MS2001

Differential Calculus

COURSE NOTES

Stephen Wills

Department of Mathematics

University College Cork

2005/06

Contents

1 Inequalities **3**
Definitions and axioms 3
Sketching functions; the quadratic 5
Inequalities involving rational functions 10
The modulus function; obtaining bounds 13

2 Limits and Continuity **18**
The intuitive idea of limits 18
The definition of a limit; basic techniques 20
Coping with infinity 27
Continuous functions 29
Continuity on intervals 32

3 Differentiation **35**
The idea and definition of derivatives 35
The Mean Value Theorem and consequences 42
Implicit differentiation 46

4 Curve Sketching and MinMax Problems **49**
Maxima and minima 49
The second derivative test 50
Applied maximum and minimum problems 54

A number of the examples covered in these notes have appeared on exam papers. This is indicated by expressions such as **Example 1.14 (S03 1(a))** which indicates that Example 1.14 was question 1(a) from the Summer paper of 2003. Similarly **Example 1.18 (A02 1(a))** was question 1(a) from the Autumn paper of 2002.

Prologue I: Functions and Fractions

These notes are designed to accompany the course MS2001 Differential Calculus, which provides a more detailed introduction than MS1001 to an area of mathematics that was developed to deal with problems in the physical sciences and now has many applications in this field and beyond.

In this course we deal with functions of one **real variable**, that is functions such as $x^7 - 4x^4 + 8x^3 - 12x^2$, $2\sin 3t + t\log t$ or $e^{\tan x}$, where the argument is a *real number* x, or t, or... Complex numbers such as $2+3i$ (where $i = \sqrt{-1}$) do not feature in this course.

Such functions will in general be denoted by lower case letters such as f, g or h, or $f(x)$, $g(t)$ or $h(x)$ if we wish to stress the argument. Note that it is not always possible to evaluate a function for all possible values of its input variable. For example $\log x$ is only defined whenever $x > 0$, and $\tan x$ is only defined for $x \neq \pm\frac{\pi}{2}, \pm\frac{3\pi}{2}, \ldots$ That is, they are only defined on a **subset** of the set of all real numbers (which is denoted by the letter $\mathbb{R}$). If we write $f : I \to \mathbb{R}$ then we mean that f is a function defined on the subset of I of $\mathbb{R}$, and taking values in the set $\mathbb{R}$ of all real numbers. That is, $f(x)$ is a real number that can be evaluated/makes sense whenever we choose any value x from the set I.

In a certain sense the main mathematical challenge posed (and solved) by this course arises from the fact that division by 0 is *not allowed*. In particular expressions such as $\frac{3}{0}$ or $\frac{-7}{0}$ are meaningless, and should not appear in your work. If we divide one function by another then we must take note of when the denominator is zero. For example consider the function

$$f(x) = \frac{x}{\sin x}.$$

This is not defined whenever $\sin x = 0$, that is when $x = 0, \pm\pi, \pm 2\pi, \ldots$ However consider what happens as x gets closer and closer to 0:

$x =$	0.5	0.2	0.1	0.01	$\cdots$
$f(x) \approx$	1.04291482	1.00669791	1.00166861	1.00001667	$\cdots$

The value of $f(x)$ gets closer and closer to 1. However it is still *incorrect* to write $f(0) = \frac{0}{\sin 0} = \frac{0}{0} = 1$, since $\frac{0}{0}$ is *undefined*. But it is precisely the *limiting behaviour* of fractions of this sort where both top and bottom approach 0 that we would like to understand and deal with rigorously.

On a related note, given two functions g and h the equation

$$\frac{g(x)}{h(x)} = 0$$

has a solution when $g(x) = 0$. The value of $h(x)$ (unless it is also 0) is immaterial. If the above equation holds there is certainly no reason to conclude that $g(x) = h(x)$ — if this latter equation is true (and both are nonzero) then $\frac{g(x)}{h(x)} = 1$.

Prologue II: Notation for Logic

A **proposition** in logic is a statement that is either *true* or *false*. For example

p_1: $x + y = y + x$ for all real numbers x and y. (TRUE)

p_2: There are 10 month in the year. (FALSE)

Often the proposition may depend on a variable, for example

p_3: It is August.

p_4: $x^2 - 4x + 3 = 0$.

p_3 is only true for 31 days each year, but never during the first semester in UCC. For p_4 note that $x^2 - 4x + 3 = (x - 1)(x - 3)$, and so p_4 is true if $x = 1$ or $x = 3$, otherwise it is false.

Given propositions p and q, we write $p \Rightarrow q$ (read "p implies q") if q is true whenever p is true. For example, if we set

p_5: Next month is September,

p_6: $x = 1$,

then $p_3 \Rightarrow p_5$, since *if* it is now August then next month it *will be* September — this logical deduction is valid no matter what month it currently is. Similarly $p_6 \Rightarrow p_4$, since if $x = 1$ then $1^2 - 4 \times 1 + 3 = 0$. Note also that if p_4 is false, i.e. x is different from 1 or 3, then p_6 must be false.

For our examples we also have $p_5 \Rightarrow p_3$, which can be written $p_3 \Leftarrow p_5$. So if p_3 is true then p_5 is true; *conversely* if p_5 is true then p_3 is also true. Such pairs of propositions are called **equivalent**, which is written $p_3 \Leftrightarrow p_5$ (read "p_3 (is true) if and only if p_5 (is true)").

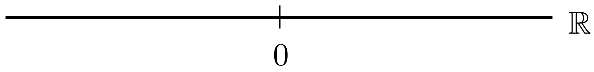
However, if p_4 is true then either $x = 1$ or $x = 3$. If $x = 3$ then p_6 is not true, i.e. $p_4 \not\Rightarrow p_6$ ("p_4 does not imply p_6"). Thus these propositions are not equivalent. All we can say is that p_6 is a sufficient condition for p_4, or p_4 is a necessary condition for p_6.

Mathematical proofs start with clearly stated assumptions and then proceed via a sequence of propositions, proving that each implies the next, until we end up with the desired conclusion. That is, our statements are usually linked by $\Rightarrow$.

1 Inequalities

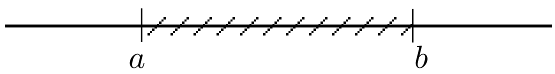
Definitions and axioms

Throughout this course the set of real numbers (that is, the points on the x-axis) will be denoted by the symbol $\mathbb{R}$. The geometrical model is a line extending infinitely far in both directions, with an arbitrary point chosen to represent 0.



Suppose we pick two points a and b on the line (written $a, b \in \mathbb{R}$). If a is to the left of b then a is less than b, or, equivalently, b is greater than a, and this is written $a < b$ or $b > a$. Similarly $a \leq b$ means that either $a < b$ *or* $a = b$. We write $a < b < c$ (or $a < b \leq c$, or ...) if $a < b$ *and* $b < c$ — we must have the inequality symbols pointing in the same direction for this to make sense. The statement $a > b < c$ is not particularly useful or meaningful as it does not indicate the relationship between a and c.

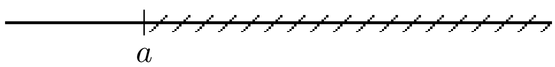
A **finite subinterval of** $\mathbb{R}$ is the collection of all points between two numbers $a < b$:



We can leave a in or out of this collection, and similarly for b, giving overall four possibilities:

$$(a, b) = \{x \in \mathbb{R} : a < x < b\} \qquad \text{neither included}$$
$$(a, b] = \{x \in \mathbb{R} : a < x \leq b\} \qquad a \text{ excluded, } b \text{ included}$$
$$[a, b) = \{x \in \mathbb{R} : a \leq x < b\} \qquad a \text{ included, } b \text{ excluded}$$
$$[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\} \qquad \text{both included.}$$

We can also consider **infinite subintervals**: given a point $a \in \mathbb{R}$ we can consider all points to the right of it, or to the left.



Again, we can leave a in or out of this collection, and so end up with the four possibilities:

$$(a, \infty) = \{x \in \mathbb{R} : x > a\}$$
$$[a, \infty) = \{x \in \mathbb{R} : x \geq a\}$$
$$(-\infty, a) = \{x \in \mathbb{R} : x < a\}$$
$$(-\infty, a] = \{x \in \mathbb{R} : x \leq a\}$$

By taking **unions** of such sets we can combine intervals to form more complicated sets. For example $(-\infty, 1) \cup [2, 3]$ consists of all those numbers that are strictly less than 1 or lie between 2 and 3 inclusive.

We will *assume* the following **axioms**, or properties, of inequalities, where $a, b, c \in \mathbb{R}$ are any three real numbers:

$\mathbf{I_1}$: Exactly one of $a < b$, $a > b$ or $a = b$ is true.

$\mathbf{I_2}$: If $a < b$ and $b < c$ then $a < c$.

$\mathbf{I_3}$: If $a < b$ then $a + c < b + c$.

$\mathbf{I_4}$: If $a < b$ and $c > 0$ then $ac < bc$; if $a < b$ and $c < 0$ then $ac > bc$.

Note. $\mathbf{I_2}$–$\mathbf{I_4}$ hold with $<$ replaced by $\leq$.

Having stated our assumptions we shall now show that various well-known results about inequalities are actually *consequences* of these assumptions.

Proposition 1.1. *Suppose that* $a, b, c, d \in \mathbb{R}$.

(i) *If* $a < b$ *and* $c < d$ *then* $a + c < b + d$. (*Addition of like inequalities*)

(ii) *If* $c > 0$ *and* $d > 0$, *or if* $c < 0$ *and* $d < 0$, *then* $cd > 0$; *if* $c < 0$ *and* $d > 0$, *or if* $c > 0$ *and* $d < 0$, *then* $cd < 0$.

(iii) $1 > 0$.

(iv) $a > 0$ *if and only if* $\dfrac{1}{a} > 0$; $a < 0$ *if and only if* $\dfrac{1}{a} < 0$.

(v) *If* $a < b$ *and* $c > 0$ *then* $\dfrac{a}{c} < \dfrac{b}{c}$; *if* $a < b$ *and* $c < 0$ *then* $\dfrac{a}{c} > \dfrac{b}{c}$.

(vi) *If* $0 < a < b$ *and* $0 < c < d$ *then* $ac < bd$ *and* $\dfrac{a}{d} < \dfrac{b}{c}$.

Proof. (i) By $\mathbf{I_3}$ we have

$$a + c < b + c \quad \text{and} \quad b + c = c + b < d + b = b + d,$$

hence, by $\mathbf{I_2}$, $a + c < b + d$.

(ii) Suppose $c > 0$ and $d > 0$. Then we can apply $\mathbf{I_4}$ with $a = 0$ and $b = d$ to get

$$ac < bc \ \Rightarrow 0 \times c = 0 < cd,$$

as required.

If $c < 0$ and $d < 0$ then we can apply $\mathbf{I_4}$ with $a = d$ and $b = 0$, this time to get $cd > 0$. The other two statements can be proved similarly.

(iii) Now $1 \neq 0$, so by $\mathbf{I_1}$ either $1 < 0$ or $1 > 0$. But, in either case, we can apply (ii) with $c = d = 1$ to get $cd = 1^2 = 1 > 0$.

(iv) Suppose $a > 0$. Again, since $\frac{1}{a} \neq 0$, we must have $\frac{1}{a} < 0$ or $\frac{1}{a} > 0$. If $\frac{1}{a} < 0$, then by (ii) (with $c = a$ and $d = \frac{1}{a}$ this time) we would have $0 > cd = a \times \frac{1}{a} = 1$, contradicting (iii). Hence we must have $\frac{1}{a} > 0$.

We have shown that if $a > 0$ then $\frac{1}{a} > 0$ — and this is true for any positive number. If, on the other hand we are given a and told $\frac{1}{a} > 0$, then we now know that $\frac{1}{1/a} = a > 0$. These two implications together give

$$a > 0 \;\Leftrightarrow\; \tfrac{1}{a} > 0.$$

That $a < 0$ is equivalent to $\frac{1}{a} < 0$ can be proved in the same way.

(v) This is immediate from $\mathbf{I_4}$ and (iv), since if $c > 0$ then $\frac{1}{c} > 0$, and if $c < 0$ then $\frac{1}{c} < 0$.

(vi) We have $a < b$ and $c > 0$, and so $ac < bc$ by $\mathbf{I_4}$. Also, $c < d$ and $b > 0$, and so $bc < bd$ by $\mathbf{I_4}$ as well. Thus, by $\mathbf{I_2}$,

$$ac < bd$$

as required. Finally, $cd > 0$ by (ii), so by (v)

$$\frac{ac}{cd} < \frac{bd}{cd} \;\Rightarrow\; \frac{a}{d} < \frac{b}{c}. \quad \square$$

Of particular importance to us are (ii), (iv) and (v). Indeed, it follows that if we take any n nonzero numbers $a_1, a_2, \ldots, a_n$, and choose any number m satisfying $1 \leq m < n$, and then calculate

$$\frac{a_1 \times a_2 \times \cdots \times a_m}{a_{m+1} \times \cdots \times a_n}$$

then the result will be *positive* if an *even* number of the original n numbers a_i are negative, otherwise it will be *negative*.

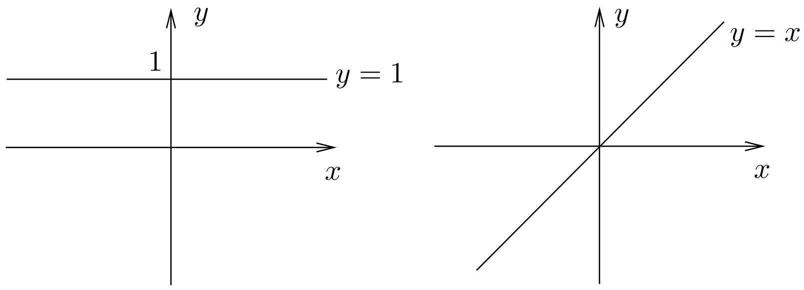
Sketching functions; the quadratic

Definition 1.2. A function $f : \mathbb{R} \to \mathbb{R}$ is **even** if $f(-x) = f(x)$ for all $x \in \mathbb{R}$. It is **odd** if $f(-x) = -f(x)$ for all $x \in \mathbb{R}$.

Thus a function is even if it is invariant under reflections in the y-axis, and odd if it is invariant under rotations through π about the origin.

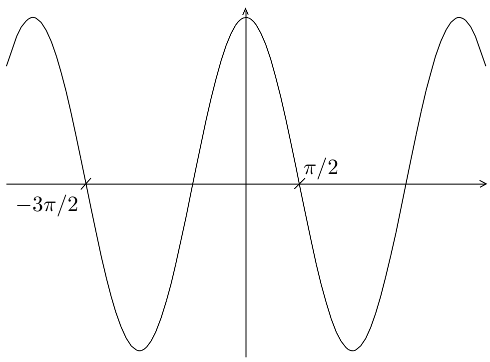
Definition 1.3. Let $I \subset \mathbb{R}$ be an interval. A function $g : I \to \mathbb{R}$ is **increasing** on I if $g(x) \leq g(y)$ whenever $x, y \in I$ such that $x \leq y$. It is **strictly increasing** if $g(x) < g(y)$ whenever $x, y \in I$ such that $x < y$. What it means for g to be (**strictly**) **decreasing** is defined similarly.

Very simple examples of even and odd functions include constant functions (i.e. $y = b$ for any $b \in \mathbb{R}$) and the function $y = x$ respectively:

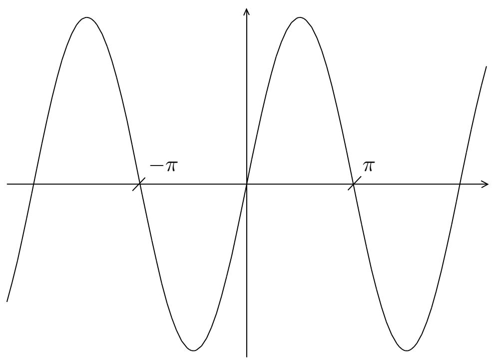


Note that the function $y = x$ is strictly increasing on the whole line $\mathbb{R}$, and the constant function is both increasing and decreasing on $\mathbb{R}$, but neither strictly increasing nor strictly decreasing.

Other important examples include $y = \cos x$, which is even, and $y = \sin x$ which is odd. Note that $y = \sin x$ is (strictly) increasing on the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$, decreasing on $[\frac{\pi}{2}, \frac{3\pi}{2}]$, and so on.



$y = \cos x$



$y = \sin x$

The behaviour of constant functions and $y = x$ generalises to higher powers of n, as can be proved using our consequences of inequalities:

Proposition 1.4. *For each positive integer $n \geq 1$ let $f_n : \mathbb{R} \to \mathbb{R}$ be the function $f_n(x) = x^n$. Then f_n is even if n is even, and odd if n is odd. Moreover each f_n is strictly increasing on the half-line $[0, \infty)$.*

Proof. To see the first part about even or oddness of f_n, we can write

$$f_n(-x) = (-x)^n = (-1 \times x)^n = (-1)^n \times x^n = \begin{cases} x^n & \text{if } n \text{ is even,} \\ -x^n & \text{if } n \text{ is odd.} \end{cases}$$

For the other part, choose any $n \geq 1$. Then note that for any $a, b \in [0, \infty)$ we have

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

In particular for $n = 2$ and $n = 3$ we have

$$a^2 - b^2 = (a - b)(a + b) \quad \text{and} \quad a^3 - b^3 = (a - b)(a^2 + ab + b^2).$$

But we have chosen $a, b \geq 0$, hence any number of the form $a^{n-k}b^{k-1}$ is nonnegative by part (ii) of Proposition 1.1. Thus, by part (i) of Proposition 1.1,

$$a^{n-1} + a^{n-2}b + a^{n-3}b^2 + \cdots + ab^{n-2} + b^{n-1} \geq 0.$$

Now if $a > b \geq 0$ then in particular $a > 0$, hence $a^{n-1} > 0$, and it follows that the above inequality is in fact a strict one. Thus $a^n - b^n$ is the product of two positive numbers (since $a - b > 0$), hence positive by part (ii) of Proposition 1.1. That is,

$$a > b \;\Rightarrow\; a^n - b^n > 0 \;\Rightarrow\; a^n > b^n \quad (\text{by } \mathbf{I_3})$$

and so f_n is strictly increasing. □

In fact the above proof shows a little more, since if we have any $a, b \in [0, \infty)$ that satisfy $a^n > b^n$, then at least one of these two numbers must be nonzero, and hence positive, and so

$$a^{n-1} + a^{n-2}b + a^{n-3}b^2 + \cdots + ab^{n-2} + b^{n-1} > 0$$

again, from which it follows that

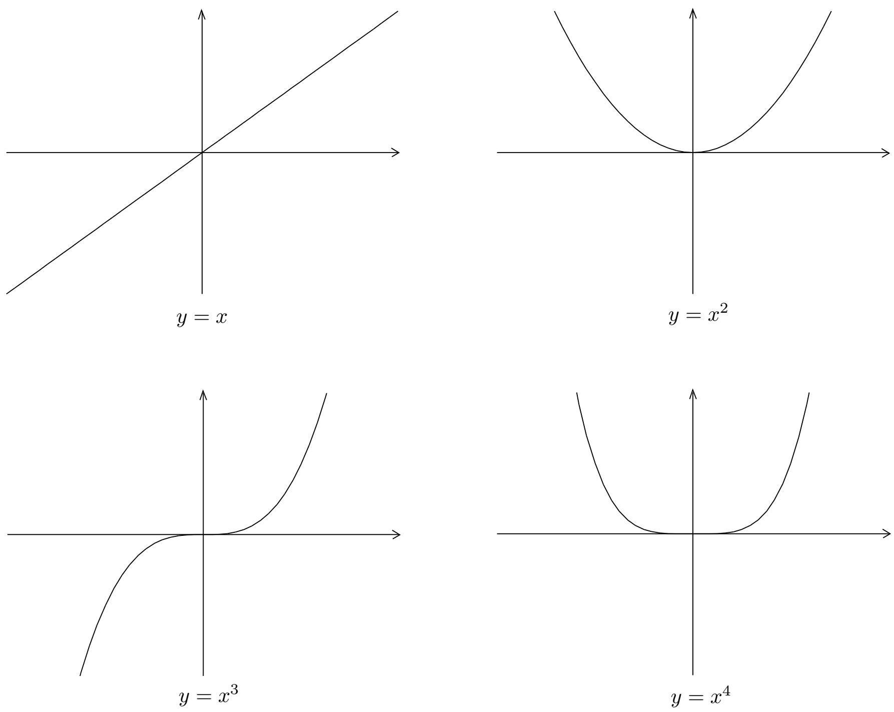
$$a - b = \frac{a^n - b^n}{a^{n-1} + a^{n-2}b + a^{n-3}b^2 + \cdots + ab^{n-2} + b^{n-1}} > 0$$

by part (v) of Proposition 1.1. Hence we have in fact shown the following:

Corollary 1.5. *For any integer $n \geq 1$ and any $a, b \in [0, \infty)$ we have*

$$a < b \;\Leftrightarrow\; a^n < b^n.$$

From the information above we can deduce some of the features of the following graphs of the powers of x:



However, when drawing the above we have not actually demonstrated that our functions are *continuous*, that is, that the graph has no gaps or jumps in it. That is the subject of the following chapter on limits.

