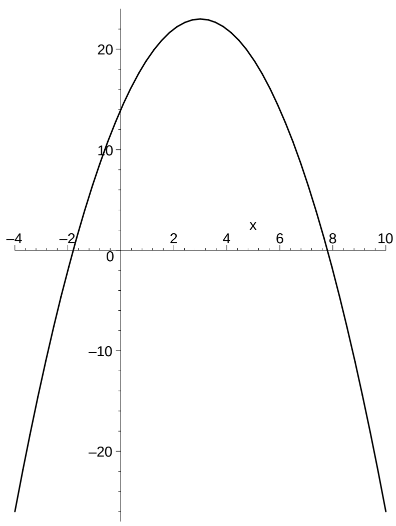


Fig. 18: The function $f = 14 - x^2 + 6x$. It has $f'(3) = 0$ and has a *maximum* at $x = 3$

Example *A critical point which is neither a maximum nor a minimum*

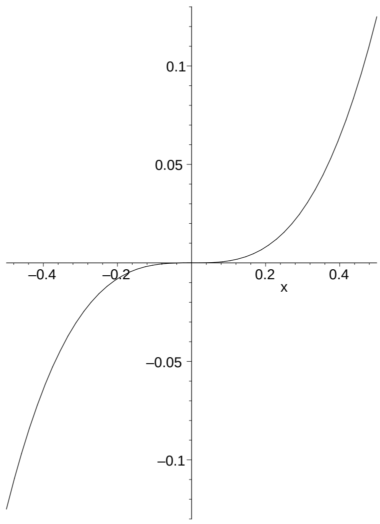


Fig. 19: The function $f = x^3$. It has $f'(0) = 0$ but has *neither maximum nor minimum* at $x = 0$.

We would like some systematic way of telling when critical points are maxima, minima or neither of the two. This is achieved by calculating the second derivative of f at the critical point. The result that we need is summarised in the following statement

$$\begin{aligned} f'(a) = 0 \text{ and } f''(a) < 0 \Rightarrow f \text{ has a } \textit{maximum} \text{ at } a \\ f'(a) = 0 \text{ and } f''(a) > 0 \Rightarrow f \text{ has a } \textit{minimum} \text{ at } a \end{aligned} \tag{2.72}$$

The reader should note carefully that 2.72 still fails to cover one case: this being the case where

$$f''(a) = 0 \tag{2.73}$$

Unfortunately when

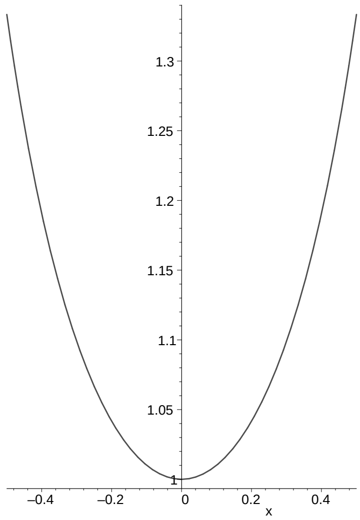
$$f'(a) = 0 \text{ and } f''(a) = 0 \tag{2.74}$$

then a may be a maximum, a minimum, or neither of the two.

The way to understand all this is to look at some examples and in particular to look at the graphs of the functions involved. We begin by analysing the critical points in the graphs of figs. 17–19.

Fig. 17 shows the function $f(x) = 1/(1-x^2)$ and so we calculate that

$$\begin{aligned} f(x) &= \frac{1}{1-x^2} \\ f'(x) &= \frac{2x}{(1-x^2)^2} \\ f''(x) &= \frac{6x^2+2}{(1-x^2)^3} \end{aligned} \tag{2.75}$$



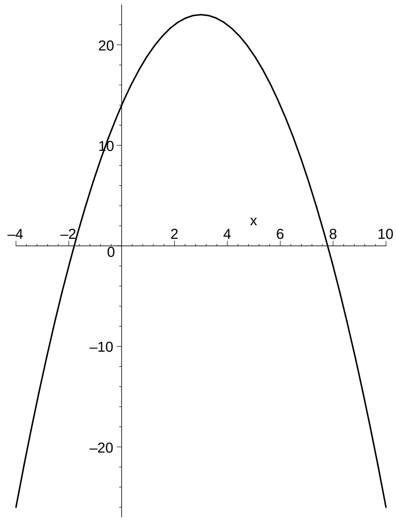
from which we find that

$$\left. \begin{aligned} f'(0) &= 0 \\ f''(0) &= 2 \end{aligned} \right\} \Rightarrow x = 0 \text{ is a minimum} \tag{2.76}$$

as we can see from the graph itself.

Fig. 18 is a graph of $f(x) = 14 - x^2 + 6x$ and we find that

$$\begin{aligned} f(x) &= 14 - x^2 + 6x \\ f'(x) &= -2x + 6 \\ f''(x) &= -2 \end{aligned} \tag{2.77}$$

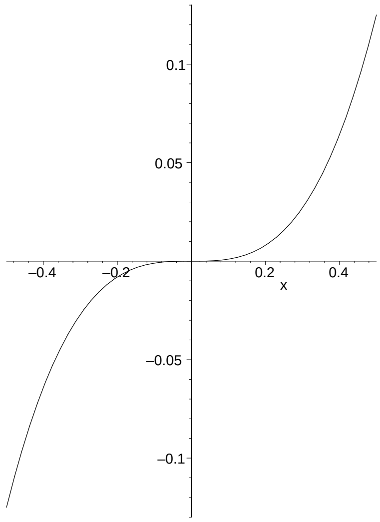


from which we find that

$$\left. \begin{aligned} f'(3) &= 0 \\ f''(3) &= -2 \end{aligned} \right\} \Rightarrow x = 3 \text{ is a maximum} \tag{2.78}$$

again as we knew already.

Now for fig. 19 for which we have



$$\begin{aligned} f(x) &= x^3 \\ f'(x) &= 3x^2 \\ f''(x) &= 6x \end{aligned} \tag{2.79}$$

so that we obtain

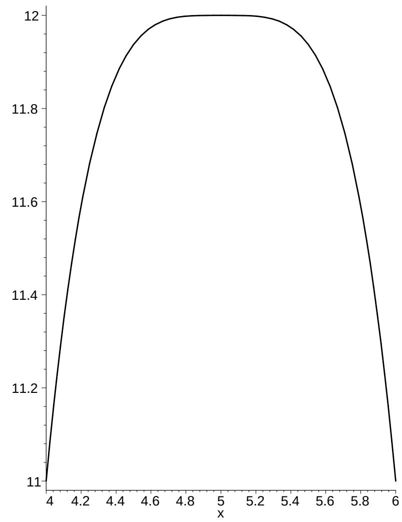
$$\left.\begin{aligned} f'(0) &= 0 \\ f''(0) &= 0 \end{aligned}\right\} \textbf{But } x = 0 \text{ is neither maximum nor minimum} \tag{2.80}$$

which again is what we can see from the graph.

Now we examine two new functions, each with a critical point, the first of which is simply

$$f(x) = 12 - (x-5)^4 \tag{2.81}$$

Routine differentiation tells us that



$$\begin{aligned} f(x) &= 12 - (x-5)^4 \\ f'(x) &= -4(x-5)^3 \\ f''(x) &= -12(x-5)^2 \end{aligned} \tag{2.82}$$

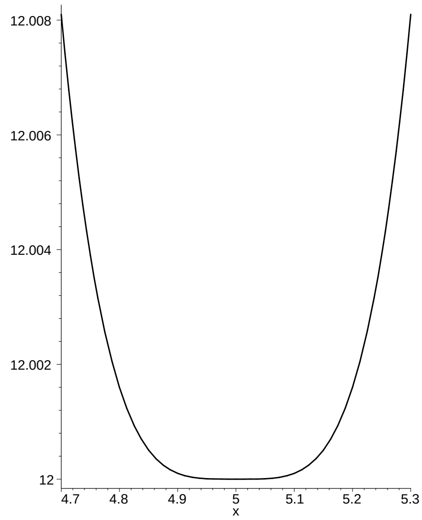
so that we obtain

$$\left.\begin{aligned} f'(5) &= 0 \\ f''(5) &= 0 \end{aligned}\right\} \textbf{But } \ x = 5 \text{ is a } \textbf{maximum} \tag{2.83}$$

as we can read off from the graph.

Now if we just change the sign of the term $-(x-5)^4$ this maximum at $x=5$ will turn into a minimum. Let's see how this works: we obtain

$$\begin{aligned} f(x) &= 12 + (x-5)^4 \\ f'(x) &= 4(x-5)^3 \\ f''(x) &= 12(x-5)^2 \end{aligned} \tag{2.84}$$



yielding

$$\left.\begin{aligned} f'(5) &= 0 \\ f''(5) &= 0 \end{aligned}\right\} \textbf{But} \quad x = 5 \text{ is a } \textbf{minimum} \tag{2.85}$$

as we can see for ourselves.

The point we are trying to emphasise in 2.79–2.84 is that when a critical point $x=a$ of a function f has $f''(a)=0$—as well as the **obligatory** $f'(a)=0$—then $x=a$ can be a maximum (as in 2.82), a minimum (as in 2.84) or neither of the two (as in 2.79).

§§ 3.2 Points of inflection

It turns out that the reason that the graph of 2.79 has a critical point which is neither maximum nor minimum can be traced to the fact that its second derivative $f''(x)$ *changes sign* as x passes through the critical point.

Let's just check this. The relevant data is

$$\begin{aligned} f(x) &= x^3 \\ f'(x) &= 3x^2 \\ f''(x) &= 6x \end{aligned}$$

so the critical point is $x=0$ we must therefore see if the sign of $f''(x)$ changes as x passes through 0. But $f''(x)$ is simply $6x$ so we have

$$\begin{aligned} x < 0 &\Rightarrow 6x < 0 \\ x > 0 &\Rightarrow 6x > 0 \end{aligned} \tag{2.86}$$

So indeed $f''(x)$ did change sign as we passed through the critical point. Any point—not necessarily a critical point—with this property is called a *point of inflection*. Hence we have the following definition.

Definition (Point of inflection) *A point of inflection of a function f is a point a for which $f''(a) = 0$ and f'' changes sign as x passes through a.*

§§ 3.3 The shape of a graph: concavity and convexity

It turns out that the value of f'' is a very useful indicator of *the shape* of the graph of f. The way this works is quite simple and is summarised in fig. 20 below.

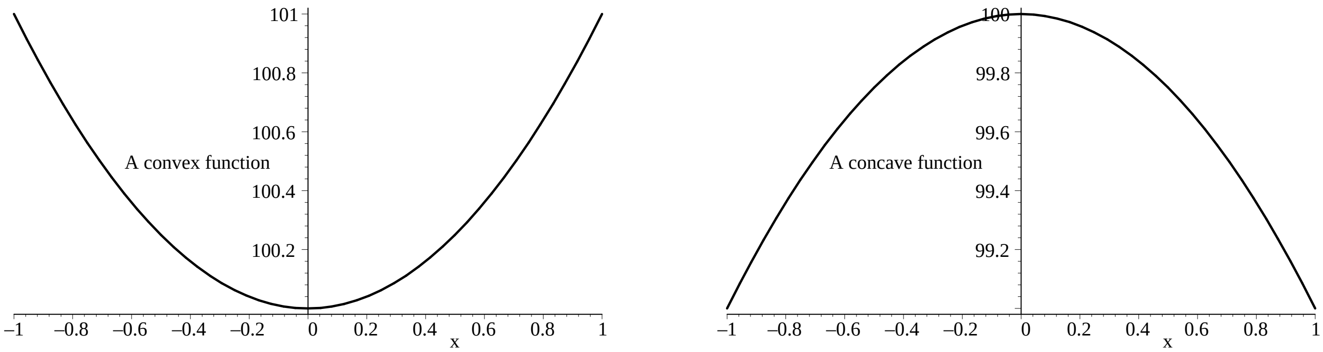


Fig. 20: The sign of f'' determines whether the graph of f is *concave* or *convex*: the graph on the left has $f'' > 0$ and the one on the right has $f'' < 0$.

In fig. 21 we see a graph which starts out concave and then becomes convex.

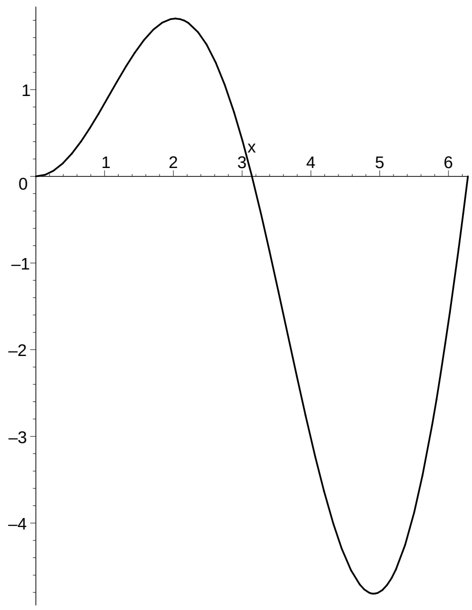


Fig. 21: A function with both concave and convex regions

This is all highly relevant for points of inflection. It is simple to see why this is so: At a point of inflection $x = a$, say, f'' changes sign (by definition), hence there are just two possibilities

(i) f'' changes from positive to negative $\Rightarrow$ the shape of f changes from convex to concave

(ii) f'' changes from negative to positive $\Rightarrow$ the shape of f changes from concave to convex

Example *A function with two points of inflection*

The function

$$f(x) = \frac{1}{1+x^2} \tag{2.87}$$

has *two* points of inflection which means that it change its shape *twice.* It is plotted in fig. 22

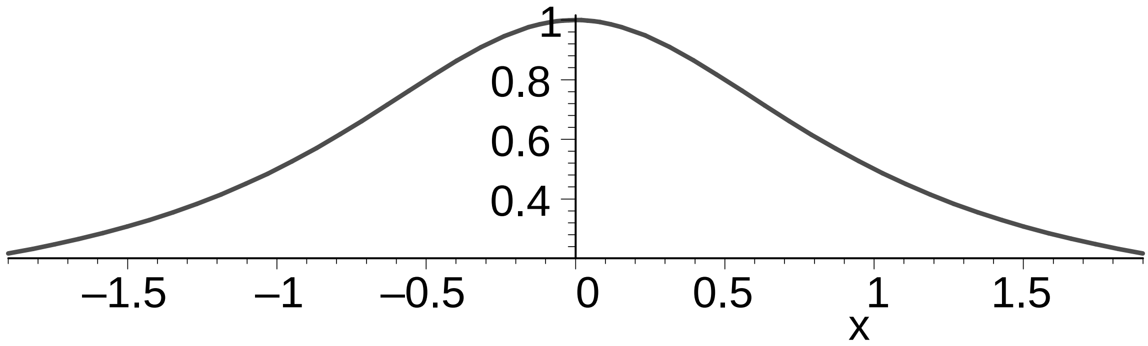


Fig. 22: This function has points of inflection at $x = \mp\frac{1}{\sqrt{3}}$

In fig. 22 $f = 1/(1+x^2)$ is convex until $x = -1/\sqrt{(3)} = -0.577$ then it becomes concave until $x = 1/\sqrt{(3)} = 0.577$ where it becomes convex again. We shall now prove this by doing a little calculation. What we need is f'' so, steaming ahead, we compute that

$$\begin{aligned} f &= \frac{1}{1+x^2} \\ \Rightarrow f' &= -\frac{2x}{(1+x^2)^2} \\ \Rightarrow f'' &= \frac{(1+x^2)^2(-2) - (-2x)(2(1+x^2)(2x)}{(1+x^2)^4}, \quad \text{(quotient rule)} \\ &= \frac{2(3x^2-1)}{(1+x^2)^3} \end{aligned} \tag{2.88}$$

The first pleasing thing to note about the formula 2.88 for f'' is that f'' *vanishes* when

$$3x^2 - 1 = 0 \Rightarrow x = \mp\frac{1}{\sqrt{3}} \tag{2.89}$$

Now, to make it really easy to see where f'' changes sign we shall simply plot it giving us fig. 23.

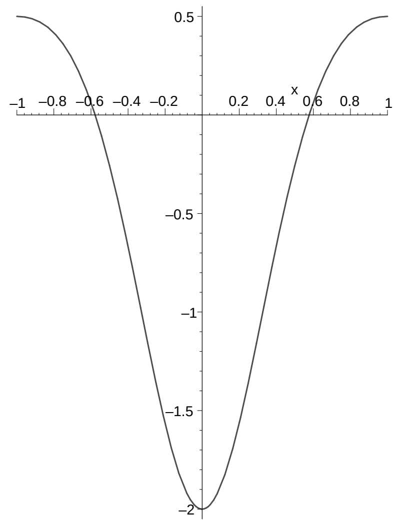


Fig. 23: The graph of f''

What we see in fig. 23 confirms our conclusions: f'' is at first positive then, at about $x = -0.57$ or so, it becomes negative and changes back to being positive at about $x = 0.57$. Thus we have confirmed all that we claimed about the shape of $f = 1/(1 + x^2)$.

§§ 3.4 Limits revisited: L'Hôpital's rule

Calculus can sometimes be quite helpful in the evaluation of limits. Perhaps the most well known and useful application of calculus to limit evaluation is called *L'Hôpital's rule* and it helps when one has to calculate a limit of *a ratio* of two functions—that is something of the form

$$\lim_{x \to a} \frac{f(x)}{g(x)} \tag{2.90}$$

Such a limit may be quite easy to calculate, it depends on the nature of f and g; L'Hôpital's rule only comes in if the numerator and denominator of f/g *both vanish* as $x \to a$. To see how this causes a problem choose the point a to be 0 and set

$$f(x) = \sin(x) \text{ and } g(x) = x \tag{2.91}$$

so that we have [10]

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to 0} \frac{\sin(x)}{x} \tag{2.92}$$

Anyhow note that because

$$\sin(0) = 0 \text{ and } x \text{ vanishes if } x = 0 \tag{2.93}$$

we can say that

$$\lim_{x \to 0} \sin(x) = 0 \text{ and } \lim_{x \to 0} x = 0 \tag{2.94}$$

[10] We already saw in 2.39 that this particular limit has the value 1, we shall confirm that again here.

so we are in trouble if we say

$$\lim_{x\to 0}\frac{\sin(x)}{x} = \frac{\lim_{x\to 0}\sin(x)}{\lim_{x\to 0} x} = \frac{0}{0} \tag{2.95}$$

The trouble comes, clearly, because we do not know what to do with the quantity

$$\frac{0}{0} \tag{2.96}$$

Now we state (without proof) L'Hôpital's rule and show how our difficulty disappears.

