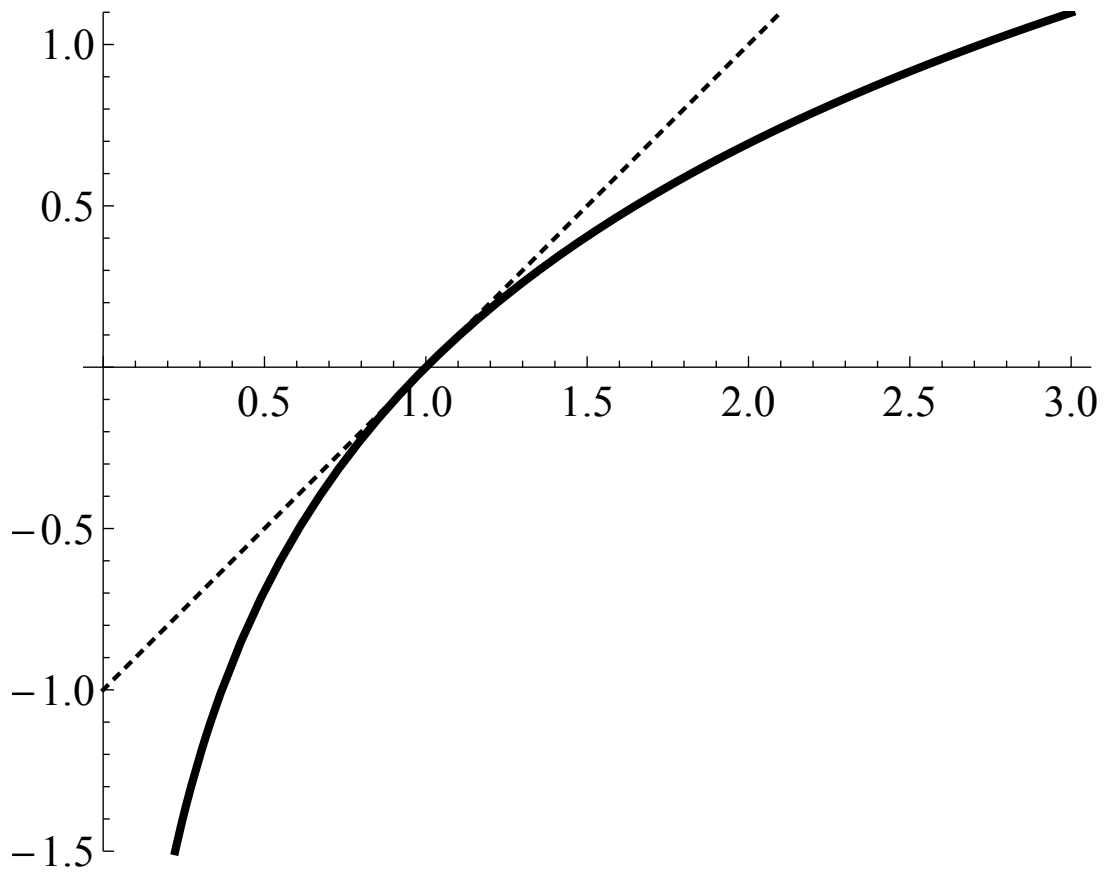


# Analysis II: MA131

Keith Ball



# Contents

<b>Continuous functions</b>	<b>3</b>
Continuity . . . . .	3
The Intermediate Value Theorem . . . . .	8
Boundedness and attainment of bounds . . . . .	12
<b>Power series I</b>	<b>14</b>
Basic properties of power series . . . . .	14
The continuity of power series . . . . .	17
The exponential . . . . .	19
Inequalities for the exponential . . . . .	23
The logarithm and powers . . . . .	24
<b>Limits and the derivative</b>	<b>27</b>
Limits . . . . .	28
One sided limits . . . . .	30
The derivative . . . . .	33
The derivative machine . . . . .	35
The Mean Value Theorem . . . . .	39
Derivatives of inverses . . . . .	44
<b>Power series II</b>	<b>47</b>
The differentiability of power series . . . . .	47
The trigonometric functions . . . . .	53
The complex exponential . . . . .	58
The tangent . . . . .	59
<b>Taylor's Theorem</b>	<b>60</b>
Cauchy's Mean Value Theorem . . . . .	60
L'Hôpital's rule . . . . .	61
Taylor's Theorem with remainder . . . . .	64
<b>The radius of convergence formula</b>	<b>73</b>

# Chapter 1. Continuous functions

## Continuity

Given two sets  $A$  and  $B$  a *function*  $f$  from  $A$  to  $B$  assigns an element of  $B$  to each element of  $A$ . Thus every function comes “equipped” with two sets:  $A$ , the set of points where the function is defined, and  $B$ , a set to which the values of the function belong. We can draw attention to these sets by writing

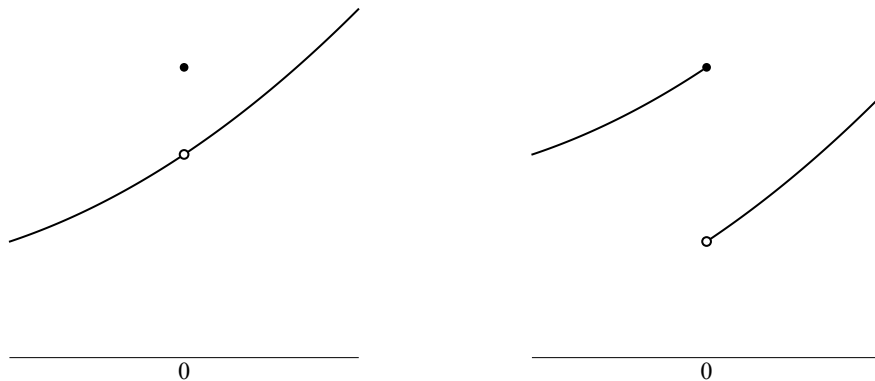
$$f : A \rightarrow B.$$

The set  $A$  is called the *domain* of  $f$ ;  $B$  is called its *codomain*. For each element  $x$  of the domain we write  $f(x)$  for the place to which  $x$  is sent: the *image* of  $x$  under  $f$ .

An interval of the real line is a subset of  $\mathbf{R}$  with the property that if  $x < y < z$  and  $x$  and  $z$  both belong to the subset then so does  $y$ . So an interval contains all the points between its ends, but the ends themselves may or may not be included. Examples are

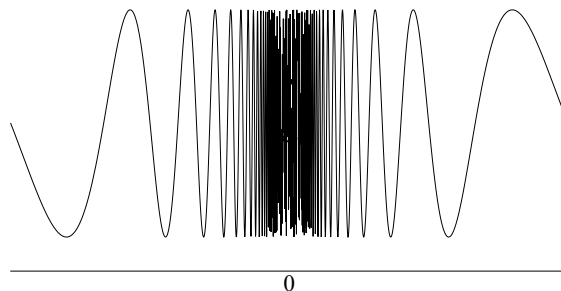
$$\begin{aligned}\{x : a \leq x \leq b\} &= [a, b] \quad \text{a closed interval} \\ \{x : a < x < b\} &= (a, b) \quad \text{an open interval} \\ \{x : a \leq x < b\} &= [a, b) \quad \text{a half-open interval} \\ \{x : a \leq x\} &= [a, \infty) \quad \text{a half-infinite interval}\end{aligned}$$

For this course we shall be considering real valued functions whose domains are intervals of the real line. We want to say what it means for such a function to be continuous. At each point of the domain there are essentially three types of problem that can occur. In each of the cases below, the function is discontinuous at 0 (but continuous elsewhere).



The first two are fairly straightforward. In the first case the function has a value at zero that doesn't "match" the rest of the function. In the second case, the function jumps in value as we pass through 0.

In the third situation  $f(x)$  oscillates infinitely often over a wide range as  $x$  approaches 0.

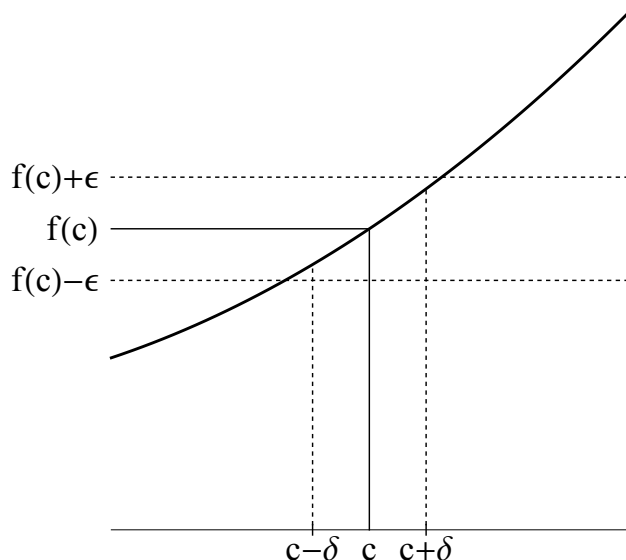


We shall define what it means for a function to be continuous at a point in such a way as to rule out these problems. Then we say that a function is continuous on an interval if it is continuous at each point of the interval.

**Definition (Continuity).** A function  $f : I \rightarrow \mathbf{R}$  defined on an interval  $I$  containing the number  $c$  is said to be continuous at  $c$  if for every  $\varepsilon > 0$ , there is a  $\delta > 0$  so that if  $x \in I$  and  $|x - c| < \delta$  then

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

The function is continuous on  $I$  if it is continuous at each point of  $I$ .



The picture shows the situation. The number  $\varepsilon$  specifies a band around the value  $f(c)$ . You want to be able to guarantee that  $f(x)$  falls into that band as long as  $x$  is close enough to  $c$ .

**Example (Continuity of  $x \mapsto x$ ).** *The function  $x \mapsto x$  is continuous at every point of the real line.*

*Proof* If  $f(x) = x$  then we can always take  $\delta = \varepsilon$  since if  $|x - c| < \varepsilon$  then

$$|f(x) - f(c)| = |x - c| < \varepsilon.$$

□

As an exercise, you can check that constant functions are continuous.

**Example (A discontinuity).** *The function  $f$  given by*

$$f(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

*is discontinuous at 0.*

*Proof* We need to find a band around  $f(0) = 0$  which we cannot guarantee to land in, merely by starting near 0. A band of width  $\varepsilon = 1/2$  will do because if  $x$  is positive, however close to zero,  $f(x) = 1$  and this is not within  $1/2$  of  $f(0)$ . □

We will do one more example which will also follow from our later arguments but will help to fix the ideas: the function  $x \mapsto x^2$ . The aim will be to show that if  $x - c$  is small then so is  $x^2 - c^2$ . Now  $x^2 - c^2 = (x - c)(x + c)$  so if  $|x - c| < \delta$  then  $|x^2 - c^2| < \delta|x + c|$  which looks good because the factor  $\delta$  is small. But the second factor  $|x + c|$  might be large (if  $c = 1,000,000$  say). However, if  $x$  is close to  $c$  then  $x + c$  is close to  $2c$  which may be large but is a fixed number. So we shall cook up  $\delta$  to compensate: roughly we want  $\delta = \varepsilon/(2|c|)$ . However we need to be a bit careful because  $x + c$  isn't actually equal to  $2c$ .

**Example (Continuity of  $x \mapsto x^2$ ).** *The function  $x \mapsto x^2$  is continuous at every point of the real line.*

*Proof* Let  $f(x) = x^2$ , let  $c \in \mathbf{R}$  and  $\varepsilon > 0$ . We want to guarantee that

$$|x^2 - c^2| < \varepsilon$$

by choosing  $x$  close enough to  $c$ . We know that  $|x^2 - c^2| = |x - c| \cdot |x + c|$ .

If we choose  $|x - c| < 1$  then  $|x|$  cannot be larger than  $|c| + 1$  and so  $|x + c|$  cannot be larger than  $2|c| + 1$ . Then if we choose

$$|x - c| < \frac{\varepsilon}{2|c| + 1}$$

we will get

$$|x^2 - c^2| = |x - c| \cdot |x + c| < \frac{\varepsilon}{2|c| + 1} (2|c| + 1) = \varepsilon$$

which is what we wanted.

We end up needing two conditions on  $|x - c|$ : namely  $|x - c| < 1$  and  $|x - c| < \frac{\varepsilon}{2|c| + 1}$ . But that is not a problem. If we choose  $\delta$  to be the smaller of 1 and  $\frac{\varepsilon}{2|c| + 1}$  the two conditions will be satisfied simultaneously as long as  $|x - c| < \delta$ .  $\square$

We wish to check the continuity of functions such as polynomials, more complicated than  $x \mapsto x^2$ . To do this we want a machine which allows us to build continuous functions from simpler ones. We want to know that when you add or multiply continuous functions the result is still continuous and also when you compose two continuous functions. There are a number of ways to do this but in view of what you already know about limits there is a particularly simple approach which depends upon rewriting the continuity property in terms of limits of sequences.

**Theorem (Sequential continuity).** *Let  $f : I \rightarrow \mathbf{R}$  be defined on the interval  $I$  and suppose that  $c \in I$ . Then  $f$  is continuous at  $c$  if and only if for every sequence  $(x_n)$  of points in  $I$  which converges to  $c$ ,*

$$f(x_n) \rightarrow f(c) \quad \text{as } n \rightarrow \infty.$$

*Proof* Suppose first that  $f$  is continuous at  $c$  and  $x_n \rightarrow c$ . Then, given  $\varepsilon > 0$  we can find  $\delta > 0$  so that if  $|x - c| < \delta$  then  $|f(x) - f(c)| < \varepsilon$ . Now choose  $N$  so that if  $n > N$ ,  $|x_n - c| < \delta$ . Then if  $n > N$  we have  $|f(x_n) - f(c)| < \varepsilon$ . So  $f(x_n) \rightarrow f(c)$ .

On the other hand suppose  $f$  is *not* continuous at  $c$  and choose  $\varepsilon > 0$  with the property that whatever  $\delta$  we pick there is a point  $x$  within  $\delta$  of  $c$  for which

$$|f(x) - f(c)| \geq \varepsilon.$$

Now build a sequence as follows. For each  $n$  choose a point  $x_n$  with  $|x_n - c| < 1/n$  but  $|f(x_n) - f(c)| \geq \varepsilon$ . Then  $x_n \rightarrow c$  as  $n \rightarrow \infty$  but  $f(x_n)$  does not converge to  $f(c)$ .  $\square$

The sequential continuity theorem immediately gives us the algebra of continuous functions.

**Theorem (Algebra of continuous functions).** *Let  $f, g : I \rightarrow \mathbf{R}$  be defined on the interval  $I$  and continuous at  $c \in I$ . Then*

1.  $f + g$  is continuous at  $c$
2.  $f \cdot g$  is continuous at  $c$
3. if  $g(c) \neq 0$  then  $f/g$  is continuous at  $c$ .

*Proof* The proofs of all 3 are essentially the same. We shall do the first. We wish to show that if  $x_n \rightarrow c$  then  $(f + g)(x_n) = f(x_n) + g(x_n) \rightarrow f(c) + g(c)$  as  $n \rightarrow \infty$ . But we know that  $f(x_n) \rightarrow f(c)$  and  $g(x_n) \rightarrow g(c)$  so we can apply the properties of limits of sequences to conclude that  $f(x_n) + g(x_n) \rightarrow f(c) + g(c)$ .  $\square$

From this we can immediately conclude that polynomials are continuous and that rational functions are continuous except where the denominator is zero.

**Corollary (Continuity of polynomials and rational functions).** *If  $p$  is a polynomial then  $p$  is continuous at every point of  $\mathbf{R}$ . If  $r = p/q$  is a ratio of two polynomials then it is continuous at every point of  $\mathbf{R}$  where  $q \neq 0$ .*

*Proof* Exercise.  $\square$

To complete the continuity *machine* we need to know that we can compose continuous functions. The statement looks a bit complicated but the idea is as simple as for the algebraic properties.

**Theorem (Composition of continuous functions).** *Let  $f : I \rightarrow \mathbf{R}$  be defined on the interval  $I$ ,  $g : J \rightarrow I$  be defined on the interval  $J$ . If  $g$  is continuous at  $c$  and  $f$  is continuous at  $g(c)$  then the composition  $f \circ g$  is continuous at  $c$ .*

*Proof* Let  $(x_n)$  be a sequence in  $J$  converging to  $c$ . Then  $g(x_n) \rightarrow g(c)$  in  $I$  and hence  $f(g(x_n)) \rightarrow f(g(c))$ .  $\square$

## The Intermediate Value Theorem

This section is devoted to a principle which has many uses and which you may have met at school. Roughly it says that if a continuous function crosses from below zero to above zero then there must be a point where it is equal to zero: it cannot *jump across* zero.

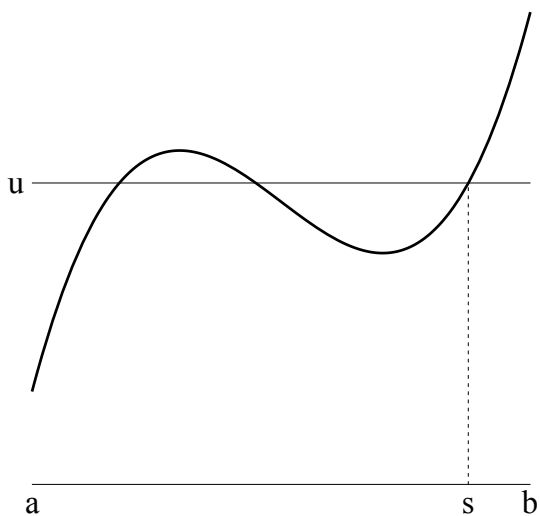
Our intuition about continuous functions is that we can draw them without taking pen from paper and therefore it should be “obvious” that a continuous function can’t skip a value. However the proof is not quite so simple because we have defined continuity at each point individually. We therefore have to find the point where the function is supposed to take the correct value and then use continuity at that point. When we are trying to demonstrate the existence of a particular real number we usually have to invoke the completeness principle. The completeness principle is our way to finger a real number.