The powers of x have been analysed and sketched above by using our axioms for inequalities between real numbers. Other functions will in general be more complicated to sketch, but often we can recognise a function as a translation or scaled version of a known function such as these powers.

Suppose we know the graph for some function $y = f(x)$, and are given a number $a \in \mathbb{R}$.

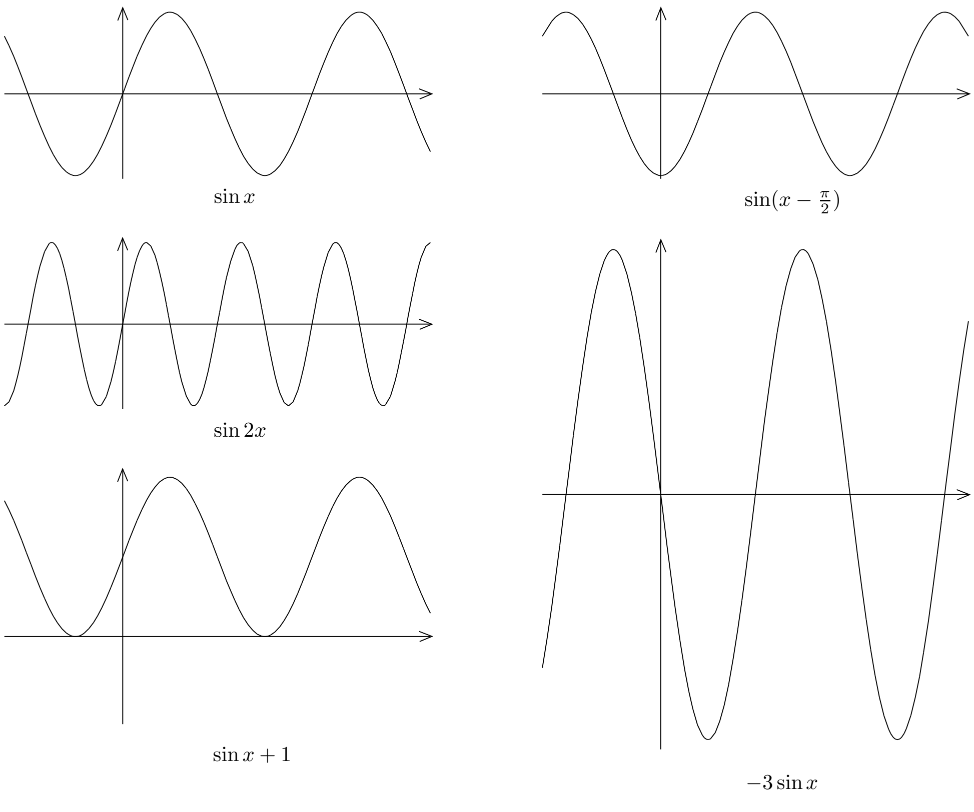
(i) The graph for $f(x-a)$ is got by translating the original a units horizontally to the right.

(ii) The graph for $f(ax)$ is got by stretching the original horizontally by a scale factor of $\frac{1}{a}$ — if $a < 0$ this involves a reflection in the y-axis and scaling by $\frac{1}{-a} > 0$.

(iii) The graph for $f(x) + a$ is obtained by translating the original vertically by a units.

(iv) The graph for $af(x)$ is obtained by stretching the original vertically by a scale factor of a — again if $a < 0$ this involves a reflection in the x-axis and then scaling by $-a > 0$.

Example 1.6. Consider the function $y = \sin x$. The functions $\sin(x - \frac{\pi}{2})$, $\sin 2x$, $\sin x + 1$ and $-3\sin x$ are obtained by applying the relevant rules from above to give:



For example $\sin(0-\frac{\pi}{2}) = \sin(-\frac{\pi}{2}) = -1$, the value of $\sin x$ taken $\frac{\pi}{2}$ units to the left, $\sin(\frac{\pi}{2} - \frac{\pi}{2}) = \sin(0) = 0$, etc. Similarly $\sin(2 \times \frac{\pi}{4}) = \sin\frac{\pi}{2} = 1$, $\sin(2 \times \frac{\pi}{2}) = \sin\pi = 0$, etc., so that the oscillations occur twice as fast. The function $-3\sin x$ can be obtained by two transformations; firstly considering $3\sin x$ which stretches the graph by a factor of 3 in the y-direction, then multiplying by -1, which is the same as reflecting in the x-axis.

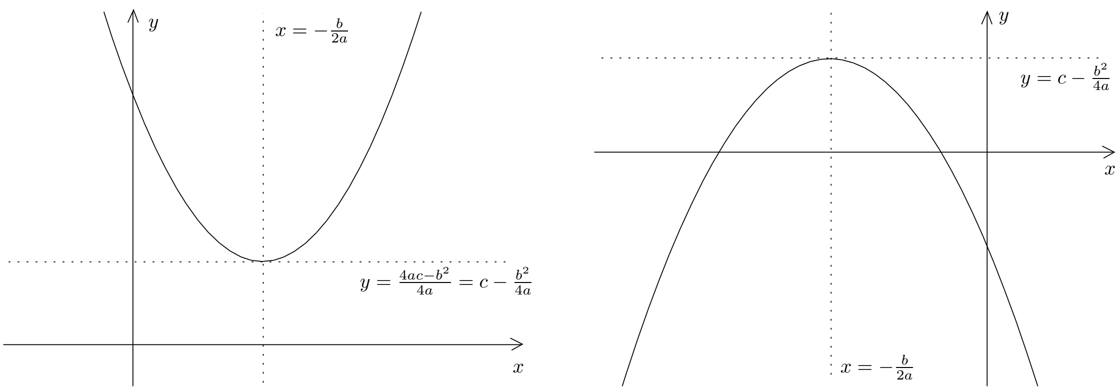
A combination of all of these rules allows us to fully understand any quadratic function, whose general form is

$$y = ax^2 + bx + c,$$

where a, b, c are constants with $a \neq 0$ to ensure that there is an x^2 term. This can be rewritten as follows:

$$\begin{aligned} y &= a\left[x^2 + \frac{b}{a}x + \frac{c}{a}\right] \\ &= a\left[\left(x^2 + 2 \times \frac{b}{2a} \times x + \left(\frac{b}{2a}\right)^2\right) - \frac{b^2}{4a^2} + \frac{c}{a}\right] \\ &= a\left[\left(x + \frac{b}{2a}\right)^2 - \frac{b^2 - 4ac}{4a^2}\right] = a\left(x + \frac{b}{2a}\right)^2 + c - \frac{b^2}{4a}. \end{aligned}$$

That is, the general quadratic can be obtained from the graph for $y = x^2$ by first translating by $-\frac{b}{2a}$ to the right, then translating $\frac{b^2-4ac}{4a}$ downwards, and finally by scaling by a factor of a in the y-direction. Alternatively the second and third steps could be thought of as scaling vertically by a, then translating up by $c - \frac{b^2}{4a}$ respectively. In either case the translations do not change the shape of the graph; the scaling by a will not change the shape if $a > 0$, but flips it over if $a < 0$. Hence every quadratic has one of the two following forms:



$a > 0$ $(b < 0)$ $\qquad$ $a < 0$ $(b < 0)$

Moreover this factorisation leads to the well-known formula for the roots of the quadratic. There is a solution to $ax^2 + bx + c = 0$ if and only if

$$\begin{aligned} & \left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2} \\ \Leftrightarrow \quad & x + \frac{b}{2a} = \pm\sqrt{\frac{b^2 - 4ac}{4a^2}} = \pm\frac{\sqrt{b^2 - 4ac}}{2a} \\ \Leftrightarrow \quad & x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \end{aligned}$$

In particular we need $b^2 - 4ac \geq 0$ if the square root is going to yield a *real* number. If $b^2 - 4ac > 0$ then $\sqrt{b^2 - 4ac} > 0$ and we have two distinct real roots — so the graph cuts the x-axis at two points. If $b^2 - 4ac = 0$ then there is a repeated real root — the graph just touches the x-axis when $x = -\frac{b}{2a}$. If $b^2 - 4ac < 0$ then there are no real solutions, only *complex* roots, and hence no intersection with the x-axis.

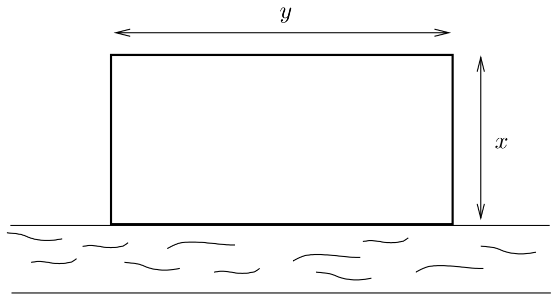
Finally, recall that $y = x^2$ is an even function, so symmetrical about the y-axis. For $y = ax^2 + bx + c$, since we translated by $-\frac{b}{2a}$ to the right, this symmetry is now about the line $x = -\frac{b}{2a}$. In particular the minimum (if $a > 0$) or maximum (of $a < 0$) values occurs for $x = -\frac{b}{2a}$, and if there are roots then they are equally spaced about this point.

Exercise 1.7. Sketch the graphs for

(i) $(x-2)^2$ (ii) $x^2 - 2x - 3$ (iii) $-2x^2 + x - 5$

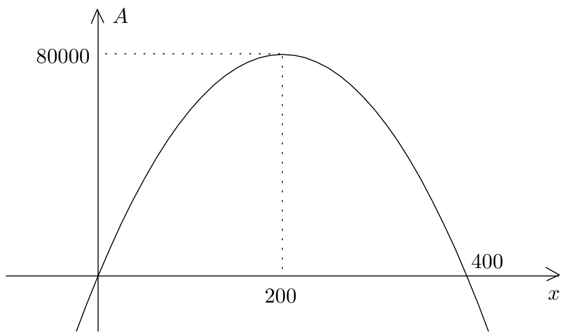
Example 1.8. A farmer has 800m of fencing and wants to fence off a rectangular field that borders a straight river. He needs no fence along the river. What are the dimensions of the field that has the largest possible area?

Solution. Let x denote the length of the side perpendicular to the river and y the length of the side parallel to the river. Thus our field looks like:



If we write down an equation that links the lengths x and y as marked to the given perimeter of 800m, solve this to get x in terms of y, and then use this expression to write the area, A, as a function x, we obtain a quadratic function. The maximum of A is thus easily found.

Fill in the details of this argument.



Inequalities involving rational functions

Definition 1.9. A **polynomial** is any function that can be written as a finite sum of multiples of positive powers of x together with constant functions. That is $f : \mathbb{R} \to \mathbb{R}$ is a polynomial if it is of the form

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

If the constant $a_n \neq 0$ in the above then we say that f is of **degree** n. So a quadratic is a polynomial of degree 2.

A **rational function** is any function g that can be written as $g(x) = p(x)/q(x)$ in terms of two polynomials p and q.

So, for example, each of the functions $f_n(x) = x^n$ that we considered above are polynomials, where f_n is of degree n. Similarly the following are all polynomials

$$x^3 - 2x + 7; \qquad -2x^7 + 3x^6 - 5x^4 - x^3 + 2x^2 + 17x - 1; \qquad x^{31} - x^{17}$$

of degree 3, 7 and 31 respectively, whereas

$$\frac{x^3 - x^2 + 3x + 8}{x^5 + 7x^4 - x^2 + 9} \quad \text{and} \quad \frac{x^{12} - 9x^9 + 7x^5 + 2x}{10x^8 + 9x^7 + 8x^6 + 7}$$

are both rational, but not polynomials. Note that the value of a polynomial function is defined for all values of x since all we are doing is multiplying and adding numbers together. The situation with rational functions is in general more complicated since division has now come into play, and the denominator may be zero. For example we have $x^3 + 2x^2 - x - 2 = (x+2)(x+1)(x-1)$

$$\Rightarrow \quad \frac{x^4 - x^2}{x^3 + 2x^2 - x - 2} = \frac{x^4 - x^2}{(x+2)(x+1)(x-1)}$$

and so the function is *not defined* when $x = -2$, -1 or 1. Indeed, *any* polynomial of degree greater than 2 can be factorised into polynomials of lower degree according to:

Theorem 1.10 (Fundamental Theorem of Algebra). *Any polynomial (over $\mathbb{R}$) can be written as a product of polynomials of degree at most* 2.

For example, our polynomial of degree 3 appearing in the denominator could be written as a product of polynomials of degree 1, each of which equals 0 for exactly one value of x. On the other hand we have

$$x^3 - 3x^2 + x - 3 = (x-3)(x^2+1),$$

and the factor $x^2 + 1$ cannot be broken up any further if we are to only use real coefficients. Indeed, since $x^2 \geq 0$ for all $x \in \mathbb{R}$ we have $x^2 + 1 \geq 1$, and so there are no (real) solutions to the equation $x^2 + 1 = 0$. Hence we cannot write it as $(x-a)(x-b)$ for some $a, b \in \mathbb{R}$. On the other hand, if we allow ourselves to use complex numbers then we have $x^2 + 1 = (x-i)(x+i)$, where $i^2 = -1$, and so

$$x^3 - 3x^2 + x - 3 = (x-3)(x-i)(x+i).$$

Another version of the Fundamental Theorem of Algebra, in which coefficients and roots are allowed to be chosen from the complex numbers $\mathbb{C}$ says that any polynomial in this setting can be written as a product of polynomials of degree 1.

We now discuss inequalities involving rational functions. So these functions involve a single real variable x, and given a particular value of x the inequality may or may not be satisfied. The **solution set of an inequality** is the set of all of those numbers x that satisfy the inequality.

Example 1.11. Find the solution set for the inequality $3x + 5 > 8x - 10$.

Solution. We have the chain of equivalences

$$\begin{array}{cll} & 3x+5 > 8x-10 & \\ \Leftrightarrow & 3x - 8x > -10 - 5 & \text{(by } \mathbf{I_3}\text{, with } c = \pm(-8x-5)) \\ \Leftrightarrow & -5x > -15 & \text{(evaluating)} \\ \Leftrightarrow & \frac{-5x}{-5} < \frac{-15}{-5} & \text{(mult. by } \tfrac{1}{-5}/-5\text{, using } \mathbf{I_4}) \\ \Leftrightarrow & x < 3 & \text{(evaluating)} \end{array}$$

Thus the equality we started with is satisfied if and only if x is less than 3, and so the solution set is $(-\infty, 3)$.

Example 1.12. Find the solution set of the inequality

$$x^2 \le 4x - 3 \qquad (\dagger)$$

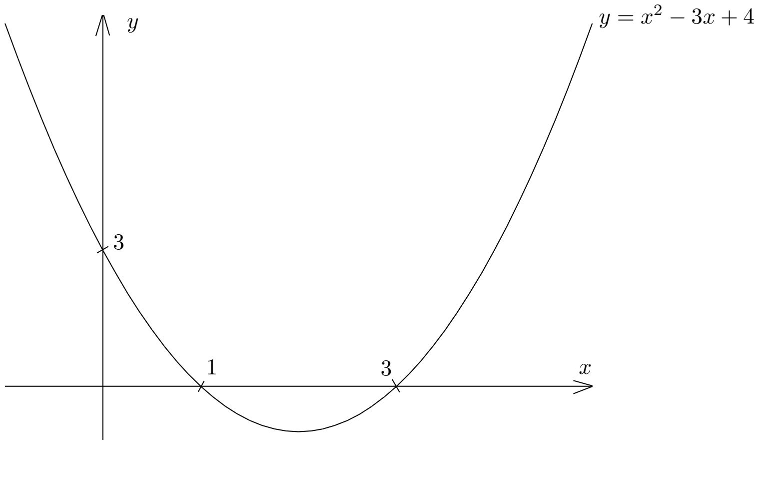
Solution. Here we can first use $\mathbf{I_3}$ to get

$$x^2 \le 4x - 3 \;\Leftrightarrow\; x^2 - 4x + 3 \le 0 \;\Leftrightarrow\; (x-1)(x-3) \le 0.$$

The factors are both of degree 1, hence both are equal to 0 for exactly one value of x, and change sign at this point. The points where the product $(x-1)(x-3)$ can change sign are thus $x = 1$ and $x = 3$, and this quadratic will stay the same sign and be nonzero in between these two values. The behaviour of $(x-1)(x-3)$ is thus summarised by

	$x < 1$	$x = 1$	$1 < x < 3$	$x = 3$	$x > 3$
$x - 1$	$-$	0	$+$	$+$	$+$
$x - 3$	$-$	$-$	$-$	0	$+$
$(x-1)(x-3)$	$+$	0	$-$	0	$+$

Thus we see that $(x-1)(x-3) \le 0$ if and only if $x \in [1, 3]$, and since this inequality is equivalent to $(\dagger)$ it follows that this is also the solution set for $(\dagger)$.



Exercise 1.13 (S02 1(a)). Find the solution set of the inequality

$$\frac{x+3}{x-1} < \frac{x}{x+2} \qquad (x \neq -2, 1) \tag{$*$}$$

and mark this set on a diagram.

Note. When doing this exercise using the method outlined in these notes you should end up with an inequality of the form $f(x) > 0$, and a number of points where the function $f(x)$ can change sign. At such points $f(x)$ will either equal 0 or be undefined, so certainly such points are not part of the solution set, and so separate columns for them in your table are not required.

Another is that you should resist the temptation to multiply either side of $(*)$ by $(x-1)(x+2)$ to get rid of the fractions. The reason for this is that the *sign* of the number $(x-1)(x+2)$ *depends* on the value of x, so we do not necessarily know if the number we are multiplying is positive or negative, and hence if we need to reverse the inequality symbol. One way to get round this problem is to treat the cases $(x-1)(x+2) < 0$ (when $-2 < x < 1$) and $(x-1)(x+2) > 0$ (when $x < -2$ or $x > 1$) separately. Another would be to multiply by the nonnegative number $(x-1)^2(x+2)^2$. In both approaches one then needs to remember to exclude the points $x = -2$ and $x = 1$ from any solution set that may be discovered, since our original inequality does not make sense for these values of x.