Theorem (L'Hôpital's rule) *Suppose that we want to evaluate*

$$\lim_{x\to a}\frac{f(x)}{g(x)} \tag{2.97}$$

and we **also** *have*

$$\lim_{x\to a} f(x) = 0 \ \textit{and} \ \lim_{x\to a} g(x) = 0$$

then

$$\lim_{x\to a}\frac{f(x)}{g(x)} = \lim_{x\to a}\frac{f'(x)}{g'(x)} \tag{2.98}$$

provided both limits exist.

In other words, if the conditions of the theorem apply, we can *replace* $f(x)$ and $g(x)$ by their derivatives. We go on immediately to show how L'Hôpital's rule removes our difficulty with $\sin(x)/x$.

Example *The evaluation of*

$$\lim_{x\to 0}\frac{\sin(x)}{x} \tag{2.99}$$

We have already noted above that

$$\lim_{x\to 0}\sin(x) = 0 \text{ and } \lim_{x\to 0} x = 0 \tag{2.100}$$

so L'Hôpital's rule gives us the result that

$$\begin{aligned}\lim_{x\to 0}\frac{\sin(x)}{x} &= \lim_{x\to 0}\frac{\cos(x)}{1}, \quad \text{using} \begin{cases}\sin'(x) = \cos(x)\\ \frac{dx}{dx} = 1\end{cases}\\ &= \cos(0)\\ &= 1, \quad \text{since } \cos(0) = 1\end{aligned} \tag{2.101}$$

So we confirm the result first encountered in 2.39 that

$$\lim_{x\to 0} \frac{\sin(x)}{x} = 1 \tag{2.102}$$

and we note that we no longer had to encounter the indeterminate quantity 0/0. We move on.

Example *The limit*

$$\lim_{x\to 1} \frac{x^3 + x - 2}{x^2 - 3x + 2} \tag{2.103}$$

We see on examining numerator and denominator that they both vanish at the limiting point $x = 1$; this means that this is a job for L'Hôpital's rule. Hence we can immediately say that

$$\begin{aligned} \lim_{x\to 1} \frac{x^3 + x - 2}{x^2 - 3x + 2} &= \lim_{x\to 1} \frac{3x^2 + 1}{2x - 3} \\ &= \frac{4}{-1} \end{aligned} \tag{2.104}$$

In other words L'Hôpital's rule has saved the day and we have found that

$$\lim_{x\to 1} \frac{x^3 + x - 2}{x^2 - 3x + 2} = -4 \tag{2.105}$$

We provide one more example.

Example *The limit*

$$\lim_{x\to 2} \frac{2 - x}{4 - x^2} \tag{2.106}$$

First we note that both numerator and denominator vanish at the limit point $x = 2$ and so, computing their derivatives, we conclude that

$$\begin{aligned} \lim_{x\to 2} \frac{2 - x}{4 - x^2} &= \lim_{x\to 2} \frac{-1}{(-2x)} \\ &= \frac{1}{4} \end{aligned} \tag{2.107}$$

and so we have our result, which is

$$\lim_{x\to 2} \frac{2 - x}{4 - x^2} = \frac{1}{4} \tag{2.108}$$

§ 4. Taylor series

We now come to a remarkable result which enables many functions to be expressed as convergent series in powers of x. These series are called *Taylor series*. Let us first quote the result and then give some examples.

If $f(x)$ is a function which satisfies certain appropriate conditions (to be elaborated on later) then $f(x)$ can be expanded in a so called *power series* given by

$$\begin{aligned} f(x) &= f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots \\ &= \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}x^n \end{aligned} \tag{2.109}$$

We have already come across an instance of a Taylor series without any comment to that effect. Let us now give the details.

Example *The exponential series*

Recall the function e^x for which we know has the properties

$$\frac{de^x}{dx} = e^x \text{ and } e^0 = 1 \tag{2.110}$$

then we can say that

$$\frac{d^n e^x}{dx^n} = e^x, \quad \text{for } n = 1, 2, 3, \ldots \tag{2.111}$$

and, setting $x = 0$, we can further say that

$$\left.\frac{d^n e^x}{dx^n}\right|_{x=0} = 1, \quad \text{for } n = 1, 2, 3, \ldots \tag{2.112}$$

Hence if we set $f = e^x$ in 2.109 above, and use these results, we find that 2.109 simply becomes the statement that

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \tag{2.113}$$

which we recognise as being the same series as we had in 2.51. So the Taylor series for e^x has just reproduced the series for e^x that we knew already.

Example *The binomial series*

Another example of a series which the reader may not have realised is a Taylor series is that given by the binomial expansion of $(1 + x)^\alpha$ where α is *not necessarily an integer.* The binomial expansion says that [11]

$$(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!}x^2 + \frac{\alpha(\alpha-1)(\alpha-2)}{3!}x^3 + \cdots \tag{2.114}$$

Now to use our formula 2.109 above we must compute

$$\frac{d^n(1+x)^\alpha}{dx^n}, \quad \text{for } n = 1, 2, 3, \ldots \tag{2.115}$$

This is easy and we find that

$$\begin{aligned} \frac{d(1+x)^\alpha}{dx} &= \alpha(1+x)^{\alpha-1} \\ \frac{d^2(1+x)^\alpha}{dx^2} &= \alpha(\alpha-1)(1+x)^{\alpha-2} \\ \vdots \quad & \quad \vdots \\ \frac{d^n(1+x)^\alpha}{dx^n} &= \alpha(\alpha-1)\cdots(\alpha-n+1)(1+x)^{\alpha-n} \end{aligned} \tag{2.116}$$

Hence if we evaluate both sides at $x = 0$ we find that

$$\left.\frac{d^n(1+x)^\alpha}{dx^n}\right|_{x=0} = \alpha(\alpha-1)\cdots(\alpha-n+1) \tag{2.117}$$

Finally if we insert this information in the Taylor series formula 2.109 above we find that it gives us

$$(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!}x^2 + \frac{\alpha(\alpha-1)(\alpha-2)}{3!}x^3 + \cdots \tag{2.118}$$

which is precisely the binomial expansion 2.114 above.

Next we need to be a bit more general because actually most suitable functions $f(x)$ can be expanded, not just in powers of x, but in powers of the quantity $(x-a)$ where a is any number. The result that describes the complete state of affairs is known as *Taylor's theorem* and we now quote it (again without proof).

[11] Note very carefully that if α **is** an integer, say $\alpha = n$, then this expansion will stop after $n+1$ terms; otherwise it does not terminate. The reader who for whom this fact is new should choose two values for α such as $\alpha = 3$ and $\alpha = 1/2$ and check that in the first case the series terminates after 4 terms but in the second case it goes on forever.

Theorem (Taylor's theorem) *Suppose that $f(x)$ is a function whose derivatives $f', f'', \ldots$ all exist then*

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \cdots \quad (2.119)$$

provided the series above converges and a certain remainder term $R_{n,a}(x) \to 0$ as $n \to \infty$.

The first thing the reader [12] should do is to notice that, if we set $a = 0$, then the formula 2.119 of Taylor's theorem reverts to the formula 2.109 that we had above.

A piece of *terminology* can be got out of the way here: this formula 2.109

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots \quad (2.120)$$

is often called the *Maclaurin series* for f. In other words a Maclaurin series is the special case of a Taylor series that results if $a = 0$. We shall just use the term Taylor series for all such series as this is the more common practice.

Another linguistic point is that the Taylor expansion 2.119 is often referred to as "the Taylor expansion of f about $x = a$".

We need a few more examples to finish off with.

Example *The Taylor series for* $\sin(x)$

If we set $a = 0$ then Taylor's theorem tells us that

$$\sin(x) = \sin(0) + \sin'(0)x + \frac{\sin''(0)}{2!}x^2 + \cdots \quad (2.121)$$

Computing the derivatives that we need gives

$$\begin{aligned} \sin'(x) = \cos(x) &\Rightarrow \sin'(0) = \cos(0) = 1 \\ \Rightarrow \sin''(x) = -\sin(x) &\Rightarrow \sin''(0) = -\sin(0) = 0 \\ \Rightarrow \sin'''(x) = -\cos(x) &\Rightarrow \sin'''(0) = -\cos(0) = -1 \\ \Rightarrow \sin''''(x) = \sin(x) &\Rightarrow \sin''''(0) = \sin(0) = 0 \end{aligned} \quad (2.122)$$

[12] The reader should not worry about this remainder term or about the convergence of the series and the proof of the theorem. We shall always assume that the series converges and that the remainder term satisfies $R_{n,a}(x) \to 0$ as $n \to \infty$ and in our examples these properties will always hold. For those who may be interested the formula for $R_{n,a}(x)$ is $R_{n,a}(x) = \int_a^x \frac{f^{(n+1)}(t)}{n!}(x-t)^n\,dt$; we note that it involves an integral—an object we do not meet officially until the next chapter.

But this pattern of 4 derivatives now repeats itself ad infinitum so that we know that the next 4 are given by

$$\begin{aligned}\sin^{(5)}(0) &= 1\\ \sin^{(6)}(0) &= 0\\ \sin^{(7)}(0) &= -1\\ \sin^{(8)}(0) &= 0\end{aligned} \tag{2.123}$$

and so on. Hence we have established that

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots \tag{2.124}$$

Example *The Taylor series for* $\cos(x)$

A similar but not quite identical calculation will give the following series for $\cos(x)$

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots \tag{2.125}$$

These series expansion for $\sin(x)$ and $\cos(x)$ can often prove extremely useful and are well worth memorising.

Example *Two expansions for* $(1+x)^{-1}$

We have already expanded $(1+x)^{\alpha}$. Now we set $\alpha = -1$ and consider the function $(1+x)^{-1}$. If we use 2.116 we obtain

$$\begin{aligned} \frac{d}{dx}\left(\frac{1}{1+x}\right) &= \frac{-1}{(1+x)^2}\\ \Rightarrow \frac{d^2}{dx^2}\left(\frac{1}{1+x}\right) &= \frac{(-1)^2\cdot 2}{(1+x)^3}\\ \Rightarrow \frac{d^3}{dx^3}\left(\frac{1}{1+x}\right) &= \frac{(-1)^3\cdot 2\cdot 3}{(1+x)^4}\\ \vdots \quad & \quad \vdots\\ \Rightarrow \frac{d^n}{dx^n}\left(\frac{1}{1+x}\right) &= \frac{(-1)^n n!}{(1+x)^{n+1}} \end{aligned} \tag{2.126}$$

Now setting $x = a$ gives

$$\left.\frac{d^n}{dx^n}\left(\frac{1}{1+x}\right)\right|_{x=a} = \frac{(-1)^n n!}{(1+a)^{n+1}} \tag{2.127}$$

and using this in 2.119 gives us the result that

$$\frac{1}{1+x} = \frac{1}{1+a} - \frac{(x-a)}{(1+a)^2} + \frac{(x-a)^2}{(1+a)^3} + \cdots + (-1)^n \frac{(x-a)^n}{(1+a)^{n+1}} + \cdots \tag{2.128}$$

Quite an interesting thing now happens and it is this: the ratio test will easily show that this series converges if

$$\begin{aligned} \left|\frac{x-a}{1+a}\right| &< 1 \\ \text{i.e.} \quad \text{if} \quad |x-a| &< |1+a| \end{aligned} \tag{2.129}$$

so by choosing *two different values of* a we can get *two different series* for $(1+x)^{-1}$ with *two different* ranges of x for which they converge.

To be specific let us choose the two values of a to be 0 and 1/2; we then get the two series

$$\begin{aligned} \frac{1}{1+x} &= 1 - x + x^2 - x^3 + \cdots \qquad \text{(convergent for } |x| < 1) \\ \frac{1}{1+x} &= \frac{1}{(3/2)} - \frac{(x-1/2)}{(3/2)^2} + \frac{(x-1/2)^2}{(3/2)^3} - \cdots \qquad \text{(convergent for } |x-1/2| < 3/2) \end{aligned} \tag{2.130}$$

So now notice that if

$$x = \frac{3}{2} \tag{2.131}$$

the first series *diverges* because $|x| > 1$ if $x = 3/2$—and so is useless—but the second series is still *convergent* because $|x - 1/2| < 3/2$ if $x = 3/2$.

Another interesting fact is that if we choose $x = 0.4$ then *both* series are convergent so does it matter which one we use. This is a practical matter which is of interest to an engineering readership; the answer is that it *does matter* if you want the series chosen to converge quickly so that you don't have to add up too many terms in order to get reasonable accuracy. The series chosen, if $x = 0.4$, should be the second one in 2.130 because it is an expansion in powers of the parameter [13]

$$x - \frac{1}{2} = 0.1, \quad \text{if } x = 0.4 \tag{2.132}$$

whereas the other series is an expansion in powers of the parameter

$$x = 0.4 \tag{2.133}$$

[13] Actually, if one looks more carefully, one sees that the expansion parameter is $(x - 1/2)/(3/2)$ which is even smaller than 0.1 when $x = 0.4$ since it has the value $0.0\dot{6}$; but this just makes things even better.

and a series in powers of a parameter will usually converge fastest when the parameter is smallest.

If the reader has access to computer mathematics packages such as *Maple*, *Mathematica* or *Matlab* all these divergence and convergence properties—as well as convergence rates—can be seen emerging numerically on the computer. Here are the results of doing such a calculation using the package *Maple*.

First of all note that if we work to 10 places of decimals then

$$\frac{1}{1+x} = 0.7142857143, \qquad \text{when } x = 0.4 \tag{2.134}$$

We shall now sum the first 11 terms of each of the two series above and then set $x = 0.4$. We find that

$$\begin{aligned}
\sum_{n=0}^{10}(-x)^n &= 1 - x + x^2 - \cdots + x^{10} \\
&= 0.7143156736, \qquad \text{when } x = 0.4 \\
\sum_{n=0}^{10}(-1)^n \frac{(x-1/2)^n}{(3/2)^{n+1}} &= \frac{1}{(3/2)} - \frac{(x-1/2)}{(3/2)^2} + \frac{(x-1/2)^2}{(3/2)^3} - \cdots + \frac{(x-1/2)^{10}}{(3/2)^{11}} \\
&= 0.7142857140, \qquad \text{when } x = 0.4
\end{aligned} \tag{2.135}$$

Now we readily see the difference in accuracy of the two series: the first one differs from the actual value quoted in eq. 2.134 in the *fourth* decimal place; but the second one does not differ until the *tenth* decimal place—a considerable improvement.

Next we have two topics which, though not calculus applications, are of universal use in many mathematical calculations. These are *polar coordinates* and *complex numbers*.

§ 5. Plane polar coordinates

The usual Cartesian coordinates (x, y) of a point in the plane are not the only coordinates one can use. In calculations with some circular symmetry it is often convenient to introduce what are called *plane polar coordinates* or simply *polar coordinates*.

Polar coordinates consist of a *distance* r and an angle θ and are denoted by

$$(r, \theta) \tag{2.136}$$

Fig. 24 illustrates how the polar coordinates of a point are obtained.

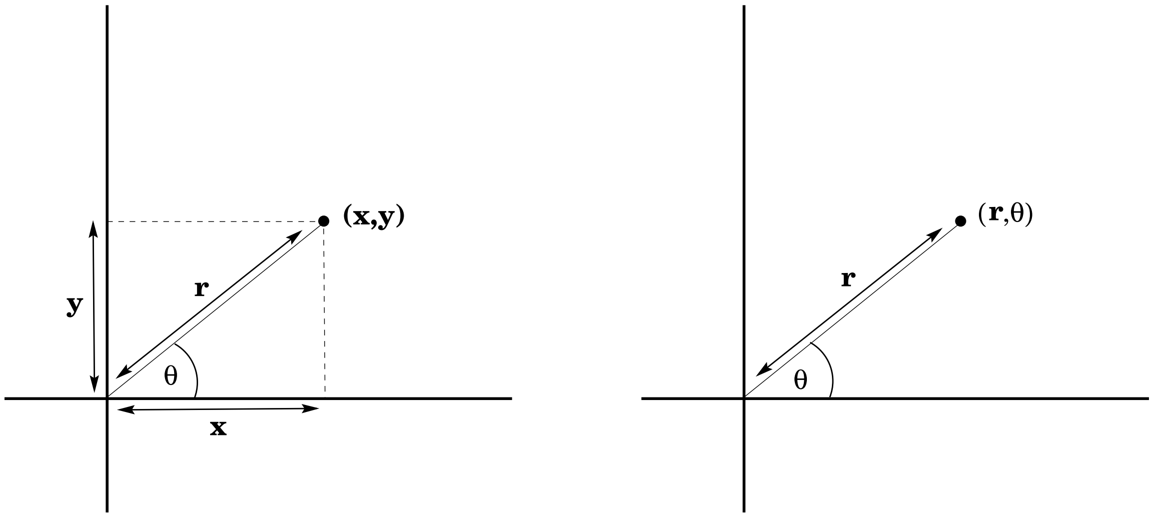


Fig. 24: The Cartesian and polar coordinates (r, θ) of a point

If we use fig. 24 we can see that the Cartesian coordinates x and y are related to r and θ by the pair of equations

$$\begin{aligned} x &= r\cos(\theta) \\ y &= r\sin(\theta) \end{aligned} \tag{2.137}$$

We can also see that $x^2 + y^2 = r^2$ and $\tan(\theta) = y/x$; from which we deduce that

$$\begin{aligned} r &= \sqrt{x^2 + y^2} \\ \theta &= \arctan\left(\frac{y}{x}\right) \end{aligned} \tag{2.138}$$

With this last pair of equations we can start with Cartesian coordinates (x, y) and compute their corresponding polar coordinates (r, θ) or start with polar coordinates (r, θ) and convert them into Cartesian coordinates (x, y).