**Theorem (Intermediate Value Theorem).** *Let  $f : [a, b] \rightarrow \mathbf{R}$  be continuous and suppose that  $u$  lies between  $f(a)$  and  $f(b)$ . Then there is a point  $c$  between  $a$  and  $b$  where  $f(c) = u$ .*

*Proof* Assume that  $f(a) < u < f(b)$  and let  $A$  be the set

$$\{x \in [a, b] : f(x) \leq u\}.$$

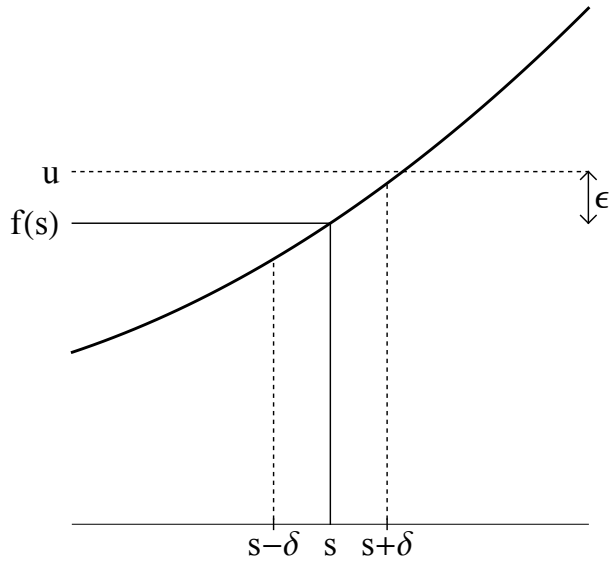
This set is non-empty since it contains  $a$  and is bounded above by  $b$ . Let  $s$  be its least upper bound. The aim is to show that  $f(s) = u$ .



We shall eliminate the two other possibilities  $f(s) < u$  and  $f(s) > u$  separately.



Suppose that we have  $f(s) < u$ .

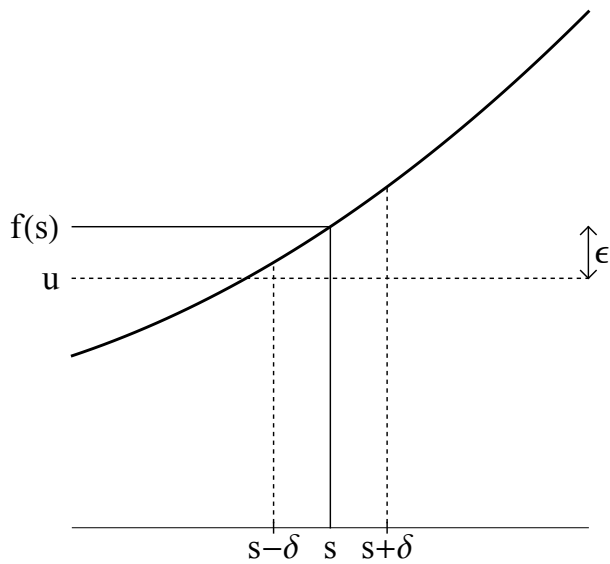


Note that that  $s \neq b$  because we know  $f(b) > u$ . If we put  $\epsilon = u - f(s)$  then for some  $\delta > 0$

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

as long as  $|x - s| < \delta$ . So in particular if  $x = s + \delta/2$  then  $f(x) < f(s) + \epsilon = u$ . This means that  $s + \delta/2 \in A$  contradicting the fact that  $s$  is an upper bound for  $A$ .

Suppose on the other hand that  $f(s) > u$ .



Note that  $s \neq a$  because  $f(a) < u$ . If we put  $\epsilon = f(s) - u$  then for some  $\delta > 0$

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

as long as  $|x - s| < \delta$ . Therefore if  $s - \delta < x \leq s$ ,  $f(x) > f(s) - \epsilon = u$  and hence  $x$  is *not* in  $A$ . This means that  $s - \delta$  is an upper bound for  $A$  contradicting the fact that  $s$  is the *least* upper bound.  $\square$

Note that there were two things we had to check about the point  $s$  that we found: that  $f(s)$  was not too small and that it was not too large. We used the two different properties of the least upper bound respectively for these two checks. This is a common feature of arguments using least upper bounds.

It is possible for a function to satisfy the intermediate value property without being continuous but it has to be a rather nasty function. Note that the IVT says that if  $u$  lies between  $f(a)$  and  $f(b)$  then there is a point  $c$  *between*  $a$  and  $b$  where  $f(c) = u$ . Try drawing a function with this property that is not continuous and you will be forced to wiggle a lot.

The IVT has many uses. The most obvious is that it guarantees the existence of solutions to equations. You already saw that each positive number has a positive square root (found as a limit). We can now deduce this very simply from the IVT.

**Corollary (The Existence of Square Roots).** *Every positive real number has an unique positive square root.*

*Proof* Let  $R$  be a positive number and consider the function  $f : x \mapsto x^2$ . Notice that  $f(0) < R$  and that for example

$$f(R + 1) = R^2 + 2R + 1 > R.$$

So there is a number  $c$  between 0 and  $R + 1$  where  $f(c) = R$ . Since  $c^2 = R$  we know that  $c$  is a positive square root of  $R$ .

To see that  $\sqrt{R}$  is unique we observe that if  $0 < x < y$  then  $x^2 < y^2$  and so two distinct positive numbers cannot have the same square.  $\square$

**Example (The existence of a solution to an equation).** *There is a solution of the equation  $x^3 + x - 1 = 0$  between 0 and 1.*

*Proof* Because if  $f(x) = x^3 + x - 1$  then  $f(0) = -1 < 0$  while  $f(1) = 1 > 0$ .

The existence of the square root is one example of a more general principle that follows from the IVT. Whenever we have a continuous, strictly increasing function we can find an inverse for it: just as  $y \mapsto \sqrt{y}$  is the inverse of the function  $x \mapsto x^2$  on  $[0, \infty)$ . An obvious question that we didn't address in the case of the square root is whether the inverse function is also continuous. In order to make this precise we define the *range* of a function  $f : A \rightarrow B$  to be the set of values that  $f$  takes:

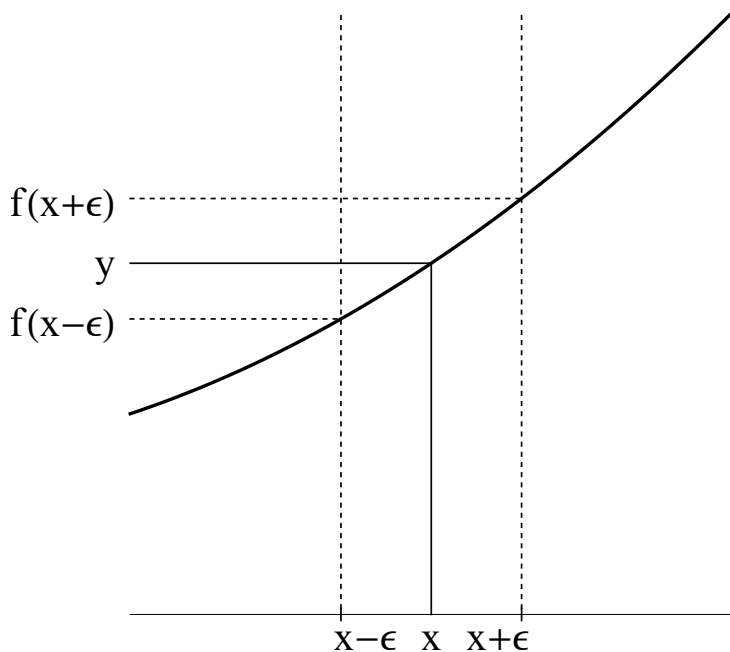
$$\{f(x) : x \in A\}.$$

**Corollary (Continuous image of an interval).** *If  $f : I \rightarrow \mathbf{R}$  is continuous on the interval  $I$  then its range is an interval.*

*Proof* If  $x$  and  $y$  are in the range then by the IVT so is any point between  $x$  and  $y$ . So the range is an interval.  $\square$

**Theorem (Existence of inverses).** *Let  $f : [a, b] \rightarrow \mathbf{R}$  be continuous and strictly increasing. Then  $f$  has an inverse defined on its range and  $f^{-1}$  is continuous.*

The same holds if  $f$  is defined on an open interval or on the whole of  $\mathbf{R}$ . The proofs are essentially the same. The picture says it all. If you want to end up within  $\varepsilon$  of  $x = g(y)$  then make sure you start between  $f(x - \varepsilon)$  and  $f(x + \varepsilon)$ .



*Proof* Let  $f(a) = c$  and  $f(b) = d$ . Since  $f$  is increasing all its values lie between  $c$  and  $d$ . So the range of  $f$  is exactly the interval  $[c, d]$ .

For each  $y$  there is a unique number  $x$  with  $f(x) = y$  because  $f$  is strictly increasing, so we may define the inverse  $g = f^{-1}$  by setting  $g(y) = x$  in each case. Clearly  $g$  is strictly increasing. The final aim is to show that  $g$  is continuous. Suppose  $y$  lies in the open interval  $(c, d)$ ,  $y = f(x)$  and  $\varepsilon > 0$ . We have  $f(x - \varepsilon) < y < f(x + \varepsilon)$ . Hence there is some  $\delta > 0$  with the property that

$$f(x - \varepsilon) < y - \delta < y < y + \delta < f(x + \varepsilon).$$

For any  $z$  between  $y - \delta$  and  $y + \delta$  the fact that  $g$  is increasing ensures that

$$x - \varepsilon < g(z) < x + \varepsilon$$

and so  $|g(z) - g(y)| < \varepsilon$ .

If  $y$  is one of the end points then the argument is the same but with one of  $f(x - \varepsilon)$  and  $f(x + \varepsilon)$  replaced by the corresponding end point of  $[c, d]$ .  $\square$

**Corollary (Existence of roots).** *For each positive  $x$  and natural number  $n$  there is a unique positive  $n^{\text{th}}$  root  $x^{1/n}$ . The map*

$$x \mapsto x^{1/n}$$

*is continuous.*

**Example (A composition).** *Let  $f : \mathbf{R} \rightarrow \mathbf{R}$  be given by*

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

*Then  $f$  is continuous on  $\mathbf{R}$ .*

## Boundedness and attainment of bounds

A continuous function on an open interval such as  $(0, 1)$  may take arbitrarily large values. For example the function  $x \mapsto 1/x$  approaches  $\infty$  as  $x$  approaches 0. This particular problem can't occur if the function is continuous on the closed interval  $[0, 1]$  since the function is supposed to approach  $f(0)$  as  $x$  approaches 0. It turns out that indeed a continuous function on a closed interval must be bounded.

**Theorem (Boundedness of continuous functions).** *Let  $f : [a, b] \rightarrow \mathbf{R}$  be continuous. Then  $f$  is bounded.*

The proof uses a method that used to be called rather appealingly “the condensation of singularities”. The idea is that if the function is unbounded we can find a sequence of points where it gets bigger and bigger. Using Bolzano-Weierstrass we find a subsequence of these points which converges. So a sequence of bad points converges to a point which we can conclude is so bad that it can’t exist.

*Proof* Suppose  $f$  is unbounded. For each  $n$  choose  $x_n \in [a, b]$  where  $|f(x_n)| \geq n$ . Now choose a subsequence  $x_{n_k}$  which converges to  $x$  say. Since the interval  $[a, b]$  is closed we must have  $x \in [a, b]$ . Then we know that  $f(x_{n_k}) \rightarrow f(x)$  but this can’t happen since the values  $f(x_{n_k})$  are becoming arbitrarily large.  $\square$

At school you were often asked to find the maximum value of a function on some interval but probably never addressed the question of whether a nice function (a continuous one say) really must have a maximum. Obviously the previous example of  $1/x$  shows that on an open interval a function need not have a maximum but this is also true even if the function is bounded. The function  $x \mapsto x$  has no maximum on  $(0, 1)$ . Its least upper bound is 1 but there is no point in the interval where the function is equal to 1. Again, this situation can’t happen on closed intervals.

**Theorem (Attainment of bounds).** *Let  $f : [a, b] \rightarrow \mathbf{R}$  be continuous. Then  $f$  has a maximum and a minimum attained in the interval.*

*Proof* Let  $M$  be the least upper bound of the set

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

If there is no point in the interval where  $f(x) = M$  then the function  $g$  given by  $g(x) = M - f(x)$  is strictly positive and continuous on the interval. Hence by the algebra of continuous functions its reciprocal is continuous and therefore bounded. Let’s say

$$\frac{1}{M - f(x)} \leq R$$

for all  $x \in [a, b]$ . Then  $1/R \leq M - f(x)$  and hence  $f(x) \leq M - 1/R$  for all  $x$  in the interval. This shows that  $M - 1/R$  is an upper bound for the set of values and contradicts the fact that  $M$  is the *least* upper bound.

A similar argument works for the minimum.  $\square$

This theorem will form the basis of our proof of the so-called Mean Value Theorem which plays a key role in many applications of the derivative.

## Chapter 2. Power series I

A power series is a series of the form

$$\sum_{n=0}^{\infty} a_n x^n$$

(or more generally

$$\sum_{n=0}^{\infty} a_n (x - x_0)^n$$

in which we centre the series at a number  $x_0$  rather than at 0). We will develop the theory just for the first type: everything transfers immediately to the more general case.

We think of a power series as a kind of infinite polynomial in which the  $(a_n)$  are coefficients. The series may or may not converge depending upon the value of  $x$ . Where it does converge it defines a function of  $x$  and we shall see that the function will automatically be continuous and differentiable. Many of the most important functions in mathematics can be written as power series: in particular the exponential function.

### Basic properties of power series

We want to know the values of  $x$  for which a power series  $\sum_0^{\infty} a_n x^n$  converges. Of course the answer will depend upon the specific  $a_n$ , but there is a fundamental feature of the answer that is common to all power series. The set of values where a power series converges will be an interval (possibly infinite) of the real line. The idea is quite simple: if  $\sum a_n t^n$  converges and we choose  $x$  to be smaller than  $t$  in size then the terms  $a_n x^n$  will be much smaller than the terms  $a_n t^n$  so we should expect  $\sum a_n x^n$  to converge as well.

We would like to prove this just by quoting the Comparison Test for series but that doesn't quite work because the series  $\sum a_n t^n$  might not converge absolutely. However we can get around this problem because  $a_n x^n$  is not just smaller than  $a_n t^n$  but is *very much* smaller when  $n$  is large.

**Theorem (Radius of convergence I).** Let  $\sum_0^\infty a_n x^n$  be a power series with  $\sum a_n t^n$  convergent. Then

$$\sum_0^\infty a_n x^n$$

converges absolutely for all  $x$  with  $|x| < |t|$ .

*Proof* Since  $\sum a_n t^n$  converges we know that  $a_n t^n \rightarrow 0$  as  $n \rightarrow \infty$  and so the sequence is bounded. There is some  $M$  for which  $|a_n t^n| < M$  for all  $n$ . Now

$$\begin{aligned} \sum_0^N |a_n x^n| &= \sum_0^N |a_n t^n| \left| \frac{x}{t} \right|^n \\ &\leq M \sum_0^N \left| \frac{x}{t} \right|^n \\ &\leq M \sum_0^\infty \left| \frac{x}{t} \right|^n \\ &= M \frac{1}{1 - |x|/|t|} \end{aligned}$$

Hence  $\sum_0^\infty |a_n x^n| < \infty$  and the series converges absolutely. □

Using this principle we can introduce what is called the radius of convergence.

**Theorem (Radius of convergence II).** Let  $\sum_0^\infty a_n x^n$  be a power series. One of the following holds.

- The series converges only if  $x = 0$ .
- The series converges for all real numbers  $x$ .
- There is a positive number  $R$  with the property that the series converges if  $|x| < R$  and diverges if  $|x| > R$ .

In the third case the number  $R$  is called the **radius of convergence**. In the first case we say that the radius of convergence is 0 and in the second that it is  $\infty$ .

*Proof* Obviously the series does converge if  $x = 0$ . If the first option does not hold than it converges for some other values of  $x$ .

The set of  $x$  for which it converges might be unbounded: it might contain arbitrarily large numbers. In that case the power series will converge absolutely for all real  $x$  and defines a function on  $\mathbf{R}$ .

Otherwise the set might contain some numbers other than 0 but be bounded. In that case we let

$$R = \sup\{|t| : \sum a_n t^n \text{ converges}\}.$$

Then by our previous theorem the series converges absolutely whenever  $|x| < R$ . It does not converge if  $|x| > R$  by definition of  $R$ .  $\square$

As an immediate corollary we get an observation that will be useful later.

**Corollary (Absolute series).** *Let  $\sum_0^\infty a_n x^n$  be a powers series with radius of convergence  $R$ . Then  $\sum_0^\infty |a_n| x^n$  also has radius of convergence  $R$ .*

**Example (The geometric series I).** *The series  $\sum_0^\infty x^n$  has radius of convergence  $R = 1$ .*

*Proof* We know that the series converges if and only if  $x \in (-1, 1)$ .  $\square$

**Example (The geometric series II).** *If  $p$  is real the series  $\sum_0^\infty p^n x^n$  has radius of convergence  $R = 1/|p|$ .*

*Proof* We know that the series converges if and only if  $px \in (-1, 1)$  which is the same as saying that  $x \in (-1/|p|, 1/|p|)$ .  $\square$

**Example (The log series).** *The series  $\sum_0^\infty \frac{x^n}{n}$  has radius of convergence  $R = 1$ .*

*Proof* From homework we know that the series converges if and only if  $x \in [-1, 1)$ .  $\square$

Notice that although we have put in the extra factor  $1/n$  in front of  $x^n$  we have not changed the radius of convergence. The point is that if we increase  $x$  beyond 1, its powers  $x^n$  increase exponentially fast and this swamps the effect of the factor  $1/n$ . However, the extra factor of  $1/n$  does make the series converge at  $x = -1$  but not absolutely.