Exercise 1.14 (S03 1(a)). Find the solution set of the inequality

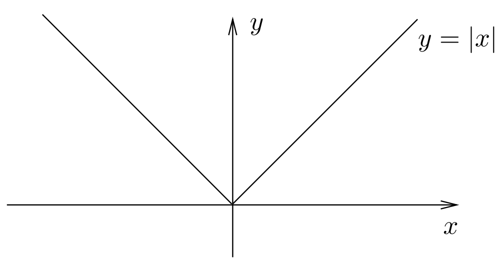
$$\frac{x}{x+2} \leq \frac{3}{x-2}$$

The modulus function; obtaining bounds

Definition 1.15. The **modulus function** is the map $|\cdot| : \mathbb{R} \to \mathbb{R}$ defined by

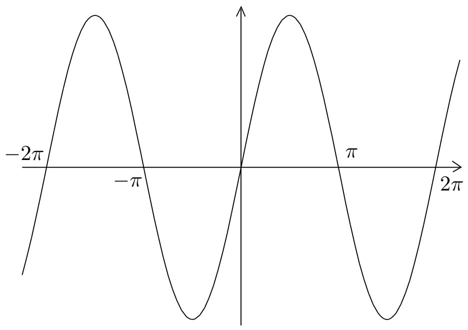
$$|x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0. \end{cases}$$

The number $|x|$ is known as the **modulus** or **absolute value of** x. The graph of the function is

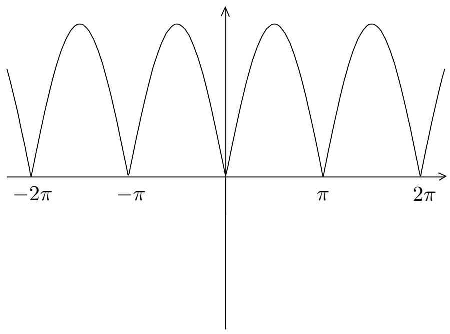


so in particular it is left unchanged by reflection in the y-axis. That is, this function is even, since $|-x| = |x|$ for all $x \in \mathbb{R}$

Basically the number $|x|$ is what we get by ignoring the sign of x if it is negative. As a result, if we have the graph of some function $y = f(x)$, then to draw the graph of the function $y = |f(x)|$ we take the part lying below the x-axis and reflect it in that line. The part lying above remains the same, since in this case $f(x) \geq 0$, hence $|f(x)| = f(x)$. For example, the graphs of $\sin x$ and $|\sin x|$ are:

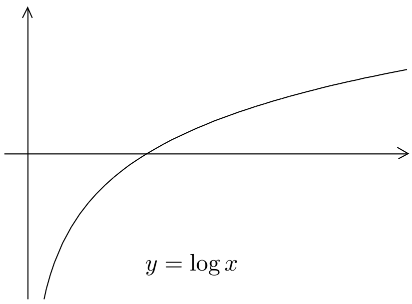


$y = \sin x$

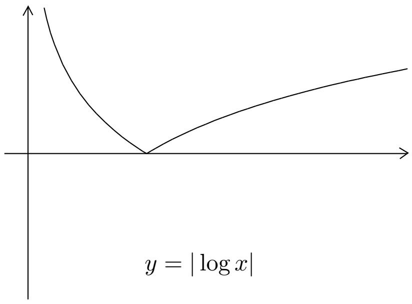


$y = |\sin x|$

Similarly the graphs of $\log x$ and $|\log x|$ are:



$y = \log x$



$y = |\log x|$

The following are properties of $|x|$ that we can prove easily from its definition:

Proposition 1.16. *Let $x, y \in \mathbb{R}$. Then*

(i) $|x|^2 = x^2$; $x \leq |x|$; $\left|\dfrac{1}{x}\right| = \dfrac{1}{|x|}$ *if* $x \neq 0$.

(ii) $|xy| = |x||y|$.

(iii) $\left|\dfrac{x}{y}\right| = \dfrac{|x|}{|y|}$ *if* $y \neq 0$.

(iv) $|x + y| \leq |x| + |y|$.

(v) $\big||x| - |y|\big| \leq |x - y|$.

Proof. (i) Now $|x| = \pm x$, so for the first part we have $|x|^2 = (\pm x)^2 = x^2$. For the other two parts we split the proof into three cases. If $x = 0$ then $|x| = x$. If $x > 0$ then $|x| = x$, and $\frac{1}{x} > 0$, hence $|\frac{1}{x}| = \frac{1}{x} = \frac{1}{|x|}$. If $x < 0$ then $|x| > 0 > x$, and $\frac{1}{x} < 0$, hence $|\frac{1}{x}| = -\frac{1}{x} = \frac{1}{-x} = \frac{1}{|x|}$.

(ii) For this note that

$$|xy|^2 = (xy)^2 = xy \times xy = x^2y^2 = |x|^2|y|^2 = (|x||y|)^2$$

where we have used the first part of (i) a number of times. But $|xy| \geq 0$ and $|x||y| \geq 0$, so by Corollary 1.5 we can take (positive) square roots to get $|xy| = |x||y|$ as required.

(iii) By (i) and (ii) we have $|\frac{x}{y}| = |x \times \frac{1}{y}| = |x||\frac{1}{y}| = |x| \times \frac{1}{|y|} = \frac{|x|}{|y|}$.

(iv) We have

$$\begin{aligned} |x+y|^2 = (x+y)^2 = x^2 + 2xy + y^2 &\leq x^2 + 2|xy| + y^2 && \text{(by (i))} \\ &= |x|^2 + 2|x||y| + |y|^2 && \text{(by (i) and (ii))} \\ &= (|x| + |y|)^2 \end{aligned}$$

However, both $|x+y| \geq 0$ and $|x|+|y| \geq 0$, and so we can conclude from Corollary 1.5 (by taking positive square roots) that $|x+y| \leq |x| + |y|$ as required.

(v) For this one note that by (iv) we get

$$|x| = |(x-y)+y| \leq |x-y| + |y| \;\Rightarrow\; |x| - |y| \leq |x-y|.$$

Similarly, by swapping the roles of x and y,

$$|y| - |x| \leq |y-x| = |-(x-y)| = |x-y|$$

But $\big||x| - |y|\big|$ is equal to $|x| - |y|$ or $-(|x| - |y|) = |y| - |x|$, and so we have shown

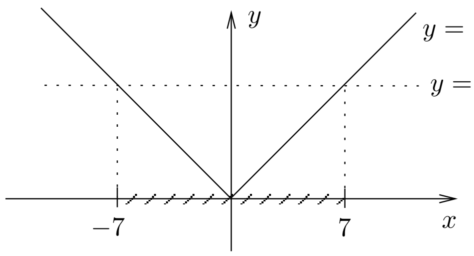
$$\big||x| - |y|\big| \leq |x-y|$$

as required. □

Now consider the inequality $|x| < 7$. To apply the definition of $|x|$ in order to find the solution set for this inequality we split the problem into two parts, depending on whether x is in $[0, \infty)$ or $(-\infty, 0)$. So we have

$$\begin{aligned} &\text{if } x \in [0,\infty) \text{ and } |x| < 7 \text{ then } 0 \leq x = |x| < 7 \\ &\text{and if } x \in (-\infty,0) \text{ and } |x| < 7 \text{ then } x < 0 \text{ and } -x = |x| < 7 \Rightarrow -7 < x < 0. \end{aligned}$$

Thus a number $x \in \mathbb{R}$ satisfies $|x| < 7$ if and only if $-7 < x < 7$, which is illustrated on the graph below:



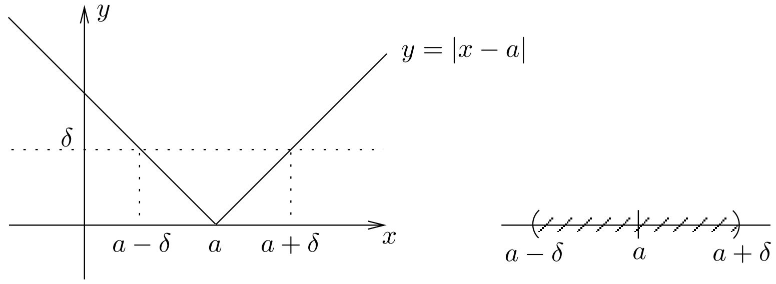
There was clearly nothing special about the choice of the number 7 in the above calculation. Indeed, for any positive number $\delta > 0$ the solution set of the inequality $|x| < \delta$ is $-\delta < x < \delta$. Furthermore, we can use this to solve the following very important inequality:

$$|x - a| < \delta$$

where $a \in \mathbb{R}$ and $\delta > 0$. Indeed, if we set $y = x - a$ then $|x - a| < \delta$ becomes $|y| < \delta$ which has solution set $-\delta < y < \delta$. But replacing y by $x - a$ gives

$$-\delta < x - a < \delta \;\Leftrightarrow\; a - \delta < x < a + \delta.$$

This is shown graphically by



noting that the graph for $y = |x - a|$ is got by shifting that for $y = |x|$ to the right by a units. Thus we see that a number x satisfies the inequality $|x - a| < \delta$ if and only if it is in the interval $(a - \delta, a + \delta)$, with centre point a and of width 2δ. That is, it satisfies the inequality if and only if its *distance* from the point a is less than δ.

One useful trick to solve inequalities involving absolute values is to recall that for any $a, b \in [0, \infty)$ we have $a < b$ if and only if $a^2 < b^2$. Moreover, $|x| \geq 0$ for all $x \in \mathbb{R}$, and $|x|^2 = x^2$. These facts help us to remove the modulus signs, without having to split everything into several cases.

Example 1.17. Find the solution set of the inequality

$$\left|\frac{2x+5}{x+4}\right| \geq 1 \qquad (x \neq -4) \tag{$*$}$$

Note that we have to exclude $x = -4$ in the above, since the fraction on the left hand side is not defined for this value of x. Also, unlike before when we were dealing with rational functions, we *are now able* to multiply by factors such as $|x + 4|$, since this number is nonnegative by definition.

Solution. We have the following equivalences (subject to $x \neq -4$):

$$\begin{aligned}
& \left|\frac{2x+5}{x+4}\right| = \frac{|2x+5|}{|x+4|} \geq 1 && \text{(by (iii) of Proposition 1.16)} \\
\Leftrightarrow\quad & |2x+5| \geq |x+4| && \text{(mult./div by } |x+4|\text{)} \\
\Leftrightarrow\quad & |2x+5|^2 \geq |x+4|^2 && \\
\Leftrightarrow\quad & 4x^2 + 20x + 25 \geq x^2 + 8x + 16 && \\
\Leftrightarrow\quad & 3x^2 + 12x + 9 \geq 0 && \text{(add/subtract } -x^2 - 8x - 16\text{)} \\
\Leftrightarrow\quad & x^2 + 4x + 3 \geq 0 && \text{(mult./div. by 3)} \\
\Leftrightarrow\quad & (x+3)(x+1) \geq 0 &&
\end{aligned}$$

We thus see that the solution set of $(*)$ is $(-\infty, -4) \cup (-4, -3] \cup [-1, \infty)$. [You could produce a table at this stage, but for solving a quadratic this may be excessive.]

Exercise 1.18 (A02 1(a)). Find the solution set of the inequality

$$|x+4| > |3x-8|$$

and mark this set on a diagram.

When applying the definition of limits in the next chapter to particular examples it is useful to find bounds for the growth of functions. That is given a function $f(x)$ and some finite subinterval $I \subseteq \mathbb{R}$, we would often like to find numbers M and/or N such that

$$|f(x)| \leq N \quad \text{or} \quad M \leq |f(x)| \leq N \quad \text{for all } x \in I.$$

It turns out that for our applications we generally only need some crude choice for N, rather than finding the best possible bound, and consequently the triangle inequality (part (iv) of Proposition 1.16) gives an efficient method. Indeed that says that

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

for all $x, y \in \mathbb{R}$. An induction argument applied to this gives

$$|x_1 + x_2 + \cdots + x_n| \leq |x_1| + |x_2| + \cdots + |x_n|$$

for any $n \geq 1$ and any $x_i \in \mathbb{R}$.

Exercise 1.19 (A04 3(b)). Find a positive number $N > 0$ such that

$$\left| x^3 - 3x\cos x + \frac{4}{x} \right| \leq N$$

for all $1 \leq x \leq 3$.

Example 1.20. Find positive numbers M and N such that $M \leq \left|\frac{x+3}{x-2}\right| \leq N$ for all $x \in [4,7]$.

Solution. By (iii) of Proposition 1.16,

$$\left|\frac{x+3}{x-2}\right| = \frac{|x+3|}{|x-2|} = \frac{a}{d}$$

where $a = |x+3|$ and $d = |x-2|$.

Now $4 \leq x \leq 7$ is equivalent to $7 \leq x+3 \leq 10$, so for such x, $|x+3| = x+3 \leq 10$. Similarly $4 \leq x \leq 7$ is equivalent to $2 \leq x-2 \leq 5$, and so $|x-2| = x-2 \geq 2$ in this case. Now if we set $b = 10$ and $c = 2$, then we have $a \leq b = 10$ and $d \geq c = 2$ and so by (vi) of Proposition 1.1

$$\frac{a}{d} \leq \frac{b}{c} = \frac{10}{2} = 5.$$

That is,

$$\left|\frac{x+3}{x-2}\right| \leq 5$$

for all $x \in [4,7]$. Hence $N = 5$ will do.

Exercise 1.21. Show that we can take $M = \frac{7}{5}$.

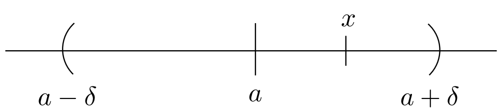
Exercise 1.22. Show that M must be no bigger than 2.

2 Limits and Continuity

The intuitive idea of limits

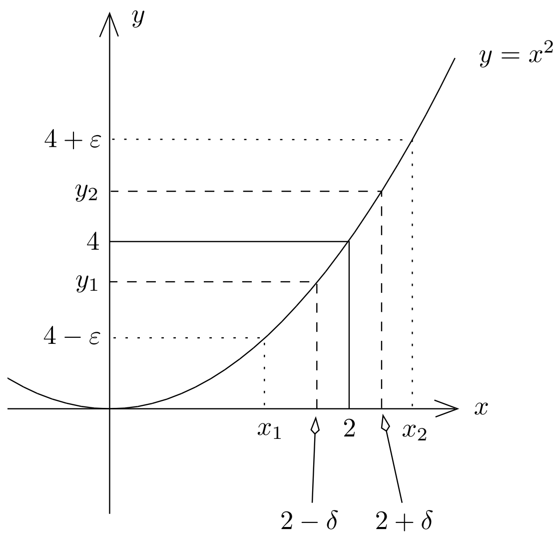
Our aim in this course is to give a rigorous introduction to the basic ideas of calculus, a branch of mathematics that has come to dominate the physical sciences. The main underlying concept in calculus is that of the limit — the behaviour of a function as the point at which you evaluate it gets closer and closer to some prescribed value. Such ideas were finally developed with sufficient rigour in the 19th Century, long after calculus had been invented in the 17th Century by Leibniz and Newton. Back then statements such as $f(x) \to f(u)$ or $\lim_{\delta x \to 0} \frac{\delta y}{\delta x}$ were written down and dealt with intuitively, but not fully explained. Indeed, the lack of clear definition or explanation often lead the practitioners to draw incorrect conclusions! Here the number δx is used to denote a small change in the variable x, and in the limit people talked of infinitesimal numbers, which were sometimes assumed to be zero and sometimes not, according to what was convenient at that stage in a calculation for the particular author...

The goal of this section is give a precise definition of limit on which we can build the appropriate definition of the derivative of a function. We shall consider the limiting behaviour of functions $f : \mathbb{R} \to \mathbb{R}$, and begin by recalling the following important result from the previous chapter: for any $a \in \mathbb{R}$ and $\delta > 0$, the inequality $|x - a| < \delta$ is equivalent to the double inequality $a - \delta < x < a + \delta$, i.e. x lies in the interval $(a - \delta, a + \delta)$:



Thus if $f : \mathbb{R} \to \mathbb{R}$ is a function and $\varepsilon > 0$ a given number, the statement that $|f(x) - l| < \varepsilon$ is equivalent to $l - \varepsilon < f(x) < l + \varepsilon$, that is, $f(x) \in (l - \varepsilon, l + \varepsilon)$. Note that a is the midpoint of the interval $(a - \delta, a + \delta)$ and l is the midpoint of the interval $(l - \varepsilon, l + \varepsilon)$.

Example 2.1. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$.



Consider the point $x = 2$, mapped by the function f to $2^2 = 4$. Choose a small positive number $\varepsilon > 0$, which should be thought of as the *acceptable degree of error*, and consider the interval $(4-\varepsilon, 4+\varepsilon)$ about 4, i.e. the set of all y such that $|y-4| < \varepsilon$.

Now put $x_1 = \sqrt{4-\varepsilon}$ and $x_2 = \sqrt{4+\varepsilon}$. Recall that $f(x) = x^2$ is strictly increasing on $[0, \infty)$ by Proposition 1.4, hence $x_1 < 2 < x_2$. Furthermore

$$x_1 < x < x_2 \quad \Rightarrow \quad x_1^2 = 4 - \varepsilon < x^2 = f(x) < 4 + \varepsilon = x_2^2 \quad \Rightarrow \quad |f(x) - 4| < \varepsilon. \; (\dagger)$$

Since $x_1 < 2 < x_2$, if we define $\delta = \frac{1}{2}\min\{2 - x_1, x_2 - 2\}$ then in particular $\delta > 0$. Also, if $|x-2| < \delta$ then $2 - \delta < x < 2 + \delta$. But

$$\begin{aligned} \delta &\le \tfrac{1}{2}(2 - x_1) \quad \Rightarrow \quad 2 - \delta \ge 2 - \tfrac{1}{2}(2 - x_1) > 2 - (2 - x_1) = x_1, \quad \text{and} \\ \delta &\le \tfrac{1}{2}(x_2 - 2) \quad \Rightarrow \quad 2 + \delta \le 2 + \tfrac{1}{2}(x_2 - 2) < 2 + (x_2 - 2) = x_2. \end{aligned}$$

Thus if $|x-2| < \delta$ then $x_1 < x < x_2$, and so $4 - \varepsilon < x^2 < 4 + \varepsilon$ by $(\dagger)$. That is

$$0 < |x-2| < \delta \;\Rightarrow\; |f(x) - 4| < \varepsilon.$$

To give an example of possible values that ε and δ might take, consider setting $\varepsilon = 0.1$. Then $x_1 = \sqrt{3.9} \approx 1.975$, and $x_2 = \sqrt{4.1} \approx 2.025$, and so $2 - \sqrt{3.9} \approx 0.025$ and $\sqrt{4.1} - 2 \approx 0.025$. Thus if we set $\delta = 0.012 \approx \frac{1}{2} \times 0.025$, then $2 - \delta = 1.988$ and $2 + \delta = 2.012$. So now if we choose any x satisfying $0 < |2 - x| < 0.012$, we will have $|f(x) - 4| < 0.1$.

Similarly if we take $\varepsilon = 0.01$ then $x_1 = \sqrt{3.99} \approx 1.9975$ and $x_2 = \sqrt{4.01} \approx 2.0025$. Thus if we set $\delta = 0.0012 \approx \frac{1}{2} \times 0.0025$ then $2 - \delta = 1.9988$ and $2 + \delta = 2.0012$, and if we choose any x satisfying $0 < |2 - x| < 0.0012$, we will have $|f(x) - 4| < 0.01$.

The important point about our calculations with ε and δ above is that *no matter how small a value of* ε we choose there is always some $\delta > 0$ for which

$$0 < |x-2| < \delta \;\Rightarrow\; |f(x) - 4| < \varepsilon.$$

For instance it will work for $\varepsilon = 0.001, 0.0000001, \ldots$ As the value of ε gets smaller and smaller, we will need to choose smaller and smaller values of δ, forcing x to be closer and closer to 2, but it is always possible to find this δ. And by choosing smaller and smaller ε we are ensuring that $f(x)$ is as close as we like to 4.

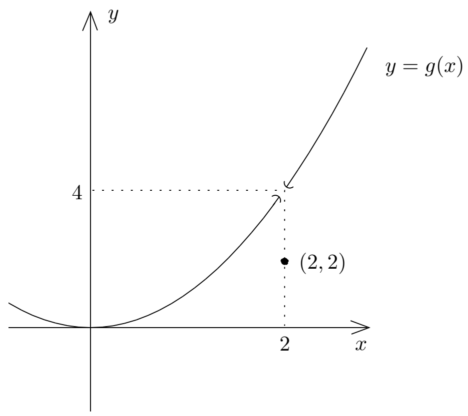
Anticipating the definition below, we write the conclusion of the above as

$$\lim_{x \to 2} f(x) = 2, \qquad \text{or} \qquad f(x) \to 4 \;\text{ as }\; x \to 2.$$

It is true that $4 = 2^2 = f(2)$, but that is not actually relevant to the above discussion. Indeed above we wrote the double inequality $0 < |x - 2| < \delta$, which implies in particular that $x \ne 2$, i.e. x does not actually take the value 2, but merely approaches it. So if we were to define a new function $g : \mathbb{R} \to \mathbb{R}$ by

$$g(x) = \begin{cases} x^2 & \text{if } x \ne 2, \\ 2 & \text{if } x = 2, \end{cases}$$

then the same calculations would apply, and we could conclude that $\lim_{x \to 2} g(x) = 4$, even though $g(2) = 2 \ne 4$.



The definition of a limit; basic techniques

Definition 2.2. Let $f : \mathbb{R} \to \mathbb{R}$ be a function and let $a, l \in \mathbb{R}$ be real numbers. Then l is called the **limit of f as x tends to a** if given *any* $\varepsilon > 0$ there is *some* $\delta > 0$ such that whenever $0 < |x - a| < \delta$ we have $|f(x) - l| < \varepsilon$. If this is the case then we write $\lim_{x \to a} f(x) = l$, or $f(x) \to l$ as $x \to a$.