§§ 5.1 Some old and new equations expressed in polar coordinates

Example *A circle*

Let us consider the Cartesian equation of a circle of radius a which is

$$x^2 + y^2 = a^2 \tag{2.139}$$

But this is the same as

$$r^2 = a^2 \tag{2.140}$$

which we might as well simplify to

$$r = a \tag{2.141}$$

which is the polar form of the equation of a circle and we see that it is simpler than the Cartesian form. This will not always be the case however—the circle is a special case.

Example *A parabola*

Recall that a parabola has the equation

$$y = Ax^2, \qquad A \text{ a constant} \tag{2.142}$$

so this becomes

$$\begin{aligned} r\sin(\theta) &= Ar^2\cos^2(\theta) \\ \Rightarrow r &= \frac{\sin(\theta)}{A\cos^2(\theta)}, \quad (r \geq 0) \end{aligned} \tag{2.143}$$

Notice that this equation has become more complicated when expressed in polar coordinates.

Example *The four leaved rose*

Now purely for fun we show the plot of a *four leaved rose* which comes from plotting a very simple function when expressed in polar coordinates. One just plots

$$r = |\cos(2\theta)| \tag{2.144}$$

and the result is shown in fig. 25, note that we have modulus signs round $\cos(2\theta)$ to make sure that r cannot go negative [14]

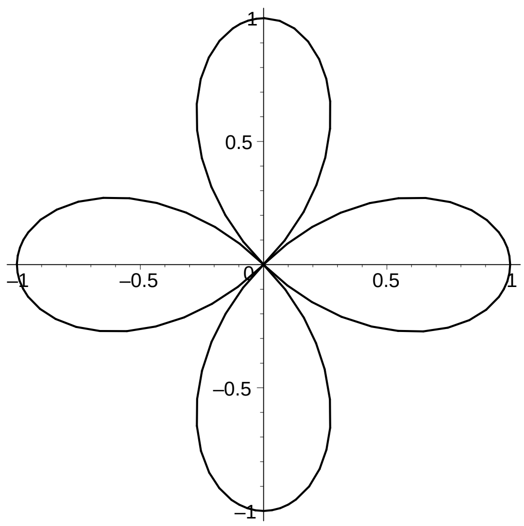


Fig. 25: A four leaved rose obtained by plotting $r = |\cos(2\theta)|$

[14] Negative r is sometimes allowed: the coordinates (r, θ) with $r < 0$ are then interpreted to mean a point with *positive* r and θ incremented by π (this incrementation reflects a point in the origin and can be derived by going back to Cartesian coordinates), i.e. to mean the point with coordinates $(-r, \theta + \pi)$ for $r < 0$. Please do not worry about this.

§ 6. Complex numbers

The reader is presumed to already have some familiarity with complex numbers so we shall only give a brief discussion of their origins and basic properties.

The first number to deal with is $\sqrt{-1}$ which is not a real number. The need to discuss $\sqrt{-1}$ arises when one tries to solve the very simple quadratic equation

$$x^2 + 1 = 0 \tag{2.145}$$

which seems to demand that we write its solution as

$$x = \mp\sqrt{-1} \tag{2.146}$$

One denotes $\sqrt{-1}$ by i so that [15]

$$\begin{aligned} i &= \sqrt{-1} \\ \Rightarrow i^2 &= -1 \end{aligned} \tag{2.147}$$

and if we include i in our calculations we can solve other quadratic equations such as, say,

$$x^2 + 9 = 0 \tag{2.148}$$

whose solutions are

$$x = \mp 3i \tag{2.149}$$

This means that we must allow $\mp 3i$ as possible numbers; more generally we allow now any number of the form

$$a + bi, \quad a, b \in \mathbf{R} \tag{2.150}$$

By the way one can write $a + bi$ or $a + ib$ or $ib + a$ etc. and they all denote the *same* complex number.

The set of all such numbers

$$a + bi, \quad a, b \in \mathbf{R} \tag{2.151}$$

is called the set of *complex numbers* which we denote by

$$\mathbf{C} \tag{2.152}$$

to distinguish it from the set $\mathbf{R}$ of real numbers.

[15] The reader who is an electrical engineer will find that, in the engineering literature, $\sqrt{-1}$ is frequently denoted by

$$j$$

instead of i. This is because i is already in such widespread use for electric current and it is felt that too much confusion would result if i was also used for $\sqrt{-1}$.

It turns out that *all* polynomial equations such as

$$x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots a_1 x + a_0 = 0 \tag{2.153}$$

where the a_i are real or complex constants, can always be solved by numbers x in $\mathbf{C}$. Hence no more new numbers need to be adjoined to $\mathbf{R}$ once we have enlarged the real numbers $\mathbf{R}$ to the complex numbers $\mathbf{C}$.

Complex numbers are often denoted by z so we may write

$$z = x + iy, \quad (x, y \in \mathbf{R}) \tag{2.154}$$

When we do this x is called the real part of z—and denoted by $Re\, z$— and y is called the imaginary part of z—and denoted by $Im\, z$—summarising the notation is

$$z = x + iy, \quad x = Re\, z,\ y = Im\, z \tag{2.155}$$

The *complex conjugate* of a complex number z is denoted by $\bar{z}$ and is defined by

$$\bar{z} = x - iy \tag{2.156}$$

The *absolute value* or *modulus* of z is denoted by $|z|$ and is defined by

$$|z| = \sqrt{x^2 + y^2} \tag{2.157}$$

It can be quite useful to notice that

$$z\bar{z} = |z|^2 \tag{2.158}$$

§§ 6.1 Complex numbers and polar coordinates

Let θ be the angle of polar coordinates and consider next the function

$$e^{i\theta} \tag{2.159}$$

Now if we use the usual series expansion 2.51 for e^x with $x = i\theta$ we find that

$$\begin{aligned} e^{i\theta} &= 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \frac{(i\theta)^5}{5!} + \cdots \\ &= 1 + i\theta - \frac{\theta^2}{2!} - \frac{i\theta^3}{3!} + \frac{\theta^4}{4!} + \frac{i\theta^5}{5!} + \cdots \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots\right) + i\left(\theta - \frac{\theta^3}{3!} + \frac{i\theta^5}{5!} - \cdots\right) \end{aligned} \tag{2.160}$$

But recall that when doing Taylor's theorem we found that (cf. 2.124 and 2.125)

$$\begin{aligned} \cos\theta &= 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots \\ \sin\theta &= \theta - \frac{\theta^3}{3!} + \frac{i\theta^5}{5!} - \cdots \end{aligned} \tag{2.161}$$

Hence we have just shown the very important and useful result that

$$e^{i\theta} = \cos\theta + i\sin\theta \tag{2.162}$$

If we replace θ by $-\theta$ and remember that $\sin(-\theta) = -\sin(\theta)$ and that $\cos(-\theta) = \cos(\theta)$ we obtain

$$e^{-i\theta} = \cos\theta - i\sin\theta \tag{2.163}$$

Now if we successively add and subtract these two formulae for $e^{i\theta}$ and $e^{-i\theta}$ we get two marvelous and very important formulae: one for $\cos\theta$ and one for $\sin\theta$. These are

$$\begin{aligned} \cos\theta &= \frac{e^{i\theta} + e^{-i\theta}}{2} \\ \sin\theta &= \frac{e^{i\theta} - e^{-i\theta}}{2i} \end{aligned} \tag{2.164}$$

We can now apply some of this information to obtain a very nice formula for complex numbers in polar coordinates. We start with

$$z = x + iy \tag{2.165}$$

and then use the polar coordinate information that $x = r\cos\theta$ and $y = r\sin\theta$ giving

$$\begin{aligned} z &= r\cos\theta + ir\sin\theta \\ &= r(\cos\theta + i\sin\theta) \\ &= re^{i\theta}, \quad \text{using 2.162} \end{aligned} \tag{2.166}$$

Hence the form of a complex number $z = x + iy$ in polar coordinates is

$$z = re^{i\theta} \tag{2.167}$$

a result of considerable use and importance. Note that the angle θ is referred to as the *argument* of z and is denoted by $\arg(z)$, i.e.

$$\text{when } z = re^{i\theta} \quad \text{then } \arg(z) = \theta \quad \text{and } |z| = r \tag{2.168}$$

Two more useful properties of complex numbers are

$$
\begin{aligned}
|zw| &= |z||w| \\
\arg(zw) &= \arg(z) + \arg(w)
\end{aligned}
\tag{2.169}
$$

The reader will find these easy to verify if polar form is used for z and w—i.e. write $z = r\exp[i\theta]$, $w = \rho\exp[i\theta]$ and substitute into 2.169 above.

§ 7. Some common differential equations

A differential equation is any equation for an unknown function f which involves at least one derivative of f for example

$$\frac{df}{dx} - kf = 0, \quad k \text{ a constant} \tag{2.170}$$

is a differential equation for f.

If a particle undergoes *simple harmonic motion* or SHM its displacement $x(t)$ at time t obeys the differential equation

$$\frac{d^2x}{dt^2} + \omega^2 x = 0 \tag{2.171}$$

where ω is the frequency of the oscillations.

Actually equation 2.170 occurs in the study or radioactive decay and also in population growth. It is worth having a look at this.

Example *Radioactive decay and population growth*

Consider a sample of radioactive material with $N(t)$ atoms at time t. Radioactive decay causes $N(t)$ to *decrease* as time goes on, hence

$$\frac{dN(t)}{dt} < 0 \tag{2.172}$$

Now *experimentally* it is found that the number of atoms decaying per unit time is doubled if the size of the sample—i.e .the number of atoms in it—is doubled. In other words the *decay rate*—which is just dN/dt—is proportional to the *number of atoms* N. This is simply the statement that

$$\frac{dN(t)}{dt} = -kN(t), \quad k \text{ a positive constant} \tag{2.173}$$

so the proportionality constant is $-k$ and we see that the minus sign is there to keep dN/dt negative. This constant k is called the *decay constant* of the radioactive element.

Now we must *solve* this differential equation. To do this we proceed informally as follows: we simply observe that if a function $f(t)$ is given by

$$f(t) = Ce^{-kt}, \qquad \text{where } C \text{ is any constant} \tag{2.174}$$

then, if we differentiate $f(t)$, we find that

$$\frac{df(t)}{dt} = -kf(t) \tag{2.175}$$

i.e. $f(t$ is a solution [16] to our differential equation and we now set

$$\begin{aligned} N(t) &= f(t) \\ \text{i. e.} \qquad N(t) &= Ce^{-kt} \end{aligned} \tag{2.177}$$

So now we know what $N(t)$ looks like; notice that if $t = 0$ then $N(t) = C$ so the constant C is the number of atoms present at the beginning $t = 0$, let us therefore *rename* C by writing

$$C = N_0$$

so that we have

$$N(t) = N_0 e^{-kt} \tag{2.178}$$

and N_0 is the number of atoms present at $t = 0$.

The well known number called the *half life* of the radioactive substance can now be calculated. The half life is the time taken for exactly half the atoms to decay. If this time is denoted by T then we have

$$N(T) = \frac{N_0}{2} \tag{2.179}$$

[16] Actually it is easy to *derive* this solution if we use just a little integration. We shall meet integration in the next chapter, but for those of you who already know some integration here is the proof:

$$\begin{aligned} \frac{dN(t)}{dt} &= -kN(t) \\ \Rightarrow \frac{dN}{N} &= -k \\ \Rightarrow \int \frac{dN}{N} &= -k \int dt \\ \Rightarrow \ln(N) &= -kt + c, \quad c \text{ a constant of integration} \\ \Rightarrow N &= e^{-kt+c} \\ \Rightarrow N &= e^{-kt} e^{c} \\ \Rightarrow N(t) &= N_0 e^{-kt}, \quad \text{where } N_0 = e^{c} \end{aligned} \tag{2.176}$$

But using our formula $N(t) = N_0 e^{-kt}$ this means that

$$\begin{aligned} N_0 e^{-kT} &= \frac{N_0}{2} \\ \Rightarrow e^{-kT} &= \frac{1}{2} \\ \Rightarrow \ln(e^{-kT}) &= \ln(1/2) = -\ln(2) \\ \Rightarrow -kT &= -\ln(2) \\ \Rightarrow T &= \frac{\ln(2)}{k} \end{aligned} \tag{2.180}$$

and we have our formula for the half life T in terms of the decay constant k.

Now for Uranium 235, Carbon 14 and Iodine 120 the decay constants k have the values

$$0.9845840634 \times 10^{-9}\, y^{-1}, \quad 1.212 \times 10^{-4}\, y^{-1}, \quad 0.513\, h^{-1} \tag{2.181}$$

respectively where y stands for years and h for hours. This means that their respective half lives are

$$7.04 \times 10^8, \text{ years} \quad 5715, \text{ years}, \quad 1.35 \text{ hours} \tag{2.182}$$

We can also use our formula for $N(t)$ to work out how long it takes for a certain amount of the material to decay: suppose there is 45% of the radioactive material left then if the age of the sample is T', say we have

$$\begin{aligned} N(T') &= 0.45 N_0 \\ \Rightarrow N_0 e^{-kT'} &= 0.45 N_0 \\ \Rightarrow T' &= -\frac{\ln(0.45)}{k} \end{aligned} \tag{2.183}$$

Hence if we were dealing with Carbon 14 we would find that the age of the sample was

$$-\frac{\ln(0.45)}{1.212 \times 10^{-4}} = 6583.69\,\text{years} \tag{2.184}$$

We can also study population growth with this our equation

$$\frac{dN(t)}{dt} = -kN(t), \quad k \text{ a positive constant} \tag{2.185}$$

we just change the sign in front of the constant k and write

$$\frac{dN(t)}{dt} = kN(t), \quad k \text{ a positive constant} \tag{2.186}$$

So, now,

$$\frac{dN(t)}{dt} > 0 \tag{2.187}$$

and $N(t)$ *increases* with time. Then, depending on the size of k, we have a certain *doubling time* for the population instead of a *half life.* Note that this population may be one of people or bacteria or anything else which obeys our differential equation.

Finally we consider differential equations which describe the behaviour of some simple electrical circuits.

Example *Two electrical circuits and their differential equations: The discharging and charging of a capacitor*

Examine the two circuits shown in fig. 26.

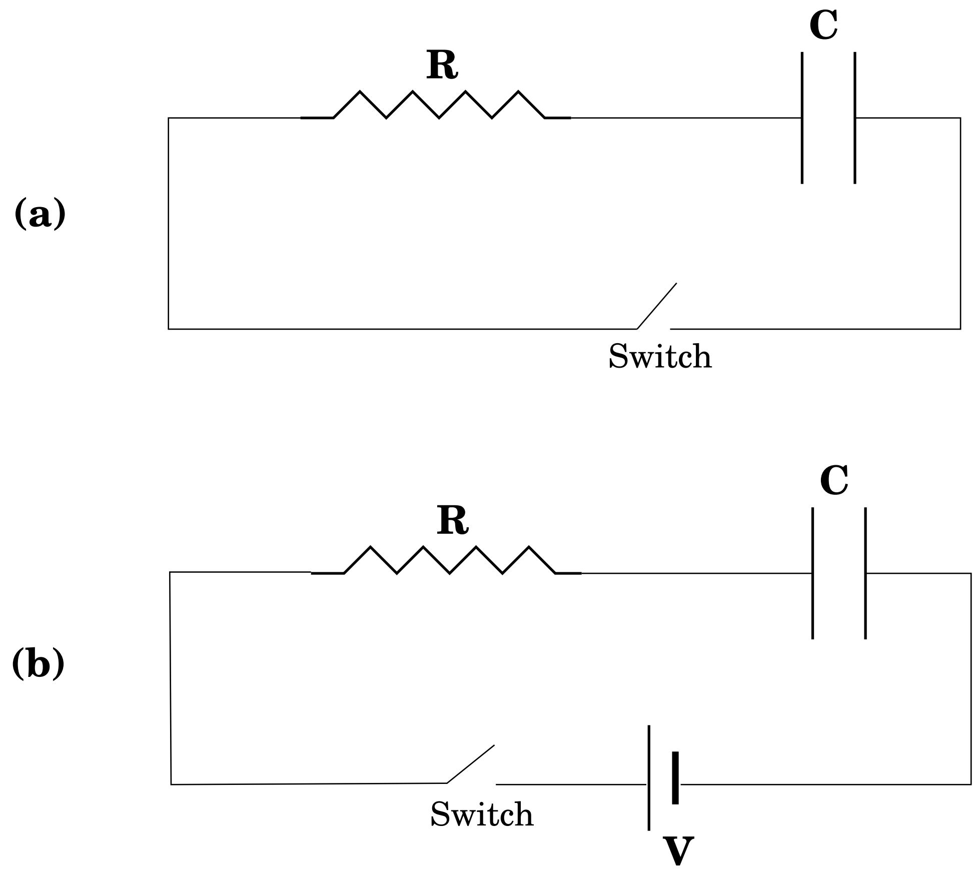


Fig. 26: The discharging and charging of a capacitor

Figure 26 (a) shows the circuit for a *discharging* capacitor while figure 26 (b) shows the circuit for a *charging* capacitor. We now examine each of these circuits in turn.

A discharging capacitor

Consider figure 26 (a) it shows a capacitor of capacitance C and a resistance R and an open switch. The resistance R represents the *necessarily non-zero* internal resistance of the capacitor, the switch and the wiring of the circuit.

We start with the *switch open* and the capacitor possessing a quantity of charge Q. Now we close the switch, thus making a closed circuit, and the capacitor begins to discharge: after t seconds the charge on the capacitor has decreased to some value $Q(t)$ and it is known that $Q(t)$ obeys the differential equation

$$R\frac{dQ(t)}{dt} + \frac{Q(t)}{C} = 0 \tag{2.188}$$

But we can rewrite this equation 2.188 as

$$\frac{dQ(t)}{dt} = -kQ(t), \qquad \text{where } k = \frac{1}{RC} \tag{2.189}$$

But we see that this is an instance of equation 2.175 above and so we immediately know that the solution to 2.189 is of the form

$$Q(t) = Ce^{-kt}, \qquad \text{where } k = \frac{1}{RC} \tag{2.190}$$

If we set $t = 0$ it easy to see that the constant C is equal to the charge on the capacitor at $t = 0$—i.e. the quantity Q—and so we have finally

$$Q(t) = Qe^{-\frac{t}{RC}} \tag{2.191}$$

but using the notation

$$\exp[x] = e^x \tag{2.192}$$

we display the solution more clearly as

$$Q(t) = Q\exp\left[-\frac{t}{RC}\right] \tag{2.193}$$

We end with figure 27 which shows the typical exponential decay shape of the graph of a discharging capacitor; the values of Q, R and C are 5, 2 and 3 respectively.