**Example.** *The series  $\sum_0^\infty nx^n$  has radius of convergence  $R = 1$ .*



*Proof* From homework we know that the series converges if and only if  $x \in (-1, 1)$ .  $\square$

For most interesting power series we can find the radius of convergence using the ratio test. Towards the end of the course we will see a formula for the radius of convergence which works for all series. As a taste let us look at another example:

**Example.** *The series  $1 + 0x + x^2 + 0x^3 + x^4 + \dots$  has radius of convergence  $R = 1$ .*

*Proof* The series is

$$1 + x^2 + x^4 + x^6 + \dots$$

which is a geometric series with ratio  $x^2$ . So it converges if and only if  $x^2 \in (-1, 1)$ . However, there is another way to do it which is a much more flexible method.

If  $x > 1$  then the terms do not tend to zero so the series diverges. If  $0 < x < 1$  then

$$1 + x^2 + x^4 + \dots < 1 + x + x^2 + x^3 + x^4 + \dots$$

which we already know converges. This means that the series converges if  $0 < x < 1$  so the radius of convergence is 1.  $\square$

We remarked that it is possible to consider power series centred at points other than 0. For example

$$\begin{aligned} \frac{1}{1-x} &= \frac{1}{2-(x+1)} = \frac{1}{2} \frac{1}{1-(x+1)/2} \\ &= \frac{1}{2} \left( 1 + \frac{x+1}{2} + \left( \frac{x+1}{2} \right)^2 + \dots \right) = \frac{1}{2} + \frac{x+1}{4} + \frac{(x+1)^2}{8} + \dots \end{aligned}$$

This converges if  $\left| \frac{x+1}{2} \right| < 1$ : in other words if  $x$  differs from  $-1$  by at most 2. The series is centred at  $-1$  and has radius of convergence 2. This means that it converges on  $(-3, 1)$ . Notice that it converges “as far as it possibly can”. The function has an asymptote at  $x = 1$  so the series cannot represent the function at  $x = 1$ .

## The continuity of power series

As was remarked several times, many of the most important functions in mathematics are given by power series so we want to know that such functions have nice properties: that they are continuous for example.

**Theorem (Continuity of power series).** Let  $\sum_0^\infty a_n x^n$  with radius of convergence  $R$ . Then the function

$$x \mapsto \sum_0^\infty a_n x^n$$

is continuous on the interval  $(-R, R)$ .

We already saw some examples in which the power series converges at one or both ends of the interval  $[-R, R]$ . If so it makes sense to ask whether the function is continuous on this larger set. Perhaps surprisingly this is a bit more subtle than the statement above. There are some quite slick proofs of the continuity of power series. We will take a very down to earth approach. Inside the radius of convergence a power series can be approximated by a polynomial: and polynomials are continuous.

*Proof* Suppose  $-R < x < R$ . We want to show that the function is continuous at  $x$ . Choose a number  $T$  with  $|x| < T < R$ . Then the series  $\sum |a_n| T^n$  converges, so for each  $\varepsilon > 0$  there is some number  $N$  for which

$$\sum_{N+1}^\infty |a_n| T^n < \varepsilon/3.$$

Now if  $|y - x| < T - |x|$  we will have  $|y| < T$  as well as  $|x| < T$ . Hence

$$\sum_{N+1}^\infty |a_n| |x|^n < \varepsilon/3 \quad \text{and} \quad \sum_{N+1}^\infty |a_n| |y|^n < \varepsilon/3.$$

The partial sum  $\sum_0^N a_n y^n$  is a polynomial in  $y$  and polynomials are continuous so there is some  $\delta_0 > 0$  with the property that if  $|y - x| < \delta_0$

$$\left| \sum_0^N a_n y^n - \sum_0^N a_n x^n \right| < \varepsilon/3.$$

Therefore if we choose  $\delta$  to be the smaller of  $\delta_0$  and  $T - |x|$  then if  $|y - x| < \delta$  we get

$$\begin{aligned} \left| \sum_0^\infty a_n y^n - \sum_0^\infty a_n x^n \right| &\leq \left| \sum_{N+1}^\infty a_n y^n \right| + \left| \sum_0^N a_n y^n - \sum_0^N a_n x^n \right| + \left| \sum_{N+1}^\infty a_n x^n \right| \\ &\leq \sum_{N+1}^\infty |a_n| |y|^n + \left| \sum_0^N a_n y^n - \sum_0^N a_n x^n \right| + \sum_{N+1}^\infty |a_n| |x|^n < \varepsilon. \end{aligned}$$

□

Armed with these properties of power series we can begin to investigate the standard functions such as the exponential. It will turn out that power series are not only continuous but also differentiable inside the radius of convergence. The derivative makes it easier to demonstrate the crucial properties of standard functions but we shall begin with the exponential before we cover derivatives so as to see some concrete applications of our theory.

## The exponential

As explained in the introduction we shall define the exponential function as a power series.

**Definition (The exponential).** If  $x \in \mathbf{R}$  the series

$$1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots + \frac{x^n}{n!} + \cdots$$

converges. We call the sum  $\exp x$ .

*Proof* The ratio of successive terms of the series is

$$\frac{x^{n+1}}{(n+1)!} \frac{n!}{x^n} = \frac{x}{n+1} \rightarrow 0$$

so the series converges by the ratio test. □

We know that the function  $x \mapsto \exp x$  is continuous on  $\mathbf{R}$  since it is a convergent power series. We would like to check that it satisfies the characteristic property  $e^{x+y} = e^x e^y$ . This will be easy to check once we have derivatives later in the course but it is very instructive to see roughly how to do it directly from the series definition.

The idea is this. When we multiply  $e^x$  by  $e^y$  we have to expand the product of two brackets

$$\left(1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots\right) \left(1 + y + \frac{y^2}{2} + \frac{y^3}{6} + \cdots\right)$$

which means we have to consider all possible products of one term from the first bracket and one term from the second bracket (and then add all of these products together). It is natural to arrange these products into a grid

	1	$x$	$x^2/2$	$x^3/6$	$\dots$
1	1	$x$	$x^2/2$	$x^3/6$	$\dots$
$y$	$y$	$xy$	$x^2y/2$	$\dots$	
$y^2/2$	$y^2/2$	$xy^2/2$	$\dots$		
$y^3/6$	$y^3/6$	$\dots$			
$\vdots$	$\vdots$				

We now want to add all the terms in this grid. One way to make sure that we include everything is to add them diagonally.

$$1 + (x + y) + (x^2/2 + xy + y^2/2) + (x^3/6 + x^2y/2 + xy^2/2 + y^3/6) + \dots$$

Now let us collect each diagonal sum over a common denominator. The first is 1 and the second is  $x + y$ . The third is

$$\frac{x^2 + 2xy + y^2}{2}$$

and the fourth is

$$\frac{x^3 + 3x^2y + 3xy^2 + y^3}{6}.$$

We immediately see that we have the binomial expansions of  $(x + y)^2$  and  $(x + y)^3$ . So the whole sum looks like

$$1 + (x + y) + \frac{(x + y)^2}{2} + \frac{(x + y)^3}{6} + \dots$$

and this is  $\exp(x + y)$ .

Because we have infinite sums we need an argument to justify this. We shall compare the sum of a certain number of the diagonals with the product of partial sums of the series for  $e^x$  and  $e^y$ .

**Theorem (The characteristic property of the exponential).** If  $x, y \in \mathbf{R}$  then

$$\exp(x + y) = \exp(x) \exp(y).$$

*Proof* By the algebra of limits it suffices to check that

$$\sum_0^{2m} \frac{(x + y)^k}{k!} - \left( \sum_0^m \frac{x^i}{i!} \right) \left( \sum_0^m \frac{y^j}{j!} \right) \rightarrow 0$$

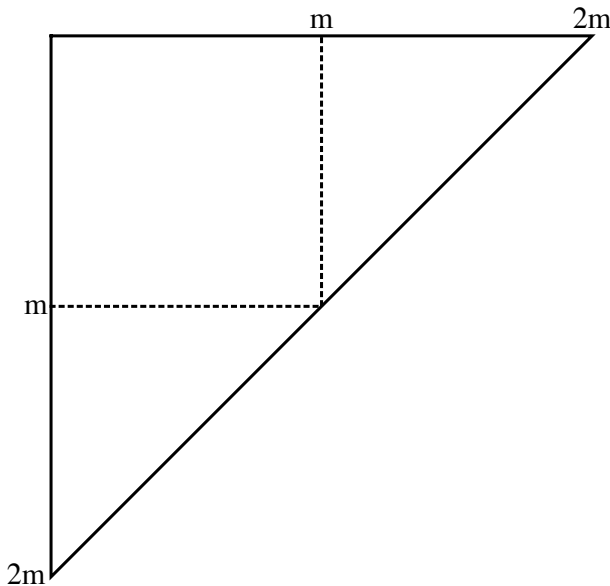
because the first term converges to  $e^{x+y}$  and the second to  $e^x e^y$ . By the binomial theorem, for each  $k$ ,

$$(x + y)^k = \sum_{i=0}^k \frac{k!}{i!(k-i)!} x^i y^{k-i} = \sum_{i+j=k} k! \frac{x^i}{i!} \frac{y^j}{j!}.$$

So

$$\begin{aligned} \sum_0^{2m} \frac{(x + y)^k}{k!} - \left( \sum_0^m \frac{x^i}{i!} \right) \left( \sum_0^m \frac{y^j}{j!} \right) \\ = \sum_{i+j \leq 2m} \frac{x^i}{i!} \frac{y^j}{j!} - \sum_{i \leq m, j \leq m} \frac{x^i}{i!} \frac{y^j}{j!}. \end{aligned}$$

We are considering the terms of the form  $\frac{x^i}{i!} \frac{y^j}{j!}$  within a triangle but with the terms in a square removed.



We are left with terms in two smaller triangles.

We want to prove that the sums in these two pieces are small.

The key is that all these terms are “far out”.

These sums are

$$\sum_{i \geq m+1, i+j \leq 2m} \frac{x^i y^j}{i! j!}$$

and a similar one with the restrictions on  $i$  and  $j$  interchanged. By the triangle inequality the absolute value of the sum is at most

$$\sum_{i \geq m+1, i+j \leq 2m} \frac{|x|^i |y|^j}{i! j!}.$$

Once we have made all the terms positive we can include other terms without decreasing the sum. So the sum of the terms in the smaller triangle has absolute value at most

$$\sum_{i \geq m+1, i+j \leq 2m} \frac{|x|^i |y|^j}{i! j!} \leq \sum_{i \geq m+1, j \geq 0} \frac{|x|^i |y|^j}{i! j!}$$

and the latter is

$$\left( \sum_{i \geq m+1} \frac{|x|^i}{i!} \right) \left( \sum_{j=0}^{\infty} \frac{|y|^j}{j!} \right).$$

The first factor tends to 0 as  $m \rightarrow \infty$  while the second is equal to  $e^{|y|}$ .  $\square$

**Theorem (The characteristic property of the exponential).** *If  $x, y \in \mathbf{R}$  then*

$$\exp(x + y) = \exp(x) \exp(y).$$

The characteristic property makes us feel comfortable writing  $\exp(x)$  as  $e^x$ . We have  $e^{x+y} = e^x e^y$  and in particular we have  $e^{-u} = 1/e^u$  for all real  $u$ .

The argument given above can be used to show that power series can be multiplied in the obvious way inside their intervals of convergence.

**Theorem (The product of power series).** *If  $\sum a_n x^n$  and  $\sum b_n x^n$  converge for  $x$  in the interval  $(-R, R)$  then so does the series  $\sum_{n=0}^{\infty} \left( \sum_{k=0}^n a_k b_{n-k} \right) x^n$  and for all such  $x$*

$$\sum_{n=0}^{\infty} \left( \sum_{k=0}^n a_k b_{n-k} \right) x^n = \left( \sum_{i=0}^{\infty} a_i x^i \right) \left( \sum_{j=0}^{\infty} b_j x^j \right).$$

*Proof Exercise.*  $\square$

## Inequalities for the exponential

For many purposes it is important to have some estimates for the exponential in terms of simpler functions. To begin with let us observe that if  $x$  is positive it is clear that  $e^x = 1 + x + x^2/2 + \dots$  is positive. For negative  $x$  it is not immediate from the power series that  $e^x$  is positive but it follows from the fact that  $e^{-u} = 1/e^u$  and this is positive if  $u$  is positive. The most useful inequalities are the following.

**Theorem (Inequalities for the exponential).** *The following estimates hold for the exponential function:*

1.  $1 + x \leq e^x$  for all real  $x$
2.  $e^x \leq 1/(1 - x)$  if  $x < 1$ .

*Proof* If  $x \geq 0$  then

$$e^x = 1 + x + \frac{x^2}{2} + \dots \geq 1 + x$$

and if  $0 \leq x < 1$

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \dots \leq 1 + x + x^2 + x^3 + \dots = \frac{1}{1 - x}.$$

So the inequalities are easy to establish if  $x \geq 0$ . To obtain them for negative  $x$  we use the characteristic property of the exponential much as we did to prove positivity.

Suppose  $x = -u$  is negative. We know that  $e^u \geq 1 + u$  and hence  $e^{-x} \geq 1 - x$ . But this implies that

$$\frac{1}{1 - x} \geq e^x$$

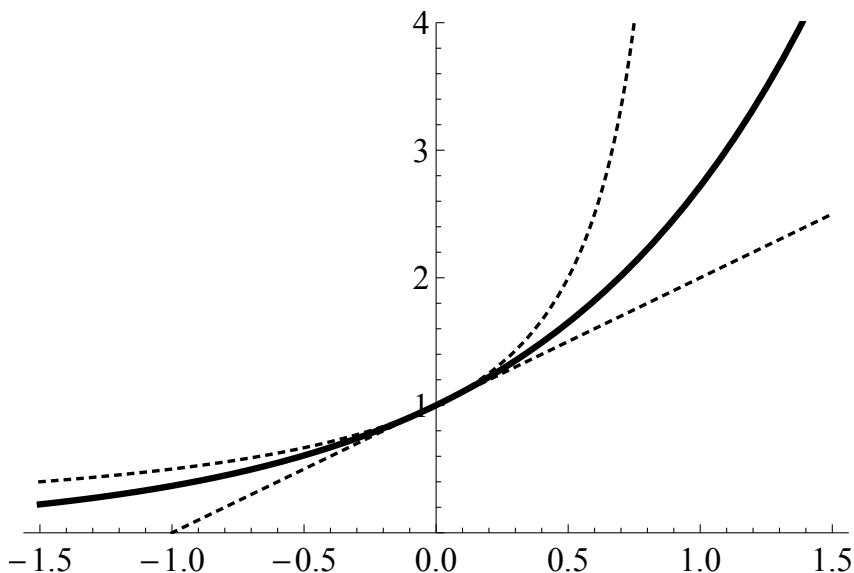
so the second inequality is now established for all  $x < 1$ .

If  $x \leq -1$  then  $1 + x \leq 0$  whereas  $e^x > 0$  so  $e^x \geq 1 + x$ . It remains to prove the first inequality for  $-1 < x < 0$ . If  $x = -u$  then  $0 < u < 1$  and so  $e^u \leq 1/(1 - u)$ . This says that  $e^{-x} \leq 1/(1 + x)$  and this implies that

$$1 + x \leq e^x.$$

□

These two inequalities sandwich the exponential rather nicely near 0 as shown in the picture.



**Corollary (The exponential increases).** *The exponential function is strictly increasing and its range is  $(0, \infty)$ .*

*Proof* Suppose  $x < y$ . Then

$$e^y = e^{y-x} e^x \geq (1 + y - x)e^x > e^x.$$

This shows that the exponential is strictly increasing.

Since  $e^x \geq 1 + x$  the exponential takes arbitrarily large values (at large  $x$ ). Since  $e^{-x} = 1/e^x$  the exponential also takes values arbitrarily close to 0 (at large negative  $x$ ). By the IVT the exponential takes all positive values.  $\square$

Since we know that the exponential increases and is continuous we know that it has an inverse, the logarithm, which will be the topic of the next section.

## The logarithm and powers

We have seen that the exponential function maps  $\mathbf{R}$  onto  $(0, \infty)$  and is continuous and strictly increasing. So we know by the IVT and its corollaries that the exponential has a continuous inverse defined on  $(0, \infty)$ : the natural logarithm.



**Theorem (The logarithm).** *There is a continuous strictly increasing function  $x \mapsto \log x$  defined on  $(0, \infty)$  satisfying*

$$e^{\log x} = x$$

*for all positive  $x$  and*

$$\log(e^y) = y$$

*for all real  $y$ . We have that for all positive  $u$  and  $v$ ,*

$$\log(uv) = \log u + \log v.$$

*Proof* We just need to check the last assertion. But

$$e^{\log u + \log v} = e^{\log u} e^{\log v} = uv.$$

Applying  $\log$  to both sides we get what we want. □

It was proved last term and deduced from the IVT above that each positive number has a positive square root. We would like to know that we can take other non-integer powers of positive numbers. The simplest way to define these is using the logarithm and exponential.

**Definition (Powers).** *If  $x > 0$  and  $p \in \mathbf{R}$  we define*

$$x^p = \exp(p \log x).$$

*We have the usual rules*

1. *If  $n$  is a positive integer then  $x^n$  as defined here is indeed the product  $x.x.\dots x$  of  $n$  copies of  $x$ .*
2.  *$x^{p+q} = x^p x^q$  for all  $x > 0$  and  $p, q \in \mathbf{R}$ .*
3.  *$\log(x^p) = p \log x$  for  $x > 0$  and  $p \in \mathbf{R}$ .*
4.  *$x^{pq} = (x^p)^q$  for all  $x > 0$  and  $p, q \in \mathbf{R}$ .*
5.  *$\exp(p) = e^p$  for all  $p \in \mathbf{R}$ .*

Notice that the last statement is not something we could have proved earlier because we did not have a definition of powers with which to make sense of the  $p^{\text{th}}$  power of  $e$ . We shall prove the last one and leave the rest as an exercise.