Note.

(i) We do not require that $f(a) = l$ — indeed we do not actually require that $f(a)$ be *defined*.

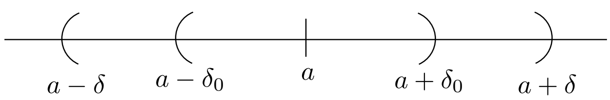
(ii) The number δ depends on ε, and is often written as $\delta(\varepsilon)$ to indicate this fact. As ε decreases in size, δ typically does so as well.

Moreover, suppose for a given $\varepsilon > 0$ we have found some $\delta > 0$ such that

$$0 < |x - a| < \delta \;\Rightarrow\; |f(x) - l| < \varepsilon.$$

Then for all other numbers δ_0 satisfying $0 < \delta_0 < \delta$ we have

$$0 < |x - a| < \delta_0 \;\Rightarrow\; 0 < |x - a| < \delta \;\Rightarrow\; |f(x) - l| < \varepsilon.$$



(iii) If the limit exists it is unique — but it *need not* exist.

In the definition of the limit above we use inequalities of the form $0 < |x - a| < \delta$, which says that distance between x and a is less than δ, but does not specify whether x should be less than or greater than a. We can give definitions of limits in which x is restricted so that it approaches a from the right or the left.

Definition 2.3. Let $f : \mathbb{R} \to \mathbb{R}$ be a function and let $a, l \in \mathbb{R}$ be real numbers.

(i) l is called the **left limit of f as x tends to a** if given *any* $\varepsilon > 0$ there is *some* $\delta > 0$ such that whenever $0 < a - x < \delta$ we have $|f(x) - l| < \varepsilon$. This is written $\lim_{x \to a-} f(x) = l$, or $f(x) \to l$ as $x \to a-$.

(ii) l is called the **right limit of f as x tends to a** if given *any* $\varepsilon > 0$ there is *some* $\delta > 0$ such that whenever $0 < x - a < \delta$ we have $|f(x) - l| < \varepsilon$. This is written $\lim_{x \to a+} f(x) = l$, or $f(x) \to l$ as $x \to a+$.

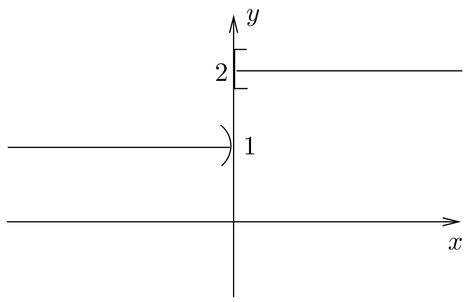
Note that for the left limit to exist we must have $0 < a - x < \delta$, which is equivalent to saying that $a - \delta < x < a$, that is x is less than a, and so must "approach from the left". Similarly, for the right limit to exist we must have $a < x < a + \delta$, and so x "approaches from the right."

Proposition 2.4. *Let f be a function $\mathbb{R} \to \mathbb{R}$ and $a \in \mathbb{R}$. Then $\lim_{x \to a} f(x)$ exists if and only if $\lim_{x \to a-} f(x)$ and $\lim_{x \to a+} f(x)$ both exist and are equal.*

Example 2.5. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} 2 & \text{if } x \geq 0, \\ 1 & \text{if } x < 0. \end{cases}$$

This has graph



and so pictorially we see that $\lim_{x \to 0+} f(x) = 2$ and $\lim_{x \to 0-} f(x) = 1$. Thus the left and right limits exist, but they are unequal. Hence $\lim_{x \to 0} f(x)$ does not exist.

Example 2.6. Let c and d be any two real numbers. Using the definition of limits show that if $f(x) = c + dx$ for all $x \in \mathbb{R}$ then for each $a \in \mathbb{R}$ we have $\lim_{x \to a} f(x) = c + da$.

Solution. For each $x \in \mathbb{R}$ and the specified point $a \in \mathbb{R}$ we have

$$|f(x) - (c + da)| = |c + dx - c - da| = |d(x - a)| = |d||x - a|. \qquad (*)$$

Case 1. $d = 0$: In this case, if we choose any positive number $\varepsilon > 0$ then by $(*)$

$$|f(x) - c| = |0||x - a| = 0 < \varepsilon$$

for *every* $x \in \mathbb{R}$. Thus if we choose *any* positive number $\delta > 0$ then

$$0 < |x - a| < \delta \;\Rightarrow\; |f(x) - c| < \varepsilon,$$

and so $\lim_{x\to a} f(x) = c$ as required.

Case 2. $d \neq 0$: Again choose some $\varepsilon > 0$, and now set $\delta = \frac{\varepsilon}{|d|} > 0$. Then by $(*)$

$$0 < |x-a| < \delta \;\Rightarrow\; |f(x) - (c+da)| < |d|\delta = |d| \times \frac{\varepsilon}{|d|} = \varepsilon,$$

and so $\lim_{x\to a} f(x) = c + da$ as required.

Example 2.7. Using the ε-δ definition of a limit, show that $\lim_{x\to 2}(x^2+5x-7) = 7$.

Solution. For convenience, define f to be the function $f(x) = x^2+5x-7$ and choose an $\varepsilon > 0$. We must find some $\delta > 0$ such that if $0 < |x-2| < \delta$, then $|f(x)-7| < \varepsilon$. Now

$$|f(x)-7| = |(x^2+5x-7)-7| = |x^2+5x-14| = |(x+7)(x-2)| = |x+7||x-2|$$

and so if we choose any $\delta > 0$, then whenever $0 < |x-2| < \delta$ we will have

$$|f(x)-7| \le \delta|x+7|$$

But what is $|x+7|$ less than? Well we are only really interested in values of x close to 2, so we can always make an *initial restriction on* δ, say by requiring that it is less than 1. So then if $|x-2| < \delta$ with such a δ, then we have restricted x to lie in $(2-\delta, 2+\delta)$, and this interval is itself contained in $(2-1, 2+1) = (1,3)$ by the assumption on δ. But if $x \in (1,3)$ then $8 = 1+7 < x+7 < 3+7 = 10$. So restricting δ to be less than 1 tells us that if $|x-2| < \delta$ then we must have $|x+7| = x+7 < 10$.

So now choose $\delta = \min\{\frac{1}{2}, \frac{\varepsilon}{10}\}$. Then $0 < \delta < 1$ (which is why we used $\frac{1}{2}$ in its definition), and whenever $0 < |x-2| < \delta$ we have, since $\delta \le \frac{\varepsilon}{10}$,

$$|f(x)-7| \le \delta|x+7| \le \frac{\varepsilon}{10}|x+7| < \frac{\varepsilon}{10} \times 10 = \varepsilon.$$

Thus we have that $\lim_{x\to 2}(x^2+5x-7) = 7$.

The above result is not overly easy to prove, certainly not as easy as it was to prove the limits in Example 2.6. However that earlier example, together with the following proposition, give a much faster route to proving the above result. The proofs given illustrate the need for, and an application of, the rigorous definition of limits.

Proposition 2.8 (Calculus of Limits). *Suppose that f and g are two functions $\mathbb{R} \to \mathbb{R}$, and that for some $a \in \mathbb{R}$ we have*

$$\lim_{x\to a} f(x) = p, \quad \textit{and} \quad \lim_{x\to a} g(x) = q$$

for some $p, q \in \mathbb{R}$. Then

(i) $\lim_{x\to a} \big(f(x) + g(x)\big) = p + q$.

(ii) $\lim_{x\to a} \big(f(x)g(x)\big) = pq$.

(iii) *If* $q \neq 0$, *then* $\lim_{x\to a} \dfrac{f(x)}{g(x)} = \dfrac{p}{q}$.

(iv) *If* n *is any positive integer and* $p > 0$ *then* $\lim_{x\to a} \sqrt[n]{f(x)} = \sqrt[n]{p}$.

Proof. (i) Fix an $\varepsilon > 0$, then $\frac{\varepsilon}{2} > 0$ as well. Since $\lim_{x\to a} f(x) = p$ and $\lim_{x\to a} g(x) = q$ there are positive numbers $\delta_f > 0$ and $\delta_g > 0$ such that

$$\begin{aligned} 0 < |x-a| < \delta_f \ &\Rightarrow\ |f(x)-p| < \frac{\varepsilon}{2}, \quad \text{and} \\ 0 < |x-a| < \delta_g \ &\Rightarrow\ |g(x)-q| < \frac{\varepsilon}{2}. \end{aligned}$$

Now if we *define* $\delta = \min\{\delta_f, \delta_g\}$ then $\delta > 0$. Moreover if $0 < |x-a| < \delta$ then $0 < |x-a| < \delta_f$ and $0 < |x-a| < \delta_g$, and so by the inequalities above we have

$$\begin{aligned} |(f(x)+g(x)) - (p+q)| &= |(f(x)-p) + (g(x)-q)| \\ &\leq |f(x)-p| + |g(x)-q| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

(iii) Fix an $\varepsilon > 0$, and assume that $p \neq 0$. Set

$$\varepsilon_f = \frac{\varepsilon|q|}{4} > 0, \quad \text{and} \quad \varepsilon_g = \min\left\{\frac{\varepsilon|q|^2}{4|p|}, \frac{|q|}{2}\right\} > 0.$$

By hypothesis there are positive numbers $\delta_f > 0$ and $\delta_g > 0$ such that

$$\begin{aligned} 0 < |x-a| < \delta_f \ &\Rightarrow\ |f(x)-p| < \varepsilon_f, \quad \text{and} \\ 0 < |x-a| < \delta_g \ &\Rightarrow\ |g(x)-q| < \varepsilon_g. \end{aligned} \tag{†}$$

By the reverse triangle inequality (part (v) of Proposition 1.16) we have

$$|g(x)| = \big|q - \big(q - g(x)\big)\big| \geq |q| - |q - g(x)|$$

and so *if* $0 < |x-a| < \delta_g$ then $|g(x) - q| < \varepsilon_g \leq \frac{1}{2}|q|$, and hence

$$|g(x)| \geq |q| - \tfrac{1}{2}|q| = \tfrac{1}{2}|q| > 0. \tag{‡}$$

Define $\delta = \min\{\delta_f, \delta_g\}$, so in particular $\delta > 0$. Now

$$\frac{f(x)}{g(x)} - \frac{p}{q} = \frac{q\big(f(x)-p\big) - p\big(g(x)-q\big)}{qg(x)}.$$

Thus if $0 < |x-a| < \delta$, then the inequalities (†) and (‡) both hold and so

$$\begin{aligned} \big|q\big(f(x)-p\big) - p\big(g(x)-q\big)\big| &\leq |q||f(x)-p| + |p||g(x)-q| \\ &< |q|\varepsilon_f + |p|\varepsilon_g \\ &\leq |q| \times \frac{\varepsilon|q|}{4} + |p| \times \frac{\varepsilon|q|^2}{4|p|} = \frac{\varepsilon|q|^2}{2} \end{aligned}$$

and

$$|qg(x)| = |q||g(x)| \geq \tfrac{1}{2}|q|^2.$$

So, finally, by part (vi) of Proposition 1.1 we have

$$0 < |x - a| < \delta \;\Rightarrow\; \left|\frac{f(x)}{g(x)} - \frac{p}{q}\right| < \frac{\frac{1}{2}\varepsilon|q|^2}{\frac{1}{2}|q|^2} = \varepsilon$$

as required.

If $p = 0$ then setting $\varepsilon_f = \frac{1}{2}\varepsilon|q|$ and $\varepsilon_g = \frac{1}{2}|q|$ at the outset of the proof, and following the same steps, will yield the desired result. □

Example 2.9. Show that $\lim_{x\to a} x^2 = a^2$ for any real number $a \in \mathbb{R}$.

Solution. To show this, let $g : \mathbb{R} \to \mathbb{R}$ be the function $g(x) = x$, then $\lim_{x\to a} g(x) = a$ as shown in Example 2.6 (taking $c = 0$, $d = 1$). Now note that $f(x) = x^2 = g(x)g(x)$, so by part (ii) of Proposition 2.8 we have

$$\lim_{x\to a} f(x) = \lim_{x\to a} \big(g(x)g(x)\big) = \big(\lim_{x\to a} g(x)\big)\big(\lim_{x\to a} g(x)\big) = a \times a = a^2.$$

Exercise 2.10 (A02 1(b)). Use the calculus of limits to evaluate the following:

$$\lim_{x\to 0} \frac{x^3 + x^2 - 5x + 3}{x^2 + x - 2}, \qquad \lim_{x\to 1} \frac{x^3 + x^2 - 5x + 3}{x^2 + x - 2}$$

Example 2.11. Find the limit $\displaystyle\lim_{x\to -1} \frac{\sqrt{2+x} - 1}{x+1}$.

Solution. A direct use of the calculus of limits to top and bottom gives the meaningless answer $\frac{0}{0}$, as would happen in the exercise above. So note that

$$\begin{aligned}
\frac{\sqrt{2+x}-1}{x+1} &= \frac{\sqrt{2+x}-1}{x+1} \times \frac{\sqrt{2+x}+1}{\sqrt{2+x}+1} \qquad (x \neq -1)\\
&= \frac{(2+x)-1}{(x+1)(\sqrt{2+x}+1)} = \frac{x+1}{(x+1)(\sqrt{2+x}+1)}\\
&= \frac{1}{\sqrt{2+x}+1}.
\end{aligned}$$

So now by Proposition 2.8(iii) and (iv) (with $n = 2$) we have

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

Exercise 2.12 (A04 1(b ii)). Use the calculus of limits to evaluate the following:

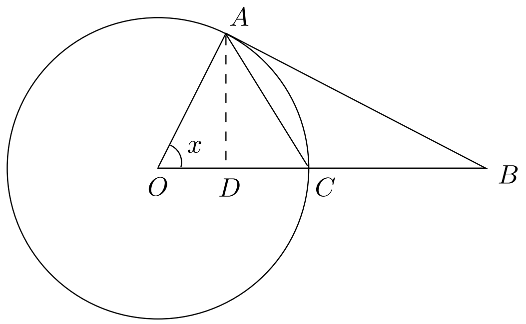
$$\lim_{x\to 2} \frac{\sqrt{2x-3} - \sqrt{x-1}}{x-2}$$

In Examples 2.10, 2.11 and 2.12 above we made use of the following fact:

Proposition 2.13. *Suppose that f and g are functions for which $f(x) = g(x)$ for all $x \neq a$. If $\lim_{x\to a} f(x)$ exists then so does $\lim_{x\to a} g(x)$, and moreover $\lim_{x\to a} f(x) = \lim_{x\to a} g(x)$.*

Example 2.14. $\lim_{x\to 0} \frac{\sin x}{x} = 1$ and $\lim_{x\to 0} \frac{\cos x - 1}{x} = 0$.

These limits involving trigonometric functions are important for the calculation of the derivatives of $\sin x$ and $\cos x$, and can be proved geometrically by considering a circle of radius 1, and the sector cut off by an angle of x radians:



OA and OC are both of length 1, so the perpendicular height OD of triangle $\triangle OAC$ is $\sin x$, and thus its area is $\frac{1}{2} \times 1 \times \sin x = \frac{1}{2}\sin x$. Similarly AB has length $\tan x$, since $O\widehat{A}B$ is a right-angle, and thus $\triangle OAB$ has area $\frac{1}{2}\tan x$. Between these two triangles is the sector OAC, whose area is $\frac{x}{2\pi} \times \pi \times 1^2 = \frac{1}{2}x$, and so we have

$$\begin{aligned} & \text{area of } \triangle OAC < \text{area of sector } OAC < \text{area of } \triangle OAB \\ \Rightarrow \quad & \tfrac{1}{2}\sin x < \tfrac{1}{2}x < \tfrac{1}{2}\tan x = \frac{\sin x}{2\cos x} \\ \Rightarrow \quad & 1 < \frac{x}{\sin x} < \frac{1}{\cos x}. \end{aligned}$$

But from the graph of $\cos x$ we know that $\cos x \to 1$ as $x \to 0$, so $\frac{1}{\cos x} \to 1$ as $x \to 0$ by the calculus of limits. This forces $\frac{x}{\sin x} \to 1$ as $x \to 0$, and so $\frac{\sin x}{x} \to 1$ as $x \to 0$ as stated.

For the other limit we have

$$\begin{aligned} \frac{\cos x - 1}{x} &= \frac{\cos x - 1}{x} \times \frac{\cos x + 1}{\cos x + 1} = \frac{\cos^2 x - 1}{x(\cos x + 1)} \\ &= \frac{-\sin^2 x}{x(\cos x + 1)} = -\frac{\sin x}{x} \times \frac{\sin x}{\cos x + 1} \end{aligned}$$

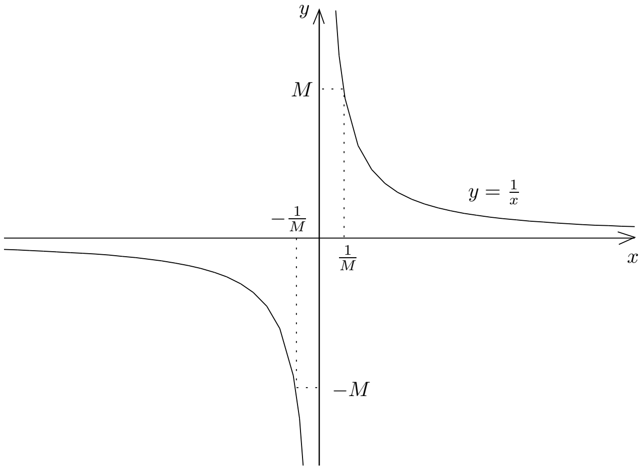
We know already that $\lim_{x\to 0} \frac{\sin x}{x} = 1$. Moreover $\lim_{x\to 0} \sin x = 0$ and $\lim_{x\to 0} (\cos x + 1) = 1 + 1 = 2$, so that

$$\lim_{x\to 0} \frac{\sin x}{\cos x + 1} = \frac{0}{2} = 0 \quad \Rightarrow \quad \lim_{x\to 0} \frac{\cos x - 1}{x} = -1 \times 0 = 0.$$

So far all but one of our functions have had limits at the given point(s), and the badly behaved one in Example 2.5 at least had (differing) left and right limits. However there are functions whose behaviour is far worse.

Example 2.15. $\lim_{x\to 0} \dfrac{1}{x}$ does not exist.

To see this consider the graph of $\frac{1}{x}$:



As x approaches 0 from the right $\frac{1}{x}$ becomes very large and positive, and as x approaches 0 from the left $\frac{1}{x}$ becomes very large and negative. More formally, if we choose any $M > 0$ then

$$0 < x < \frac{1}{M} \;\Rightarrow\; \frac{1}{x} > M \quad \text{and} \quad -\frac{1}{M} < x < 0 \;\Rightarrow\; \frac{1}{x} < -M$$

That is $\frac{1}{x}$ becomes greater (or less) than any bound M $(-M)$ we choose for all values of x sufficiently close to 0.

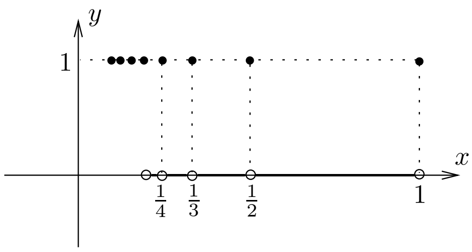
We say that $\lim_{x\to 0} \dfrac{1}{x}$ does not exist because the function **diverges** or **blows up** at the origin. Other functions fail to have limits despite not blowing up at the point of interest:

Example 2.16. $\lim_{x\to 0} f(x)$ does not exist for $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1 & \text{if } x = \frac{1}{n} \text{ for some } n \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

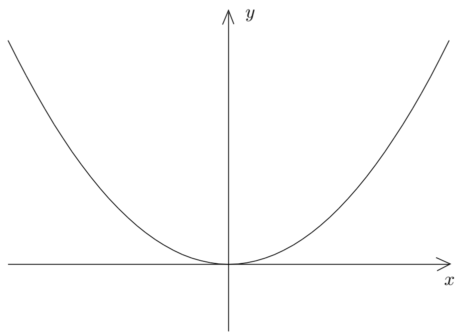
To see this note that for any choice of $\delta > 0$ there are values of x satisfying $0 < x < \delta$ for which $f(x) = 1$ (take any $x = \frac{1}{n}$ for $n > \frac{1}{\delta}$) and values of x for which $f(x) = 0$. Thus $f(x)$ cannot approach a *single* value as $x \to 0+$.