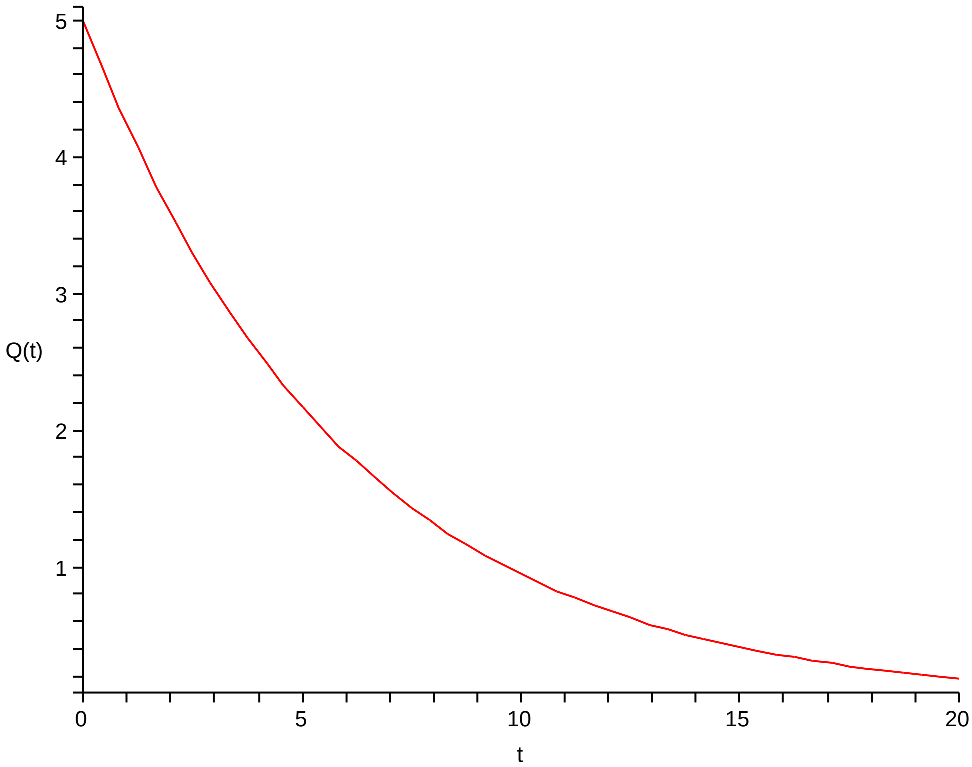


Fig. 27: A discharging capacitor

Now it is time to move on to the charging case.

A charging capacitor

Figure (b) shows the circuit relevant for a charging capacitor: one sees a capacitor of capacitance C, a battery supplying a *constant* voltage V and the usual resistance R representing the combined internal resistance (*necessarily non-zero*) of the battery, the capacitor and the wiring.

We suppose that we begin with a completely *discharged* capacitor: i.e. before the the switch in the circuit is closed the charge Q in the capacitor is *zero*. However, after the switch has been closed for t seconds, the charge $Q(t)$ in the capacitor grows to some non-zero value and obeys the differential equation

$$R\frac{dQ(t)}{dt} + \frac{Q(t)}{C} = V \tag{2.194}$$

To solve this differential equation 2.194 requires only a small tweaking of what we have done above, here is how it goes. First we note that

$$\begin{aligned} R\frac{dQ(t)}{dt} + \frac{Q(t)}{C} &= V \\ \Rightarrow \frac{dQ(t)}{dt} &= -\frac{1}{RC}Q(t) + \frac{V}{R} \\ \Rightarrow \frac{dQ(t)}{dt} &= -\frac{1}{RC}\left(Q(t) - CV\right) \end{aligned} \tag{2.195}$$

Now introduce the function $F(t)$ defined by

$$F(t) = Q(t) - CV \tag{2.196}$$

and notice that

$$\begin{aligned} \frac{dF(t)}{dt} &= \frac{d}{dt}\left(Q(t) - CV\right) \\ &= \frac{dQ(t)}{dt}, \qquad \text{since } CV \text{ is a constant} \end{aligned} \tag{2.197}$$

Next substitute 2.196 and 2.197 into the last line of 2.195 thereby obtaining

$$\frac{dF(t)}{dt} = -\frac{1}{RC}F(t) \tag{2.198}$$

which we recognise as 2.175 yet again and so the solution is

$$F(t) = F\exp\left[-\frac{t}{RC}\right] \tag{2.199}$$

where F is the value of $F(t)$ at $t = 0$; and one can check by substituting $t = 0$ into 2.196 above that this means that

$$F = -CV \tag{2.200}$$

But since

$$F(t) = Q(t) - CV \tag{2.201}$$

the final solution is given by

$$Q(t) - CV = F\exp\left[-\frac{t}{RC}\right], \qquad F = -CV \tag{2.202}$$

which, after a little tidying and rearranging, can be written as

$$Q(t) = CV\left(1 - \exp\left[-\frac{t}{RC}\right]\right) \tag{2.203}$$

The reader should use the formula 2.203 above to check that the *initial* charge on the capacitor is indeed 0: i.e. check that at $t = 0$ we do have $Q(t) = 0$; it is also of interest to note that

$$\lim_{t \to \infty} Q(t) = CV \tag{2.204}$$

In other words the *final* charge on the capacitor is CV. The graph of a charging capacitor is one we have already displayed in figure 28 but here is is again (it has been obtained from formula 2.203 by using the values $C = 2$, $V = 1/2$ and $R = 0.111$).

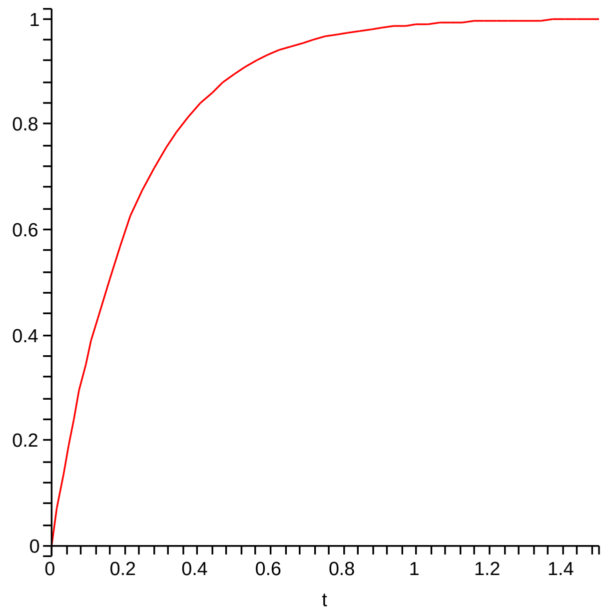


Fig. 28: A charging capacitor

Example *A second order differential equation*

Consider the differential equation

$$a\frac{d^2 f(x)}{dx^2} + b\frac{df(x)}{dx} + cf(x) = 0, \qquad a,\ b,\ c \text{ all constants} \tag{2.205}$$

This is called a *second order* differential equation because of the presence of the term $d^2 f/dx^2$—in general the order of any differential equation is called n when $d^n f/dx^n$ is the highest derivative appearing in it. We shall now learn how to solve this equation.

All one has to do is to substitute

$$f(x) = \exp[rx], \qquad r \text{ a constant} \tag{2.206}$$

into the differential equation 2.205. Doing this and demanding that the RHS be zero yields

$$\begin{aligned} ar^2 \exp[rx] + br\exp[rx] + c\exp[rx] &= 0 \\ \Rightarrow (ar^2 + br + c)\exp[rx] &= 0 \\ \Rightarrow (ar^2 + br + c) &= 0 \end{aligned} \tag{2.207}$$

Thus we have a solution of the form

$$f(x) = \exp[rx] \tag{2.208}$$

if r satisfies the quadratic equation

$$ar^2 + br + c = 0 \tag{2.209}$$

We immediately know that there are two solutions for r given by the standard formula

$$r = \frac{-b \mp \sqrt{b^2 - 4ac}}{2a} \tag{2.210}$$

For convenience let us denote these two solutions by r_+ and r_- where

$$r_+ = \frac{-b + \sqrt{b^2 - 4ac}}{2a}, \qquad r_- = \frac{-b - \sqrt{b^2 - 4ac}}{2a} \tag{2.211}$$

Hence we have two solutions to our differential equation and these are

$$f)x) = \exp[r_+ x] \quad \text{and} \quad f(x) = \exp[r_- x] \tag{2.212}$$

In fact the *general solution* to our differential equation is got by talking a linear combination of these two solutions by which we mean that *all solutions* to

$$a\frac{d^2 f(x)}{dx^2} + b\frac{df(x)}{dx} + cf(x) = 0, \qquad a,\ b,\ c \text{ all constants}$$

are given by

$$f(x) = A\exp[r_+ x] + B\exp[r_- x], \qquad \begin{cases} A \text{ and } B \text{ constants} \\ r_\mp = \dfrac{-b \mp \sqrt{b^2 - 4ac}}{2a} \end{cases} \tag{2.213}$$

All that one has to do to find the solution in a particular case is to give the constants A and B the right values.

Here is a concrete worked example

Example *The solutions to the differential equation*

$$\frac{d^2 f(x)}{dx^2} - 7\frac{df(x)}{dx} + 12 f(x) = 0 \tag{2.214}$$

We see that

$$a = 1, \quad b = -7, \quad c = 12 \tag{2.215}$$

and so 2.213 tells us that

$$\begin{aligned} r_{\mp} &== \frac{7 \mp \sqrt{7^2 - 4 \cdot 12}}{2} \\ \Rightarrow r_{+} &= 4, \quad r_{-} = 3 \end{aligned} \tag{2.216}$$

and so we have the two independent solutions

$$f(x) = \exp[4x] \qquad \text{and } f(x) = \exp[3x] \tag{2.217}$$

and the general solution

$$f(x) = A\exp[4x] + B\exp[3x], \qquad A \text{ and } B \text{ any two constants} \tag{2.218}$$

Our next concrete example is the differential equation for simple harmonic motion or SHM for short.

Example *Simple harmonic motion or SHM*

The differential equation for a quantity y undergoing simple harmonic motion is

$$\frac{d^2 y(t)}{dt^2} + \omega^2 y = 0 \tag{2.219}$$

Indeed we have already quoted this equation in 2.171 above. Note that the notation for unknown function of the differential equation has changed from $f(x)$ to $y(t)$, this is *only a notational change* and the reader must get used to such changes.

In any case, using our solution formula 2.213 above, we see that

$$a = 1, \quad b = 0, \quad c = \omega^2 \tag{2.220}$$

which gives the result that

$$\begin{aligned} r_{\mp} &= \mp\sqrt{-\omega^2} \\ \Rightarrow r_{\mp} &= \mp i\omega \\ \Rightarrow y(t) &= A\exp[i\omega t] + B\exp[-i\omega t] \end{aligned} \tag{2.221}$$

But now we utilise eq. 2.162 which says that

$$\exp[i\theta] = \cos(\theta) + i\sin(\theta) \tag{2.222}$$

and, on using this fact in 2.221 we find that

$$\begin{aligned} y(t) &= A\left(\cos(\omega t) + i\sin(\omega t) + B\left(\cos(\omega t) - i\sin(\omega t)\right.\right. \\ &= (A+B)\cos(\omega t) + i(A-B)\sin(\omega t) \\ \Rightarrow y(t) &= C\cos(\omega t) + D\sin(\omega t), \qquad \text{where } C = A+B, \quad D = i(A-B) \end{aligned} \tag{2.223}$$

Hence we see that the two independent solutions to the SHM differential equation are

$$\cos(\omega t) \qquad \text{and} \qquad \sin(\omega t) \tag{2.224}$$

and these functions have the well known oscillatory behaviour illustrated in fig. 29 below.

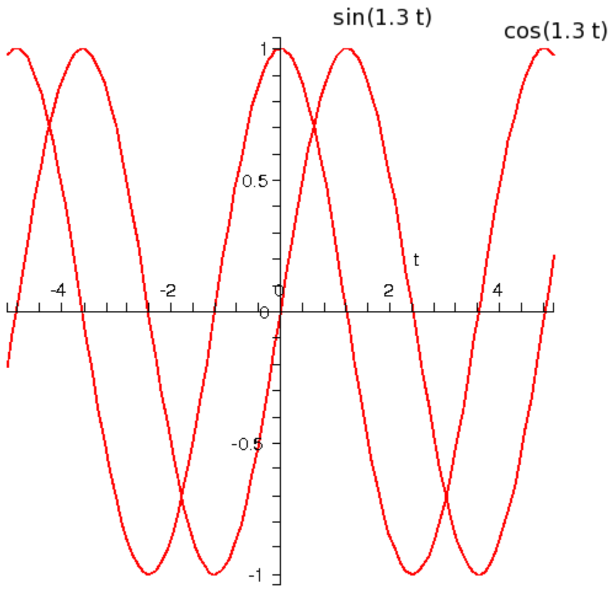


Fig. 29: The functions $\cos(\omega t)$ and $\sin(\omega t)$ $(\omega = 1.3)$

We end our section on differential equations with the differential equation for an electric circuit which possesses the fundamental trio of electrical properties which are resistance, capacitance and induction.

Example *Another electrical circuit and its differential equation: An LRC circuit*

The charge $Q(t)$ at time t on the capacitor for the circuit shown in fig. 30 obeys the differential equation

$$L\frac{d^2Q(t)}{dt^2} + R\frac{dQ(t)}{dt} + \frac{Q(t)}{C} = V(t) \tag{2.225}$$

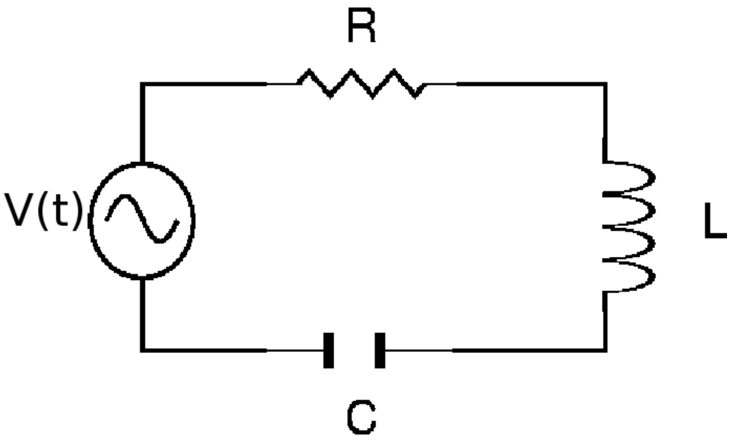


Fig. 30: An LRC circuit

$$L\frac{d^2Q(t)}{dt^2} + R\frac{dQ(t)}{dt} + \frac{Q(t)}{C} = V(t) \tag{2.226}$$

where L, R and C denote inductance, resistance and capacitance respectively.

The term $V(t)$ is called a *forcing term* and represents an applied voltage so let us take $V(t)$ to be an oscillatory, or AC, voltage given by

$$V(t) = V_0\cos(\omega t) \tag{2.227}$$

First of all if

$$V_0 = 0 \tag{2.228}$$

so that there is *no applied voltage* then the nature of the solution $Q(t)$ is controlled by the sign of the parameter [17]

$$R^2 - \frac{4L}{C} \tag{2.229}$$

If we use the data in 2.213 we can verify that the possible behaviours are shown in fig. 31.

[17] The interested student may like to check that since $a = L$, $b = R$ and $c = 1/C$ then

$$r_\mp = \frac{-R \mp \sqrt{R^2 - \frac{4L}{C}}}{2L} = -\frac{R}{2L} \mp \frac{1}{2L}\sqrt{R^2 - \frac{4L}{C}}$$

and it is just when $R^2 - \frac{4L}{C}$ becomes *negative* that the term $\sqrt{R^2 - \frac{4L}{C}}$ becomes pure imaginary: this is what gives an oscillatory part to the solution via the cosines and sines in the formula $\exp[i\theta] = \cos(\theta) + i\sin(\theta)$ above—indeed this happened in the SHM equation which corresponds here to the special case where $R = 0$ and $\omega^2 = 1/(LC)$.

Fig. 31: $Q(t)$ for $V_0 = 0$ with $R^2 - \frac{4L}{C} < 0$ on the left and $R^2 - \frac{4L}{C} > 0$ on the right

We see that for $R^2 - \frac{4L}{C} < 0$ we have a *damped oscillation* while for $R^2 - \frac{4L}{C} > 0$ we just have a decaying solution; if $R^2 - \frac{4L}{C} = 0$ the solution may grow a little at first but then also decays as time progresses.

Now we require $V_0 \neq 0$ so that the applied voltage *is present.*[18] In this case a principle property of the solution is that it can exhibit what is called *resonance*. This means an enhancement of the size of the solution when a certain parameter has a specified value cf. fig. 32

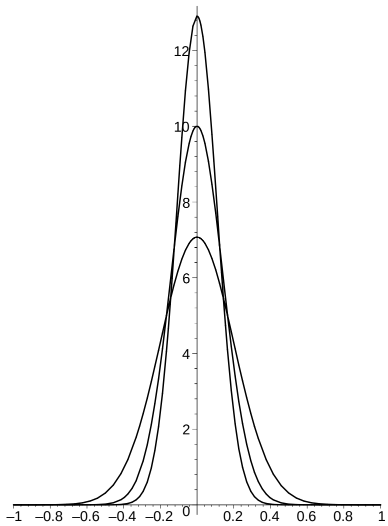


Fig. 32: Resonance: $Q(t)$ is more and more peaked as ω approaches $1/\sqrt{LC}$

Fig. 32 shows several graphs of $Q(t)$ each with a peak and a different value of ω: the more pronounced the peak the closer the driving frequency ω is to the value $1/\sqrt{LC}$. This value

[18] We have not shown how to derive a solution for such an equation with a *varying non zero* RHS so the reader must just accept what follows below without proof.