*Proof* By definition of the power,

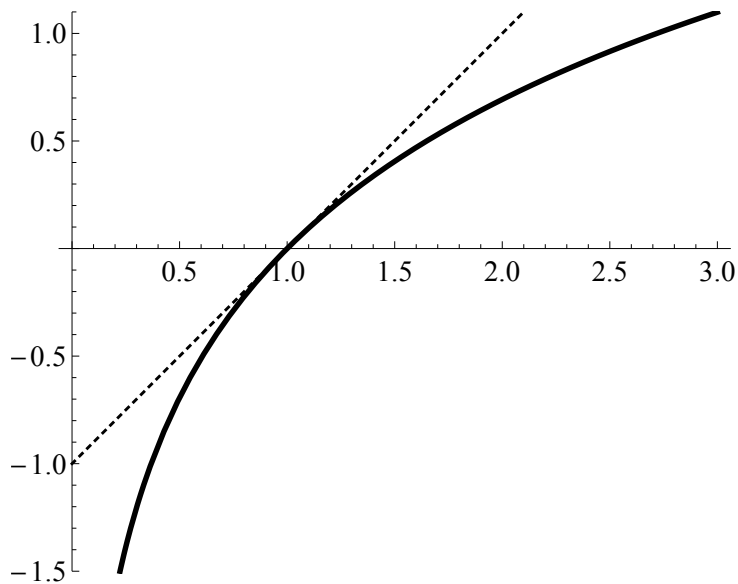
$$e^p = \exp(p \log(e)).$$

Now we know that  $\log(e) = 1$  so the second expression is  $\exp(p)$ . □

The inequalities we proved for the exponential immediately give us inequalities for the logarithm. You are asked to demonstrate these in the Homework. The most crucial one is this:

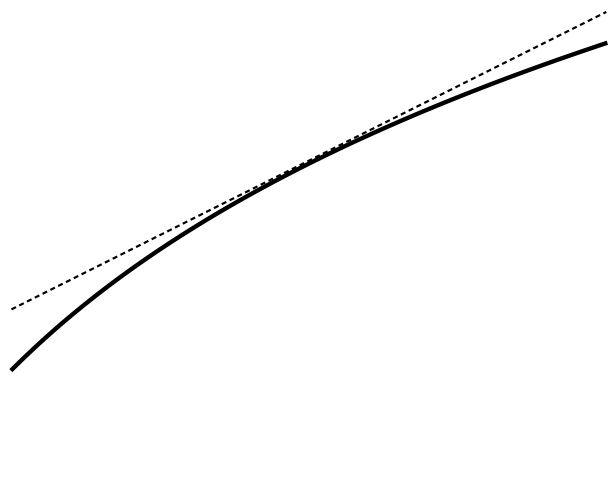
**Theorem (The tangent to the logarithm).** *If  $x > 0$  then  $\log x \leq x - 1$ .*

This says that the graph of  $y = \log x$  lies below the line which is its tangent at the point  $(1, 0)$ .

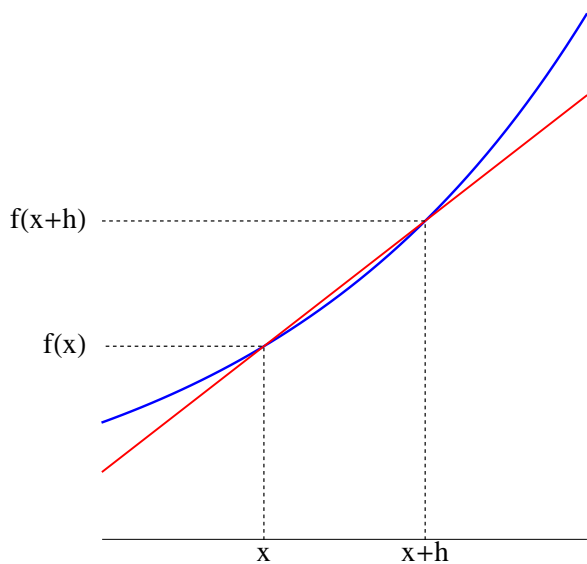


## Chapter 3. Limits and the derivative

In this chapter we shall develop the basic theory of derivatives. The problem that we start with is to find the instantaneous slope of a curve: the slope of its tangent line at a point  $(x, f(x))$  on the curve  $y = f(x)$ .



The geometric picture of our method is this: we consider the point  $(x, f(x))$  and a nearby point  $(x + h, f(x + h))$  and look at the chord joining the two points.



The slope of this chord is

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

We now ask what happens to this slope as the nearby point gets closer and closer to the point we care about: what happens as  $h \rightarrow 0$ . We ask whether the quotient  $\frac{f(x+h)-f(x)}{h}$  approaches a limit. If so we define this to be the slope of the curve at  $(x, f(x))$  or the derivative  $f'(x)$ .

In order to carry out this process and analyse it we need to have a definition of the limit in question. This will be the first part of the chapter.

## Limits

**Definition (Limits of functions).** Let  $I$  be an open interval,  $c \in I$  and  $f$  a real valued function defined on  $I$  except possibly at  $c$ . We say that

$$\lim_{x \rightarrow c} f(x) = L$$

if for every  $\varepsilon > 0$  there is a number  $\delta > 0$  so that if  $0 < |x - c| < \delta$  then

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

Thus we can guarantee that  $f(x)$  is close to  $L$  by insisting that  $x$  is close to  $c$ , but we exclude the possibility that  $x = c$ : we don't care what  $f$  does at  $c$  itself nor even whether  $f$  is defined at  $c$ . The reason for this freedom is that we want to discuss limits like

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

in which the function is not defined at  $h = 0$ .

Let's look at a couple of examples.

**Example.** Let  $f : \mathbf{R} \rightarrow \mathbf{R}$  be defined by

$$f(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x \neq 0 \end{cases}$$

Then

$$\lim_{x \rightarrow 0} f(x) = 0.$$

**Example.**

$$\lim_{x \rightarrow 1} \frac{x^2 - 1}{x - 1} = 2.$$

*Proof* As long as  $x \neq 1$

$$\frac{x^2 - 1}{x - 1} = x + 1$$

and as  $x \rightarrow 1$  this approaches 2. □

Strictly speaking the last statement is something we haven't yet proved: we shall do so now. By comparing the definitions of continuity and limits we can immediately prove the following lemma.

**Lemma (Limits and continuity).** *If  $f : I \rightarrow \mathbf{R}$  is defined on the open interval  $I$  and  $c \in I$  then  $f$  is continuous at  $c$  if and only if*

$$\lim_{x \rightarrow c} f(x) = f(c).$$

*Proof* Exercise. □

The last line really contains two pieces of information: that the limit exists and that it equals  $f(c)$ . In the previous example the function  $x \mapsto x + 1$  is continuous everywhere and hence at  $c = 1$  so we get  $\lim_{x \rightarrow 1} (x + 1) = 1 + 1 = 2$ .

We want to have a machine for calculating limits like the continuity machine: for example we want to know that

$$\lim_{x \rightarrow c} (f(x) + g(x)) = \lim_{x \rightarrow c} f(x) + \lim_{x \rightarrow c} g(x)$$

and

$$\lim_{x \rightarrow c} f(x)g(x) = \left( \lim_{x \rightarrow c} f(x) \right) \left( \lim_{x \rightarrow c} g(x) \right)$$

whenever  $f$  and  $g$  have limits at  $c$ . One way to prove these would be to relate limits of functions to limits of sequences just as we did for continuity. This will be an exercise in the homework.

An alternative approach would be to use a slightly weird trick to deduce what we want from what we already know about continuity. The trick is that if  $\lim_{x \rightarrow c} f(x) = L$  we can define a new function

$$F(x) = \begin{cases} f(x) & \text{if } x \neq c \\ L & \text{if } x = c \end{cases}$$

and this new function will be continuous at  $c$ . Then we just use the rules for continuity to deduce the same rules for limits.

We will take for granted the following theorems.

**Theorem (Continuous and sequential limits).** *If  $f : I \setminus \{c\} \rightarrow \mathbf{R}$  is defined on the interval  $I$  except at  $c \in I$  then*

$$\lim_{x \rightarrow c} f(x) = L$$

*if and only if for every sequence  $(x_n)$  in  $I \setminus \{c\}$  with  $x_n \rightarrow c$  we have*

$$f(x_n) \rightarrow L.$$

*Proof* Homework □

**Theorem (Algebra of limits).** *If  $f, g : I \setminus \{c\} \rightarrow \mathbf{R}$  are defined on the interval  $I$  except at  $c \in I$  and  $\lim_{x \rightarrow c} f(x)$  and  $\lim_{x \rightarrow c} g(x)$  exist then*

1.  $\lim_{x \rightarrow c} (f(x) + g(x)) = \lim_{x \rightarrow c} f(x) + \lim_{x \rightarrow c} g(x)$

2.  $\lim_{x \rightarrow c} f(x)g(x) = \lim_{x \rightarrow c} f(x) \lim_{x \rightarrow c} g(x)$

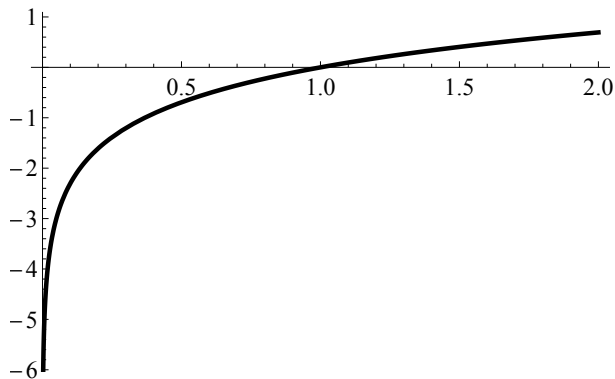
3. *if  $\lim_{x \rightarrow c} g(x) \neq 0$  then*

$$\lim_{x \rightarrow c} \frac{f(x)}{g(x)} = \frac{\lim_{x \rightarrow c} f(x)}{\lim_{x \rightarrow c} g(x)}.$$

*Proof* Exercise □

## One sided limits

It often happens that we wish to understand the behaviour of  $f(x)$  as approaches the end of the interval where  $f$  is defined. For example we want to know what happens to  $\log x$  as  $x$  approaches 0 from the right.



So we define one sided limits.

**Definition (One sided limits).** Let  $f$  a real valued function defined on the open interval  $(c, d)$ . We say that

$$\lim_{x \rightarrow c^+} f(x) = L$$

if for every  $\varepsilon > 0$  there is a number  $\delta > 0$  so that if  $c < x < c + \delta$  then

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

We read the expression as “The limit of  $f(x)$  as  $x$  approaches  $c$  from the right is  $L$ ”.

We define the limit from the left

$$\lim_{x \rightarrow c^-} f(x)$$

similarly.

**Example.** Let  $f : \mathbf{R} \rightarrow \mathbf{R}$  be defined by

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/2 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases}$$

Then

$$\lim_{x \rightarrow 0^+} f(x) = 1$$

and

$$\lim_{x \rightarrow 0^-} f(x) = 0.$$

At this point we digress slightly to discuss limits at infinity and infinite limits. At times it is useful to have a notation indicating that a function behaves something like  $1/x^2$  as  $x \rightarrow 0$ .

**Definition (Infinite limits).** If  $f : I \setminus \{c\} \rightarrow \mathbf{R}$  is defined on an interval  $I$  except perhaps at  $c \in I$  we write

$$\lim_{x \rightarrow c} f(x) = \infty$$

if for every  $M > 0$  there is a  $\delta > 0$  so that if  $0 < |x - c| < \delta$  then  $f(x) > M$ . The limit  $-\infty$

$$\lim_{x \rightarrow c} f(x) = -\infty$$

is defined similarly.

The limit is infinite if we can make  $f(x)$  as large as we please by insisting that  $x$  is close to  $c$  (but not equal to  $c$ ).

**Example.**

$$\lim_{x \rightarrow 0} \frac{1}{x^2} = \infty.$$

*Proof* Given  $M > 0$  choose  $\delta = 1/\sqrt{M}$ . Then if  $0 < |x| < \delta = 1/\sqrt{M}$  we have  $0 < x^2 < 1/M$  and hence

$$\frac{1}{x^2} > M.$$

□

There is also a one sided version.

**Example.**

$$\lim_{x \rightarrow 0^+} \frac{1}{x} = \infty, \quad \lim_{x \rightarrow 0^-} \frac{1}{x} = -\infty.$$

We also on occasions wish to study the behaviour of functions as the variable becomes large.

**Definition (Limits at infinity).** If  $f : \mathbf{R} \rightarrow \mathbf{R}$  we write

$$\lim_{x \rightarrow \infty} f(x) = L$$

if for every  $\varepsilon > 0$  there is an  $N$  so that if  $x > N$  then  $|f(x) - L| < \varepsilon$ .



This looks very much like the definition of convergence of a sequence. The only difference is that we now consider arbitrary real  $x$  instead of just natural numbers  $n$ . The algebra of limits applies equally well to limits at infinity.

**Example.**

$$\lim_{x \rightarrow \infty} \frac{1}{x} = 0.$$

*Proof* Exercise.

**Example.**

$$\lim_{x \rightarrow \infty} \frac{x}{e^x} = 0.$$

*Proof* If  $x > 0$  we have  $e^x \geq 1 + x + x^2/2 > x^2/2$ . So

$$0 < \frac{x}{e^x} < \frac{x}{x^2/2} = \frac{2}{x} \rightarrow 0$$

as  $x \rightarrow \infty$ . □

## The derivative

The idea of calculating the slope of a curve existed before Newton and Leibniz and the derivatives of certain functions were already known: in particular the derivatives of the monomials  $x \mapsto x^n$  for positive integers  $n$ . It was also known how to calculate areas under certain curves. Three key points made up the invention of what we call calculus.

The first key point was the creation of a derivative “machine” to enable us to calculate derivatives of all the standard functions: polynomials, rational functions, the exponential and trigonometric functions and anything we can build by adding multiplying or composing these functions.

The second key point of calculus is the recognition that the slope of a curve, the derivative, can be considered as a function and then differentiated again. For Newton this was crucial since his aim was to derive Kepler’s Laws of planetary motion from the inverse square law of gravitation. The inverse square law tells you the *acceleration* of a planet: the second derivative of its position.

The third key point of calculus was the realisation that differentiation and integration are opposites of one another. This made it possible to simplify the process of integration enormously by using the rules for derivatives.

**Definition (The derivative).** Suppose  $f : I \rightarrow \mathbf{R}$  is defined on the open interval  $I$  and  $c \in I$ . We say that  $f$  is differentiable at  $c$  if

$$\lim_{h \rightarrow 0} \frac{f(c+h) - f(c)}{h}$$

exists. If so we call the limit  $f'(c)$ .

**Example (The derivative of  $x$ ).** If  $f(x) = x$  then  $f$  is differentiable at every point of  $\mathbf{R}$  and  $f'(c) = 1$  for all  $c$ .

*Proof* For every  $c$  and  $h \neq 0$

$$\frac{f(c+h) - f(c)}{h} = \frac{c+h-c}{h} = 1$$

and so

$$\lim_{h \rightarrow 0} \frac{f(c+h) - f(c)}{h} = \lim_{h \rightarrow 0} 1 = 1.$$

□

**Example (The derivative of  $x^2$ ).** If  $f(x) = x^2$  then  $f$  is differentiable at every point of  $\mathbf{R}$  and  $f'(c) = 2c$  for all  $c$ .

*Proof* For every  $c$  and  $h \neq 0$

$$\frac{f(c+h) - f(c)}{h} = \frac{(c+h)^2 - c^2}{h} = \frac{c^2 + 2ch + h^2 - c^2}{h} = 2c + h$$

and so

$$\lim_{h \rightarrow 0} \frac{f(c+h) - f(c)}{h} = \lim_{h \rightarrow 0} (2c + h) = 2c.$$

□

**Example (The derivative of  $1/x$ ).** If  $f(x) = 1/x$  then  $f$  is differentiable at every point of  $\mathbf{R}$  except 0 and  $f'(c) = -1/c^2$  for all  $c \neq 0$ .

*Proof* For every  $c \neq 0$  and  $h$  satisfying  $0 < |h| < |c|$

$$\frac{f(c+h) - f(c)}{h} = \frac{1/(c+h) - 1/c}{h} = \frac{c - (c+h)}{h(c+h)c} = \frac{-1}{(c+h)c}$$

and so

$$\lim_{h \rightarrow 0} \frac{f(c+h) - f(c)}{h} = \lim_{h \rightarrow 0} \frac{-1}{(c+h)c} = \frac{-1}{c^2}.$$

□

## The derivative machine

Obviously we don't want to carry out this process for complicated functions so we need to build a machine to do it for us. The derivative machine is like a food mixer. The mixer has 3 basic parts: the motor, jug and lid (say) and separate blades to handle different foods. The derivative machine has 3 basic parts: the sum rule, the product rule and the chain rule and special rules to handle powers, the exponential and the trigonometric functions.

In order to start building the machine we need to know that if a function is differentiable at a point  $c$  then it is continuous there. To make the picture clearer let us start by observing that we can rewrite the derivative

$$f'(c) = \lim_{x \rightarrow c} \frac{f(x) - f(c)}{x - c}$$

since if we put  $x = c + h$  and  $h \rightarrow 0$  we have  $x \rightarrow c$ .

**Lemma (Differentiability and continuity).** *If  $I$  is an open interval,  $f : I \rightarrow \mathbf{R}$  is differentiable at  $c \in I$  then  $f$  is continuous at  $c$ .*

*Proof* We know that

$$\frac{f(x) - f(c)}{x - c} \rightarrow f'(c)$$

as  $x \rightarrow c$ . Hence

$$f(x) - f(c) = \frac{f(x) - f(c)}{x - c} (x - c) \rightarrow f'(c) \cdot 0 = 0$$

as  $x \rightarrow c$  which implies that  $f(x) \rightarrow f(c)$  as required. □

The sum and product rules follow easily from the algebra of limits.