Note however that $f(x) = 0$ for all $x < 0$, and so $\lim_{x\to 0-} f(x) = 0$.

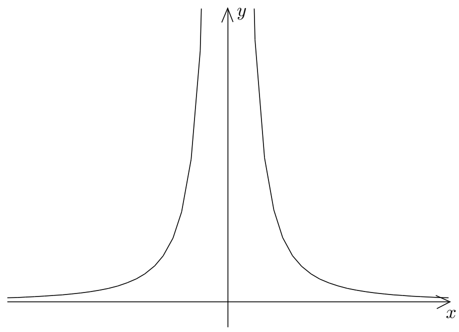


Coping with infinity

Consider the function $f(x) = \frac{1}{x^2}$. We can obtain a sketch the graph of this function by looking at the graph for $x \mapsto x^2$:



$y = x^2$



$y = \frac{1}{x^2}$

The function blows up at $x = 0$ but unlike $\frac{1}{x}$ it goes to $+\infty$ as x converges to 0 from either side. Moreover, note that as x gets very large, in either direction, the value of $\frac{1}{x^2}$ is always positive but gets closer and closer to 0. The notation and definitions for these sorts of behaviour are as follows:

Definition 2.17. Let f be a function on $\mathbb{R}$ and $a \in \mathbb{R}$. Then f **diverges to infinity as x tends to a** if for *every* $M > 0$ there is *some* $\delta > 0$ such that

$$0 < |x - a| < \delta \;\Rightarrow\; f(x) > M.$$

This is written $f(x) \to +\infty$ as $x \to a$.

Similarly, f **diverges to minus infinity as x tends to a** if for *every* $M > 0$ there is *some* $\delta > 0$ such that

$$0 < |x - a| < \delta \;\Rightarrow\; f(x) < -M,$$

and this is written $f(x) \to -\infty$ as $x \to a$.

We can adjust the definition accordingly to take care to left and right limits, and when we do this can give meaning to the statements $\frac{1}{x} \to +\infty$ as $x \to 0+$, and $\frac{1}{x} \to -\infty$ as $x \to 0-$. Indeed, this was *proved* in Example 2.15.

For the limit of a function as the variable x gets large, the appropriate definitions are as follows:

Definition 2.18. Let f be a function on $\mathbb{R}$ and let $l_1, l_2 \in \mathbb{R}$. Then f **converges to l_1 as x tends to $+\infty$** if for *every* $\varepsilon > 0$ there is *some* $M_1 > 0$ such that

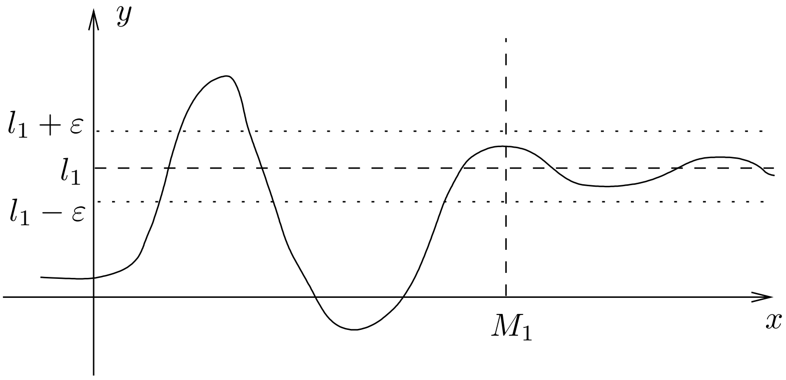
$$x > M_1 \;\Rightarrow\; |f(x) - l_1| < \varepsilon.$$

This is written $f(x) \to l_1$ as $x \to +\infty$.

Similarly f **converges to l_2 as x tends to $-\infty$** if for *every* $\varepsilon > 0$ there is *some* $M_2 > 0$ such that

$$x < -M_2 \;\Rightarrow\; |f(x) - l_2| < \varepsilon,$$

which is written $f(x) \to l_2$ as $x \to -\infty$.



If we combine elements of the two definitions we can then give a rigorous meaning to statements such as "$f(x) \to +\infty$ as $x \to +\infty$".

Example 2.19.

(a) For each $n \geq 1$ let $f_n(x) = x^n$ for each $x \in \mathbb{R}$. Then (recalling Proposition 1.4)

$$f_n(x) \to +\infty \text{ as } x \to +\infty$$
$$f_n(x) \to -\infty \text{ as } x \to -\infty \text{ if } n \text{ is odd}$$
$$f_n(x) \to +\infty \text{ as } x \to -\infty \text{ if } n \text{ is even}$$

That this is the case follows since $f_n(x) \geq x$ for all $x \geq 1$, and if $x > 0$ then

$$(-x)^n = (-1)^n x^n = \begin{cases} x^n & \text{if } n \text{ is even,} \\ -x^n & \text{if } n \text{ odd.} \end{cases}$$

(b) Let

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

Now $x^2 + 3x + 2 \to (-1)^2 + 3 \times (-1) + 2 = 0$ as $x \to -1$ by the calculus of limits, and $x - 3 \to (-1) - 3 = -4 \neq 0$. Moreover $x^2 + 3x + 2 = (x + 1)(x + 2)$, so

$$-1 < x < 3 \;\Rightarrow\; x^2 + 3x + 2 > 0 \text{ and } x - 3 < 0 \;\Rightarrow f(x) < 0, \text{ and}$$
$$-2 < x < -1 \;\Rightarrow\; x^2 + 3x + 2 < 0 \text{ and } x - 3 < 0 \;\Rightarrow f(x) > 0.$$

It follows that

$$f(x) \to +\infty \text{ as } x \to -1-, \quad \text{and} \quad f(x) \to -\infty \text{ as } x \to -1+.$$

For the behaviour as $x \to \pm\infty$, divide top and bottom by x^2. This gives

$$f(x) = \frac{x-3}{x^2+3x+2} = \frac{\frac{1}{x} - \frac{3}{x^2}}{1 + \frac{3}{x} + \frac{2}{x^2}} \to \frac{0}{1} = 0$$

as $x \to +\infty$ and as $x \to -\infty$, since $\frac{1}{x^n} \to 0$ as $x \to \pm\infty$ for each $n \geq 1$.

Exercise 2.20. Carry out a similar analysis as done for f in part (b) of Example 2.19 above for the functions $g(x) = \dfrac{x^5 - 4x^2 + 2}{5 + 2x^4 - 7x^5}$ and $h(x) = \dfrac{x^2 - x + 1}{x - 2}$.

Continuous functions

The intuitive idea of a continuous function is one whose graph has 'no jumps', or 'can be drawn without taking the pen off of the paper.' Consequently it should have a limit at each point a, that is $f(x)$ should converge to some value as x tends to a, and furthermore that value should be equal to the value of the function at that point. This idea is captured by the following:

Definition 2.21. Let $f : \mathbb{R} \to \mathbb{R}$ be a function and let $a \in \mathbb{R}$. We say that f **is continuous at the point** a if

(i) $\lim_{x \to a} f(x)$ exists, *and*

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

f **is continuous on** $\mathbb{R}$ if it is continuous at *every* point $a \in \mathbb{R}$.

A function that is not continuous at a point a is called **discontinuous**. Intuitively speaking it has a break in the graph at this point.

If we rewrite this using the definition of limits, we see that a function $f : \mathbb{R} \to \mathbb{R}$ is continuous at a point a if for *every* choice of $\varepsilon > 0$ there is *some* $\delta > 0$ such that

$$0 < |x - a| < \delta \;\Rightarrow\; |f(x) - f(a)| < \varepsilon.$$

Now we must insist that f is defined at the point $x = a$ unlike when we were considering limits earlier, and since $|f(x) - f(a)| = 0$ when $x = a$, we can remove the inequality $0 < |x - a|$ from the definition. Thus the function is continuous at a if for every $\varepsilon > 0$ there is some $\delta > 0$ such that

$$|x - a| < \delta \;\Rightarrow\; |f(x) - f(a)| < \varepsilon.$$

Example 2.22. For every choice of $c, d \in \mathbb{R}$ we know that the function $f(x) = c + dx$ is continuous since by Example 2.6 we have that $\lim_{x \to a} f(x) = c + da = f(a)$.

Again, this one result in conjunction with the calculus of limits can be used to give many more examples of continuous functions.

Proposition 2.23. *If $f(x)$ and $g(x)$ are functions $\mathbb{R} \to \mathbb{R}$ that are continuous at $x = a$, then so are the functions $f(x) + g(x)$ and $f(x)g(x)$. Moreover if $g(a) \neq 0$ then the function $\frac{f(x)}{g(x)}$ is continuous at $x = a$, as is the function $\sqrt[n]{f(x)}$ for each $n \geq 2$ if $f(a) > 0$.*

Proof. These are all immediate consequences of the calculus of limits, as given in Proposition 2.8. For instance, consider the function fg which maps the point x to $f(x)g(x)$. Since f and g are continuous at a, f and g have limits there, and moreover

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

Thus, by Proposition 2.8, $\lim_{x \to a} \big(f(x)g(x)\big)$ exists, and is equal to the product of the above limits, hence equals $f(a)g(a) = (fg)(a)$ as required. □

Since any polynomial can be written as a sum of products of the functions considered in Example 2.22, the above result leads immediately to the following:

Corollary 2.24. *Every polynomial is continuous on* $\mathbb{R}$*. Every rational function is continuous wherever it is defined.*

Definition 2.25. Let f and g be functions from $\mathbb{R}$ to $\mathbb{R}$. Their **composition** is the function $g \circ f : \mathbb{R} \to \mathbb{R}$ is defined by

$$(g \circ f)(x) = g\big(f(x)\big).$$

That is, apply f to the point x and then apply g to the result.

Example 2.26. Consider the functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2 + 3$ and $g(x) = 1 - x$. Calculate $f \circ g(x)$ and $g \circ f(x)$.

Note in particular that $f \circ g$ and $g \circ f$ are different functions in this example; this is usually the case for compositions.

Proposition 2.27. *Let* f *and* g *be functions* $\mathbb{R} \to \mathbb{R}$ *with* f *continuous at some point* a*, and* g *continuous at the image point* $f(a)$*. Then* $g \circ f$ *is continuous at* a*.*

Proof. For each $\varepsilon > 0$ we must show that there is a $\delta > 0$ such that

$$|x - a| < \delta \;\Rightarrow\; |(g \circ f)(x) - (g \circ f)(a)| < \varepsilon.$$

So fix an $\varepsilon > 0$, then since g is continuous at the point $f(a)$, there is some $\delta_g > 0$ such that

$$|t - f(a)| < \delta_g \;\Rightarrow\; |g(t) - g(f(a))| < \varepsilon. \tag{$*$}$$

But in turn we know that f is continuous at a, so given this positive number $\delta_g > 0$ we know that there is some $\delta > 0$ such that

$$|x - a| < \delta \;\Rightarrow\; |f(x) - f(a)| < \delta_g. \tag{$\dagger$}$$

So if we take any $x \in \mathbb{R}$ that satisfies $|x - a| < \delta$ then $|f(x) - f(a)| < \delta_g$ by ($\dagger$). But this means that we can apply ($*$) with $t = f(x)$ to get

$$|x - a| < \delta \;\Rightarrow\; |f(x) - f(a)| < \delta_g \;\Rightarrow\; |g(f(x)) - g(f(a))| < \varepsilon,$$

which is precisely what we needed to show. □

So far we have only considered continuous functions that have been defined at every point of $\mathbb{R}$. This may not always be the case, as in the following examples.

Example 2.28. Define a function f on $\mathbb{R}$ by

$$f(x) = \frac{x^2 - 9}{x - 3}, \quad x \neq 3.$$

Since $x^2 - 9 = (x - 3)(x + 3)$ we have that $f(x) = x + 3$ for all $x \neq 3$, and so for $a \neq 3$ we have $\lim_{x \to a} f(x) = a + 3 = f(a)$, hence f is continuous at every $a \neq 3$.

Moreover by Proposition 2.13 and Example 2.6 $\lim_{x\to 3} f(x) = 6$. So if we were to extend the domain of definition of f by *setting* $f(3) = 6$ then we will obtain a continuous function.

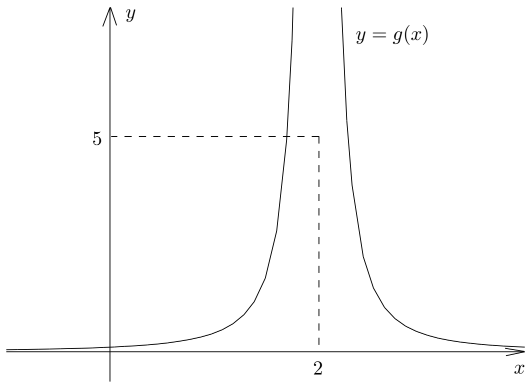
On the other hand if we define f to be *any other value* at $x = 3$ then it would still be true that $f(x) \to 6$ as $x \to 3$, but the redefined function will not be continuous in this case since we will have $f(3) \neq 6$.

Example 2.29. The function $f(x) = \frac{1}{x}$ is discontinuous at $x = 0$ since we have shown that $\lim_{x\to 0} f(x)$ does not exist.

Similarly consider the function g defined on all of $\mathbb{R}$ by

$$g(x) = \begin{cases} \dfrac{1}{(x-2)^2} & \text{if } x \neq 2, \\ 5 & \text{if } x = 2. \end{cases}$$

As x tends to 2 we see that $(x-2)^2$ converges to 0 and is positive, and so $f(x)$ diverges to $+\infty$. Hence $\lim_{x\to 2} f(x)$ cannot exist, and so f is not continuous at 2, despite the fact that it is defined. Moreover, any change in the definition of f at $x = 2$ will not alter this fact.



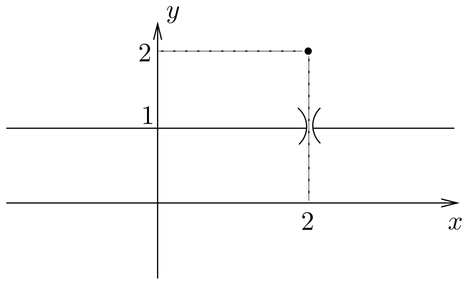
The behaviour in Example 2.28 is different from that in Example 2.29. In the case of Example 2.28 the point $x = 3$ is known as a **removable discontinuity** since the function does have a limit there, and so redefining the function will produce a continuous function. In the case of Example 2.29 the points of discontinuity of f and g are **essential discontinuities** since they do not have limits, and so no matter how we define the functions at these points, they will never be continuous there.

Another example of a removable discontinuity is the following:

Example 2.30. Let f be the function $\mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1 & \text{if } x \neq 1, \\ 2 & \text{if } x = 1. \end{cases}$$

Then $\lim_{x\to 1} f(x) = 1 \neq f(1)$, and so f is not continuous at $x = 1$. If we redefine f to be 1 at $x = 1$ then we will remove the singularity.



Exercise 2.31 (S03 2(b)). Consider the following function on the real line $\mathbb{R}$:

$$f(x) = \frac{x^2 - 5x - 14}{2x^2 + 3x - 2}$$

At which point(s) is f continuous? At which point(s) is f undefined?

Determine whether the point(s) at which f is undefined are essential or removable singularities.

Describe the behaviour of $f(x)$ as $x \to +\infty$

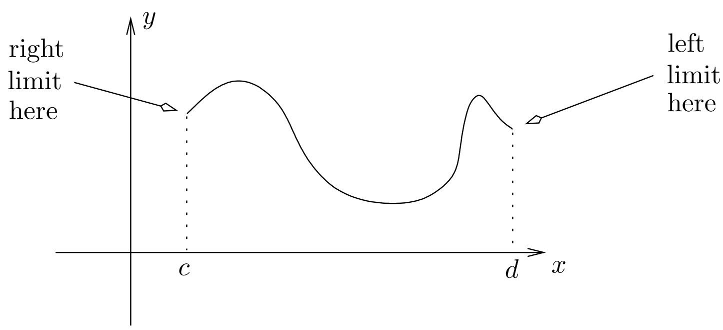
Continuity on intervals

Often a function may be defined only on some interval I of the real line and in this case we must make some changes to our definition of continuity. If $a \in I$ is not an endpoint then x can approach a from either side and so it still makes sense to ask if $f(a) = \lim_{x \to a} f(x)$. If $a \in I$ is an endpoint then x can only approach this value from one side and still be in I, and so we must use one sided limits in our modified definition of continuity.

Thus if $I = [c, d]$ for some $c < d$, then a function $f : I \to \mathbb{R}$ is *continuous on* I if it is continuous at each point of I, which means that at a point $a \in I$

$$\begin{aligned}
&\lim_{x \to a} f(x) \text{ exists and is equal to } f(a) \text{ if } c < a < d, \\
&\lim_{x \to a+} f(x) \text{ exists and is equal to } f(a) \text{ if } a = c, \quad \text{and} \\
&\lim_{x \to a-} f(x) \text{ exists and is equal to } f(a) \text{ if } a = d.
\end{aligned}$$

Again, this just means that the graph has no breaks over $[c, d]$.



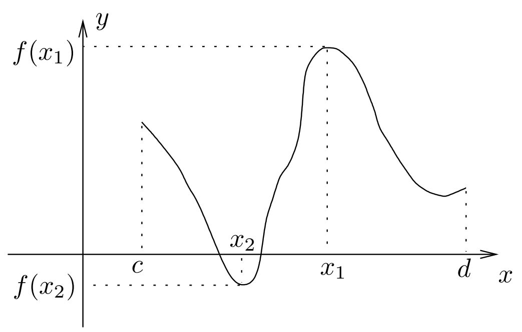
A moments thought shows that if we start with a function that is continuous on all of $\mathbb{R}$ and restrict it so that we only consider values of x from some subinterval I then that function will also be continuous. This idea provides a ready supply of continuous functions on subintervals of $\mathbb{R}$ to which we can apply the following result.

Theorem 2.32 (Intermediate Value Theorem). *Let $c, d \in \mathbb{R}$ with $c < d$, and let $f : [c, d] \to \mathbb{R}$ be continuous. Then there are points $x_1, x_2 \in [c, d]$ such that*

$$f(x_2) \leq f(x) \leq f(x_1) \quad \text{for all } x \in [c, d].$$

Moreover f takes all values between $f(x_2)$ and $f(x_1)$. That is, if $y \in \mathbb{R}$ satisfies $f(x_2) \leq y \leq f(x_1)$ then there is some $x \in [c, d]$ such that $f(x) = y$.

The number $f(x_1)$ is called the **maximum** of f on $[c, d]$ and the number $f(x_1)$ is called the **minimum** of f on $[c, d]$. The result is intuitively obvious, but not so easy to prove. It depends on the *completeness* of the real number system.



Thus we see that any continuous function on an interval of the form $[c, d]$ is **bounded**, that is the values $f(x)$ that the function takes for $x \in [c, d]$ lie between two numbers. Conversely if we have an **unbounded** function g on $[c, d]$, it follows that it cannot be continuous on $[c, d]$.

Also, the Intermediate Value Theorem is not valid if we replace the interval $[c, d]$ with any of the other three intervals having c and d as end points. For example the function $f(x) = \frac{1}{x}$ is well-defined and continuous on the open interval $(0, 1)$ by Theorem 2.23. However $f(x) \to +\infty$ as $x \to 0+$: there is no number M such that $f(x) \leq M$ for all $x \in (0, 1)$. Also $f(x) > 1$ for all $x \in (0, 1)$ and $\lim_{x \to 1} f(x) = 1$, so that $f(x)$ gets as close to 1 as we like. But there is no point $x \in (0, 1)$ for which $f(x) = 1$.

One thing that the Intermediate Value Theorem can tell us is if there is a solution to a given equation in a given interval. For instance consider the polynomial $f(x) = x^5 - 2x^2 + 4x - 2$. Then f is continuous on the interval $[0, 1]$, since it is continuous on all of $\mathbb{R}$. Moreover $f(0) = -2$ and $f(1) = 1$. Thus there must be some $x_0 \in (0, 1)$ such that $f(x_0) = 0$ since $f(x)$ assumes all values between -2 and 1, and so there is a solution to $x^5 - 2x^2 + 4x - 2 = 0$ in this interval. Unlike the case of quadratics, cubics and quartics, there is no formula for finding the roots of quintics (and polynomials of higher degree), and so reasoning like the above gives a first step to locating them.

This use of the Intermediate Value Theorem also has a similar, but more theoretical application.