$1/\sqrt{LC}$ is called the *natural frequency*[19] of the circuit and the phenomenon of enhancement as $\omega \to 1/\sqrt{LC}$ is called *resonance*. The circuit designer usually chooses to adjust the value of the natural frequency so as to deliberately enhance or reject the applied voltage depending on his or her needs.

[19] This is the frequency that the circuit would oscillate with if there was no applied voltage and R were zero—the differential equation would then be that of simple harmonic motion.

CHAPTER III
Integration

§ 1. Integrals and areas under curves

INTEGRATION, in its most limited sense, is just the reverse of differentiation. However integration is also a way of calculating areas and when this understood there is seen to be a much deeper link between the two operations—much deeper, that is, than the elementary observation that one is the reverse of the other. This point will be explained when we deal with what is called the *fundamental theorem of calculus*, cf. 3.5 below.

So, to begin with, we shall consider the problem of finding the area of region under a curve, cf. fig. 33.

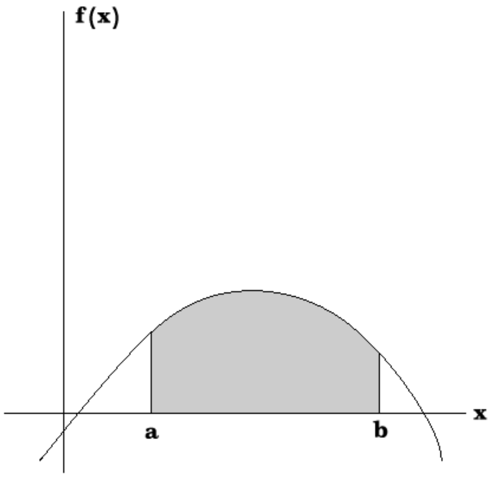


Fig. 33: The area under a curve

The shaded region in fig. 33 is the area that we wish to find: it lies between the vertical lines $x = a$ and $x = b$ and the notation used for this area is that it is denoted by

$$\int_a^b f(x)\,dx \tag{3.1}$$

and eq. 3.1 is referred to as "the integral of f from a to b".

What we do to calculate this area is to use rectangles as shown in fig. 34

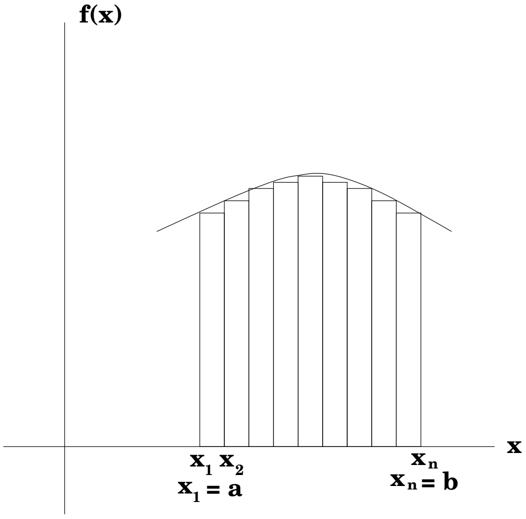


Fig. 34: Rectangles being used to calculate the area

Fig. 34 shows a series of $n-1$ rectangles whose collective area *approximates* the area between $x = a$ and $x = b$. The idea is that, as the number of these rectangles goes to infinity, the approximation becomes *exact*.

Now since fig. 34 shows that the first rectangle has base $(x_2 - x_1)$ and height $f(x_1)$ and the i^{th} rectangle has base $(x_{i+1} - x_i)$ and height $f(x_i)$ and so on. Hence the the area of all the rectangles is the sum

$$\sum_{i=1}^{n-1} f(x_i)(x_{i+1} - x_i) \tag{3.2}$$

Then to define the integral of f from a to b we take the limit as $n \to \infty$ giving

Definition (The integral $\int_a^b f(x)\,dx$)

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \sum_{i=1}^{n-1} f(x_i)(x_{i+1} - x_i) \tag{3.3}$$

Having defined $\int_a^b f(x)\,dx$ the really important thing is the result one obtains when $f(x)$ is already the derivative of some other function, say

$$f(x) = \frac{dF(x)}{dx}, \qquad \text{for some function } F(x) \tag{3.4}$$

This result is what is given in the fundamental theorem of calculus which we now quote.

Theorem (The fundamental theorem of calculus) *If a function of the form dF/dx is integrated then its integral is given by the formula*

$$\int_a^b \frac{dF(x)}{dx}\,dx = F(b) - F(a) \tag{3.5}$$

We shall sketch a proof of this theorem but first a piece of notation: the quantity $F(b) - F(a)$ is often denoted by $[F(x)]_a^b$ that is

$$[F(x)]_a^b = F(b) - F(a) \tag{3.6}$$

where $F(x)$ is any function. Now for the sketch of the proof.

Proof: First we adjust the bases of the rectangles in fig. 34 to have the same size—which we denote by Δx—that is we have

$$(x_1 - x_2) = (x_2 - x_3) = \cdots = (x_n - x_{n-1}) = \Delta x \tag{3.7}$$

Using this this sum in $\sum_{i=1}^{n-1} f(x_i)(x_{i+1} - x_i)$ of the integral definition 3.3 simplifies: we find that

$$\sum_{i=1}^{n-1} f(x_i)(x_{i+1} - x_i) = \sum_{i=1}^{n-1} f(x_i)\Delta x \tag{3.8}$$

But since $f = dF/dx$ we must now consider the sum

$$\sum_{i=1}^{n-1} \frac{dF(x_i)}{dx}\Delta x \tag{3.9}$$

Now the definition of the derivative is $dF(x_i)/dx$ is

$$\begin{aligned} \frac{dF(x_i)}{dx} &= \lim_{h\to 0} \frac{F(x_i + h) - F(x_i)}{h} \\ &= \lim_{\Delta x\to 0} \frac{F(x_i + \Delta x) - F(x_i)}{\Delta x}, \quad \text{on setting } h = \Delta x \end{aligned} \tag{3.10}$$

But

$$\begin{aligned} \Delta x &= (x_{i+1} - x_i) \\ \Rightarrow x_i + \Delta x &= x_i + x_{i+1} - x_i \\ &= x_{i+1} \end{aligned} \tag{3.11}$$

and using this fact in 3.10 we obtain

$$\frac{dF(x_i)}{dx} = \lim_{\Delta x \to 0} \frac{F(x_{i+1}) - F(x_i)}{\Delta x} \tag{3.12}$$

Putting together 3.3, 3.11 and 3.12 we have

$$\begin{aligned} \int_a^b \frac{dF(x)}{dx}\,dx &= \lim_{n\to\infty} \sum_{i=1}^{n-1} \lim_{\Delta x \to 0} \frac{F(x_{i+1}) - F(x_i)}{\Delta x} \Delta x \\ &= \lim_{n\to\infty} \sum_{i=1}^{n-1} (F(x_{i+1}) - F(x_i)), \quad \text{since } \Delta x \text{ cancels} \end{aligned} \tag{3.13}$$

But if we write out the sum $\sum_{i=1}^{n-1}(F(x_{i+1}) - F(x_i))$ we saw that all, except for two terms, cancel for we have

$$\begin{aligned} \sum_{i=1}^{n-1} (F(x_{i+1}) - F(x_i)) = (F(x_n) &- F(x_{n-1})) \\ + (F(x_{n-1}) &- F(x_{n-2})) \\ + (F(x_{n-2}) &- F(x_{n-3})) \\ \vdots \qquad & \qquad \vdots \\ + (F(x_3) &- F(x_2)) \\ + (F(x_2) &- F(x_1)) \end{aligned} \tag{3.14}$$

If we now look at 3.14 we see that all the terms on the RHS cancel except for the first and the last and so

$$\begin{aligned} \sum_{i=1}^{n-1} (F(x_{i+1}) - F(x_i)) &= F(x_n) - F(x_1) \\ &= F(b) - Fa), \quad \text{recalling from fig. 34 that } \begin{cases} x_n = b \\ x_1 = a \end{cases} \end{aligned} \tag{3.15}$$

So putting this result back into 3.12 we see that we have achieved an informal proof of the *fundamental theorem of calculus* which asserts—cf. 3.5—that

$$\int_a^b \frac{dF(x)}{dx}\,dx = F(b) - F(a) \tag{3.16}$$

§§ 1.1 Some notation and terminology

Some perfectly straightforward terminology can now be given: If we integrate any function f from a to b—that is calculate

$$\int_a^b f(x)\,dx \tag{3.17}$$

then the numbers a and b are called the *limits of integration*. One sometimes omits the limits of integration and writes

$$\int f(x)\,dx \tag{3.18}$$

The notation $\int f(x)\,dx$ simply denotes a function F whose derivative is f; hence we have

$$\int f(x)\,dx = F(x), \qquad \text{means that} \qquad \frac{dF(x)}{dx} = f(x) \tag{3.19}$$

F is then called the *integral* [1] *of* f.

For example if $f(x) = x^3$ then we could write

$$\int x^3\,dx = \frac{x^4}{4}, \qquad \text{because} \qquad \frac{d}{dx}\left(\frac{x^4}{4}\right) = x^3 \tag{3.20}$$

However notice that, if C is any constant, we could also write

$$\int x^3\,dx = \frac{x^4}{4} + C, \qquad \text{because} \qquad \frac{d}{dx}\left(\frac{x^4}{4} + C\right) = x^3 \tag{3.21}$$

Such a constant C is called a *constant of integration*.

This example shows that the integral F of any function is not unique since one can add a constant of integration to F; more concisely one can say: if F is an integral of f then so is $F + C$ for any constant C.

Finally to distinguish integrals with and without limits one has the following terminology. An integral *with limits* such as

$$\int_a^b f(x)\,dx \tag{3.22}$$

is called a *definite integral*, while an integral *without limits* like

$$\int f(x)\,dx \tag{3.23}$$

is called an *indefinite integral*.

§§ 1.2 Some examples

[1] Alternative terms are often used for F. An integral of F is sometimes called a *primitive of* F or an *anti-derivative of* F—we shall use neither term since the word *integral* is the more common usage.

Example *The area under the parabola* $f(x) = x^2$ *between* $x = 3$ *and* $x = 9$.
With the fundamental theorem of calculus under our belt it is a simple matter to calculate the area shown in the fig. 35.

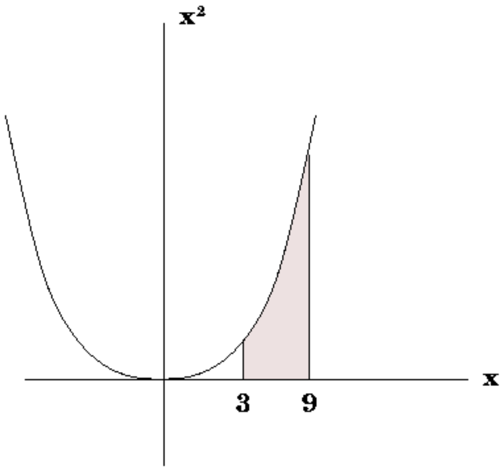


Fig. 35: The area under the parabola between $x = 3$ and $x = 9$

The area A that we want is given by

$$A = \int_3^9 x^2 \, dx \tag{3.24}$$

and so we can straightaway compute that

$$\begin{aligned} \int_3^9 x^2 \, dx &= \left[\frac{x^3}{3}\right]_3^9 \\ &= \left[\frac{9^3}{3} - \frac{3^3}{3}\right] \\ &= \left[\frac{729}{3} - \frac{27}{3}\right] \\ &= 234 \end{aligned} \tag{3.25}$$

and so we have our area A. That example was a definite integral so next we consider an indefinite integral.
Example *The integral*

$$\int \sin^2(x) \cos(x) \, dx \tag{3.26}$$

So this time we just want a function F that satisfies

$$\frac{dF}{dx} = \sin^2(x) \cos(x) \tag{3.27}$$

After some tinkering with various combinations of trigonometric functions we should be able to find that a solution is

$$F(x) = \frac{\sin^3(x)}{3} \tag{3.28}$$

which certainly does the trick. Thus we can write

$$\int \sin^2(x)\cos(x)\,dx = \frac{\sin^3(x)}{3} \tag{3.29}$$

and if we want to be really precise we add on a constant of integration and write the *most general* statement of the result which is

$$\int \sin^2(x)\cos(x)\,dx = \frac{\sin^3(x)}{3} + C, \qquad \text{for any constant } C \tag{3.30}$$

The next example illustrates a very important point. It is that if the area being calculated *lies under* the x-axis, rather than above it, then the area A calculated by the integral will be *negative.*

Example *Areas under the x-axis count negatively in integrals*

$$\int_{\pi/2}^{3\pi/2} \cos(x)\,dx \tag{3.31}$$

Fig. 36 shows the graph of $\cos(x)$ between $x = \pi/2$ and $3\pi/2$; this can be seen to be an interval on which $\cos(x)$ is *negative.*

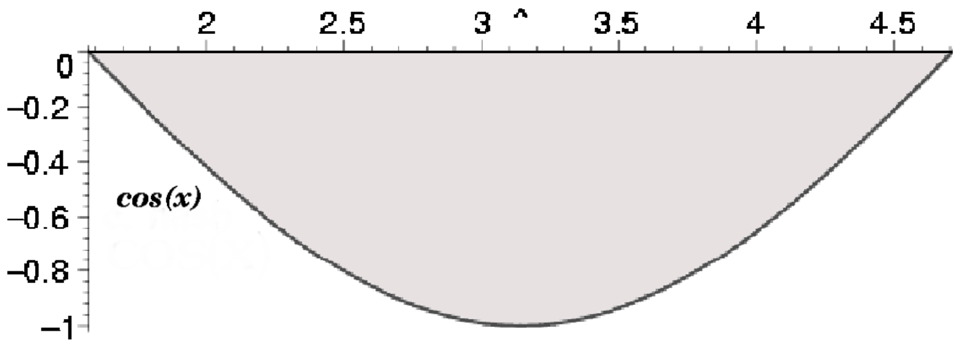


Fig. 36: The graph of $\cos(x)$ for the interval $[\pi/2, 3\pi/2]$

The integral we want is $\int_{\pi/2}^{3\pi/2} \cos(x)\,dx$ and we compute that

$$\begin{aligned}\int_{\pi/2}^{3\pi/2} \cos(x)\,dx &= [\sin(x)]_{\pi/2}^{3\pi/2} \\ &= [\sin(3\pi/2) - (\sin(\pi/2))] \\ &= [-1-1], \quad \text{since } \begin{cases}\sin(\pi/2) = 1 \\ \sin(3\pi/2) = -1\end{cases} \\ &= -2\end{aligned} \tag{3.32}$$

Hence the integral has the value

$$-2 \tag{3.33}$$

and is negative as claimed. The *actual shaded area* shown in fig. 36 is of course

$$+2 \tag{3.34}$$

One should not be disturbed by this: all that is happening is that an integral $\int_a^b f(x)\,dx$ can have both *negative and positive* contributions depending on whether the function f is *negative or positive* for a given part of the integration interval $[a, b]$.

This can lead to an integral being zero because the positive and negative contributions exactly cancel. This happens in the next example.

Example *The integral*

$$\int_0^{2\pi} \sin(x)\,dx \tag{3.35}$$

First let's get the computation of the integral out of the way. We find that

$$\begin{aligned}\int_0^{2\pi} \sin(x)\,dx &= -\left[\cos(x)\right]_0^{2\pi} \\ &= -\left[\cos(2\pi) - \cos(0)\right] \\ &= -\left[1 - 1\right], \quad \text{since} \begin{cases} \cos(2\pi) = 1 \\ \cos(0) = 1 \end{cases} \\ &= 0 \end{aligned} \tag{3.36}$$

So the integral does indeed vanish. Now if we look at fig. 37 we see two shaded areas of differing densities and it is clear that what has happened is that these two areas have simply cancelled.

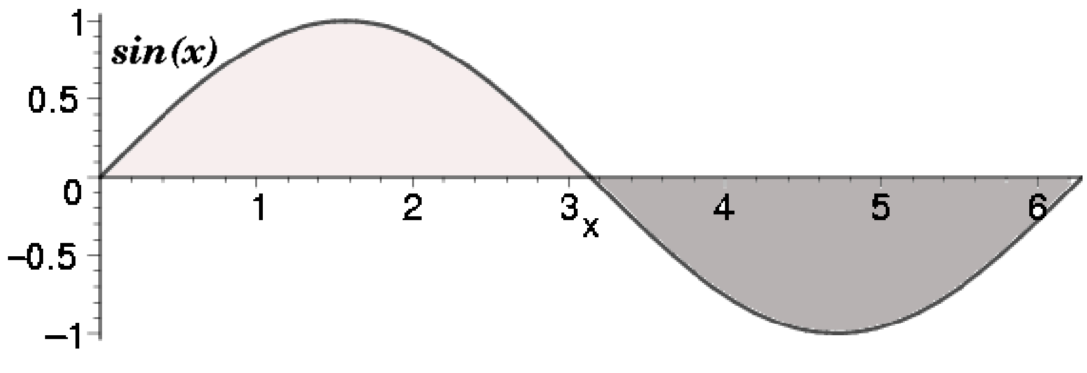


Fig. 37: The two cancelling areas for $\sin(x)$ on the interval $[0, 2\pi]$

If one really wants to prove this then one can calculate the two areas using separate integrals. In other words one calculates the two integrals

$$\int_0^{\pi} \sin(x)\,dx, \quad \int_{\pi}^{2\pi} \sin(x)\,dx \tag{3.37}$$

One then readily verifies that the first integral is positive and equal to

$$+2 \tag{3.38}$$

while the second is equal to

$$-2 \tag{3.39}$$

A useful fact about a definite integral such as

$$\int_a^c f(x)\,dx \tag{3.40}$$

is that one can choose a number b between a and c and split the integral up into two pieces. What one obtains is just

$$\int_a^c f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx, \quad a < b < c \tag{3.41}$$

All that we are doing is observing that the area given by $\int_a^c f(x)\,dx$ is the sum of the two smaller areas $\int_a^b f(x)\,dx$ and $\int_b^c f(x)\,dx$.