**Theorem (The sum and product rules).** Suppose  $f, g : I \rightarrow \mathbf{R}$  are defined on the open interval  $I$  and are differentiable at  $c \in I$ . Then  $f + g$  and  $fg$  are differentiable at  $c$  and

$$(f + g)'(c) = f'(c) + g'(c)$$

and

$$(fg)'(c) = f'(c)g(c) + f(c)g'(c).$$

*Proof* The first will be left as an exercise. For the second

$$\begin{aligned} \frac{f(x)g(x) - f(c)g(c)}{x - c} &= \frac{(f(x) - f(c))g(x) + f(c)(g(x) - g(c))}{x - c} \\ &= \frac{f(x) - f(c)}{x - c}g(x) + f(c)\frac{g(x) - g(c)}{x - c}. \end{aligned}$$

The algebra of limits tells us that as  $x \rightarrow c$  this expression approaches

$$f'(c)g(c) + f(c)g'(c).$$

□

Note that to use the algebra of limits for the product rule we needed to use the fact that  $g(x) \rightarrow g(c)$  as  $x \rightarrow c$ : in other words, the continuity of  $g$  at  $c$ .

We are now in a position to prove the differentiability of all polynomials. To begin with let's observe that if  $f$  is a constant function then its derivative is clearly 0. By the product rule or just a direct check we can see that if we multiply a function  $f$  by a constant  $C$  then we also multiply the derivative by  $C$ . Using the sum rule we can then handle all polynomials as long as we check the derivative of each power  $x \mapsto x^n$ . We shall do this by induction.

**Lemma (The derivatives of the monomials).** If  $n$  is a positive integer then the derivative of  $x \mapsto x^n$  is  $x \mapsto nx^{n-1}$ .

*Proof* We already saw this for  $n = 1$ . Assume inductively that we have the result for  $f(x) = x^n$ . Then  $x^{n+1} = xf(x)$  so by the product rule its derivative is

$$1.f(x) + xf'(x) = x^n + nx^{n-1} = (n + 1)x^n$$

completing the inductive step.

□

At school you learned the quotient rule for derivatives. I have not included it in the machine since it follows from the product rule, the chain rule and the derivative of the function  $x \mapsto 1/x$ .

The third part of the derivative machine is the chain rule. If we form the composition of two functions  $f$  and  $g$

$$x \mapsto f(g(x))$$

we want to know that we can differentiate it and obtain the correct formula for the derivative.

**Theorem (Chain rule).** *Suppose  $I$  and  $J$  are open intervals,  $f : I \rightarrow \mathbf{R}$  and  $g : J \rightarrow I$ , that  $g$  is differentiable at  $c$  and  $f$  is differentiable at  $g(c)$ . Then the composition  $f \circ g$  is differentiable at  $c$  and*

$$(f \circ g)'(c) = f'(g(c)) \cdot g'(c).$$

This may look a bit different from the chain rule that you know but in fact it is exactly the one that you are accustomed to using: let's see why. Let's take  $g(x) = 1 + x^2$  and  $f(u) = u^3$ . Then the composition is  $f \circ g : x \mapsto (1 + x^2)^3$ . We have  $g'(x) = 2x$  and  $f'(u) = 3u^2$ . According to the theorem the derivative of the composition is

$$f'(g(c)) \cdot g'(c) = 3g(c)^2 \cdot g'(c) = 3(1 + c^2)^2 \cdot 2c = \dots$$

Note that the derivative of  $f$  is evaluated at  $g(c)$ : the only place that makes any sense because  $f$  is defined on an interval containing  $g(c)$ . The chain rule is a bit harder to prove than the other two rules. How could we try to prove the theorem? We want to investigate the ratio

$$\frac{f(g(x)) - f(g(c))}{x - c}$$

as  $x \rightarrow c$ . The obvious thing to do is to write this as

$$\frac{f(g(x)) - f(g(c))}{g(x) - g(c)} \times \frac{g(x) - g(c)}{x - c}.$$

The second factor converges to  $g'(c)$  as  $x \rightarrow c$ .

For the first factor we note that  $g(x) \rightarrow g(c)$  as  $x \rightarrow c$  so the limit of the first factor is therefore

$$\lim_{g(x) \rightarrow g(c)} \frac{f(g(x)) - f(g(c))}{g(x) - g(c)}$$

which appears to be  $f'(g(c))$ . So we appear to have proved that

$$(f \circ g)'(c) = f'(g(c))g'(c).$$

There is a problem however. The quantity  $g(x) - g(c)$  could be equal to zero for lots of values of  $x$ : perhaps even all of them. In that case we can't divide by this quantity. For this reason we need to start with a new way to express the differentiability of a function.

**Lemma (Local linearisation).** *Suppose  $I$  is an open interval,  $f : I \rightarrow \mathbf{R}$  and  $c \in I$ . Then  $f$  is differentiable at  $c$  if and only if there is a number  $A$  and a function  $\varepsilon$  with the properties that for all  $x$*

$$f(x) - f(c) = A(x - c) + \varepsilon(x)(x - c),$$

$\varepsilon(c) = 0$  and  $\varepsilon$  is continuous at  $c$ : ( $\varepsilon(x) \rightarrow 0$  as  $x \rightarrow c$ ). If this happens  $A = f'(c)$ .

This is sometimes called the Weierstrass-Caratheodory criterion. The lemma says that if  $x$  is close to  $c$  then  $f(x)$  is approximately given by the linear function

$$x \mapsto f(c) + f'(c)(x - c)$$

which you recognise as the first Taylor approximation to  $f$ .

*Proof* If the condition holds then

$$\frac{f(x) - f(c)}{x - c} = A + \varepsilon(x)$$

and this approaches  $A$  as  $x \rightarrow c$ . Hence  $f$  is differentiable with derivative  $f'(c) = A$ .

On the other hand suppose  $f$  is differentiable at  $c$ . Set  $A = f'(c)$  and define  $\varepsilon$  as follows

$$\varepsilon(x) = \begin{cases} \frac{f(x) - f(c)}{x - c} - A & \text{if } x \neq c \\ 0 & \text{if } x = c. \end{cases}$$

If  $x \neq c$  then  $f(x) - f(c) = A(x - c) + \varepsilon(x)(x - c)$  holds because of the way  $\varepsilon$  is defined, while if  $x = c$  the formula is obvious. To check that  $\varepsilon(x) \rightarrow 0$  as  $x \rightarrow c$  observe that

$$\lim_{x \rightarrow c} \varepsilon(x) = \lim_{x \rightarrow c} \left( \frac{f(x) - f(c)}{x - c} - f'(c) \right) = 0.$$

□

We are now ready to prove the chain rule.

*Proof (of the chain rule)*  $f$  is differentiable at  $g(c)$  so for all  $y$

$$f(y) - f(g(c)) = f'(g(c))(y - g(c)) + \varepsilon(y)(y - g(c)).$$

where  $\varepsilon(g(c)) = 0$  and  $\varepsilon$  is continuous at  $g(c)$ . Hence

$$f(g(x)) - f(g(c)) = f'(g(c))(g(x) - g(c)) + \varepsilon(g(x))(g(x) - g(c)).$$

Consequently if  $x \neq c$

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

As  $x \rightarrow c$ ,

$$\frac{g(x) - g(c)}{x - c} \rightarrow g'(c)$$

while  $\varepsilon \circ g$  is continuous at  $c$  so  $\varepsilon(g(x)) \rightarrow \varepsilon(g(c)) = 0$ . Hence

$$\frac{f(g(x)) - f(g(c))}{x - c} \rightarrow f'(g(c)) \cdot g'(c)$$

as required. □

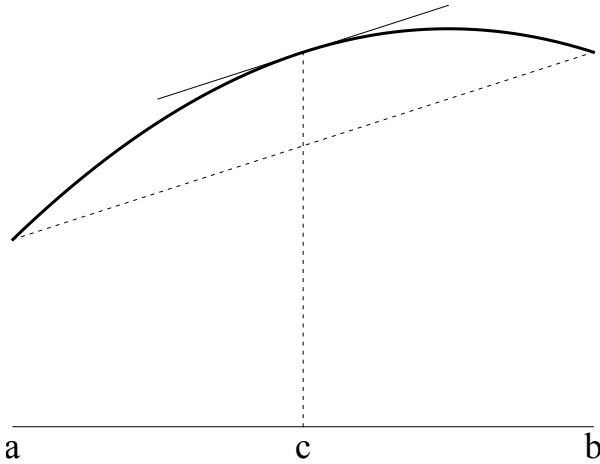
## The Mean Value Theorem

In this section we prove one of the most useful facts about derivatives: the so-called Mean Value Theorem (MVT). It provides a way to relate the values of a function to the values of its derivative.

**Theorem (Mean Value Theorem).** *Suppose  $f : [a, b] \rightarrow \mathbf{R}$  is continuous on the closed interval  $[a, b]$  and differentiable on the open interval  $(a, b)$ . Then there is a point  $c \in (a, b)$  where*

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

The theorem says that between any pair of points  $a$  and  $b$ , there is a third point  $c$  where the slope of the curve  $y = f(x)$  is equal to the slope of the chord joining  $(a, f(a))$  and  $(b, f(b))$ . Geometrically this is intuitively obvious as can be seen from the picture below.



Another way to interpret the theorem, that explains its name, is this. Suppose you drive a distance of 30 miles in one hour. Then your average (or mean) speed for the trip is 30mph. The theorem says that at some point in your trip your speed will be exactly 30mph: at some point the needle of your speedometer will point to 30.

To see why the MVT might be useful let's deduce some immediate consequences.

**Corollary (Functions with positive derivative).** *If  $f : I \rightarrow \mathbf{R}$  is differentiable on the open interval  $I$  and  $f'(x) > 0$  for all  $x$  in the interval then  $f$  is strictly increasing on the interval.*

*Proof* If there were two points  $a$  and  $b$  with  $a < b$  but  $f(a) \geq f(b)$  then we could find a point  $c$  where

$$f'(c) = \frac{f(b) - f(a)}{b - a} \leq 0$$

contradicting the hypothesis. □

**Corollary (Functions with zero derivative).** *If  $f : I \rightarrow \mathbf{R}$  is differentiable on the open interval  $I$  and  $f'(x) = 0$  for all  $x$  in the interval then  $f$  is constant on the interval.*

*Proof* Exercise. □

The MVT has many uses in the theory of integration and the study of differential equations. The previous corollary can be regarded as the statement that the only solutions of the differential equation  $f'(x) = 0$  are the constant functions.



In a similar way we can obtain uniqueness of solutions to other differential equations.

**Example (Uniqueness of solution to a Diff. Eq. I).** *The only functions  $f : \mathbf{R} \rightarrow \mathbf{R}$  satisfying*

$$f'(x) = f(x)$$

*are the functions  $f(x) = Ae^x$  for some constant  $A$ .*

*Proof* We shall assume that  $e^x$  is differentiable and is its own derivative (a fact that will be proved in the next chapter). Suppose  $f$  is such a solution and let  $g(x) = e^{-x}f(x)$ . Then by the chain rule we have

$$g'(x) = -e^{-x}f(x) + e^{-x}f'(x) = e^{-x}(-f(x) + f'(x)) = 0.$$

By the corollary  $g$  is a constant function with value  $A$  (say). Then  $f(x) = e^xg(x) = Ae^x$ . □

**Example (Uniqueness of solution to a Diff. Eq. II).** *The only functions  $f : (0, \infty) \rightarrow \mathbf{R}$  satisfying*

$$f'(x) + \frac{1}{x}f(x) = 2$$

*are the functions  $f(x) = A/x + x$  for some constant  $A$ .*

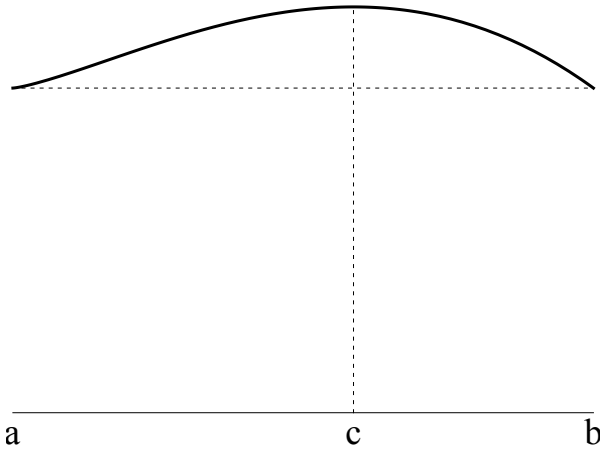
*Proof* Let  $g(x) = xf(x) - x^2$  for positive  $x$ . Then

$$g'(x) = xf'(x) + f(x) - 2x = x \left( f'(x) + \frac{1}{x}f(x) - 2 \right) = 0.$$

So  $g$  is constant  $A$  (say), and the formula for  $f$  follows. □

In order to prove the MVT we shall start by proving the special case in which the slope is 0.

**Theorem (Rolle).** *Suppose  $f : [a, b] \rightarrow \mathbf{R}$  is continuous on the closed interval  $[a, b]$  and differentiable on the open interval  $(a, b)$  and that  $f(a) = f(b)$ . Then there is a point  $c$  in the open interval where  $f'(c) = 0$ .*



*Proof* If  $f$  is constant on the interval then its derivative is zero everywhere. If not it takes values different from  $f(a) = f(b)$ . Assume it is somewhere larger than  $f(a)$ .

Since  $f$  is continuous on the closed interval it attains its maximum value at some point  $c$  and this cannot be  $a$  or  $b$ : so  $c$  lies in  $(a, b)$ . If  $x > c$  then  $f(x) - f(c) \leq 0$  while  $x - c > 0$  so the ratio

$$\frac{f(x) - f(c)}{x - c} \leq 0.$$

So  $f'(c)$  is a limit of non-positive values and so is not positive.

On the other hand if  $x < c$  then  $f(x) - f(c) \leq 0$  while  $x - c < 0$  so the ratio

$$\frac{f(x) - f(c)}{x - c} \geq 0.$$

So  $f'(c)$  is a limit of non-negative values and so is not negative. Therefore  $f'(c) = 0$ .  $\square$

Notice that we found the point  $c$  without having any formula for  $f$  or its derivative. We simply used a fact which has nothing to do with derivatives: that  $f$  attains its maximum. This fact was proved in the first section using the Bolzano-Weierstrass Theorem which does not construct the point in any computable way.

We now come to the proof of the MVT. We shall modify the function  $f$  by subtracting a linear function whose slope is  $\frac{f(b)-f(a)}{b-a}$ . This new function will take the same values at the two ends and so have a point with zero slope by Rolle's Theorem. But at this point the slope of  $f$  must be the same as the slope of the linear function that we subtracted from it.

*Proof (of the MVT)* Consider the function given by

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

Then

$$g(b) - g(a) = f(b) - f(a) - (b - a) \frac{f(b) - f(a)}{b - a} = 0.$$

So by Rolle's Theorem there is a point  $c$  where  $g'(c) = 0$ . But this implies

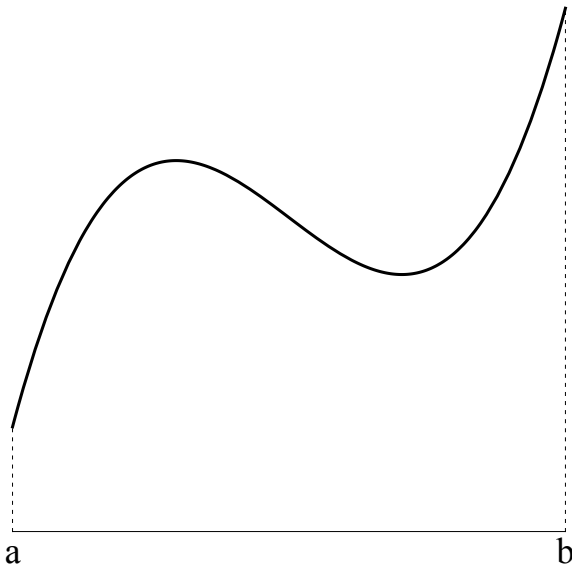
$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

□

In the proof of Rolle's Theorem we used the fact that if a differentiable function  $f : [a, b] \rightarrow \mathbf{R}$  attains its maximum at a point of the open interval  $(a, b)$  then its derivative must be zero at that point. This principle underlies the familiar method for finding maxima and minima. For a function on a closed interval the cleanest statement of the principle is this.

**Theorem (Extrema and derivatives).** *Suppose  $f : [a, b] \rightarrow \mathbf{R}$  is continuous and that it is differentiable on the open interval. Then  $f$  attains its maximum and minimum either at points in  $(a, b)$  where  $f' = 0$  or at one of the ends  $a$  or  $b$ .*

The second possibility can of course happen.



Finding where  $f' = 0$  doesn't tell you where the maximum is but it narrows down the options very considerably. Normally there will only be a few points where  $f' = 0$  and so

you only have to check those and the ends. Once you have found all these possibilities ( $a$ ,  $b$  and some points  $x_1, x_2, \dots$  where the derivative vanishes) the most reliable way to find the maximum is just to calculate the values  $f(a)$ ,  $f(b)$ ,  $f(x_1)$ ,  $f(x_2)$  and so on. You then just check which one is greatest and which is least.

It is customary in elementary calculus texts to suggest the use of the second derivative to try to find maxima and minima. This method is sometimes useful but has two drawbacks:

- If  $f'(c) = 0$  and  $f''(c) < 0$  it tells you that  $c$  is a *local* maximum but the function may have several local maxima many of which are not the maximum.
- If  $f''(c) = 0$  then it tells you nothing.

(On the other hand, the idea can often be useful in theoretical situations in the other direction. If the maximum occurs at  $c$  in the open interval then you can conclude that  $f'(c) = 0$  and  $f''(c) \leq 0$  which may be valuable information.)

If you have a function defined on an open interval, on  $(0, \infty)$  or the whole of  $\mathbf{R}$  there are a number of options. Sometimes it is easy to see that the function has its maximum in a certain closed interval and then use the closed interval version. Sometimes it is better to use the derivative to see that the function increases up to the maximum and decreases after it.