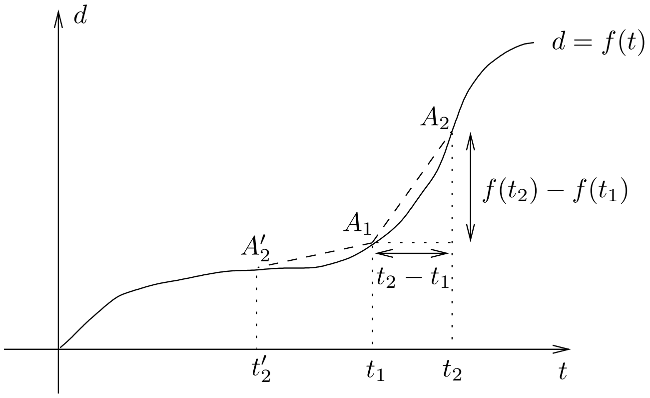
Proposition 2.33. *Let $n \geq 1$ be an integer and $a \geq 0$ any real number. There exists a unique real number $b \geq 0$ that satisfies $b^n = a$; that is, there is a unique positive nth root of a.*

Proof. Recall from Proposition 1.4 that the function $f : [0, \infty) \to \mathbb{R}$ given by $f(x) = x^n$ is strictly increasing, and is a polynomial hence continuous. Moreover we know that $f(x) \to +\infty$ as $x \to \infty$ (cf. part (a) of Example 2.19). In particular there must be some $c > 0$ such that $f(c) = c^n > a$. Thus, by the Intermediate Value Theorem, f restricted to the interval $[0, c]$ must take all values between $f(0) = 0$ and $f(c) > a$, and so there must be some $b \geq 0$ that satisfies $f(b) = b^n = a$. That is, there is *an* nth root. That it is unique follows from the fact that f is strictly increasing, since if $d \geq 0$ such that $d \neq b$ then either $d < b$ or $d > b$, and in either case $f(d) \neq f(b)$. □

3 Differentiation

The idea and definition of derivatives

Consider a distance-time graph for some object (car, bicycle, atom,...) moving in a 'continuous manner' through 'one-dimensional space'. We plot the displacement from the starting point as a function of time, starting at $t = 0$.



Thus the function $d = f(t)$ is continuous. The average velocity $V(t_1, t_2)$ of our body as t varies from time t_1 to time t_2 is given by

$$V(t_1, t_2) = \frac{f(t_2) - f(t_1)}{t_2 - t_1} = \text{slope of the line segment } A_1A_2.$$

If we set $h = t_2 - t_1$, so that $t_2 = t_1 + h$, this can be rewritten as

$$V(t_1, t_2) = \frac{f(t_1 + h) - f(t_1)}{h}.$$

Here h is the 'change in time'. We could also take the second time to be t_2', an earlier time than t_1. That is $t_2' < t_1$, which is equivalent to saying that $h = t_2' - t_1 < 0$. But in either case as we take smaller and smaller values of h (i.e. as h tends to 0), we hope that this number will converge to a limit that we can call the *actual* or *instantaneous* velocity at time $t = t_1$.

We are trying to approximate the possibly complicated curve $d = f(t)$ by a straight line at each point, that is, find the slope of the straight line that just touches the curve $d = f(t)$ at the point $(t_1, f(t_1))$. Thus this straight line should meet the curve at this point and go in the same direction — it is the **tangent** to the curve.

Definition 3.1. A function $f : \mathbb{R} \to \mathbb{R}$ is **differentiable at** $a \in \mathbb{R}$ if

$$\lim_{h \to 0} \frac{f(a + h) - f(a)}{h} \text{ exists.}$$

The value of this limit is the **derivative of** f **at** a and is denoted $f'(a)$. The function f is **differentiable on** $\mathbb{R}$ if it is differentiable at every $a \in \mathbb{R}$. If this is the case, then $f' : a \mapsto f'(a)$ is a new function $\mathbb{R} \to \mathbb{R}$.

Remarks. (i) Some authors will write

$$f'(a) = \lim_{\Delta a \to 0} \frac{f(a + \Delta a) - f(a)}{\Delta a} \quad \text{(when it exists)}$$

where Δa stands for the *change in the variable a*

Also, often people use notation of the form $y = f(x)$ for a function, in which case if y (or f) is differentiable everywhere then the function f' is denoted $\frac{dy}{dx}$

(ii) When defining the derivative we choose a function f and a point a, and from these construct a new function g in the variable h by

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

for all $h \neq 0$. Note that g depends on the variable h. We then ask if $\lim_{h \to 0} g(h)$ exists. The derivative *cannot* be calculated by substituting $h = 0$ or $\Delta a = 0$ directly into these formulae. It is for this reason that our definition of a limit of a function $g(x)$ as x tends to some number b did not actually depend on $g(b)$, and indeed did not require that this even be defined.

Since we hope to use the derivative to find straight line approximations to more complicated curves, we ought to check that it behaves correctly if we consider a straight line.

Proposition 3.2. *Let $c, d \in \mathbb{R}$ and define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = c + dx$. Then f is differentiable on $\mathbb{R}$ with $f' : \mathbb{R} \to \mathbb{R}$ given by $f'(x) = d$ for all $x \in \mathbb{R}$. That is, the derivative at each point is equal to the gradient of this straight line.*

Proof. Choose an $x \in \mathbb{R}$, then for any $h \neq 0$

$$\begin{aligned} \frac{f(x + h) - f(x)}{h} &= \frac{[c + d(x + h)] - [c + dx]}{h} \\ &= \frac{dh}{h} = d \end{aligned}$$

and so clearly

$$\lim_{h \to 0} \frac{f(x + h) - f(x)}{h} = d.$$

Thus the limit exists for all $x \in \mathbb{R}$ (and is independent of x), with $f'(x) = d$. □

In particular we see that if f is a constant function, i.e. $d = 0$, then its derivative $f'(x)$ is zero everywhere.

Example 3.3. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2 + 5x + 2$. Show that f is differentiable on $\mathbb{R}$ and find f'.

Solution. For each $x \in \mathbb{R}$ and $h \neq 0$

$$\begin{aligned} \frac{f(x + h) - f(x)}{h} &= \frac{[(x + h)^2 + 5(x + h) + 2] - [x^2 + 5x + 2]}{h} \\ &= \frac{x^2 + 2xh + h^2 + 5x + 5h + 2 - x^2 - 5x - 2}{h} \\ &= \frac{(2x + 5)h + h^2}{h} \\ &= 2x + 5 + h \end{aligned}$$

and so

$$\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}=\lim_{h\to 0}(2x+5+h)=2x+5$$

by applying Example 2.6 (with $c=2x+5$ and $d=1$) to this function in the variable h. Hence f is differentiable with $f'(x)=2x+5$.

More generally using this method one can show that for any $a,b,c\in\mathbb{R}$ that $f(x)=ax^2+bx+c$ is differentiable with $f'(x)=2ax+b$.

However, we might ask which functions in general are differentiable, and the next result shows that we can restrict our search somewhat.

Proposition 3.4. *Let $f:\mathbb{R}\to\mathbb{R}$ be a function and let $a\in\mathbb{R}$. Suppose that f is differentiable at a, then f is continuous at a.*

Proof. Since f is assumed to be differentiable at a we are assuming that

$$f'(a)=\lim_{h\to 0}\frac{f(a+h)-f(a)}{h}\quad\text{exists.}$$

So in particular $f(a)$ must be defined. For any $h\neq 0$

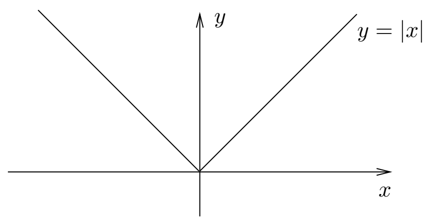
$$f(a+h)-f(a)=h\times\frac{f(a+h)-f(a)}{h}$$

and thus, by the calculus of limits,

$$\begin{aligned}\lim_{h\to 0}(f(a+h)-f(a))&=\left(\lim_{h\to 0}h\right)\left(\lim_{h\to 0}\frac{f(a+h)-f(a)}{h}\right)\\&=0\times f'(a)=0.\end{aligned}$$

Thus, by the calculus of limits once more, $\lim_{h\to 0}f(a+h)=f(a)$, which is equivalent to saying that $\lim_{x\to a}f(x)=f(a)$ (why?), so that f is continuous at a as required. □

Taking the *contrapositive* of this result we see that if a function f is not continuous at a given point a then it cannot be differentiable there. However the *converse* to the proposition is not true: there are functions that are continuous at some point but fail to be differentiable there. For example consider the behaviour of function $f(x)=|x|$ at $x=0$. It is continuous there but not differentiable:



From the graph we see that $f(x)$ converges to $|0|=0$ as $x\to 0$, so that it is continuous there. However, if $h>0$ then

$$\frac{f(h)-f(0)}{h}=\frac{|h|-|0|}{h}=\frac{h}{h}=1\ \Rightarrow\ \lim_{h\to 0+}\frac{f(h)-f(0)}{h}=1,$$

and if $h < 0$ then

$$\frac{f(h)-f(0)}{h} = \frac{|h|-|0|}{h} = \frac{-h}{h} = -1 \Rightarrow \lim_{h\to 0-} \frac{f(h)-f(0)}{h} = -1.$$

Thus we do have left and right limits but they are different, hence the required *limit* cannot exist by Proposition 2.4. For this f we can say that it has a **left derivative at** 0 and a **right derivative at** 0 (denoted $f'_-(0)$ and $f'_+(0)$ respectively), but that they are different.

More generally, the following is a direct consequence of Proposition 2.4:

Proposition 3.5. *A function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at a point a if and only if it has left and right derivatives at that point, and they are equal.*

Remark. The official definition of left and right derivatives are

$$f'_-(a) = \lim_{h\to 0-} \frac{f(a+h)-f(a)}{h} \quad \text{and} \quad f'_+(a) = \lim_{h\to 0+} \frac{f(a+h)-f(a)}{h},$$

whenever these limits exist.

If we are dealing with a function that is defined on a subinterval of $\mathbb{R}$ of the form $[c, d]$ then we could use left and right derivatives to define what it means for f to be differentiable on $[c, d]$. However in practice we are only really interested in differentiability of f on *open* intervals, that is intervals of the form (c, d), in which case the usual definition involving two-sided limits applies. This follows since if $a \in (c, d)$ then the fraction

$$\frac{f(a+h)-f(a)}{h}$$

is well-defined for all h satisfying $|h| < \min\{a-c, d-a\}$ (that is for all *sufficiently small* h), and so h can tend to 0 while assuming both positive and negative values.

As with limits and continuity there are rules that allow us to break down the problem of finding a derivative into simpler parts, such as those calculated in Proposition 3.2.

Proposition 3.6. *Let $f, g : \mathbb{R} \to \mathbb{R}$ be functions that are differentiable at some $a \in \mathbb{R}$. Then*

(i) *$f + g$ is differentiable at a, with $(f + g)'(a) = f'(a) + g'(a)$.*

(ii) *fg is differentiable at a, with $(fg)'(a) = f'(a)g(a) + f(a)g'(a)$.* [**Product Rule**]

(iii) *If $g(a) \neq 0$ then $\frac{f}{g}$ is differentiable at a with*

$$\left(\frac{f}{g}\right)'(a) = \frac{f'(a)g(a) - f(a)g'(a)}{g(a)^2}.$$ [**Quotient Rule**]

Remark. If we use the $\frac{d}{dx}$ notation, then the product and quotient rules are usually written

$$\frac{d}{dx}(uv) = \frac{du}{dx}v + u\frac{dv}{dx}, \quad \text{and} \quad \frac{d}{dx}\left(\frac{u}{v}\right) = \frac{\frac{du}{dx}v - u\frac{dv}{dx}}{v^2}.$$

Proof. All of these essentially follow by careful application of the calculus of limits. For example, for the product rule (ii) we are assuming that both of the limits

$$f'(a) = \lim_{h\to 0} \frac{f(a+h) - f(a)}{h} \quad \text{and} \quad g'(a) = \lim_{h\to 0} \frac{g(a+h) - g(a)}{h}$$

exist, and want to show that the limit

$$\lim_{h\to 0} \frac{(fg)(a+h) - (fg)(a)}{h}$$

exists and equals the specified value. Now

$$\begin{aligned}(fg)(a+h) - (fg)(a) &= f(a+h)g(a+h) - f(a)g(a) \\ &= f(a+h)[g(a+h) - g(a)] + [f(a+h) - f(a)]g(a)\end{aligned}$$

and so

$$\frac{(fg)(a+h) - (fg)(a)}{h} = f(a+h) \times \frac{g(a+h) - g(a)}{h} + \frac{f(a+h) - f(a)}{h} \times g(a).$$

But f is continuous at a, hence $\lim_{h\to 0} f(a+h) = f(a)$, and so we may apply the calculus of limits to the above to get

$$\begin{aligned}\lim_{h\to 0} \frac{(fg)(a+h) - (fg)(a)}{h} &= \lim_{h\to 0} f(a+h) \times \lim_{h\to 0} \frac{g(a+h) - g(a)}{h} \\ &\quad + \left(\lim_{h\to 0} \frac{f(a+h) - f(a)}{h}\right) \times g(a) \\ &= f(a)g'(a) + f'(a)g(a)\end{aligned}$$

as required.

Similar manipulations will give the quotient rule (iii). To carry this out we will need to make use of the fact that since we are assuming g is differentiable (and hence continuous) at a with $g(a) \neq 0$, then the function is nonzero for all values of x sufficiently close to a. □

Induction and the techniques of the above proposition give the following:

Proposition 3.7. *For each integer $n \geq 0$ define $f_n : \mathbb{R} \to \mathbb{R}$ by $f_n(x) = x^n$. Then f_n is differentiable on $\mathbb{R}$ with $f_n'(x) = nx^{n-1}$.*

Proof. From Proposition 3.2 we already know that $f_0(x) = 1$ and $f_1(x) = x$ are differentiable on $\mathbb{R}$ with $f_0'(x) = 0$ and $f_1'(x) = 1$ (take $c = 1, d = 0$ and $c = 0, d = 1$ respectively).

Now suppose that there is some $N \geq 1$ such that f_n is differentiable with $f_n'(x) = nx^{n-1}$ for all $0 \leq n \leq N$. Since $f_{N+1}(x) = x^{N+1} = x^N x = f_N(x)f_1(x)$, we see that f_{N+1} is the product of two functions that are differentiable on $\mathbb{R}$ by our assumption, and so by the product rule we know that f_{N+1} is also differentiable on $\mathbb{R}$, with

$$f_{N+1}'(x) = f_N'(x)f_1(x) + f_N(x)f_1'(x) = Nx^{N-1} \times x + x^N \times 1 = (N+1)x^N$$

as required. So now, by induction, the stated formula holds for all integers $n \geq 0$. □

Remarks.

(i) The Binomial Theorem can be used to give an alternative route to proving this result that avoids using induction.

(ii) Combining this proposition with another induction argument shows that every polynomial is differentiable at each point of $\mathbb{R}$.

Example 3.8. If we have $f(x) = x^{17} - 3x^8 + 4x^5$ then it is a sum of three functions that are each differentiable on $\mathbb{R}$, hence is itself differentiable on $\mathbb{R}$ by part (i) of Proposition 3.6. Moreover we have

$$f'(x) = 17x^{16} - 3 \times 8x^7 + 4 \times 5x^4 = 17x^{16} - 24x^7 + 20x^4.$$

Proposition 3.9. *For each integer $n \geq 1$ define the function g_n by $g_n(x) = \dfrac{1}{x^n}$. Then each g_n is differentiable at every $x \neq 0$, with $g_n'(x) = -\dfrac{n}{x^{n+1}}$.*

Proof. Using the notation of the previous result, note that for every $x \neq 0$, $f_n(x) = x^n \neq 0$ is differentiable and non-zero. Thus, since $g_n(x) = \dfrac{f_0(x)}{f_n(x)}$, we can apply the quotient rule to deduce that g_n is differentiable on $(-\infty, 0) \cup (0, \infty)$. Moreover

$$g_n'(x) = \frac{f_0'(x)f_n(x) - f_0(x)f_n'(x)}{f_n(x)^2}$$

for all $x \neq 0$. But we know that $f_0'(x) = 0$ and $f_n'(x) = nx^{n-1}$, and so

$$g_n'(x) = \frac{0 \times x^n - 1 \times nx^{n-1}}{x^{2n}} = -\frac{n}{x^{n+1}}$$

as required. □

Remark. These two propositions can be summarised by saying that for any integer m the function $x \mapsto x^m$ is differentiable wherever it is defined, with

$$\frac{d}{dx}x^m = mx^{m-1}.$$

Here, by definition, $x^m = \dfrac{1}{x^{-m}}$ if $m < 0$ (that is $x^{-1} = \dfrac{1}{x}, x^{-2} = \dfrac{1}{x^2}$ etc.).

Example 3.10. Consider the function $g(x) = x^7 + \frac{1}{x^2} - \frac{5}{x^{10}}$. Writing this as $g(x) = x^7 + x^{-2} - 5x^{-10}$ we see that g is the sum of three functions each of which is differentiable whenever $x \neq 0$. Thus $g(x)$ is differentiable whenever $x \neq 0$ and

$$g'(x) = 7x^6 - 2x^{-3} + 50x^{-11} = 7x^6 - \frac{2}{x^3} + \frac{50}{x^{11}}.$$

The final rule for calculating derivatives, known as the Chain Rule, involves composition of functions. Consider the function $f(x) = (x + 1)^2$. Expanding this gives $f(x) = x^2 + 2x + 1$, and so f is differentiable on $\mathbb{R}$ with $f'(x) = 2x + 2 = 2(x + 1)$. Our next result gives us another means of calculating this identity.

Proposition 3.11 (Chain Rule). *Let $f, g : \mathbb{R} \to \mathbb{R}$ be functions, and let F denote the composition $F = g \circ f$ (that is, $F(x) = g(f(x))$ for each $x \in \mathbb{R}$). If $a \in \mathbb{R}$ such that f is differentiable at a and g is differentiable at $f(a)$, then F is differentiable at a with*

$$F'(a) = g'(f(a))f'(a).$$

Proof. Define a function $G : \mathbb{R} \to \mathbb{R}$ by

$$G(k) = \begin{cases} \frac{g(f(a)+k)-g(f(a))}{k} - g'(f(a)) & \text{if } k \neq 0, \\ 0 & \text{if } k = 0. \end{cases}$$

Then $\lim_{k\to 0} G(k) = 0$, since we assumed that g is differentiable at $f(a)$. Hence G is continuous at 0. Rearranging the equation above we see that

$$g(f(a) + k) - g(f(a)) = k[G(k) + g'(f(a))] \quad \text{for all} \quad k \in \mathbb{R}. \tag{$*$}$$

Now for any $h \neq 0$

$$\begin{aligned} F(a + h) - F(a) &= g\big(f(a + h) - f(a) + f(a)\big) - g(f(a)) \\ &= g(r(h) + f(a)) - g(f(a)) \end{aligned}$$

where we define $r(h) = f(a + h) - f(a)$ for all $h \in \mathbb{R}$. The right hand side of this last equation is of the form of the left hand side of $(*)$, with $k = r(h)$, and so

$$F(a + h) - F(a) = r(h)[G(r(h)) + g'(f(a))].$$

But f is differentiable at a, hence continuous there as well, thus

$$\lim_{h\to 0} \frac{r(h)}{h} = f'(a) \quad \text{and} \quad \lim_{h\to 0} r(h) = r(0) = 0.$$

That is, r is a function on $\mathbb{R}$ that is continuous at 0 and satisfies $r(0) = 0$, and G is continuous at 0 and satisfies $G(0) = 0$. Hence $x \mapsto (G \circ r)(x)$ is continuous at 0, with $\lim_{h\to 0} (G \circ r)(h) = (G \circ r)(0) = 0$ by Proposition 2.27. So now

$$\begin{aligned} \lim_{h\to 0} \frac{F(a + h) - F(a)}{h} &= \lim_{h\to 0} \frac{r(h)}{h} \times \lim_{h\to 0} [G(r(h)) + g'(f(a))] \\ &= f'(a) \times [0 + g'(f(a))] = g'(f(a))f'(a) \end{aligned}$$

as required. □

Example 3.12. Define a function F on $\mathbb{R}$ by $F(x) = (3x^2 + x + 4)^{50}$. Deduce that F is differentiable on $\mathbb{R}$ and find $F'(x)$.