It is also useful to observe that if one *interchanges* a *and* b in the integral $\int_a^b f(x)\,dx$ then the integral *changes sign*. In other words we have

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx \tag{3.42}$$

This fact is evident already in the statement 3.5 of the fundamental theorem of calculus where we see at once that the RHS changes sign if a and b are interchanged.

§ 2. Difficult integrals...?

Unlike derivatives, integrals can often be impossible to compute in terms of well known functions [2] and then one has to resort to numerical methods to deal with them.

[2] This may seem a slightly vague statement since the phrase "well known functions" is certainly not precise. However mathematicians do have a more precise set of functions in mind. This sets consists of all functions that can be obtained by addition, multiplication, division and composition of: polynomials, trigonometric functions and their inverses, and the functions ln and exp. These functions are sometimes then referred to as the *elementary functions*.

For example, there is no elementary function F which satisfies

$$\frac{dF(x)}{dx} = e^{-x^2} \tag{3.43}$$

In everyday language we say *we can't do the integral*

$$\int e^{-x^2}\,dx \tag{3.44}$$

Another example of an integral *"we can't do"* is

$$\int \sqrt{x}\sin(x)\,dx \tag{3.45}$$

Despite this we *can do* the very similar integral

$$\int x\sin(x)\,dx \tag{3.46}$$

for differentiation readily verifies the correctness of the statement that

$$\int x\sin(x)\,dx = \sin(x) - x\cos(x) \tag{3.47}$$

Thus a small change in the function to be integrated can make a big difference.

§ 3. Some well known integrals

It is always useful to have a list of integrals for functions that one comes across often. For this reason we provide a short table of integrals on p. 86.

§ 4. Integration techniques

There are two main techniques for doing integrals and these are
(a) Integration by change of variables or substitution.
(b) Integration by parts. We shall study both of these methods beginning with (a).

§§ 4.1 Integration by change of variable or integration by substitution

Next we illustrate the technique of integration by *change of variable* (also called integration by *substitution*). Consider then the following example.

Example *The integral*

$$\int_0^{\pi/2} \sin^5(x)\cos(x)\,dx \tag{3.48}$$

$f(x)$	$\int f(x)\,dx$
$x^n,\ (n \neq -1)$	$\frac{x^{n+1}}{n+1}$
$\frac{1}{x}$	$\ln(x)$
e^x	e^x
a^x	$\frac{a^x}{\ln(a)}$
$\ln(x)$	$x\ln(x) - x$
$\sin(x)$	$-\cos(x)$
$\cos(x)$	$\sin(x)$
$\tan(x)$	$-\ln(\cos(x))$
$\sec(x)$	$\ln(\sec(x) + \tan(x))$
$\mathrm{cosec}(x)$	$-\ln(\mathrm{cosec}(x) + \cot(x))$
$\cot(x)$	$\ln(\sin(x))$
$\sec^2(x)$	$\tan(x)$
$\sec(x)\tan(x)$	$\sec(x)$
$\sinh(x)$	$\cosh(x)$
$\cosh(x)$	$\sinh(x)$
$\tanh(x)$	$\ln(\cosh(x))$
$\frac{1}{1+x^2}$	$\arctan(x)$
$\frac{1}{\sqrt{1-x^2}}$	$\arcsin(x)$
$-\frac{1}{\sqrt{1-x^2}}$	$\arccos(x)$

A table of useful integrals

We change variable from x to u where

$$u = \sin(x) \tag{3.49}$$

So that we have

$$\begin{aligned} \frac{du}{dx} &= \cos(x) \\ \Rightarrow du &= \cos(x)dx \\ \Rightarrow dx &= \frac{du}{\cos(x)} \end{aligned} \tag{3.50}$$

So now, if we *temporarily* omit the limits of integration, we can substitute this information into the integral giving us the equation

$$\begin{aligned} \int \sin^5(x)\cos(x)\,dx &= \int u^5 \cos(x)\frac{du}{\cos(x)} \\ &= \int u^5 du \end{aligned} \tag{3.51}$$

Finally to restore the limit of integration we notice that, since $u = \sin(x)$, we have

$$\begin{aligned} x = 0 &\Rightarrow u = 0 \\ x = \frac{\pi}{2} &\Rightarrow u = \sin(\pi/2) = 1 \end{aligned} \tag{3.52}$$

Hence our limits in terms of u are 0 and 1 and, restoring the limits to both integrals, we have the equality

$$\begin{aligned} \int_0^{\pi/2} \sin^5(x)\cos(x)\,dx &= \int_0^1 u^5\,du \\ &= \left[\frac{u^6}{6}\right]_0^1 \\ &= \left[\frac{1}{6} - 0\right] \\ &= \frac{1}{6} \end{aligned} \tag{3.53}$$

and our integral has been completed. We move on.

Example *The integral*

$$\int \frac{dx}{x\ln(x)} \tag{3.54}$$

This time we set

$$\begin{aligned} u &= \ln(x) \\ \Rightarrow \frac{du}{dx} &= \frac{1}{x} \\ \Rightarrow dx &= x\,du \end{aligned} \tag{3.55}$$

Now we put this information into our integral and find that

$$\begin{aligned}\int \frac{dx}{x\ln(x)} &= \int \frac{xdu}{xu}\\ &= \int \frac{du}{u}\\ &= \ln(u)\\ &= \ln(\ln(x))\end{aligned} \tag{3.56}$$

and so we have established that

$$\int \frac{dx}{x\ln(x)} = \ln(\ln(x)) \tag{3.57}$$

Time for our next integral which is

Example *The integral*

$$\int_0^a \sqrt{a^2 - x^2}\, dx \tag{3.58}$$

The trick here is to use a trigonometric substitution or change of variable; the one that works for this case [3] is

$$\begin{aligned}x &= a\sin(\theta)\\ \Rightarrow dx &= a\cos(\theta)\, d\theta\end{aligned} \tag{3.59}$$

We must also change the limits of integration to accommodate the new variable θ; to this end note that

$$\begin{aligned}x = 0 &\Rightarrow \theta = 0\\ x = 1 &\Rightarrow \theta = \frac{\pi}{2}\end{aligned} \tag{3.60}$$

[3] For other closely related integrals such as

$$\int \left(a^2 \mp x^2\right)^{\mp 1/2}\, dx$$

one should also try a trigonometric substitution such as

$$x = a\sin(\theta),\ x = a\cos(\theta),\ x = a\tan(\theta)$$

$x = a\tan(\theta)$ being the one to use when $(a^2 + x^2)$, rather than $(a^2 - x^2)$, occurs in the integrand.

so the new limits are 0 and $\pi/2$ and we obtain

$$\begin{aligned}
\int_0^a \sqrt{a^2 - x^2}\, dx &= \int_0^{\pi/2} \sqrt{a^2 - a^2 \sin(\theta)}\, a\cos(\theta)\, d\theta \\
&= \int_0^{\pi/2} \sqrt{a^2(1-\sin^2(\theta))}\, a\cos(\theta)\, d\theta \\
&= \int_0^{\pi/2} a\cos(\theta) a\cos(\theta)\, d\theta, \quad \text{using } (1-\sin^2(\theta)) = \cos^2(\theta) \\
&= a^2 \int_0^{\pi/2} \cos^2(\theta)\, d\theta \\
&= a^2 \int_0^{\pi/2} \frac{1+\cos(2\theta)}{2}\, d\theta, \quad \text{using } \cos^2(\theta) = \frac{1+\cos(2\theta)}{2} \\
&= a^2 \left[\frac{\theta}{2} + \frac{\sin(2\theta)}{4}\right]_0^{\pi/2} = a^2\left[\frac{\pi}{4} + \frac{\sin(\pi)}{4}\right] = \frac{a^2\pi}{4}, \quad \text{since } \sin(\pi) = 0
\end{aligned} \tag{3.61}$$

Now we consider

Example *The integral*

$$\int \frac{e^{2x}}{\sqrt{e^x+1}}\, dx \tag{3.62}$$

This time we shall try

$$\begin{aligned}
u &= \sqrt{e^x+1} \\
\Rightarrow du &= \frac{1}{2}\left(e^x+1\right)^{-1/2} e^x dx
\end{aligned} \tag{3.63}$$

Also

$$\begin{aligned}
u &= \sqrt{e^x+1} \\
\Rightarrow e^x &= u^2 - 1 \\
&\Rightarrow \begin{cases} e^x + 1 = u^2 \\ e^{2x} = (u^2-1)^2 \\ dx = \dfrac{2u}{u^2-1} du \end{cases}
\end{aligned} \tag{3.64}$$

Our integral now displays the following transformation

$$\begin{aligned}
\int \frac{e^{2x}}{\sqrt{e^x+1}}\, dx &= \int \frac{(u^2-1)^2}{u} \frac{2u}{u^2-1}\, du \\
&= \int 2(u^2-1)\, du \\
&= \frac{2}{3}u^3 - 2u
\end{aligned} \tag{3.65}$$

But $u = \sqrt{e^x + 1}$ so we have deduced that

$$\int \frac{e^{2x}}{\sqrt{e^x + 1}}\,dx = \frac{2}{3}\left(e^x + 1\right)^{3/2} - 2\left(e^x + 1\right)^{1/2} \tag{3.66}$$

We are now ready for the other main integration technique which is integration by parts.

§§ 4.2 Integration by parts

All integrations by parts rest on a clever use of the same formula. This formula is very simply obtained: one just integrates the formula for the derivative of a product. More precisely we begin with the formula

$$\frac{d}{dx}\left(f(x)g(x)\right) = f(x)\frac{dg(x)}{dx} + \frac{df(x)}{dx}g(x) \tag{3.67}$$

which we rewrite as

$$f(x)\frac{dg(x)}{dx} = \frac{d}{dx}\left(f(x)g(x)\right) - \frac{df(x)}{dx}g(x) \tag{3.68}$$

and then we integrate both sides from a to b yielding

$$\int_a^b f(x)\frac{dg(x)}{dx}\,dx = \int_a^b \frac{d}{dx}\left(f(x)g(x)\right)\,dx - \int_a^b \frac{df(x)}{dx}g(x)\,dx$$

$$\Rightarrow \int_a^b f(x)\frac{dg(x)}{dx}\,dx = \left[f(x)g(x)\right]_a^b - \int_a^b \frac{df(x)}{dx}g(x)\,dx$$

Or more compactly

$$\int_a^b fg'\,dx = \left[fg\right]_a^b - \int_a^b f'g\,dx \tag{3.69}$$

This last formula is the one we want and we emphasise its importance by quoting it as a theorem.

Theorem (Integration by parts) *If f and g are two differentiable functions then*

$$\int_a^b fg'\,dx = \left[fg\right]_a^b - \int_a^b f'g\,dx$$

or

$$\int fg'\,dx = fg - \int f'g\,dx, \quad \textit{without limits} \tag{3.70}$$

Now we need some examples of integration by parts in action. We begin with something simple.

Example *The integral*

$$\int xe^x \, dx \tag{3.71}$$

The method of integration by parts then consists of equating the integrand to fg' and then using 3.70. So we write

$$xe^x = fg' \tag{3.72}$$

and immediately we are faced with the task of deciding which part of xe^x should we equate to f and which to g'. The answer to this problem is that one proceeds partially by trial and error and partially by previous experience. This time we choose to set

$$f = x \tag{3.73}$$

which forces

$$g' = e^x \tag{3.74}$$

With this choice for f and g we use 3.70 giving us

$$\begin{aligned} \int xe^x \, dx &= xe^x - \int 1 \cdot e^x \, dx \\ &= xe^x - e^x \end{aligned} \tag{3.75}$$

and so we are finished.

A favourite trick in integration by parts is to set g' equal to 1 and then use the formula 3.70. This can be seen at work in our next calculation.

Example *The integral*

$$\int \ln(x) \, dx \tag{3.76}$$

We set

$$g' = 1 \tag{3.77}$$

We must therefore set $f = \ln(x)$ and 3.70 gives

$$\begin{aligned} \int \ln(x) \cdot 1 \, dx &= \ln(x) \cdot x - \int \frac{1}{x} \cdot x \, dx \\ &= x\ln(x) - x \end{aligned} \tag{3.78}$$

and so we have our integral and the result agrees, as it must, with that quoted in our integral table on p. 86.

A second popular trick is to try to express $\int f\,dx$ in terms of itself and solve the resulting formula. We illustrate this next.

Example *The integral*

$$\int \frac{\ln(x)}{x}\,dx \tag{3.79}$$

Setting $f = \ln(x)$, and thus $g' = 1/x$, we obtain the equation

$$\begin{aligned} \int \frac{\ln(x)}{x}\,dx &= \ln(x)\ln(x) - \int \frac{1}{x}\ln(x)\,dx \\ \Rightarrow 2\int \frac{\ln(x)}{x}\,dx &= \ln(x)\ln(x) \\ \text{or} \qquad \int \frac{\ln(x)}{x}\,dx &= \frac{\ln^2(x)}{2} \end{aligned} \tag{3.80}$$

and we note the appearance of $\int \ln(x)/x\,dx$ on both sides of the equation in the first line of 3.80. Another integral of this type is

Example *The integral*

$$\int \sin^2(x)\,dx \tag{3.81}$$

Choosing $f = \sin(x)$, and therefore $g' = \sin(x)$ also, gives

$$\begin{aligned} \int \sin^2(x)\,dx &= \sin(x)(-\cos(x)) - \int \cos(x)(-\cos(x))\,dx \\ &= -\sin(x)\cos(x) + \int \cos^2(x)\,dx \\ &= -\sin(x)\cos(x) + \int (1-\sin^2(x))\,dx, \quad \text{using } \cos^2(x) = 1-\sin^2(x) \\ \Rightarrow 2\int \sin^2(x)\,dx &= -\sin(x)\cos(x) + \int dx \\ \Rightarrow \int \sin^2(x)\,dx &= \frac{x - \sin(x)\cos(x)}{2} \end{aligned} \tag{3.82}$$

§ 5. A little more integration

We shall finish this chapter with a bit more integration practice.

Example *The integral*

$$\int \frac{e^{\sqrt{x}}}{\sqrt{x}}\,dx \tag{3.83}$$

We use the substitution $u = \sqrt{x}$ and so find that

$$\begin{aligned} \frac{du}{dx} &= \frac{1}{2}x^{-1/2} \\ \Rightarrow dx &= 2\,x^{1/2}du \end{aligned} \tag{3.84}$$

Hence we obtain

$$\begin{aligned} \int \frac{e^{\sqrt{x}}}{\sqrt{x}}\,dx &= \int e^u \frac{2\,x^{1/2}du}{\sqrt{x}} \\ &= 2\int e^u\,du \\ &= 2e^u = 2e^{\sqrt{x}} \end{aligned} \tag{3.85}$$

and so we have shown that

$$\int \frac{e^{\sqrt{x}}}{\sqrt{x}}\,dx = 2e^{\sqrt{x}} \tag{3.86}$$

Example *The integral*

$$\int x^2 e^x\,dx \tag{3.87}$$

We can do this by parts but we will apply the method twice to get to the end. Setting

$$g' = e^x \tag{3.88}$$

we get

$$\int x^2 e^x\,dx = x^2 e^x - \int 2xe^x\,dx \tag{3.89}$$

This leaves us with the integral $\int xe^x\,dx$ still to do; so for this integral we again set

$$g' = e^x \tag{3.90}$$

and obtain

$$\begin{aligned} \int xe^x\,dx &= xe^x - \int 1\cdot e^x\,dx \\ &= xe^x - e^x \end{aligned} \tag{3.91}$$

Substituting 3.91 into 3.89 we find the result that

$$\int x^2 e^x \, dx = x^2 e^x - 2xe^x + 2e^x \tag{3.92}$$

Example *The integral*

$$\int \frac{e^x}{e^{2x} + 2e^x + 1} \, dx \tag{3.93}$$

For this integral note that the denominator of the integrand is a perfect square—that is

$$e^{2x} + 2e^x + 1 = (e^x + 1)^2 \tag{3.94}$$

This means that we can write

$$\int \frac{e^x}{e^{2x} + 2e^x + 1} \, dx = \int \frac{e^x}{(e^x + 1)^2} \, dx \tag{3.95}$$

and now if we set

$$u = e^x + 1 \tag{3.96}$$

we find that

$$\begin{aligned} \int \frac{e^x}{(e^x + 1)^2} \, dx &= \int \frac{e^x}{u^2} \frac{du}{e^x} \\ &= \int \frac{du}{u^2} \\ &= -\frac{1}{u} = -\frac{1}{e^x + 1} \end{aligned} \tag{3.97}$$

So we have deduced that

$$\int \frac{e^x}{e^{2x} + 2e^x + 1} \, dx = -\frac{1}{e^x + 1} \tag{3.98}$$

Example *The integral*

$$\int \sqrt{x} \ln(x) \, dx \tag{3.99}$$

For this example we use integration by parts and set

$$g' = \sqrt{x} \tag{3.100}$$

This yields the equation

$$\begin{aligned}\int \sqrt{x}\ln(x)\,dx &= \frac{2}{3}x^{3/2}\ln(x) - \int \frac{2}{3}x^{3/2}\frac{1}{x}\,dx \\ &= \frac{2}{3}x^{3/2}\ln(x) - \frac{2}{3}\int x^{1/2}\,dx \\ &= \frac{2}{3}x^{3/2}\ln(x) - \left(\frac{2}{3}\right)^2 x^{3/2}\end{aligned} \tag{3.101}$$

So we have found that

$$\int \sqrt{x}\ln(x)\,dx = \frac{2}{3}x^{3/2}\left(\ln(x) - \frac{2}{3}\right) \tag{3.102}$$

CHAPTER IV
Miscellaneous topics

§ 1. Volumes, surfaces and curves

WE shall use this chapter to discuss an assorted list of topics most of them applications of calculus. We begin with an important use of integration to calculate *volumes*, *surface areas* and *lengths of curves* rather than just the *area* under a curve. We shall deal first with volumes. To this end we introduce what is called a *volume of revolution* and then calculate its volume.

§§ 1.1 Volumes of revolution and their volumes

A volume of revolution is made by rotating a curve through one complete revolution—i.e. 2π radians—about an axis. In fig. 38 we show what is needed to understand what is going on. The first part of the figure simply show the graph of the curve $f(x)$ while the second illustration shows the solid obtained by rotating this curve once about the x-axis. Such a solid is called a *volume of revolution.* We shall now see how to calculate the volume of this solid.