**Example.** Find the maximum of  $xe^{-x}$  on  $\mathbf{R}$ .

The derivative is  $x \mapsto (1-x)e^{-x}$  which is positive if  $x < 1$  and negative if  $x > 1$ . By the MVT the function increases until  $x = 1$  and then decreases. So the maximum occurs at  $x = 1$  where the function is equal to  $e^{-1}$ .  $\square$

As a consequence we have proved that  $xe^{-x} \leq e^{-1}$  for all  $x$ . Actually we already knew this.

$$x = 1 + x - 1 \leq e^{x-1}.$$

## Derivatives of inverses

We saw earlier that if  $f$  is a continuous strictly increasing function then it has a continuous inverse. An obvious question is whether the inverse is differentiable whenever the original function is differentiable and how the derivatives are related.

Actually the second question is not too hard to answer. Suppose  $f$  and  $g$  are inverses and we know that they are differentiable: then

$$f(g(x)) = x$$

on the domain of  $g$ . We can differentiate this equation using the chain rule to get

$$f'(g(x))g'(x) = 1$$

and thus we conclude that

$$g'(x) = \frac{1}{f'(g(x))}.$$

Notice that the derivative of  $f$  is evaluated at  $g(x)$  which looks a bit complicated but as we discussed when looking at the chain rule, it is the only thing that makes sense. It also gives the right answers. Let's see how this works in practice (ignoring for the moment the fact that we don't yet know that the exponential and trig. functions are differentiable).

**Example (The derivative of  $x \mapsto \sqrt{x}$ ).** Let  $f : x \mapsto x^2$  be the squaring function and  $g$  the square root. Then

$$g'(x) = \frac{1}{f'(g(x))} = \frac{1}{2g(x)} = \frac{1}{2\sqrt{x}}.$$

**Example (The derivative of log).** Let  $f$  be the exponential function and  $g$  the logarithm. Then

$$g'(x) = \frac{1}{f'(g(x))} = \frac{1}{\exp(\log x)} = \frac{1}{x}.$$

**Example (The derivative of  $\sin^{-1}$ ).** Let  $f : [-\pi/2, \pi/2] \rightarrow [-1, 1]$  be the function  $x \mapsto \sin x$  and  $g$  the inverse sine. Then

$$g'(x) = \frac{1}{f'(g(x))} = \frac{1}{\cos(\sin^{-1} x)}.$$

This is not the usual way you are accustomed to writing the derivative. In the homework you are asked to confirm that it is  $\frac{1}{\sqrt{1-x^2}}$ . Now let's prove the general theorem.

**Theorem (Derivatives of inverses).** Let  $f : (a, b) \rightarrow \mathbf{R}$  be differentiable with positive derivative. Then  $g = f^{-1}$  is differentiable and

$$g'(x) = \frac{1}{f'(g(x))}.$$

*Proof* Since  $f$  has positive derivative it is continuous and strictly increasing. Therefore it has a continuous inverse. Let  $(c, d)$  be the range of  $f$ , let  $x$  be in the interval  $(c, d)$  and  $g(x) = y$ . We want to calculate

$$\lim_{u \rightarrow x} \frac{g(u) - g(x)}{u - x}.$$

Let  $v = g(u)$  so that  $u = f(v)$ . Then the quotient is

$$\frac{v - y}{f(v) - f(y)}.$$

As  $u \rightarrow x$  we know that  $v = g(u) \rightarrow y = g(x)$  because  $g$  is continuous at  $x$ . So we want to calculate

$$\lim_{v \rightarrow y} \frac{v - y}{f(v) - f(y)}.$$

We know that

$$\lim_{v \rightarrow y} \frac{f(v) - f(y)}{v - y} = f'(y)$$

and that this limit is positive. So by the properties of limits we have

$$\lim_{u \rightarrow x} \frac{g(u) - g(x)}{u - x} = \lim_{v \rightarrow y} \frac{v - y}{f(v) - f(y)} = \frac{1}{f'(y)} = \frac{1}{f'(g(x))}.$$

□

## Chapter 4. Power series II

### The differentiability of power series

We start this chapter by checking that power series are differentiable inside the radius of convergence. Naturally the proof of this is more difficult than the proof of continuity. Suppose  $f(x) = \sum a_n x^n$  is a power series with radius of convergence  $R$ . We want to show that for  $|x| < R$  the derivative exists and

$$f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}.$$

In other words we want to know that we can differentiate the series term by term as if it were a polynomial. The sum rule doesn't tell us that we can, because an infinite sum is not the same as a finite one.

We want to show that if  $|x| < R$ , the sum  $\sum_{n=1}^{\infty} n a_n x^{n-1}$  converges and

$$\frac{\sum_{n=0}^{\infty} a_n y^n - \sum_{n=0}^{\infty} a_n x^n}{y - x} \rightarrow \sum_{n=1}^{\infty} n a_n x^{n-1}$$

as  $y \rightarrow x$ . The left side is

$$\sum_{n=0}^{\infty} a_n \frac{y^n - x^n}{y - x} = \sum_{n=1}^{\infty} a_n (y^{n-1} + y^{n-2}x + \cdots + x^{n-1}).$$

This looks promising because if  $y$  is close to  $x$  the sum

$$y^{n-1} + y^{n-2}x + \cdots + x^{n-1}$$

is close to  $n x^{n-1}$  because we know that polynomials are continuous.

We want to conclude that

$$\sum_{n=1}^{\infty} a_n (y^{n-1} + y^{n-2}x + \cdots + x^{n-1}) \rightarrow \sum_{n=1}^{\infty} n a_n x^{n-1}.$$

The trouble is that because we have an infinite sum we want the two things to be close together for every  $n$  at the same time and this is not something that follows from what we already know. Let's start gently.

**Lemma (The differentiability of power series I).** Let  $\sum a_n x^n$  be a power series with radius of convergence  $R$ . Then the series  $\sum n a_n x^{n-1}$  has the same radius of convergence.

*Proof* We know that the absolute series  $\sum |a_n| x^n$  has the same radius of convergence as  $\sum a_n x^n$ . Now if  $0 < x < R$  choose  $y$  with  $x < y < R$ . Then  $\sum |a_n| x^n$  and  $\sum |a_n| y^n$  both converge and hence so does

$$\sum_{n=0}^{\infty} |a_n| \frac{y^n - x^n}{y - x} = \sum_{n=1}^{\infty} |a_n| (y^{n-1} + y^{n-2}x + \cdots + x^{n-1}).$$

But the last sum is larger than  $\sum_{n=1}^{\infty} |a_n| n x^{n-1}$  so the latter also converges. This means that  $\sum n a_n x^{n-1}$  converges absolutely as required.  $\square$

We now move on to the full theorem and use the existence of the derivative to control the limits we are trying to evaluate.

**Theorem (The differentiability of power series II).** Let  $f(x) = \sum a_n x^n$  be a power series with radius of convergence  $R$ . Then  $f$  is differentiable on  $(-R, R)$  and

$$f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}.$$

*Proof* Choose  $T$  with  $|x| < T < R$ . We know from the lemma that the series  $\sum n |a_n| T^{n-1}$  converges so given  $\varepsilon > 0$  there is a number  $N$  so that

$$\sum_{n=N+1}^{\infty} n |a_n| T^{n-1} < \frac{\varepsilon}{3}.$$

Now if  $0 < |y - x| < T - |x|$  we have  $|y| < T$  as well as  $|x| < T$  and so

$$\left| \sum_{n=N+1}^{\infty} n a_n x^{n-1} \right| \leq \sum_{n=N+1}^{\infty} n |a_n| |x|^{n-1} < \frac{\varepsilon}{3}$$

and also

$$\begin{aligned} \left| \sum_{N+1}^{\infty} a_n \frac{y^n - x^n}{y - x} \right| &= \left| \sum_{N+1}^{\infty} a_n (y^{n-1} + y^{n-2}x + \cdots + x^{n-1}) \right| \\ &\leq \sum_{N+1}^{\infty} |a_n| (|y|^{n-1} + \cdots + |x|^{n-1}) \\ &\leq \sum_{N+1}^{\infty} n |a_n| T^{n-1} < \frac{\varepsilon}{3}. \end{aligned}$$



The sum

$$\sum_{n=1}^N a_n (y^{n-1} + y^{n-2}x + \cdots + x^{n-1})$$

is a polynomial in  $y$  whose value at  $x$  is  $\sum_1^N na_n x^{n-1}$  so there is a  $\delta_0 > 0$  with the property that if  $0 < |y - x| < \delta_0$

$$\begin{aligned} & \left| \sum_1^N a_n \frac{y^n - x^n}{y - x} - \sum_1^N na_n x^{n-1} \right| \\ &= \left| \sum_1^N a_n (y^{n-1} + y^{n-2}x + \cdots + x^{n-1}) - \sum_1^N na_n x^{n-1} \right| < \frac{\varepsilon}{3}. \end{aligned}$$

So if we choose  $\delta$  to be the smaller of  $\delta_0$  and  $T - |x|$  then whenever  $0 < |y - x| < \delta$

$$\begin{aligned} & \left| \sum_1^\infty a_n \frac{y^n - x^n}{y - x} - \sum_1^\infty na_n x^{n-1} \right| \\ & \leq \left| \sum_{N+1}^\infty a_n \frac{y^n - x^n}{y - x} \right| + \left| \sum_1^N a_n \frac{y^n - x^n}{y - x} - \sum_1^N na_n x^{n-1} \right| + \left| \sum_{N+1}^\infty na_n x^{n-1} \right| < \varepsilon. \end{aligned}$$

□

**Corollary (The derivative of the exponential).**

$$\exp'(x) = \exp(x).$$

*Proof*

$$\exp(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots .$$

We can differentiate term by term to get

$$\exp'(x) = 0 + 1 + x + \frac{x^2}{2} + \cdots = \exp(x).$$

□

As promised we can now give a simple proof of the characteristic property of the exponential.

**Corollary (The characteristic property of the exponential).** *If  $x$  and  $y$  are real numbers then  $\exp(x + y) = \exp(x) \exp(y)$ .*

*Proof* For a fixed number  $z$  consider the function

$$x \mapsto \exp(x) \exp(z - x).$$

We may differentiate this with respect to  $x$  using the product rule and the chain rule to get

$$\exp(x) \exp(z - x) - \exp(x) \exp(z - x) = 0.$$

By the MVT the function is constant. At  $x = 0$  the function is  $\exp z$  so we know that for all  $x$

$$\exp(x) \exp(z - x) = \exp(z).$$

Now if we set  $z = x + y$  we get the conclusion we want. □

We now know that the exponential function is differentiable, has the correct derivative and satisfies the characteristic property. Using what we did earlier on the derivatives of inverses we can also conclude that the logarithm has the correct derivative.

**Corollary (The derivative of log).** *If  $f : x \mapsto \log x$  then*

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

**Example (Power series solution of a Diff Eq. I).** *Let  $f$  be defined on  $(-1, 1)$  by the series*

$$f(x) = \sum_{n=1}^{\infty} \frac{x^n}{n^2}.$$

*Then  $f$  satisfies the differential equation*

$$xf''(x) + f'(x) = \frac{1}{1-x}.$$

*Proof* The series has radius of convergence 1 as you saw in the homework.

$$f'(x) = \sum_{n=1}^{\infty} \frac{x^{n-1}}{n}$$
$$f''(x) = \sum_{n=1}^{\infty} \frac{(n-1)x^{n-2}}{n}$$

and so

$$xf''(x) = \sum_{n=1}^{\infty} \frac{(n-1)x^{n-1}}{n}.$$

So

$$xf''(x) + f'(x) = \sum_{n=1}^{\infty} \frac{nx^{n-1}}{n} = \sum_{n=1}^{\infty} x^{n-1} = \frac{1}{1-x}.$$

□

**Example (Power series solution of a Diff Eq. II).** Consider the differential equation

$$xy'' + y' + xy = 0.$$

Suppose there were a solution given by a power series

$$y = \sum_0^{\infty} a_n x^n.$$

Then

$$xy'' = \sum_2^{\infty} n(n-1)a_n x^{n-1} \quad \text{and} \quad y' = \sum_1^{\infty} na_n x^{n-1}$$

while

$$xy = \sum_2^{\infty} a_{n-2} x^{n-1}.$$

So

$$a_1 + \sum_2^{\infty} (n^2 a_n + a_{n-2}) x^{n-1} = 0.$$

We can look at the coefficients of each power and set them to 0

$$a_1, \quad a_0 + 4a_2, \quad a_1 + 9a_3, \quad a_2 + 16a_4, \quad \dots$$

We can make these all zero by choosing all the odd numbered  $a_i$  to be zero but then take

$$a_0 = 1, \quad a_2 = -\frac{1}{4}, \quad a_4 = \frac{1}{64}, \quad \dots$$

and more generally

$$a_{2m} = \frac{(-1)^m}{4^m (m!)^2}.$$

It is quite easy to see that if we define a function  $J_0$  by

$$J_0(x) = \sum_0^{\infty} \frac{(-1)^m}{4^m(m!)^2} x^{2m}$$

then the series has infinite radius of convergence and the function satisfies the differential equation.

$$xy'' + y' + xy = 0.$$

This function is called the Bessel function of order zero. Bessel functions turn up in the modelling of the vibration of a drumhead.

The last item in this section will explain how we can determine the coefficients of a power series if we know the function. Suppose I know that

$$f(x) = \sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + \dots$$

on the interval  $(-R, R)$ . From the coefficients I can build the function by carrying out the sum. How can I go the other way? If I know the function  $f$  how can I find its coefficients?

The first one is easy  $a_0 = f(0)$ . The next one,  $a_1$  is not so obvious but we know that

$$f'(x) = a_1 + 2a_2 x + 3a_3 x^2 + \dots$$

and so we can calculate  $a_1$  by differentiating  $f$ : we have  $a_1 = f'(0)$ .

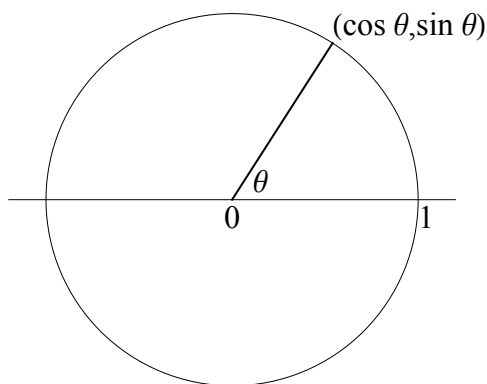
We can continue in this way

$$\begin{aligned} a_0 &= f(0) \\ a_1 &= f'(0) \\ a_2 &= \frac{f''(0)}{2} \\ &\vdots \end{aligned}$$

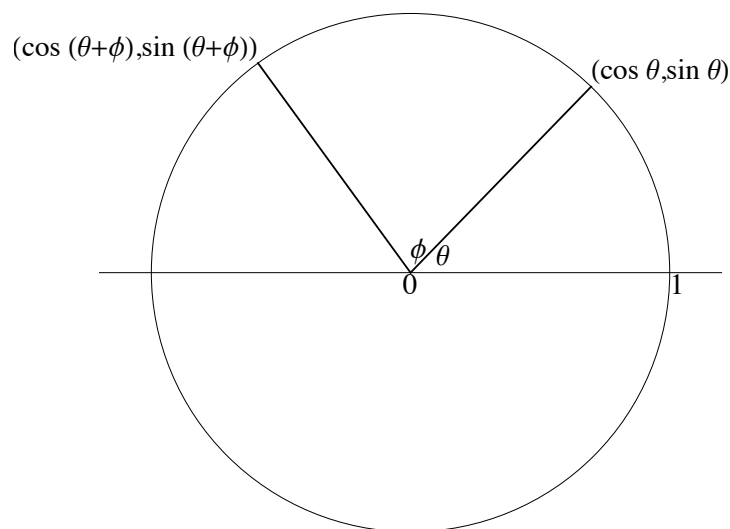
You will recognise these numbers as the coefficients in the Taylor expansion for  $f$  to which we shall return in the last chapter.

## The trigonometric functions

For the purposes of this course we shall define the trigonometric functions sine and cosine by power series and then check that they have the properties we expect them to have. Our aim will be to show that for each  $\theta$  the point  $(\cos \theta, \sin \theta)$  lies on the circle of radius one with centre 0 and that the radius through this point makes an angle  $\theta$  with the horizontal.

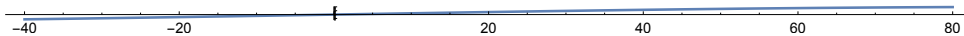


The first statement is just that  $\cos^2 \theta + \sin^2 \theta = 1$ . For the second we need to be clear that we are measuring angle in radians so let us recall what that means. The Babylonians were originally responsible for the division of the circle into 360 equal parts: what we now call degrees. Measuring angle in degrees has the property that adding two angles corresponds to the geometric process of rotating through one angle followed by another.



This is a desirable property (also possessed by radians). However, degrees have a property

that is far from desirable. Below is the graph of  $y = \sin x^\circ$  drawn with the same scale on both axes.



The slope of the graph at  $x = 0$  is  $\pi/180 \approx 0.0174533$ . So if we differentiate the function we get  $0.0174\dots \cos x^\circ$ : we get a funny multiple of  $\cos$  rather than  $\cos$  itself. The choice of radians removes this problem.

How do we measure the size of an angle in radians? We draw the circle of radius 1 and then for a given angle we use the length of the circular arc that it spans as the measure of the angle.

The size of the angle is the length of the arc it spans: simple.

So the second thing we will need to check is that for each  $t$ , the point on the circle whose distance from the horizontal measured around the circle is  $t$ , has coordinates  $(\cos t, \sin t)$ .

**Definition (The trig functions).** For  $x \in \mathbf{R}$  we define

$$\begin{aligned}\cos x &= 1 - \frac{x^2}{2} + \frac{x^4}{24} - \dots + (-1)^k \frac{x^{2k}}{(2k)!} + \dots \\ \sin x &= x - \frac{x^3}{6} + \frac{x^5}{120} - \dots + (-1)^k \frac{x^{2k+1}}{(2k+1)!} + \dots\end{aligned}$$

It is easy to check that these series converge everywhere by the ratio test. Cosine is obviously an even function and sine is obviously odd. From our general theory we know that both functions are differentiable and we can check that

$$\frac{d}{dx} \cos x = -\sin x$$

and

$$\frac{d}{dx} \sin x = \cos x.$$

Using these derivatives we can check the addition formulae quite easily.