Solution. Define functions f and g on $\mathbb{R}$ by $f(x) = 3x^2 + x + 4$ and $g(x) = x^{50}$. Then $F = g \circ f$, since $F(x) = f(x)^{50}$. Now f and g are both polynomials, hence differentiable on $\mathbb{R}$, and so F is differentiable on $\mathbb{R}$ with $F'(x) = g'(f(x))f'(x)$. From our previous results about differentiating polynomials we have

$$f'(x) = 3 \times 2x + 1 = 6x + 1, \quad \text{and} \quad g'(x) = 50x^{49}.$$

Thus

$$F'(x) = 50f(x)^{49} \times (6x + 1) = 50(3x^2 + x + 4)^{49}(6x + 1).$$

Exercise 3.13. Define a function G by $G(x) = \dfrac{1}{(4x^3+7x^2)^{10}}$. Deduce that G is differentiable whenever $x \neq 0, -\frac{7}{4}$ and find $G'(x)$

Note that it would be foolish to try and expand out the function F and use the usual rules for differentiating polynomials. Moreover for G such a technique is no longer available to us.

With the results proved so far we can only differentiate polynomials and rational functions. The following are the derivatives of some other useful functions:

- If $f(x) = \sin x$ then $f'(x) = \cos x$ for all $x \in \mathbb{R}$.
- If $f(x) = \cos x$ then $f'(x) = -\sin x$ for all $x \in \mathbb{R}$.
- If $f(x) = \tan x$ then $f'(x) = \sec^2 x = (\sec x)^2$ for all $x \in \mathbb{R}$ such that $x \neq \frac{(2n+1)\pi}{2}$ for $n = 0, \pm 1, \pm 2, \ldots$, and where $\sec x = (\cos x)^{-1}$.
- If $f(x) = e^x$ then $f'(x) = e^x$ for all $x \in \mathbb{R}$. This is the only nonzero function that satisfies both $f'(x) = f(x)$ (i.e. is equal to its own derivative) and $f(0) = 1$.
- If $f(x) = \log x$ then $f'(x) = \dfrac{1}{x}$ for all $x > 0$.

The derivative for $\sin x$ follows since for any $x \in \mathbb{R}$ and $h \neq 0$

$$\begin{aligned}\frac{\sin(x+h) - \sin x}{h} &= \frac{\sin x \cos h + \sin h \cos x - \sin x}{h} \\ &= \frac{\sin h}{h}\cos x + \sin x \frac{\cos h - 1}{h} \\ &\to 1 \times \cos x + \sin x \times 0 = \cos x\end{aligned}$$

by Example 2.14. For $\cos x$ we can do a similar calculation, or use the chain rule since $\cos x = \sin(x + \frac{\pi}{2})$. For $\tan x$ we use the quotient rule since $\tan x = \frac{\sin x}{\cos x}$.

Exercise 3.14 (S02 4). Find the derivatives of the following functions by using the product, quotient and chain rules, along with known derivatives:

$$\begin{aligned} f(x) &= e^x \sin 4x + x^2 \\ h(x) &= \frac{(2x+1)^4}{x-3} \qquad (x \neq 3)\end{aligned}$$

Find the equation of the tangent to the graph of h at the point $(-\frac{1}{2}, 0)$.

The Mean Value Theorem and consequences

Theorem 3.15 (Rolle's Theorem). *Let f be a function from $[a,b]$ to $\mathbb{R}$ such that f is continuous on $[a,b]$ and differentiable on (a,b). Furthermore suppose that $f(a) = f(b)$. Then there is some $c \in (a,b)$ such that $f'(c) = 0$.*

Proof. If f is constant, that is if $f(x) = f(a)$ for all $x \in [a, b]$, then the theorem follows immediately from Proposition 3.2. So suppose that $f(x) \neq f(a)$ for some $x \in (a, b)$. Then we know from the Intermediate Value Theorem (Theorem 2.32) that there numbers $m, M \in \mathbb{R}$ and points $x_1, x_2 \in [a, b]$ such that $f(x_1) = M$ and $f(x_2) = m$, and that $m \leq f(x) \leq M$ for all $x \in [a, b]$.

Since f is not constant one of m or M must differ from $f(a)$; suppose it is M, then $x_1 \neq a$ and $x_1 \neq b$, so that $x_1 \in (a, b)$. Thus f is differentiable at this point, and so must have left and right derivatives that are equal. But for all (small) $h > 0$

$$\frac{f(x_1+h) - f(x_1)}{h} = \frac{f(x_1+h) - M}{h} \leq 0 \ \Rightarrow\ f'_+(x_1) = \lim_{h \to 0+} \frac{f(x_1+h) - f(x_1)}{h} \leq 0,$$

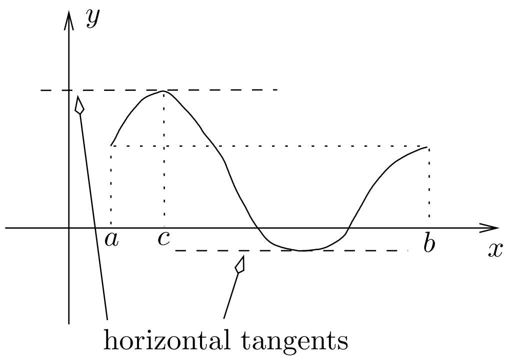
and for all (small) $h < 0$

$$\frac{f(x_1+h) - f(x_1)}{h} = \frac{f(x_1+h) - M}{h} \geq 0 \ \Rightarrow\ f'_-(x_1) = \lim_{h \to 0-} \frac{f(x_1+h) - f(x_1)}{h} \geq 0.$$

Thus we have $f'(x_1) = f'_+(x_1) \leq 0$ and $f'(x_1) = f'_-(x_1) \geq 0$, and the only way for this to be true is if $f'(x_1) = 0$.

If in fact $M = f(a)$ then we must have $m = f(x_2) < f(a)$ and a similar argument will give that $f'(x_2) = 0$. □

This result just says that if a function f defined on a closed interval of the form $[a, b]$ is sufficiently well-behaved to have a tangent at each point in (a, b), then at least one of these tangents must be horizontal. However it gives no information about precisely how many solutions there are to the equation $f'(c) = 0$.



Theorem 3.16 (Mean Value Theorem). *Let f be a function from $[a, b]$ to $\mathbb{R}$ such that f is continuous on $[a, b]$ and differentiable on (a, b). Then there is some $c \in (a, b)$ such that*

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

Proof. Define $F : [a, b] \to \mathbb{R}$ by

$$F(x) = f(x) - (x - a)\left(\frac{f(b) - f(a)}{b - a}\right).$$

Then F is the sum of two functions both of which are continuous on $[a, b]$ and differentiable on (a, b). Hence F is also continuous on $[a, b]$ and differentiable on (a, b).

Moreover $F(a) = F(b) = f(a)$. Thus F satisfies the hypotheses of Rolle's Theorem, and so there must be some $c \in (a, b)$ such that $F'(c) = 0$.

But now note that

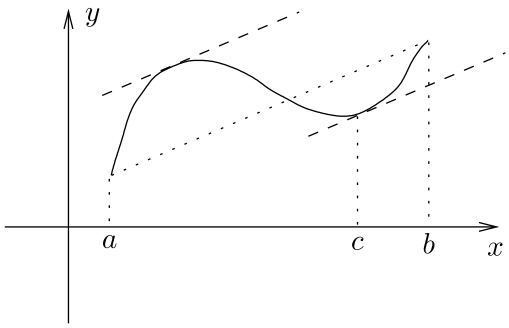
$$F'(x) = f'(x) - \frac{f(b) - f(a)}{b - a},$$

and so

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

as required. □

This theorem says that if the function f defined on the interval $[a, b]$ is suitably well behaved then there is a point in the interval where the tangent to f has the same slope as the line segment joining $(a, f(a))$ to $(b, f(b))$.



Exercise 3.17 (S04 4(b)). Verify that the function $f : [0, \pi] \to \mathbb{R}$ defined by

$$f(x) = x^2 \sin x$$

satisfies the hypotheses of Rolle's Theorem. Hence show that there is some $x \in (0, \pi)$ that satisfies the equation

$$x = -2 \tan x.$$

Example 3.18. Verify the Mean Value Theorem for $f(x) = 3x^2 - 5x + 4$ on $[1, 4]$.

Solution. First note that f is a polynomial hence continuous and differentiable on all of $\mathbb{R}$. Thus it is continuous on $[1, 4]$ and differentiable on $(1, 4)$. Also

$$\frac{f(4) - f(1)}{4 - 1} = \frac{32 - 2}{4 - 1} = 10,$$

and $f'(x) = 6x - 5$. So we must find some $c \in (1, 4)$ such that $f'(c) = 10$, i.e. some $c \in (1, 4)$ such that

$$6c - 5 = 10.$$

But the only real number satisfying this equation is $c = \frac{5}{2}$, which does indeed lie in our interval.

One important use of the Mean Value Theorem is that it allows us to *prove* that if a function has positive derivative in some interval then it has to be strictly increasing — a result that is intuitively obvious since all the tangents in this interval are pointing up and to the right.

Proposition 3.19. *Suppose that the function f is differentiable on (a, b) for some $a, b \in \mathbb{R}$ such that $a < b$.*

(a) *If $f'(x) = 0$ for all $x \in (a, b)$ then f is constant.*

(b) *If $f'(x) > 0$ for all $x \in (a, b)$ then f is strictly increasing.*

(c) *If $f'(x) < 0$ for all $x \in (a, b)$ then f is strictly decreasing.*

Proof. To show that f is constant if it has zero derivative fix any two points $x_1, x_2 \in (a, b)$ with $x_1 < x_2$. Then f restricted to $[x_1, x_2]$ satisfies the hypotheses of the Mean Value Theorem and so there is some $c \in (x_1, x_2)$ such that

$$f'(c) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}. \tag{†}$$

But by assumption $f'(c) = 0$, hence we must have $f(x_1) - f(x_2) = 0$, so that $f(x_1) = f(x_2)$ as required.

If, instead, $f'(x) > 0$ for all $x \in (a, b)$, then the same argument produces a $c \in (x_1, x_2)$ such that (†) holds, and now our assumption is that $f'(c) > 0$. But this in turn implies that $f(x_2) - f(x_1) > 0$ (since $x_2 - x_1 > 0$) and so $f(x_2) > f(x_1)$ as required. The proof of part (c) is similar. □

The techniques we have developed so far allow us to prove the existence of nth roots of nonnegative real numbers. In fact our work can extended as follows. For any positive *rational* number r (that is any number of the form $r = \frac{m}{n}$ for integers $m \geq 1$ and $n \geq 1$) we can make the following definition

$$x^r = (x^{1/n})^m \quad \text{for all } x > 0.$$

That is, take the nth root of x then raise it to the mth power. Using the uniqueness aspect of Proposition 2.33 we can show that this definition does not depend on the way we write r (for example if we took $r = \frac{1}{2}$, then we also have $r = \frac{3}{6}$, and so we should check that $x^{1/2} = (x^{1/6})^3$), and also that $x^r = (x^m)^{1/n}$. Moreover the uniqueness implies that the following formulae hold:

$$(x^r)^s = x^{rs}, \quad x^r x^s = x^{r+s}$$

for all $x > 0$ and positive rationals r and s. Furthermore, if we define x^{-r} to be $\dfrac{1}{x^r}$ and $x^0 = 1$ for all $x > 0$ then the above formulae remain true for any choice of rationals r and s.

Similarly we can use aspects of the Mean Value Theorem etc. to actually *define* what we mean by $\log x$ and e^x. Indeed, one way to define the natural logarithm is to set

$$\log x = \int_1^x \frac{1}{t}\, dt \qquad \text{for all } x > 0.$$

That is $\log x$ is the integral of $\dfrac{1}{t}$ from 1 to x — the theory of integration is covered in detail in MS2002. Immediate consequences of this definition are that $\log 1 = 0$

and that $\log x$ is differentiable with $\frac{d}{dx}\log x = \frac{1}{x}$. These can then be used to prove well-known formulae such as

$$\log xy = \log x + \log y \quad \text{and} \quad \log x^n = n \log x \tag{$*$}$$

for any real numbers $x, y > 0$ and integer $n \geq 0$. Also, since $\frac{d}{dx}\log x = \frac{1}{x} > 0$ for all $x > 0$, we see that $\log x$ is strictly increasing on $(0, \infty)$, and from $(*)$ it follows that $\log x \to -\infty$ as $x \to 0+$ and $\log x \to +\infty$ as $x \to +\infty$. As a result for each $y \in \mathbb{R}$ there is a unique solution x to the equation $y = \log x$ (the proof of this is essentially the same as the proof of Proposition 2.33). This solution x is denoted e^y, and this is one way to *define* the exponential function. It follows that

$$y = \log(e^y) \quad \text{for all } y \in \mathbb{R} \quad \text{and} \quad x = e^{\log x} \quad \text{for all } x > 0.$$

Then, from $(*)$, we get that $e^{x+y} = e^x e^y$ for all $x, y \in \mathbb{R}$.

Implicit differentiation

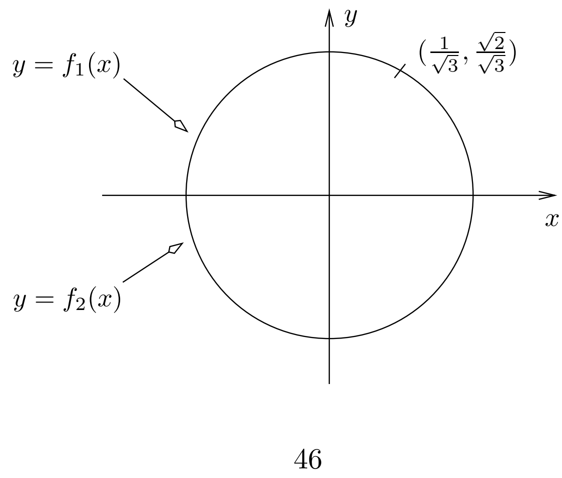
Thus we can use calculus to define roots, logarithms and exponentials, and from the definition of $\log x$ we get immediately that $\frac{d}{dx}\log x = \frac{1}{x}$. But what about differentiability of x^r and e^x? This is can be established using a technique known as **implicit differentiation**.

Example 3.20. Consider the equation

$$x^2 + y^2 = 1. \tag{C}$$

This is the equation of the circle of radius 1 whose centre is $(0, 0)$. It does not define a function $\mathbb{R} \to \mathbb{R}$ for two important reasons. First note that if $x > 1$ or $x < -1$ then $x^2 > 1$ hence $y^2 = 1 - x^2 < 0$, and there are no *real* numbers y that satisfy this inequality. So in order to have $x, y \in \mathbb{R}$ satisfying (C) we must have $x \in [-1, 1]$. But there is perhaps a more important reason that (C) does not define a function which is that for any $x \in (-1, 1)$ there are *two solutions* y to the equation (C), namely $\sqrt{1 - x^2}$ and $-\sqrt{1 - x^2}$.

Note that $f_1(x) = \sqrt{1 - x^2}$ and $f_2(x) = -\sqrt{1 - x^2}$ *are* both functions $[-1, 1] \to \mathbb{R}$. They are the two **branches** of the **relation** $x^2 + y^2 = 1$.



Suppose we wanted to find the equation of the tangent to the point $(\frac{1}{\sqrt{3}}, \frac{\sqrt{2}}{\sqrt{3}})$ on (C). This lies on the branch f_1, and so we could calculate $f_1'(\frac{1}{\sqrt{3}})$ to find the slope of the tangent.

However, an alternative is to instead differentiate the equation (C) directly. We think of y as a function of x, then y^2 is the product of y with itself and so the product rule tells us that

$$\frac{d}{dx}(y^2) = \frac{dy}{dx} \times y + y \times \frac{dy}{dx} = 2y\frac{dy}{dx}.$$

So then differentiating (C) gives

$$\frac{d}{dx}(x^2 + y^2) = \frac{d}{dx}(1) \quad \Rightarrow \quad 2x + 2y\frac{dy}{dx} = 0 \quad \Rightarrow \quad \frac{dy}{dx} = -\frac{2x}{2y} = -\frac{x}{y} \quad \text{if } y \neq 0.$$

Thus when $x = \frac{1}{\sqrt{3}}$ and $y = \frac{\sqrt{2}}{\sqrt{3}}$ we have

$$\frac{dy}{dx} = -\frac{\frac{1}{\sqrt{3}}}{\frac{\sqrt{2}}{\sqrt{3}}} = -\frac{1}{\sqrt{2}},$$

and so the tangent is

$$y - \frac{\sqrt{2}}{\sqrt{3}} = -\frac{1}{\sqrt{2}}(x - \frac{1}{\sqrt{3}}) \;\Leftrightarrow\; \sqrt{2}y + x = \sqrt{3}.$$

If we wanted to use the more traditional route then we would have to differentiate $f_1(x) = \sqrt{1 - x^2}$. We know how to differentiate the function $f(x) = 1 - x^2$ inside the square root ($f'(x) = -2x$ for all $x \in \mathbb{R}$), and so if we wanted to apply the chain rule to find $f_1'(x)$ then we need to know how to differentiate square roots. The following answers this problem:

Proposition 3.21. *For any rational number r the function $f(x) = x^r$ is differentiable on $(0, \infty)$ with $f'(x) = rx^{r-1}$ (cf. the case when r is an integer).*

Proof. Suppose we have $r = \frac{m}{n}$ for integers m and n, and let $y = x^r = (x^m)^{1/n}$. Raising both sides to the nth power gives

$$y^n = \left((x^m)^{1/n}\right)^n = (x^m)^1 = x^m.$$

Thus we can differentiate both sides, using the chain rule on the left hand side, to get

$$ny^{n-1} \times \frac{dy}{dx} = mx^{m-1}$$

Since $y > 0$ we can divide both sides by y^{n-1} to get

$$\frac{dy}{dx} = \frac{mx^{m-1}}{ny^{n-1}} = \frac{m}{n} \times \frac{x^{m-1}}{\left((x^{1/n})^m\right)^{n-1}} = rx^{m-1-\frac{m(n-1)}{n}}$$

and

$$m - 1 - \frac{m(n-1)}{n} = \frac{n(m-1) - m(n-1)}{n} = \frac{m-n}{n} = r - 1$$

as required. □

Recall that for each $x \in \mathbb{R}$ the number e^x is the unique solution y to the equation $x = \log y$. Differentiating this equation with respect to x and noting that $\frac{d}{dt} \log t = \frac{1}{t}$ we can arrive at the formula $\frac{d}{dx} e^x = e^x$ in a similar way to the proposition above. Indeed, using exponentials and logarithms we can actually define x^r for any $x > 0$ and *all* $r \in \mathbb{R}$, and establish the formula $\frac{d}{dx} x^r = r x^{r-1}$ by an alternative route that involves the chain rule.

Exercise 3.22 (A03 6(a)). Find the tangent to the implicitly defined curve

$$x \sin y + \cos x + e^{-\sin x} = e^{-1}$$

at the point $(\frac{\pi}{2}, \pi)$. Where does this tangent intersect the x-axis?

4 Curve Sketching and MinMax Problems

Maxima and minima

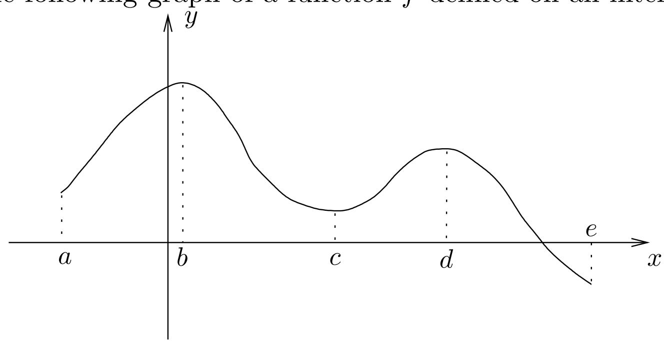
Definition 4.1. Let $f : \mathbb{R} \to \mathbb{R}$ be a function and $a, b \in \mathbb{R}$. Then f has **a local maximum at** a if there is some interval $(c_1, d_1) \subset \mathbb{R}$ such that $a \in (c_1, d_1)$ and

$$f(x) \leq f(a) \quad \text{for all } x \in (c_1, d_1).$$

Similarly, f has **a local minimum at** b if there is some interval $(c_2, d_2) \subset \mathbb{R}$ such that $b \in (c_2, d_2)$ and

$$f(x) \geq f(b) \quad \text{for all } x \in (c_2, d_2).$$

Note that we do not require that $f(a)$ be the largest value that the function ever takes, only that it is greater than $f(x)$ for all x (sufficiently) close to a. Similarly for $f(b)$. Consider the following graph of a function f defined on an interval $[a, e]$:



Here b is the **absolute** maximum (and also a local maximum), a and c are local minima, d a local maximum and e the **absolute minimum**.