All we have to do is to divide the solid up into cylinders or disks: fig. 38 shows a typical disk, with *centre* x, *radius* $f(x)$ and thickness dx.

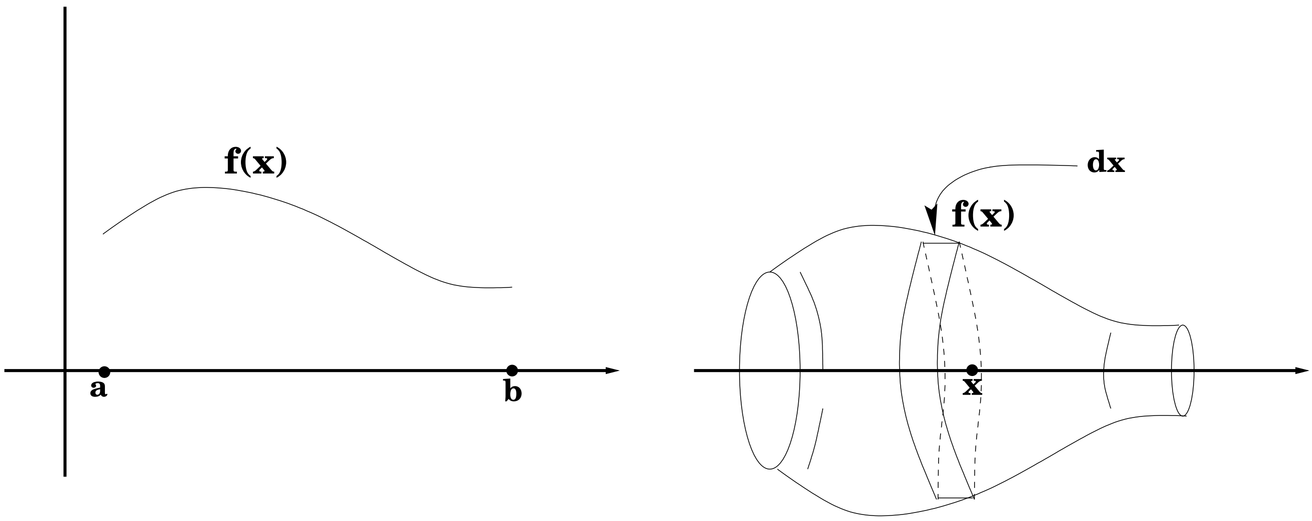


Fig. 38: A volume of revolution made from the function $f(x)$

Now a disk of radius r and thickness h is just a cylinder and so has volume

$$\pi r^2 h \tag{4.1}$$

Hence the disk shown in fig. 38 has volume

$$\pi(f(x))^2 dx \tag{4.2}$$

The entire volume V of the solid will be obtained by summing over all disks as their thickness dx tends to zero; but this is just the integral

$$\int_a^b \pi f^2(x)\, dx \tag{4.3}$$

where a and b are the points where the curve $f(x)$ begins and ends. Thus we have a formula for the volume V of the surface of revolution namely

$$V = \int_a^b \pi f^2(x)\, dx \tag{4.4}$$

Let us use this formula in an actual calculation.

Example *The volume of an ellipsoid of revolution*

An ellipse centered at the origin has the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{4.5}$$

and is shown in fig. 39.

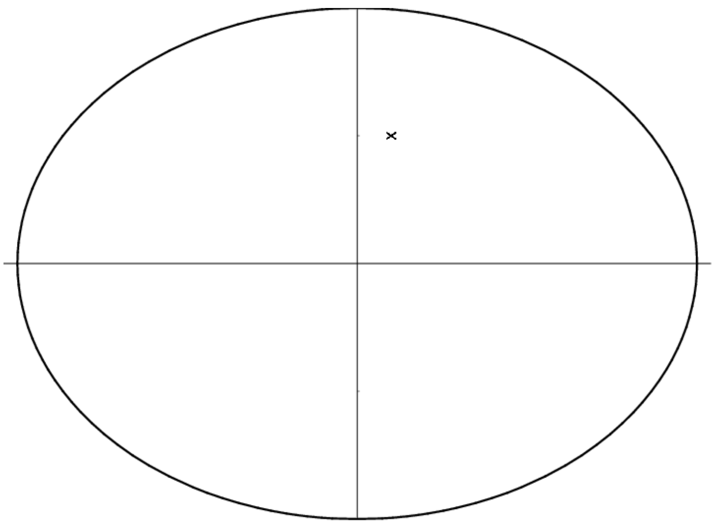


Fig. 39: The ellipse $x^2/a^2 + y^2/b^2 = 1$

If we take *half* of this ellipse—as shown in fig. 40—and rotate it about the x axis we obtain a solid called an *ellipsoid of revolution*.

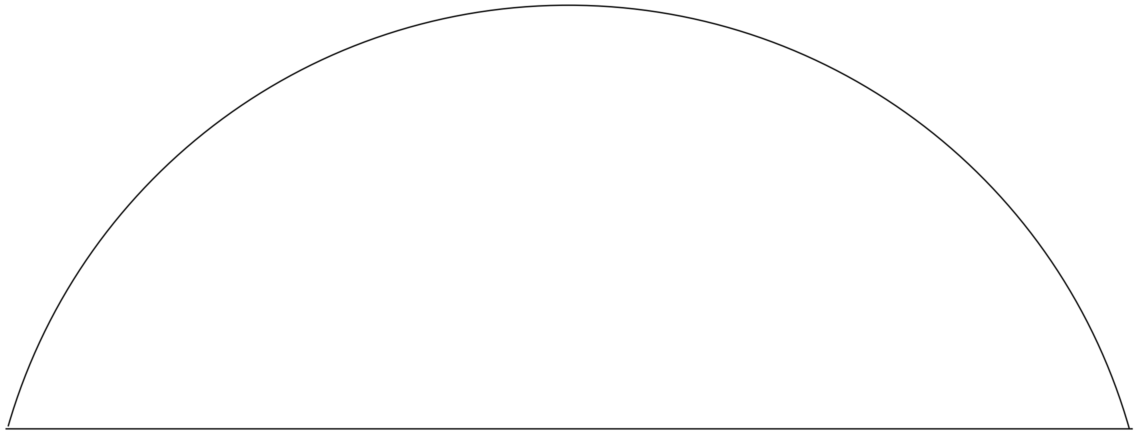


Fig. 40: The upper half of the ellipse $x^2/a^2 + y^2/b^2 = 1$

Now we need to find the appropriate function $f(x)$ so that we can use formula 4.4; but $f(x)$ is just the quantity y so this will come at once from the equation of the ellipse: We have

$$\begin{aligned}\frac{x^2}{a^2} + \frac{y^2}{b^2} &= 1 \\ \Rightarrow y^2 &= b^2\left(1 - \frac{x^2}{a^2}\right) \\ \Rightarrow y &= b\sqrt{1 - \frac{x^2}{a^2}}\end{aligned} \tag{4.6}$$

Hence $f(x)$ is given by the equation

$$f(x) = b\sqrt{1 - \frac{x^2}{a^2}} \tag{4.7}$$

Now we can use formula 4.4 to compute the volume V of this ellipsoid. This gives us that

$$\begin{aligned}V &= \int_a^b \pi f^2(x)\,dx \\ &= \pi \int_{-a}^{a} b^2\left(1 - \frac{x^2}{a^2}\right)\,dx \\ &= \pi b^2\left[x - \frac{x^3}{3a^2}\right]_{-a}^{a} \\ &= \pi b^2\left[a - \frac{a^3}{3a^2} - (-a) + \frac{(-a)^3}{3a^2}\right] \\ &= \pi b^2\left[\frac{4a}{3}\right]\end{aligned} \tag{4.8}$$

So the volume of the ellipsoid of revolution is given by

$$V = \frac{4\pi ab^2}{3} \tag{4.9}$$

Notice that, since a circle is a special case of an ellipse, we should be able to reproduce the formula for the volume of a sphere by making our ellipse a circle. This will provide a check on our calculation and so we shall do it. All we have to do is to set

$$b = a \tag{4.10}$$

and then the equation of the ellipse becomes

$$\begin{aligned} \frac{x^2}{a^2} + \frac{y^2}{a^2} &= 1 \\ \Rightarrow x^2 + y^2 &= a^2 \end{aligned} \tag{4.11}$$

i.e. the ellipse has a become a circle of radius a. Such a circle would, if we rotated its upper half, produce a sphere of radius a which we know has a volume

$$\frac{4\pi a^3}{3} \tag{4.12}$$

But if we set $b = a$ in the formula 4.9 for the volume of our ellipsoid we obtain

$$\frac{4\pi a^3}{3} \tag{4.13}$$

which is indeed the volume of the sphere, and our check has succeeded.

§§ **1.2 Surfaces of revolution and their areas**

We can also use integration to find the area of the *surface* of one of these solids. Such a surface is called, not surprisingly, a *surface of revolution.*

Now to calculate the area of this surface one can again use disks but this time they the disks are not small cylinders of height dx but *skew disks* of *slanted height dl*. An example of a skew disk is shown in fig. 41.

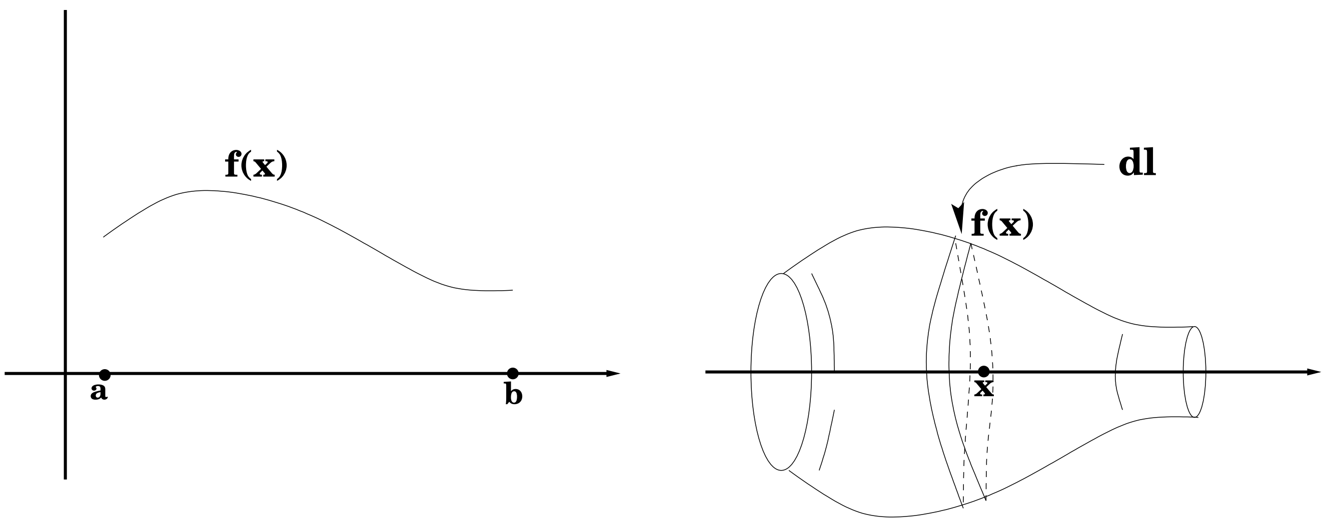


Fig. 41: A surface of revolution and its division into *skew* disks

The reader should carefully compare figs. 38 and 41 until he or she notices that the disk of 38 have *straight sides* of height dx while those of 41 have *slanted sides* of length dl.

The important point is that dl is, in general, not equal to dx because dl points along the direction of the tangent to f at x while dx is always parallel to the x-axis. But if the tangent makes an angle θ with the x-axis at the point x then, by definition of the derivative, we know that

$$f'(x) = \tan(\theta) \tag{4.14}$$

It is also true that dx and dl are at an angle θ relative to one another so that we have

$$dx = dl\cos(\theta) \tag{4.15}$$

Now we are ready to calculate the area: a typical skew disk of radius $f(x)$ and skew side of slanted length dl has *surface area*

$$\begin{aligned} &2\pi f(x)dl \\ &= 2\pi f(x)\frac{dx}{\cos(\theta)}, \qquad \text{using } 4.15 \end{aligned} \tag{4.16}$$

But we can relate $\cos(\theta)$ to $f'(x)$ by using 4.14. We have the fact that

$$\begin{aligned} 1+\tan^2(\theta) &= \sec^2(\theta) \\ &= \frac{1}{\cos^2(\theta)} \\ \Rightarrow \frac{1}{\cos(\theta)} &= \sqrt{1+\tan^2(\theta)} \\ &= \sqrt{1+f'(x)^2}, \qquad \text{using } 4.14 \end{aligned} \tag{4.17}$$

So our surface area formula now becomes

$$2\pi f(x)\frac{dx}{\cos(\theta)} = 2\pi f(x)\sqrt{1+f'(x)^2}\,dx \tag{4.18}$$

The final task to get the surface area S of the surface of revolution is to sum over all such skew disks and this gives us an integral. The resulting expression for S

$$S = \int_a^b 2\pi f(x)\sqrt{1+f'(x)^2}\,dx \tag{4.19}$$

It is now time to look at an example.

Example *The surface area of the ellipsoid*

The calculation is rather straightforward since we have already had the experience of the volume computation.

Recall that 4.7 says that

$$\begin{aligned} f(x) &= b\sqrt{1-\frac{x^2}{a^2}} \\ \Rightarrow f'(x) &= b\left(\frac{1}{2}\right)\left(1-\frac{x^2}{a^2}\right)^{-1/2}\left(\frac{-2x}{a^2}\right) \\ &= \frac{-bx}{a^2}\frac{1}{\sqrt{1-\frac{x^2}{a^2}}} \\ \Rightarrow 1+f'(x)^2 &= 1+\frac{b^2x^2}{a^4}\frac{1}{1-\frac{x^2}{a^2}} \\ &= \frac{a^4+(b^2-a^2)x^2}{a^4(1-\frac{x^2}{a^2})} \end{aligned} \tag{4.20}$$

Substituting this information back into the expression 4.19 for the area S gives

$$S = 2\pi\int_{-a}^{a}\frac{b}{a^2}\sqrt{a^4+(b^2-a^2)x^2}\,dx \tag{4.21}$$

Now to speed up the calculation [1] we set

$$b = a \tag{4.23}$$

so that we are now computing the surface area of a sphere. With this simplification we find that

$$\begin{aligned} S &= 2\pi a\int_{-a}^{a} dx \\ &= 2\pi a\,[x]_{-a}^{a} \\ &= 4\pi a^2 \end{aligned} \tag{4.24}$$

[1] We don't want to bother the reader with the expression for the surface area when $b \neq a$, but for the curious we give it anyway in this footnote. If we use the fact that

$$\int\sqrt{1-x^2}\,dx = \frac{1}{2}\left\{x\sqrt{1-x^2}+\arcsin(x)\right\}, \quad \text{(integrate by parts twice to get this)}$$

then we find that we can do the integral for S after a suitable change of variable. After doing this, and putting in the limits, we find that

$$S = 2\pi\left\{b^2+\frac{a^2b}{\sqrt{a^2-b^2}}\arcsin\left(\frac{\sqrt{a^2-b^2}}{a}\right)\right\} \tag{4.22}$$

and this is indeed the surface area of a sphere of radius a.

We can now turn to the matter of calculating the length of an arbitrary curve.

§§ 1.3 Curves and their lengths

If $f(x)$ is any curve, such as that depicted in fig. 41, then we can easily derive an integral formula for the length of f. If the endpoints of the curve are, as in fig. 41, at $x = a$ and $x = b$ then the length L of the curve is given by

$$L = \int_a^b dl \tag{4.25}$$

where dl is an infinitesimal piece of the curve exactly as it is in 41. But 4.15 gives the relation between dx and dl as

$$dx = dl \cos(\theta) \tag{4.26}$$

and 4.17 tells us that

$$\cos(\theta) = \frac{1}{\sqrt{1 + f'(x)^2}} \tag{4.27}$$

Hence $dl = \sqrt{1 + f'(x)^2}\, dx$ giving us our final formula for the curve length which is

$$L = \int_a^b \sqrt{1 + f'(x)^2}\, dx \tag{4.28}$$

We are now ready to do a calculation.

Example *The length of an arc of a circle*

If the circle has radius A then its equation is

$$\begin{aligned} x^2 + y^2 &= A^2 \\ \Rightarrow y &= \sqrt{A^2 - x^2} \end{aligned} \tag{4.29}$$

This means we set

$$f(x) = \sqrt{A^2 - x^2} \tag{4.30}$$

and so

$$f'(x) = -\frac{x}{\sqrt{A^2 - x^2}} \tag{4.31}$$

and thus

$$\begin{aligned} L &= \int_a^b \sqrt{1 + \frac{x^2}{A^2 - x^2}}\, dx \\ &= \int_a^b \sqrt{\frac{x^2 - A^2 + x^2}{A^2 - x^2}}\, dx \\ &= A \int_a^b \frac{dx}{\sqrt{A^2 - x^2}} \end{aligned} \tag{4.32}$$

Let us take a circle of unit radius so that $A = 1$ and then we have

$$\begin{aligned} L &= \int_a^b \frac{dx}{\sqrt{1-x^2}} \\ &= [\arcsin(x)]_a^b\,, \quad \text{using our table of integrals} \end{aligned} \tag{4.33}$$

Finally let us decide to compute the length of a quadrant of this unit circle—cf. fig 42; this means that we must choose the interval $[a, b]$ to be given by

$$[a, b] = [0, 1] \tag{4.34}$$

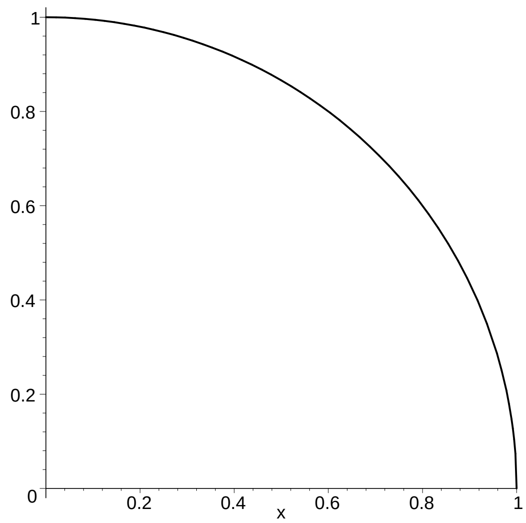


Fig. 42: A quadrant of a circle of radius 1

We then find that

$$\begin{aligned} L &= [\arcsin(x)]_0^1 \\ &= [\arcsin(1) - \arcsin(0)] \\ &= \left[\frac{\pi}{2} - 0\right] \\ &= \frac{\pi}{2} \end{aligned} \tag{4.35}$$

and so $L = \pi/2$; a fact which we can easily deduce without integration since the arc is precisely a quarter of the circle's circumference.