**Theorem (The addition formulae for the trig functions).** For all real  $x$  and  $y$

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

*Proof* We shall check the first: the second is similar. For a fixed  $z$  let

$$f(x) = \cos x \cos(z - x) - \sin x \sin(z - x).$$

Using the derivatives above it is easy to check that  $f'(x) = 0$  for all  $x$  and so  $f$  is constant. When  $x = 0$  we have

$$f(0) = \cos 0 \cos z - \sin 0 \sin z = \cos z.$$

Hence  $f(x) = \cos z$  for all  $x$ . Now if we set  $z = x + y$  we get

$$\cos x \cos y - \sin x \sin y = f(x) = \cos z = \cos(x + y).$$

□

This immediately shows us that the point  $(\cos x, \sin x)$  lies on the circle for each  $x$ .

**Corollary (The circular property).** *For all real  $x$*

$$\cos^2 x + \sin^2 x = 1.$$

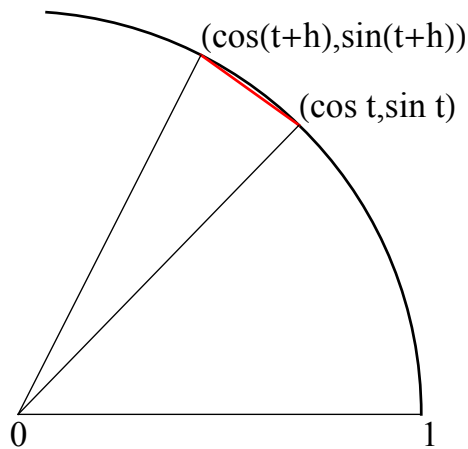
*Proof* Set  $y = -x$  in the addition formula for  $\cos$ . We know that  $\cos(-x) = \cos x$  and  $\sin(-x) = -\sin x$  and so we get

$$1 = \cos 0 = \cos(x - x) = \cos x \cos(-x) - \sin x \sin(-x) = \cos^2 x + \sin^2 x.$$

□

The remaining thing we need to check is that as  $t$  increases the point  $(\cos t, \sin t)$  traces out the circle at rate 1. Let  $L(t)$  be the length of the circular arc from  $(1, 0)$  to the point  $(\cos t, \sin t)$ . We want to show that  $L(t) = t$  for each  $t$ . By the MVT it suffices to show that  $L'(t) = 1$  at each point.

Consider the point  $(\cos t, \sin t)$  and a nearby point  $(\cos(t + h), \sin(t + h))$ .



When  $h$  is very small, the (straight line) distance between these two points is approximately the same as the length of the circular arc between them. The straight line distance is

$$\sqrt{(\cos(t+h) - \cos t)^2 + (\sin(t+h) - \sin t)^2}.$$

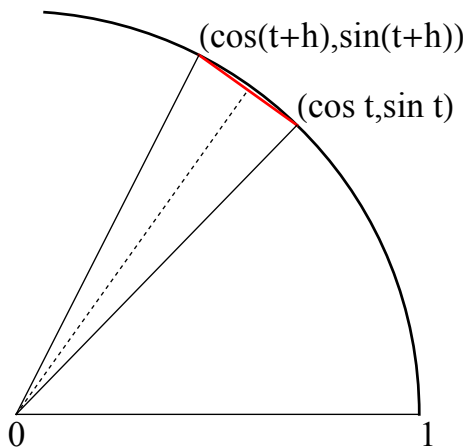
So our aim is to show that

$$\lim_{h \rightarrow 0^+} \frac{1}{h} \sqrt{(\cos(t+h) - \cos t)^2 + (\sin(t+h) - \sin t)^2} = 1$$

for every  $t$ .

If we knew that our power series *do* give the point at the correct angle then we could use geometry to calculate the length:





The length is supposed to be  $2 \sin(h/2)$ . That would be good because then

$$\begin{aligned} & \lim_{h \rightarrow 0^+} \frac{1}{h} \sqrt{(\cos(t+h) - \cos t)^2 + (\sin(t+h) - \sin t)^2} \\ &= \lim_{h \rightarrow 0^+} \frac{2 \sin(h/2)}{h} = \lim_{h \rightarrow 0^+} \frac{\sin(h/2)}{h/2} = \lim_{p \rightarrow 0^+} \frac{\sin p}{p} = 1 \end{aligned}$$

because the last expression is just the derivative of  $\sin$  at 0.

However we don't yet know that our functions cosine and sine correspond to the geometry of the circle. So we have to check that for  $h > 0$  (and no bigger than  $2\pi$ )

$$\sqrt{(\cos(t+h) - \cos t)^2 + (\sin(t+h) - \sin t)^2} = 2 \sin\left(\frac{h}{2}\right)$$

using the properties that we do know: the addition formulae. But

$$\begin{aligned} & (\cos(t+h) - \cos t)^2 + (\sin(t+h) - \sin t)^2 \\ &= \cos^2(t+h) - 2 \cos(t+h) \cos t + \cos^2 t + \sin^2(t+h) - 2 \sin(t+h) \sin t + \sin^2 t \\ &= 2 - 2 \cos(t+h) \cos t - 2 \sin(t+h) \sin t \\ &= 2 - 2 \cos(t+h-t) = 2 - 2 \cos h = 4 \sin^2(h/2) \end{aligned}$$

by repeated use of the addition formulae. When  $h$  is positive (and less than  $2\pi$ ) then  $\sin(h/2)$  is positive so the square root is  $2 \sin(h/2)$ . This completes the proof.

We have now checked that our power series do produce the  $x$  and  $y$  coordinates of the correct point on the circle. If we use the symbol  $\pi$  to denote half the circumference of the circle then we know that  $\cos \pi = -1$ ,  $\sin \pi = 0$  and so on. We can also see that the trig functions are periodic with period  $2\pi$  although we could deduce that directly from the addition formulae once we know that  $\cos 2\pi = 1$  and  $\sin 2\pi = 0$ . For example

$$\cos(x + 2\pi) = \cos x \cos 2\pi - \sin x \sin 2\pi = \cos x.$$

There are a number of special values that one has to know.

$$\begin{array}{ll} \sin 0 = 0 & \cos 0 = 1 \\ \sin \pi/6 = 1/2 & \cos \pi/6 = \sqrt{3}/2 \\ \sin \pi/4 = 1/\sqrt{2} & \cos \pi/4 = 1/\sqrt{2} \\ \sin \pi/3 = \sqrt{3}/2 & \cos \pi/3 = 1/2 \end{array}$$

## The complex exponential

The proof of the addition formulae for the trig functions looks very much like the proof of the characteristic property of the exponential and the addition formulae have the same “shape”. The cosine of the sum  $\cos(x + y)$  is a product of the cosines of  $x$  and  $y$  together with a product of sines. This is not coincidence. If we introduce the complex number  $i$  whose square is  $-1$  we can write

$$e^{it} = 1 + it - \frac{t^2}{2} - i\frac{t^3}{6} + \frac{t^4}{24} + \dots = \cos t + i \sin t.$$

This formula

$$e^{it} = \cos t + i \sin t$$

linking the exponential and trig functions was described by Feynman as “our jewel”. Now we have

$$\begin{aligned} \cos(x + y) + i \sin(x + y) &= e^{i(x+y)} = e^{ix} e^{iy} \\ &= (\cos x + i \sin x)(\cos y + i \sin y). \end{aligned}$$

The last can be expanded as

$$\cos x \cos y - \sin x \sin y + i(\sin x \cos y + \cos x \sin y)$$

from which we can read off the addition formulae for cosine and sine.

In a similar way we can relate the derivatives of  $\cos$  and  $\sin$  to the derivative of the exponential function.

We chose the exponential function and we chose to measure angle in radians in order to make the derivatives work out right. We can now see that these two choices are really the same choice. The formula

$$e^{it} = \cos t + i \sin t$$

only works if the angles are in radians and we use the correct exponential.

## The tangent

Once we have defined  $\cos$  and  $\sin$  we can define the tangent. We know that  $\cos x = 0$  whenever  $x$  is an odd multiple of  $\pi/2$ . For all other points we can define

$$\tan x = \frac{\sin x}{\cos x}.$$

Using our knowledge of the derivatives of  $\cos$  and  $\sin$  we can find the derivative of  $\tan$ . In the HW you are asked to find the derivative of the inverse,  $\tan^{-1}$  and from this to find a power series for  $\tan^{-1}$ . You are also asked to derive the addition formula for  $\tan$  from those for  $\cos$  and  $\sin$ .

An obvious question: what is the power series for  $\tan x$  for  $x$  close to 0? Assuming that the power series exists you can find it by repeatedly differentiating  $\tan$ . Here are the first few terms

$$x + \frac{x^3}{3} + \frac{2x^5}{15} + \frac{17x^7}{315} + \frac{62x^9}{2835} + \frac{1382x^{11}}{155925} + \dots$$

The numerator 1382 has a prime factor 691. This suggests that you cannot write down a simple formula for the coefficients: and indeed you can't. They are related to values of what is known as the  $\zeta$  function.

## Chapter 5. Taylor's Theorem

### Cauchy's Mean Value Theorem

We start this section with a more general form of the MVT known as Cauchy's MVT. The MVT says that under suitable conditions there is a point  $t$  between  $a$  and  $b$  with

$$f'(t) = \frac{f(b) - f(a)}{b - a}.$$

This can be interpreted as saying that if  $g$  is the function given by  $g(x) = x$  then

$$\frac{f'(t)}{g'(t)} = \frac{f(b) - f(a)}{g(b) - g(a)}$$

since  $g'(t) = 1$  for every  $t$ . Cauchy's MVT says that the same thing holds for any differentiable function  $g$  for which the statement makes sense. This theorem will be of interest to us solely in order to prove something called l'Hôpital's Rule for finding limits.

**Theorem (Cauchy's Mean Value Theorem).** *If  $f, g : [a, b] \rightarrow \mathbf{R}$  are continuous, are differentiable on  $(a, b)$  and  $g'(t) \neq 0$  for  $t$  between  $a$  and  $b$  then there is a point  $t$  where*

$$\frac{f'(t)}{g'(t)} = \frac{f(b) - f(a)}{g(b) - g(a)}.$$

*Proof* Consider the function

$$x \mapsto h(x) = f(x)(g(b) - g(a)) - (f(b) - f(a))g(x).$$

At  $x = a$  the value of the function is

$$h(a) = f(a)g(b) - f(a)g(a) - f(b)g(a) + f(a)g(a) = f(a)g(b) - f(b)g(a).$$

Similarly

$$h(b) = f(b)g(b) - f(b)g(a) - f(b)g(b) + f(a)g(b) = f(a)g(b) - f(b)g(a).$$

So  $h(b) = h(a)$  and by Rolle's Theorem there is a point  $t$  between  $a$  and  $b$  where  $h'(t) = 0$ .

Thus we have

$$h(x) = f(x)(g(b) - g(a)) - (f(b) - f(a))g(x)$$

and there is a point  $t$  where  $h'(t) = 0$ . But this means that

$$f'(t)(g(b) - g(a)) = (f(b) - f(a))g'(t).$$

Since  $g'$  is non-zero on  $(a, b)$  Rolle's Theorem applied to  $g$  shows that  $g(b) - g(a) \neq 0$  as well and we can rearrange to get the conclusion of the theorem.  $\square$

## L'Hôpital's rule

It often happens that we wish to calculate limits such as

$$\lim_{x \rightarrow 0} \frac{\sin x}{1 - \sqrt{1 - x}}$$

in which both the numerator and denominator converge to 0. So we can't calculate the limit just by substituting  $x = 0$ . An obvious example that we have already considered and understood is the definition of  $f'(c)$

$$\lim_{x \rightarrow c} \frac{f(x) - f(c)}{x - c}.$$

It is not surprising that in the more general situation, derivatives can often provide an answer. In the example above

$$\lim_{x \rightarrow 0} \frac{\sin x}{1 - \sqrt{1 - x}}$$

we know that the derivative of  $\sin$  at 0 is 1 and therefore that

$$\sin x \approx x$$

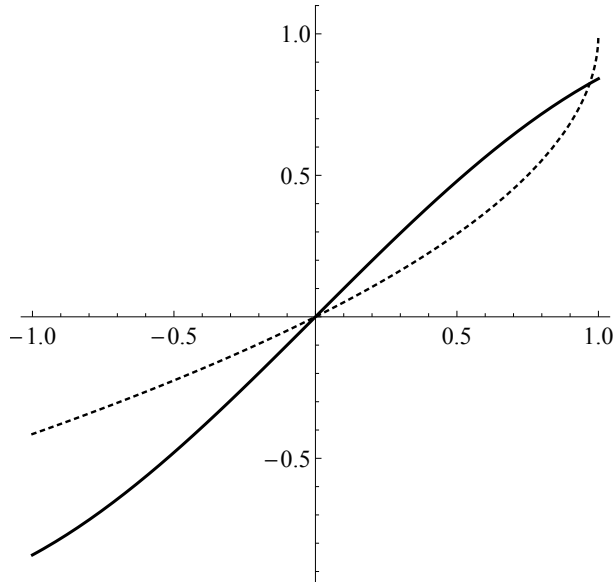
when  $x$  is close to 0. The derivative of  $x \mapsto \sqrt{1 - x}$  at 0 is  $-1/2$ . Therefore

$$\sqrt{1 - x} \approx 1 - x/2$$

when  $x$  is close to 0. This in turn means that  $1 - \sqrt{1 - x} \approx x/2$  and so it looks as though

$$\frac{\sin x}{1 - \sqrt{1 - x}} \approx \frac{x}{x/2} = 2.$$

The figure shows the graphs of  $\sin x$  (solid) and  $1 - \sqrt{1 - x}$  (dashed). Both pass through the origin. The slope of  $\sin$  at the origin is 1. The slope of the other graph is  $1/2$ . So close to the origin the value of the sine graph is about twice as large as that of the other graph.



We can make this argument rigorous and streamline the calculation: the upshot is a principle known as l'Hôpital's rule.

**Theorem (l'Hôpital's rule).** *If  $f, g : I \rightarrow \mathbf{R}$  are differentiable on the open interval  $I$  containing  $c$  and  $f(c) = g(c) = 0$  then*

$$\lim_{x \rightarrow c} \frac{f(x)}{g(x)} = \lim_{x \rightarrow c} \frac{f'(x)}{g'(x)}$$

*provided the second limit exists.*

In the example above we take

$$f(x) = \sin x \quad \text{for which} \quad f'(x) = \cos x$$

and

$$g(x) = 1 - (1 - x)^{1/2} \quad \text{for which} \quad g'(x) = 1/2(1 - x)^{-1/2}.$$

In order to compute the limit

$$\lim_{x \rightarrow 0} \frac{\sin x}{1 - \sqrt{1 - x}}$$

we consider the ratio of the derivatives

$$\lim_{x \rightarrow 0} \frac{\cos x}{1/2(1 - x)^{-1/2}} = \lim_{x \rightarrow 0} \frac{2 \cos x}{(1 - x)^{-1/2}} = 2.$$

Therefore according to l'Hôpital

$$\lim_{x \rightarrow 0} \frac{\sin x}{1 - \sqrt{1 - x}} = 2.$$

The advantage of expressing the theorem in the way we did, with a limit on the right, is that it may be possible to use it repeatedly. If we find that  $f'(c) = g'(c) = 0$  then we cannot calculate the limit of the derivatives just by substitution as we did with the example. But we can apply the theorem a second time to get

$$\lim_{x \rightarrow c} \frac{f''(x)}{g''(x)}$$

provided that the functions are twice differentiable.

For example, suppose we want

$$\lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2}.$$

The numerator and denominator both approach 0 as  $x \rightarrow 0$ . Differentiating both with respect to  $x$  we are led to consider the limit

$$\lim_{x \rightarrow 0} \frac{\sin x}{2x}.$$

The numerator and denominator of this fraction also approach 0 as  $x \rightarrow 0$  so we can differentiate again and consider

$$\lim_{x \rightarrow 0} \frac{\cos x}{2} = \frac{1}{2}.$$

*Proof (of l'Hôpital's rule).* Suppose that

$$\lim_{x \rightarrow c} \frac{f'(x)}{g'(x)}$$

does indeed exist. Then it cannot be that  $g'(x) = 0$  at a sequence of points converging to  $c$ . So there is some interval around  $c$  on which  $g'$  is non-zero (except perhaps at  $c$  itself). So  $g'$  is non-zero on an interval each side of  $c$ . This enables us to apply Cauchy's MVT. Because  $f(c) = g(c) = 0$

$$\lim_{x \rightarrow c} \frac{f(x)}{g(x)} = \lim_{x \rightarrow c} \frac{f(x) - f(c)}{g(x) - g(c)}.$$

As long as  $x$  is in the region around  $c$  where  $g' \neq 0$  Cauchy's MVT ensures that there is a point  $t$  (depending upon  $x$ ) between  $c$  and  $x$  where

$$\frac{f(x) - f(c)}{g(x) - g(c)} = \frac{f'(t)}{g'(t)}.$$

As  $x \rightarrow c$  the corresponding  $t$  is forced to approach  $c$  as well and so

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(c)}{g(x) - g(c)} \rightarrow \lim_{t \rightarrow c} \frac{f'(t)}{g'(t)}.$$

□

There is also a version of l'Hôpital at infinity.

**Theorem (l'Hôpital's rule at infinity).** *If  $f, g : \mathbf{R} \rightarrow \mathbf{R}$  are differentiable and*

$$\lim_{x \rightarrow \infty} f(x) = 0 \quad \text{and} \quad \lim_{x \rightarrow \infty} g(x) = 0$$

or

$$\lim_{x \rightarrow \infty} f(x) = \infty \quad \text{and} \quad \lim_{x \rightarrow \infty} g(x) = \infty$$

then

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

provided the second limit exists.

**Example.**

$$\lim_{x \rightarrow \infty} \frac{x}{e^x} = \lim_{x \rightarrow \infty} \frac{1}{e^x} = 0.$$

l'Hôpital's rule is included in this chapter because it takes the form of an extended mean value theorem. However the main topic of the section will be Taylor's Theorem.

## Taylor's Theorem with remainder

At school you met Taylor expansions. You saw a statement something along the lines of

$$f(x) \approx f(a) + f'(a)(x - a) + \frac{f''(a)}{2}(x - a)^2 + \dots$$

For most purposes we need something a bit more precise. It isn't clear from the statement above how good the approximation is. Let us recall the example from the introduction.

$$(1 + x)^{1/2} \approx 1 + \frac{x}{2} - \frac{x^2}{8}.$$



If  $x = 1$  for example we can check that the error is

$$\sqrt{2} - (1 + 1/2 - 1/8) = \sqrt{2} - 11/8 = 0.0392\dots$$

But this depends upon knowing the value  $\sqrt{2}$  which is the number we are trying to approximate. Can we estimate the error without actually calculating the thing we want to approximate? The error will depend in a complicated way on the particular function that we are approximating. So we will have to express the error in terms of the function: the aim is to find an expression which we can (usually) estimate.

The first such expression is contained in the following theorem.

**Theorem (Taylor's Theorem, Lagrange Remainder).** *If  $f : I \rightarrow \mathbf{R}$  is  $n$  times differentiable on the open interval  $I$  containing  $a$  and  $b$  then*

$$\begin{aligned} f(b) = & f(a) + f'(a)(b-a) + \frac{f''(a)}{2}(b-a)^2 + \dots \\ & + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1} + \frac{f^{(n)}(t)}{n!}(b-a)^n \end{aligned}$$

for some point  $t$  between  $a$  and  $b$ .

The number  $t$  depends upon the function  $f$  as well as upon  $a$  and  $b$  so we don't have any way of determining what it is in general. The error is given in terms of the  $n^{\text{th}}$  derivative of the function. If we can calculate this derivative we may be able to show that it can't be too big anywhere between  $a$  and  $b$  and so it won't matter that we don't know the exact value of  $t$ .

We will give two proofs of this theorem. The first is the "natural" proof. The second is a trick proof which allows us to prove several different versions of the theorem which are useful for different purposes. The first proof is an extension of our original proof for the MVT in which we modified  $f$  by a linear function. In this argument we shall modify  $f$  (to get a new function  $h$ ) by a polynomial of degree  $n$ .

*Proof* The function  $g$  given by

$$g(x) = f(x) - \left( f(a) + f'(a)(x-a) + \dots + \frac{f^{(n-1)}(a)}{(n-1)!}(x-a)^{n-1} \right)$$

satisfies  $g(a) = 0$ ,  $g'(a) = 0$  and so on up to  $g^{(n-1)}(a) = 0$ . It also satisfies  $g^{(n)}(x) = f^{(n)}(x)$  for all  $x$  because  $f$  and  $g$  differ by a polynomial of degree only  $n-1$ .

If we put

$$h(x) = g(x) - g(b) \frac{(x-a)^n}{(b-a)^n}$$

then  $h$  also has its first  $n-1$  derivatives vanishing at  $a$  but in addition it satisfies  $h(b) = 0$ .

We now proceed inductively. Since  $h(b) = h(a) = 0$  there is a point  $t_1$  in  $(a, b)$  where  $h'(t_1) = 0$  by Rolle's Theorem. Since  $h'(t_1) = h'(a) = 0$  there is a point  $t_2$  in  $(a, t_1)$  with  $h''(t_2) = 0$ . Continuing in this way we eventually get a point  $t = t_n$  where  $h^{(n)}(t) = 0$ . In terms of  $g$  this says that

$$g^{(n)}(t) = g(b) \frac{n!}{(b-a)^n}.$$

$$g(x) = f(x) - \left( f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!} (x-a)^{n-1} \right)$$

so

$$g(b) = \frac{g^{(n)}(t)}{n!} (b-a)^n = \frac{f^{(n)}(t)}{n!} (b-a)^n$$

or

$$f(b) - \left( f(a) + f'(a)(b-a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!} (b-a)^{n-1} \right) = \frac{f^{(n)}(t)}{n!} (b-a)^n$$

which is exactly the statement of the theorem. □

### Example.

Let us try to estimate the error in the Taylor approximation at 0

$$(1+x)^{1/2} \approx 1 + \frac{x}{2} - \frac{x^2}{8}$$

for  $x = 1/2$  say. If  $f(u) = (1+u)^{1/2}$  then

$$\begin{aligned} f'(u) &= 1/2(1+u)^{-1/2} \\ f''(u) &= -1/4(1+u)^{-3/2} \\ f'''(u) &= 3/8(1+u)^{-5/2} \end{aligned}$$

so when  $u$  is zero we get

$$\begin{aligned} f(0) &= 1 \\ f'(0) &= 1/2 \\ f''(0) &= -1/4. \end{aligned}$$

The theorem tells us that

$$(1+x)^{1/2} = f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \frac{f'''(t)}{6}x^3 = 1 + \frac{x}{2} - \frac{x^2}{8} + \frac{f'''(t)}{6}x^3$$

for some  $t$  between 0 and  $x$ . Now

$$f'''(u) = 3/8(1+u)^{-5/2}$$

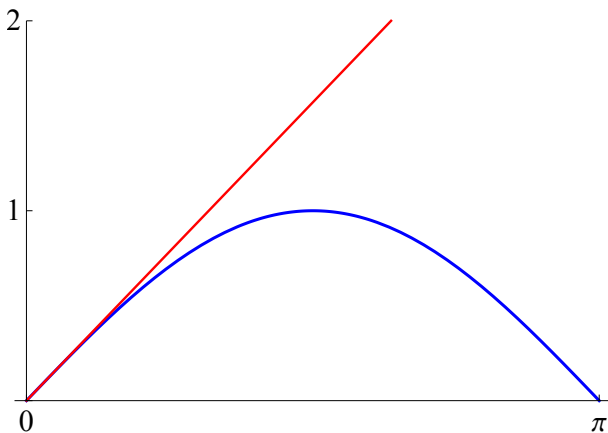
so when  $x = 1/2$  the error is

$$\frac{f'''(t)}{6}x^3 = \frac{3/8(1+t)^{-5/2}}{6} \left(\frac{1}{2}\right)^3 = \frac{1}{16 \times 8(1+t)^{5/2}} = \frac{1}{128(1+t)^{5/2}}.$$

We don't know the value of  $t$  but we do know that it lies between 0 and  $1/2$ . So  $1+t > 1$  and hence the error cannot be more than  $1/128$ . In fact the error is about 0.006 so our estimate  $1/128 = 0.0078\dots$  is quite good.

Taylor's Theorem can be used to prove inequalities like the ones we have seen before:  $e^x \geq 1+x$  and so on.

**Example.** If  $0 \leq x \leq \pi$  then  $\sin x \leq x$ .



*Proof* Let  $f(x) = \sin x$ . We have  $f'(x) = \cos x$  and  $f''(x) = -\sin x$ . So

$$f(0) = 0, \quad f'(0) = 1$$

and by Taylor's Theorem with  $n = 2$

$$f(x) = f(0) + f'(0)x + \frac{f''(t)}{2}x^2.$$

This says

$$\sin x = 0 + x - \frac{\sin t}{2}x^2$$

for some  $t$  between 0 and  $x$ . As long as  $0 \leq x \leq \pi$  we have  $0 < t < \pi$  and so  $\sin t > 0$ .  $\square$

In the Homework you showed that for  $|x| < 1$

$$-\log(1-x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \dots$$

which can be rewritten as a series for the logarithm near 1:

$$\log x = (x-1) - \frac{(x-1)^2}{2} + \dots$$

In our next example we shall derive this using Taylor's Theorem.

**Example (Taylor Series for log).**

Let  $f$  be the function  $f : x \mapsto \log x$ . Then we know

$$\begin{aligned} f'(x) &= x^{-1} \\ f''(x) &= -x^{-2} \\ f'''(x) &= 2x^{-3} \\ &\vdots \\ f^{(n)}(x) &= (-1)^{n-1} (n-1)! x^{-n}. \end{aligned}$$

Let us find the Taylor approximations at  $x = 1$ .

$$\begin{aligned} f(1) &= 0 \\ f'(1) &= 1 \\ f''(1) &= -1 \\ f'''(1) &= 2 \\ &\vdots \\ f^{(n)}(1) &= (-1)^{n-1} (n-1)!. \end{aligned}$$

The  $n^{\text{th}}$  Taylor formula at 1 is therefore

$$\begin{aligned} \log x &= 0 + (x-1) - \frac{(x-1)^2}{2} + \cdots + (-1)^{n-2} \frac{(x-1)^{n-1}}{n-1} \\ &\quad + (-1)^{n-1} t^{-n} \frac{(x-1)^n}{n} \\ &= (x-1) - \frac{(x-1)^2}{2} + \cdots + (-1)^{n-2} \frac{(x-1)^{n-1}}{n-1} \\ &\quad + (-1)^{n-1} \frac{1}{n} \left( \frac{x-1}{t} \right)^n \end{aligned}$$

where  $t$  is a number between 1 and  $x$ .

If  $1 \leq x \leq 2$  then  $0 < x-1 \leq 1 \leq t$ . Hence the error term is at most  $\frac{1}{n}$  which tends to 0 as  $n \rightarrow \infty$ . This means that we can take a limit and conclude that we have an infinite series for  $\log x$

$$\log x = (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \cdots$$

as long as  $1 \leq x \leq 2$ . If we set  $x = 2$  then the series becomes

$$1 - \frac{1}{2} + \frac{1}{3} - \cdots$$

which converges by the alternating series theorem. In the homework you will see another proof that it converges to  $\log 2$  which the earlier proof didn't give.

The same argument works with a bit more care if  $1/2 \leq x < 1$ . But if  $x < 1/2$  then it might go wrong: for example if  $x = 1/3$  and  $t = 1/2$  then

$$\frac{x-1}{t} = -\frac{4}{3}$$

and so the powers of this number increase very fast. It turns out that the series does converge to the logarithm for  $0 < x < 1/2$  but the Lagrange Remainder is the wrong tool for proving this. (To put it another way, the error term is in fact small and so the number  $t$  that actually gives the remainder will be bigger than  $x-1$  but the theorem doesn't tell us this.)

Cauchy came up with a different form of the remainder that works nicely in some cases. We will deduce it from the following version of Taylor's Theorem.

**Theorem (Taylor's Theorem, Trick Proof).** If  $f : I \rightarrow \mathbf{R}$  is  $n$  times differentiable on the open interval  $I$  containing  $a$  and  $b$  and  $0 \leq k \leq n - 1$  then

$$f(b) = f(a) + f'(a)(b-a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1} \\ + \frac{f^{(n)}(t)}{(n-1)!(n-k)}(b-t)^k(b-a)^{n-k}$$

for some point  $t$  between  $a$  and  $b$ .

*Proof* Let  $R$  be the remainder for which we are trying to find a formula

$$R = f(b) - \left( f(a) + f'(a)(b-a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1} \right).$$

This time we define

$$h(x) = f(b) - f(x) - f'(x)(b-x) - \cdots - \frac{f^{(n-1)}(x)}{(n-1)!}(b-x)^{n-1} - R \frac{(b-x)^{n-k}}{(b-a)^{n-k}}.$$

Obviously  $h(b) = 0$ . Also  $h(a) = 0$  because of the definition of  $R$ . So by Rolle's Theorem there is a number  $t$  between  $a$  and  $b$  where  $h'(t) = 0$ . So

$$h'(x) = -f'(x) + f'(x) - f''(x)(b-x) + \frac{f''(x)}{2}2(b-x) - \frac{f'''(x)}{2}(b-x)^2 \\ + \cdots - \frac{f^{(n)}(x)}{(n-1)!}(b-x)^{n-1} + (n-k)R \frac{(b-x)^{n-k-1}}{(b-a)^{n-k}} \\ = -\frac{f^{(n)}(x)}{(n-1)!}(b-x)^{n-1} + (n-k)R \frac{(b-x)^{n-k-1}}{(b-a)^{n-k}}$$

where for each term we use the product rule, and most of the sum cancels out.

If we now substitute  $t$  and set the expression equal to 0 we get

$$R = \frac{f^{(n)}(t)}{(n-1)!(n-k)}(b-t)^k(b-a)^{n-k}$$

which is the statement we wanted. □

The Lagrange remainder form of Taylor's Theorem follows immediately by setting  $k = 0$ . The most useful other consequence of the theorem is the case  $k = n - 1$  which is known as Taylor's Theorem with Cauchy remainder.

**Theorem (Taylor's Theorem, Cauchy Remainder).** If  $f : I \rightarrow \mathbf{R}$  is  $n$  times differentiable on the open interval  $I$  containing  $a$  and  $b$  then

$$f(b) = f(a) + f'(a)(b-a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1} + \frac{f^{(n)}(t)}{(n-1)!}(b-t)^{n-1}(b-a)$$

for some point  $t$  between  $a$  and  $b$ .

The point is that we have replaced  $(b-a)^n$  by  $(b-t)^{n-1}(b-a)$  and  $t$  is closer to  $b$  than  $a$  is.

Let us return to the logarithm and try to check the error for  $0 < x < 1$ . We have  $f(x) = \log x$  and we set  $a = 1$  and  $b = x$ . The Taylor formula is

$$\log x = (x-1) - \frac{(x-1)^2}{2} + \cdots + (-1)^{n-2} \frac{(x-1)^{n-1}}{n-1} + (-1)^{n-1} \frac{(x-t)^{n-1}(x-1)}{t^n}$$

We now have an error term which apart from  $(-1)^{n-1}$  is

$$\frac{(x-t)^{n-1}(x-1)}{t^n} = \frac{x-1}{t} \left(\frac{x-t}{t}\right)^{n-1} = \frac{x-1}{t} (-1)^{n-1} \left(1 - \frac{x}{t}\right)^{n-1}$$

where  $0 < x < t < 1$ . Since  $x < t < 1$  we have that  $0 < 1 - x/t < 1 - x$  and so the error tends to zero as  $n \rightarrow \infty$  as fast as the exponential  $(1-x)^{n-1}$ . Therefore the infinite series converges for  $0 < x < 1$ .

**Theorem (The log series).** If  $0 < x \leq 2$

$$\log x = (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \cdots$$

Equivalently, if  $-1 \leq t < 1$

$$-\log(1-t) = t + \frac{t^2}{2} + \frac{t^3}{3} + \cdots$$

The second form tends to be the one we use most often. You already saw a different proof of this in the homework.

**Example (Binomial Series).**

We know the Binomial Theorem which allows us to expand a power

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2}x^2 + \dots + x^n$$

if  $n$  is a non-negative whole number. We also know that

$$(1+x)^{-1} = 1 - x + x^2 - x^3 + \dots$$

provided  $-1 < x < 1$ .

It is not altogether surprising that we can find a series expansion for every power

$$(1+x)^s = 1 + sx + \frac{s(s-1)}{2}x^2 + \frac{s(s-1)(s-2)}{6}x^3 + \dots$$

provided  $|x| < 1$ . (The series will be infinite unless  $n$  is a non-negative integer.) In the homework you are asked to prove this (without using Taylor series).



## Appendix. The radius of convergence formula

In this last section we shall produce a formula for the radius of convergence of a general power series. In order to do so we shall need to recall the definition of the lim sup. If  $(x_n)$  is a sequence of real numbers which is bounded above we can examine the suprema of the tails

$$u_m = \sup \{x_n : n \geq m\}.$$

The further out the starting point  $m$  the fewer terms we are looking at. So these suprema decrease. So either they converge to some number  $L$  or they decrease to  $-\infty$ . In the first case we say that

$$\limsup x_n = L$$

and in the second that

$$\limsup x_n = -\infty.$$

If the sequence is *not* bounded above we write

$$\limsup x_n = \infty.$$

Whatever the value of the limsup it can be characterised in the following way. If  $\limsup x_n = L$  then for every  $t < L$  there are infinitely many terms of the sequence above  $t$ , while if  $t > L$  there is some point beyond which all terms are below  $t$ . Thus the lim sup tells you how high the sequence reaches over and over again.

Now suppose that we have a power series  $\sum a_n x^n$  and consider the sequence whose terms are  $|a_n|^{1/n}$ . Let its lim sup be  $L$  and note that  $L$  is non-negative (or  $\infty$ ) because the terms are non-negative. Suppose  $|x| > 1/L$  and hence  $1/|x| < L$ . Then for infinitely many values of  $n$

$$|a_n|^{1/n} > 1/|x|$$

and hence

$$|a_n| > 1/|x|^n.$$

So for infinitely many values of  $n$  the terms of the power series satisfy  $|a_n x^n| > 1$ . So the terms do not tend to zero and the series diverges.

On the other hand suppose  $|x| < 1/L$ . Choose a number  $t$  with  $|x| < 1/t < 1/L$ . Since  $t > L$  we know that from some point onward

$$|a_n|^{1/n} < t$$

and hence

$$|a_n| < t^n.$$

From this point on the terms of the series satisfy  $|a_n x^n| < (t|x|)^n$ . We chose  $t$  in such a way that  $t|x| < 1$  so the power series has terms which are dominated by a convergent geometric series and hence it converges.

The upshot is that the series diverges if  $|x| > 1/L$  and converges if  $|x| < 1/L$ . So the radius of convergence is  $1/L$ . The formula we get is sometimes called the Hadamard formula.

**Theorem (The radius of convergence formula).** *The power series  $\sum a_n x^n$  has radius of convergence*

$$\frac{1}{\limsup_{n \rightarrow \infty} |a_n|^{1/n}}.$$

**Example.** *The series  $\sum 2^n x^n$  has radius of convergence  $1/2$ .*

**Example.** *The series  $\sum x^{k^2}$  has radius of convergence  $1$ .*

*Proof* The coefficients are given by

$$a_n = \begin{cases} 1 & \text{if } n \text{ is a square} \\ 0 & \text{if not.} \end{cases}$$

Then

$$|a_n|^{1/n} = \begin{cases} 1 & \text{if } n \text{ is a square} \\ 0 & \text{if not.} \end{cases}$$

So

$$\limsup_{n \rightarrow \infty} |a_n|^{1/n} = 1.$$

□