Example *Find the length of the curve* $f(x) = 1/x + x^3/12$ *between* $x = 2$ *and* $x = 5$

The curve $f(x)$ is shown in fig. 43.

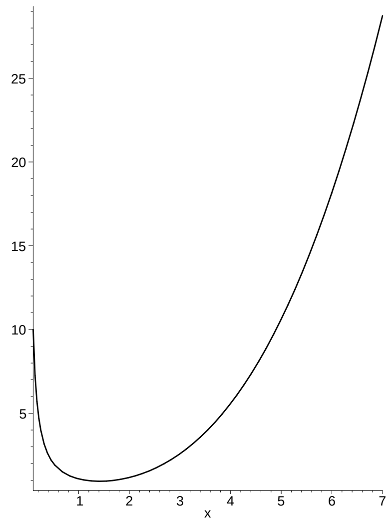


Fig. 43: The curve $f(x) = 1/x + x^3/12$

With

$$f(x) = \frac{1}{x} + \frac{x^3}{12} \tag{4.36}$$

we find that

$$f'(x) = -\frac{1}{x^2} + \frac{3x^2}{12} \tag{4.37}$$

and this gives us

$$\begin{aligned} 1 + f'(x)^2 &= 1 + \left(-\frac{1}{x^2} + \frac{3x^2}{12}\right)^2 \\ &= 1 + \frac{1}{x^4} - \frac{1}{2} + \frac{x^4}{16} \\ &= \frac{1}{x^4} + \frac{1}{2} + \frac{x^4}{16} \\ &= \left(\frac{1}{x^2} + \frac{x^2}{4}\right)^2 \end{aligned} \tag{4.38}$$

This means that the length L in question is given by

$$\begin{aligned} L &= \int_2^5 \sqrt{\left(\frac{1}{x^2} + \frac{x^2}{4}\right)^2}\, dx \\ &= \int_2^5 \left(\frac{1}{x^2} + \frac{x^2}{4}\right) dx \\ &= \left[-\frac{1}{x} + \frac{x^3}{12}\right]_2^5 \\ &= \left[-\frac{1}{5} + \frac{5^3}{12} + \frac{1}{2} - \frac{2^3}{12}\right] \\ &= \frac{201}{20} \end{aligned} \tag{4.39}$$

and so L is found.

Example *The length of a piece of a parabola*

For this example we take

$$f(x) = x^2 \tag{4.40}$$

which is a parabola and is displayed in fig. 44.

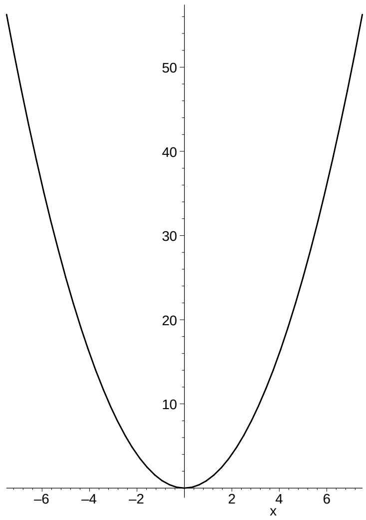


Fig. 44: The parabola $f(x) = x^2$

We readily compute that

$$\begin{aligned} f'(x) &= 2x \\ \Rightarrow \sqrt{1 + f'(x)^2} &= \sqrt{1 + 4x^2} \end{aligned} \tag{4.41}$$

Hence the length L of the section of this parabola between $x = a$ and $x = b$ is given by

$$L = \int_a^b \sqrt{1 + 4x^2}\, dx \tag{4.42}$$

If we choose, say, $a = 0$ and $b = 7$ then we find that

$$L = \int_0^7 \sqrt{1 + 4x^2}\, dx \tag{4.43}$$

Now if we simply *accept* that

$$\int \sqrt{1 + 4x^2}\, dx = \frac{1}{2}x\sqrt{1 + 4x^2} + \frac{1}{4}\operatorname{arcsinh}(2x) \tag{4.44}$$

a fact that we do *not* expect the reader to be able to derive. Then we find that

$$\begin{aligned} L &= \left[\frac{1}{2}x\sqrt{1 + 4x^2} + \frac{1}{4}\operatorname{arcsinh}(2x)\right]_0^7 \\ &= \left[\frac{1}{2}7\sqrt{1 + 4 \cdot 49} + \frac{1}{4}\operatorname{arcsinh}(14) - 0\right] \\ &= 49.95821036 \end{aligned} \tag{4.45}$$

where we used a calculator to evaluate the final expression.

§ 2. The Mean value of a function

Suppose we have a function f(x) whose value we measure n times thereby obtaining the series of values

$$f(x_1), f(x_2), \ldots f(x_n) \tag{4.46}$$

The mean, or average, of these n measurements is

$$\frac{f(x_1) + f(x_2) + \cdots + f(x_n)}{n} \tag{4.47}$$

We shall denote[2] this mean by $< f >$ so that

$$< f >= \frac{f(x_1) + f(x_2) + \cdots + f(x_n)}{n} \tag{4.48}$$

[2] The mean is also sometimes denoted by $\bar{f}$.

If n becomes very large—or indeed tends to infinity—then the resulting mean, which we shall still denote by $< f >$, can be expressed as an integral. We shall now explain this.

Suppose that the possible values of x all lie within the interval $[a, b]$ and suppose that we consider, for a moment the integral of f over this interval. Returning to 3.3 for the definition of an integral we have [3]

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \left(f(x_1) + f(x_2) + \cdots + f(x_n)\right) \Delta x \tag{4.49}$$

But since the interval $[a, b]$ is divided into precisely n pieces, each of size Δx, we have

$$\begin{aligned} n\Delta x &= b - a \\ \Rightarrow \Delta x &= \frac{(b-a)}{n} \end{aligned} \tag{4.50}$$

Substituting this into our expression for the integral gives

$$\begin{aligned} \int_a^b f(x)\,dx &= \lim_{n\to\infty} \left(f(x_1) + f(x_2) + \cdots + f(x_n)\right) \frac{(b-a)}{n} \\ &= \lim_{n\to\infty} \frac{\left(f(x_1) + f(x_2) + \cdots + f(x_n)\right)}{n} (b-a) \\ &= < f > (b-a) \end{aligned} \tag{4.51}$$

In other words we have shown that the mean $< f >$ of f over the interval $[a, b]$ is just $\int_a^b f\,dx$ divided by the length of the interval; that is

$$< f > = \frac{\displaystyle\int_a^b f(x)\,dx}{b-a} \tag{4.52}$$

The importance of this way of defining the mean is that it is very well suited to quantities that vary *continuously* with time such as the voltage output from an electrical device. Here is an example.

Example *The mean voltage produced by an oscillator*

An oscillator produces a current output $I(t)$ given by

$$I(t) = a\cos(10\omega t) + b\cos(\omega t) \tag{4.53}$$

where t stands for time—cf. fig. 45.

[3] Note when comparing with 3.3 that we have used n instead of $n-1$, and written Δx for $(x_{i+1} - x_i)$

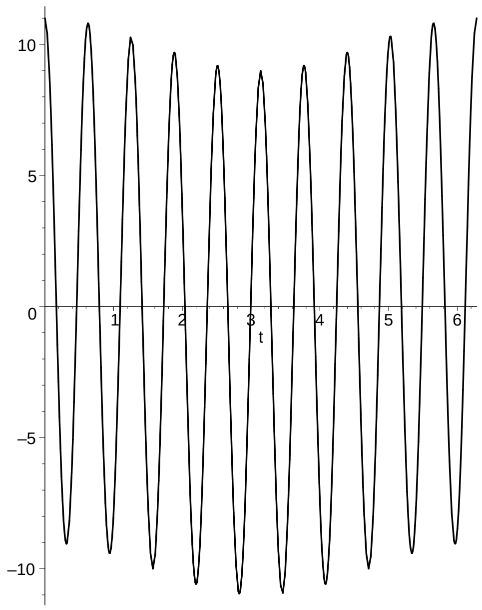


Fig. 45: The oscillator current $I(t) = a\cos(10\omega t) + b\cos(\omega t)$ for $a = 10$, $b = 1$, $\omega = 1$

Notice that $I(T)$ is a sum of a very rapidly varying term $a\cos(10\omega t)$ and a term of slower variation $b\cos(\omega t)$. The term of rapid variation has a periodic time T_{rapid} given by

$$T_{rapid} = \frac{2\pi}{10\omega} \tag{4.54}$$

and we shall now calculate the mean of $I(t)$ over this period. We have

$$< I >= \frac{1}{T_{rapid}} \int_0^{T_{rapid}} (a\cos(10\omega t) + b\cos(\omega t))\, dt, \quad T_{rapid} = \frac{2\pi}{10\omega} \tag{4.55}$$

But

$$\begin{aligned} \int_0^{T_{rapid}} (a\cos(10\omega t) + b\cos(\omega t))\, dt &= \left[\frac{a}{10\omega}\sin(10\omega t) + \frac{b}{\omega}\sin(\omega t)\right]_0^{T_{rapid}} \\ &= \left[\frac{a}{10\omega}\sin(10\omega T_{rapid}) + \frac{b}{\omega}\sin(\omega T_{rapid}) - 0\right] \\ &= \frac{a}{10\omega}\sin\left(10\omega\frac{2\pi}{10\omega}\right) + \frac{b}{\omega}\sin\left(\omega\frac{2\pi}{10\omega}\right) \\ &= \frac{a}{10\omega}\sin(2\pi) + \frac{b}{\omega}\sin\left(\frac{2\pi}{10}\right) \\ &= 0 + \frac{b}{\omega}\sin\left(\frac{2\pi}{10}\right) \end{aligned} \tag{4.56}$$

Hence we find that

$$\begin{aligned} < I > &= \frac{1}{T_{rapid}}\frac{b}{\omega}\sin\left(\frac{2\pi}{10}\right) \\ &= \frac{10\omega}{2\pi}\frac{b}{\omega}\sin\left(\frac{2\pi}{10}\right) \qquad (4.57) \\ \text{i.e.} \qquad < I > &= \frac{10b}{2\pi}\sin\left(\frac{2\pi}{10}\right) \end{aligned}$$

Note that the periodic time T of the *combined* system of both terms is just the periodic time of the slowest varying term; so in this case it is

$$T = \frac{2\pi}{\omega} \qquad (4.58)$$

and if we calculate the mean $< I >$ over this time T we shall find that

$$< I >= 0 \qquad (4.59)$$

because, during this time T, the graph of $I(T)$ spends exactly the same amount of time above the t axis as below (cf. fig. 45) and this causes the integral $\int_0^T I(t)\,dt$ to vanish.

This vanishing of the mean of a function is illustrated in the next example.

Example *The mean of the simple sinusoidal voltage* $V(t) = a\sin(\omega t)$

In this example we have a variable voltage which $V(t)$ is just a sine wave and so has periodic time $T = 2\pi/\omega$ and we can see at once that its mean is zero. Calculating the mean we have

$$\begin{aligned} < V > &= \frac{1}{T}\int_0^T a\sin(\omega t)\,dt, \quad T = \frac{2\pi}{\omega} \\ &= -\frac{1}{T}\left[\frac{a\cos(\omega t)}{\omega}\right]_0^T \\ &= -\frac{1}{T}\left[\frac{a\cos(\omega T)}{\omega} - \frac{a\cos(0)}{\omega}\right] \qquad (4.60) \\ &= -\frac{\omega}{2\pi}\left[\frac{a\cos(2\pi)}{\omega} - \frac{a}{\omega}\right] \\ &= -\frac{\omega}{2\pi}\left[\frac{a}{\omega} - \frac{a}{\omega}\right] \\ \text{i.e.} \qquad < V > &= 0 \end{aligned}$$

§ 3. The Root Mean Square or RMS value of a function

Since the mean of a voltage such as $V(t) = a\sin(\omega t)$ vanishes and so $< V >$ tells us little about such a voltage it is useful to introduce another related mean which gives more information about $V(t)$. This is called the *Root Mean Square value* or the *RMS value* of the function in question.

If f is any function then we shall denotes its RMS value on an interval $[a, b]$ by

$$f_{RMS} \tag{4.61}$$

where the definition of f_{RMS} is that

$$f_{RMS} = \sqrt{< f^2 >}$$

So f^2_{RMS} is the average of f^2 rather than f. The important point to note then is that since f^2 is *always positive*—being the square of something—then f_{RMS} will never be zero unless f itself is identically zero which is a trivial case.

In any case we have

$$f_{RMS} = \sqrt{< f^2 >} \tag{4.62}$$

and it is time to do an example.

Example *The RMS value of the voltage $V(t) = V_0\sin(\omega t)$ over its period T*

First we note from our previous work above that the period $T = 2\pi/\omega$ and then we calculate

$$< V^2 > \tag{4.63}$$

This is given by

$$\begin{aligned}
< V^2 > &= \frac{1}{T}\int_0^T V^2(t)\,dt \\
&= \frac{1}{T}\int_0^T V_0^2 \sin^2(\omega t)\,dt \\
&= \frac{1}{T}\int_0^T V_0^2 \frac{1}{2}\left(1+\cos(2\omega t)\right)\,dt, \qquad \text{using } \sin^2(t) = \frac{1}{2}\left(1+\cos(2t)\right) \\
&= \frac{1}{T}\frac{V_0^2}{2}\left[t + \frac{1}{2\omega}\sin(2\omega t)\right]_0^T \\
&= \frac{1}{T}\frac{V_0^2}{2}\left[T + \frac{1}{2\omega}\sin(2\omega T) - 0\right] \\
&= \frac{\omega}{2\pi}\frac{V_0^2}{2}\left[\frac{2\pi}{\omega} + \frac{1}{2\omega}\sin(4\pi)\right] \\
&= \frac{\omega}{2\pi}\frac{V_0^2}{2}\left[\frac{2\pi}{\omega} + 0\right] \\
&= \frac{V_0^2}{2}
\end{aligned} \tag{4.64}$$

So $V_{RMS} = \sqrt{< V^2 >}$ is given by

$$V_{RMS} = \frac{V_0}{\sqrt{2}} \tag{4.65}$$

This is a well known result in electrical engineering.

The reader should note that, since $V(t) = V_0 \sin(\omega t)$, then the *largest value* that $V(t)$ can have is when $\sin(\omega t)$ has its largest value which is unity; this in turn means that the maximum value of $V(t)$ is V_0 and V_0 is often called the *peak voltage*. So it is both common and useful to think of V_{RMS} as being $1/\sqrt{2}$ times the peak voltage; it is worth memorising the approximate value of $1/\sqrt{2}$ so we quote it here

$$\frac{1}{\sqrt{2}} = 0.707\ldots \tag{4.66}$$

We see that V_{RMS} is about 70% of the peak voltage V_0. If a country uses an AC system of voltage supply the voltage usually quoted is the RMS value—230 volts for Ireland.

§ 4. Numerical Integration

We have seen already that many integrals are hard to do in closed from and so have to be done numerically. there are many techniques for doing this. Two simple ones which are often introduced at the beginning of a discussion on numerical integration are the *trapezoidal rule* and *Simpson's rule*; the latter being superior to the former. We shall have a very brief look at Simpson's rule.

§§ 4.1 Simpson's rule

Suppose we want to evaluate

$$\int_a^b f(x)\, dx \tag{4.67}$$

then we divide the interval $[a, b]$ up into

$$2n \tag{4.68}$$

intervals so that each has size

$$h = \frac{b-a}{2n} \tag{4.69}$$

Now we select $2n + 1$ values of x in $[a, b]$ given by $x_j = a + jh$ for $j = 0, \ldots 2n$, i.e.

$$\begin{aligned} x_0 &= a \\ x_1 &= a + h \\ x_2 &= a + 2h \\ \vdots \quad & \quad \vdots \\ x_{2n-1} &= x + 2(n-1)h \\ x_{2n} &= b \end{aligned} \tag{4.70}$$

With this accomplished *Simpson's rule* gives the following *approximate* value for the integral

$$\int_a^b f(x)\,dx \sim \frac{b-a}{6n}\{f(a) + 2(f(x_1) + f(x_3) + \cdots + f(x_{2n-1})) \\ + 4(f(x_2) + f(x_4) + \cdots + f(x_{2n-2})) + f(b)\} \tag{4.71}$$

and the *error* is proportional to

$$\frac{1}{n^4} \tag{4.72}$$

We simply quote this formula 4.71 as we have no space left to give its (quite straightforward) derivation. We finish by using Simpson's rule in an example.

Example *The integral*

$$\int_0^2 \sqrt{x}\sin(x)\,dx \tag{4.73}$$

First if we use a computer package such as *Maple*, *Mathematica* or *Matlab* then they have very good in built numerical integration routines and using Maple we can ask for, say, 10 significant figures of accuracy and then we obtain the result that

$$\int_0^2 \sqrt{x}\sin(x)\,dx = 1.235911456 \tag{4.74}$$

Now if we use Simpson's rule, and use Maple to evaluate 4.71 for $n = 10$, 20, 30, 40, 50, 60 and 70, we obtain the set of values

$$\begin{array}{c} 1.235915368 \\ 1.235911716 \\ 1.235911508 \\ 1.235911474 \\ 1.235911468 \\ 1.235911459 \\ 1.235911456 \end{array} \tag{4.75}$$

and the reader can see the sort of accuracy of Simpson's rule. We obtain the correct answer to 10 significant figures with $n = 70$. It is now easy to treat many more examples and we leave this task to the reader and bring these lectures to a close here.

Contents

CHAPTER III

CHAPTER IV