The next result gives a method for locating local maxima and minima.

Proposition 4.2. *Let f be a continuous function on the interval $[c, d]$. If f attains its maximum value at some $x_1 \in (c, d)$ and is differentiable at this point, then $f'(x_1) = 0$. Similarly if f attains its minimum value at some $x_2 \in (c, d)$ and is differentiable there, then $f'(x_2) = 0$.*

Proof. This is essentially a rehash of the proof of Rolle's Theorem. Since $f(x_1)$ is the maximum value that f takes on this interval, $f(x_1 + h) \leq f(x_1)$ for all $h \neq 0$, so that

$$f(x_1 + h) - f(x_1) \leq 0 \quad \text{for all } h \neq 0.$$

But f is differentiable at x_1 so the left and right derivatives must exist, and they must be equal to $f'(x_1)$. Hence

$$\frac{f(x_1 + h) - f(x_1)}{h} \geq 0 \quad \text{for all} \quad h < 0$$

and so

$$f'(x_1) = f'_-(x_1) = \lim_{h \to 0-} \frac{f(x_1 + h) - f(x_1)}{h} \geq 0.$$

Similarly

$$f'(x_1) = f'_+(x_1) = \lim_{h \to 0+} \frac{f(x_1 + h) - f(x_1)}{h} \leq 0,$$

since now we require $h > 0$ in the limit. Thus we must have $f'(x_1) = 0$, as required. The proof for the point x_2 runs along the same lines. □

So to look for the local maxima and minima of a *differentiable* function f we should calculate its derivative f' and look for solutions of the equation $f'(x) = 0$, since if f has a local maximum or minimum at some point a then clearly we can restrict f to an interval of the form $[c, d]$ such that $a \in (c, d)$ and $f(a)$ is the absolute maximum or minimum for this restriction. The points where $f'(x) = 0$ are known as **critical** or **stationary points** of the function f.

Exercise 4.3. Examine the critical points of the function $f : [-3, 3] \to \mathbb{R}$ defined by $f(x) = x^3 - 3x$ and sketch its graph.

The second derivative test

In the previous example it was easy to determine which critical point was a local maximum and which a local minimum since the function was relatively simple. This may not always be the case and so we now develop a more methodical technique for analysing this question. Suppose $f : \mathbb{R} \to \mathbb{R}$ is a function that is differentiable everywhere on $\mathbb{R}$, then we have a new function $f' : \mathbb{R} \to \mathbb{R}$ defined by setting

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

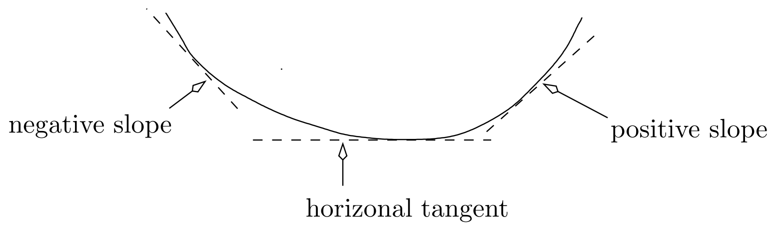
for each $x \in \mathbb{R}$. The number $f'(x)$ is a measure of the rate at which the function is changing at the point x — it is the slope of the tangent to the graph of f at this point. If it is positive then f is increasing and if it is negative then f is decreasing. Moreover if f has a local maximum or minimum at x then $f'(x) = 0$.

So now suppose that the function f' is itself differentiable. That is for each x the following limit exists:

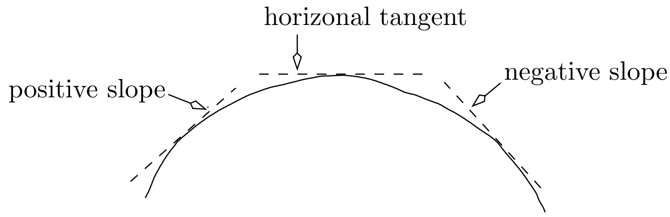
$$\lim_{h \to 0} \frac{f'(x+h) - f'(x)}{h}.$$

The resulting function is called the **second derivative** of f and is denoted f'', that is, $f'' = (f')'$. If we are writing $y = f(x)$ then the derivative and second derivative are often denoted $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$ respectively.

Suppose f and f' are differentiable (with f'' continuous) and that $f'(a) = 0$ and $f''(a) > 0$ for some $a \in \mathbb{R}$. Then the function $f'(x)$ is increasing at $x = a$, that is the gradient of f at $x = a$ is increasing. Since it is zero at $x = a$ it must be negative to the immediate left of a and positive to the immediate right. That is, the graph must take the form



and so this point is a local minimum. If, conversely, $f'(b) = 0$ and $f''(b) < 0$ then the gradient $f'(x)$ is decreasing around $x = b$. Thus it must be positive to the immediate left of $x = b$ and negative to the immediate right. That is, the graph must take the form



and so this point is a local maximum.

This is summarised in the following:

Theorem 4.4. *Suppose the function* $f : \mathbb{R} \to \mathbb{R}$ *is twice differentiable and that* $f'(a) = 0$ *for some* $a \in \mathbb{R}$.

(a) *If* $f''(a) < 0$ *then* $x = a$ *is a local maximum.*

(b) *If* $f''(a) > 0$ *then* $x = a$ *is a local minimum.*

(c) *If* $f''(a) = 0$ *then we have no information.*

Remarks. (i) A function f is **twice differentiable** if it is differentiable on $\mathbb{R}$ and if the resulting function $f' : \mathbb{R} \to \mathbb{R}$ is also differentiable. There are functions that are differentiable but not twice differentiable.

A function is **infinitely differentiable** if we can carry on repeatedly differentiating each subsequent derivative, i.e. f is differentiable, as is f', and so is f'' and so on. It is not hard to check that all polynomials are infinitely differentiable.

(ii) For part (c) consider the functions $f(x) = x^4$, $g(x) = -x^4$ and $h(x) = x^3$.

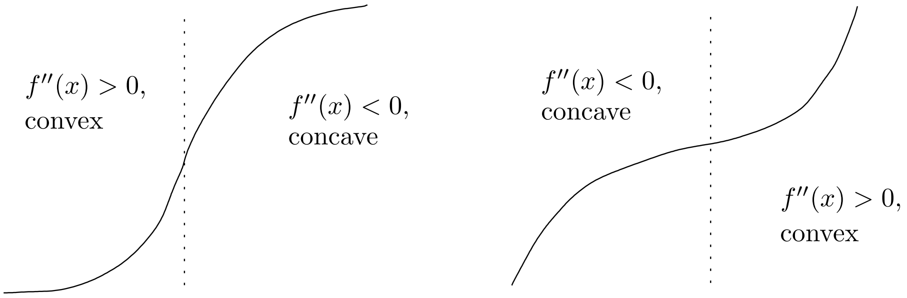
Definition 4.5. Let $f : \mathbb{R} \to \mathbb{R}$ be a twice differentiable function and let $a \in \mathbb{R}$. If there are $c, d \in \mathbb{R}$ satisfying $c < a < d$ such that $f''(a) = 0$ and either

(i) $f''(x) < 0$ for $x \in (c, a)$ and $f''(x) > 0$ for $x \in (a, d)$, or

(ii) $f''(x) > 0$ for $x \in (c, a)$ and $f''(x) < 0$ for $x \in (a, d)$

then a is called a **point of inflection** of f. That is, a is a point of inflection if f'' changes sign at that point.

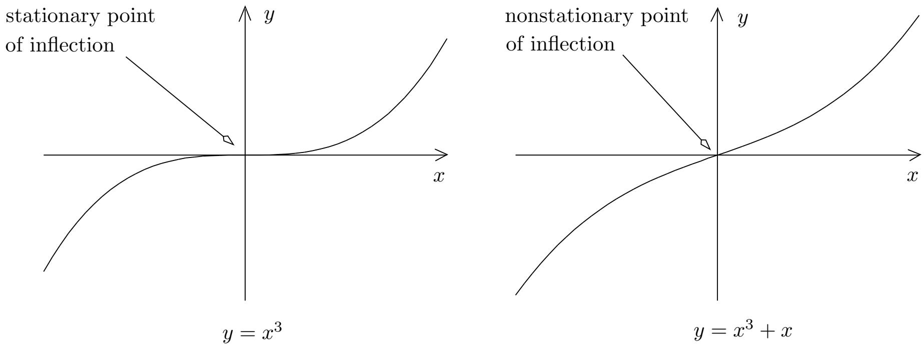
Note that if $f''(x) > 0$ on an interval, then $f'(x)$ is increasing on that interval and so the graph of f is **convex**. If, on the other hand, $f''(x) < 0$ on an interval then $f'(x)$ is decreasing on that interval and so the graph of f is **concave**. Thus at a point of inflection the graph of f changes from being convex to being concave, or vice versa. It is important to note that a need not be a critical point of f.



Example 4.6. Consider the functions $f(x) = x^3$ and $g(x) = x^3 + x$ on $\mathbb{R}$. Both are polynomials and so infinitely differentiable, with

$$f'(x) = 3x^2 \geq 0,\ f''(x) = 6x, \quad \text{and} \quad g'(x) = 3x^2 + 1 > 0,\ g''(x) = 6x.$$

Thus f has a critical point at $x = 0$, but g has no critical points since $3x^2 + 1 \geq 1$ for all $x \in \mathbb{R}$. However we have $f''(0) = g''(0) = 0$, $f''(x) = g''(x) < 0$ when $x < 0$ and $f''(x) = g''(x) > 0$ when $x > 0$. So for both of these functions we have that $x = 0$ is a point of inflection since the second derivatives change sign. Note that the only solution to $f(x) = 0$ is $x = 0$, which is also the case for $g(x) = 0$. Their respective graphs are thus



Exercise 4.7 (S03 5(b)). Consider the function

$$y = (x-1)^2(x+2)^2.$$

Find the critical points of y and determine their nature. Where does the graph of y meet the x-axis and the y-axis?

Use this information to sketch the graph of y.

Example 4.8. Consider the function $f(x) = x^2(x^2 + 4)$. Find its critical points and determine their nature. Find any other points of inflection. Find the sets on which it is strictly increasing and decreasing. Describe the behaviour of the function as $x \to +\infty$ and $x \to -\infty$

Use all of this information to draw a sketch of the function.

Solution. By the product rule

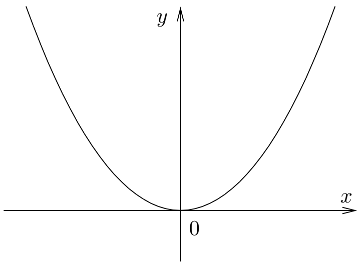
$$f'(x) = 2x(x^2 + 4) + x^2 \times 2x = 2x(2x^2 + 4) = 4x(x^2 + 2)$$

which is equal to zero precisely when $x = 0$ since $x^2 + 2 \geq 2$ for all $x \in \mathbb{R}$. Also,

$$f''(x) = 4(x^2 + 2) + 4x \times 2x = 12x^2 + 8$$

which is greater than zero for all x. In particular $f''(0) = 8 > 0$, and so there is a minimum when $x = 0$ and $f(0) = 0$. Also, there are no points of inflection. Moreover, because $x^2 + 2 > 0$ for all $x \in \mathbb{R}$, f is strictly increasing on $(0, \infty)$ and strictly decreasing on $(-\infty, 0)$.

Since $f(x) = x^4(1 + \frac{4}{x^2})$, we have that $f(x) \approx x^4$ for large x, since $\frac{4}{x^2} \to 0$ as $x \to \pm\infty$. Thus $f(x) \to +\infty$ as $x \to +\infty$ and as $x \to -\infty$. Hence the graph of f is of the form:



Exercise 4.9 (S02 5(b)). For the function

$$g(x) = \frac{2x}{3x^2 + 1}$$

find the critical points and their nature, find the sets on which g is strictly increasing and strictly decreasing, and find the points of inflection. What happens as $x \to +\infty$ and $x \to -\infty$?

Use this information to sketch the graph of g.

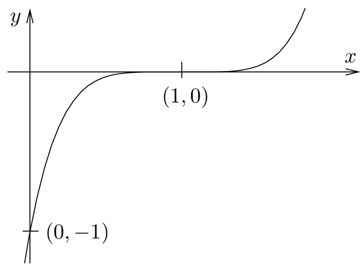
Example 4.10. Consider the function $f(x) = (x - 1)^5$. Find its critical points and determine their nature. Find any other points of inflection. Find the sets on which it is strictly increasing and decreasing. Describe the behaviour of the function as $x \to +\infty$ and $x \to -\infty$

Use all of this information to draw a sketch of the function.

Solution. Applying the chain rule twice we get

$$f'(x) = 5(x - 1)^4 \quad \text{and} \quad f''(x) = 20(x - 1)^3.$$

Thus $f'(x) = 0$ precisely when $x = 1$, and f'' changes sign at this point. For all other x we have that $f'(x) > 0$, so that f is strictly increasing on $(-\infty, 1)$ and $(1, \infty)$, and there is one critical point, but it is a point of inflection rather than being a local maximum or minimum. Note that $f(x) = 0$ exactly when $x = 1$. Finally, we have $f(x) = [x(1 - \frac{1}{x})]^5 = x^5(1 - \frac{1}{x})^5$, and since $\frac{1}{x} \to 0$ as $x \to \pm\infty$, we have that $(1 - \frac{1}{x})^5 \to 1$, and so $f(x) \approx x^5$ for large x. Thus $f(x) \to +\infty$ as $x \to +\infty$ and $f(x) \to -\infty$ as $x \to -\infty$. Hence the graph of f is of the form:



Applied maximum and minimum problems

The techniques we have learned in this course can be used to solve problems that involve some sort of optimisation. The basic strategy for such a problem is the following:

Step 1: Identify the quantity to be maximised/minimised, and all of the variables it depends on.

Step 2: Use (hopefully obvious) constraints to eliminate all but one of the variables, and determine the range of values this variable should take to be meaningful.

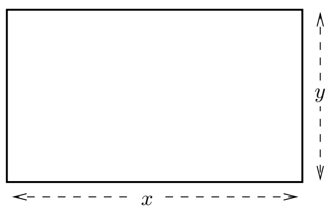
Step 3: Find the absolute maximum and/or minimum by differentiating with respect to the one variable, making sure to check the values of the function at the end points (if these exist).

Exercise 4.11. If we want to make a cylindrical tin can to hold 250cm^3 of baked beans, what radius and height will minimise the cost of materials?

Example 4.12. Show that of all the rectangles of a given perimeter, the one with the greatest area is a square.

Solution. Let the perimeter of the rectangle be l, and suppose that the sides have lengths x and y. So then if A denotes the area of our rectangle we have

$$l = 2x + 2y \quad \text{and} \quad A = xy$$



We can use the first equation rewrite A as a function of the one variable x, since

$$y = \frac{l}{2} - x. \tag{†}$$

So then

$$A(x) = x\Big(\frac{l}{2} - x\Big) = \frac{lx}{2} - x^2,$$

and we must have $0 \le x \le l/2$ to ensure that the lengths x and y are both nonnegative. Differentiating with respect to x gives

$$A'(x) = \frac{l}{2} - 2x$$

which is equal to 0 when $x = l/4$. That is, there is a stationary point when $x = l/4$. But since $A(0) = 0$, $A(l/4) = l^2/8 - l^2/16 = l^2/8 > 0$ and $A(l/2) = 0$ we see that the maximum area occurs when $x = l/4$. When this happens we have from (†) that

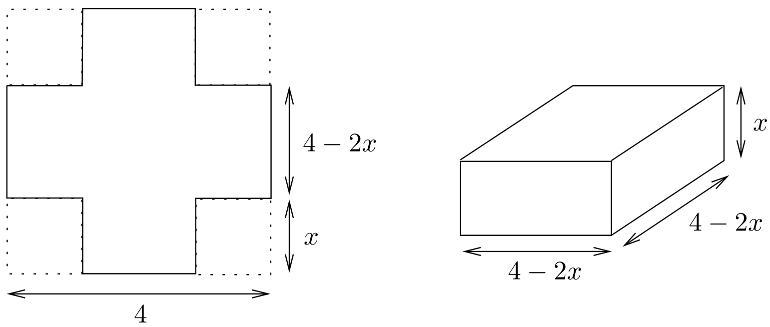
$$y = \frac{l}{2} - \frac{l}{4} = \frac{l}{4} = x$$

and so our rectangle is in fact a square as required.

Exercise 4.13. A woman at a point A on the shore of a circular lake with radius 2km wants to arrive at the point C diametrically opposite A on the other side of the lake in the shortest possible time. She can walk at the rate of 4km/h, and row a boat at 2km/h. How should she proceed?

Example 4.14. A box with an open top is to be constructed from a square piece of cardboard 4m wide by cutting out a square from each of the four corners and bending up the sides. Find the largest volume such a box can have.

Solution. The diagram for this situation is:



where x is the length of the side of the square removed from each corner. Since the total length is 4m, the length of a side of the base is $4 - 2 \times x = 2(2 - x)$m. Thus the area of the base will be $[2(2-x)]^2 = 4(2-x)^2\,\text{m}^2$. Moreover, the height of the box will be $x\,$m, and so the overall volume is

$$V(x) = x \times 4(2-x)^2 = 4x(2-x)^2$$

and note that we must have $x \ge 0$, and $4 - 2x \ge 0$, so that we must only consider $x \in [0, 2]$. But differentiating with respect to x gives

$$\begin{aligned} V'(x) &= 4 \times (2-x)^2 + 4x \times 2(2-x) \times (-1) = 4(2-x)\big[(2-x) - 2x\big] \\ &= 4(2-x)(2-3x) \end{aligned}$$

which is zero precisely when $x = 2$ and $x = \frac{2}{3}$. The maximum value of V will occur at either one of these values, or when $x = 0$, which is the other endpoint for the interval of values x may assume. But

$$V(0) = V(2) = 0, \qquad \text{and} \qquad V\Big(\frac{2}{3}\Big) = 4 \times \frac{2}{3} \times \Big(\frac{4}{3}\Big)^2 = \frac{128}{27}.$$

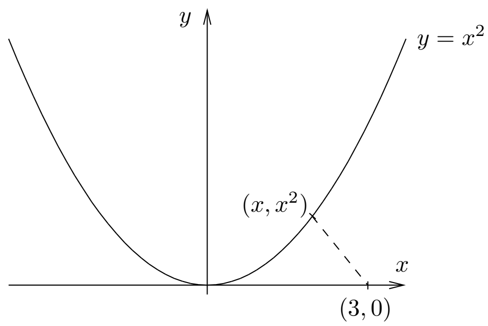
Thus the largest volume is $\frac{3}{2}\text{m}^3$.

Exercise 4.15 (S02 6(b)). A Norman window has the shape of a rectangle surmounted by a semicircle (Thus the diameter of the semicircle is equal to the width of the rectangle). If the perimeter of the window is 10m, find the dimensions of the window so that the greatest possible amount of light is admitted.

What is the area of the window when this achieved?

Example 4.16. Find the point on the parabola $y = x^2$ that is closest to the point $(3, 0)$.

Solution. Consider the following diagram representing this situation:



If d denotes the distance of the point $(3, 0)$ from a general point on the parabola (x, x^2) then we have

$$d(x) = \sqrt{(x-3)^2 + (x^2-0)^2}$$

by Pythagoras' Theorem. We wish to minimise the value of d as x varies, but the minimum value of d will occur for the same value of x as the minimum value of the function $D = d^2$, and it is easier to minimise D since it does not involve square roots. So we have

$$D(x) = (x-3)^2 + x^4 = x^4 + x^2 - 6x + 9$$

and hence

$$D'(x) = 4x^3 + 2x - 6 = 2(2x^3 + x - 3).$$

By inspection we see that $x = 1$ is a root of $D'(x) = 0$, and in fact

$$D'(x) = 2(x-1)(2x^2 + 2x + 3).$$

But $2x^2 + 2x + 3 = 2(x + \frac{1}{2})^2 + \frac{5}{2} > 0$ for all x, and so $D'(x) = 0$ only when $x = 1$. Also,

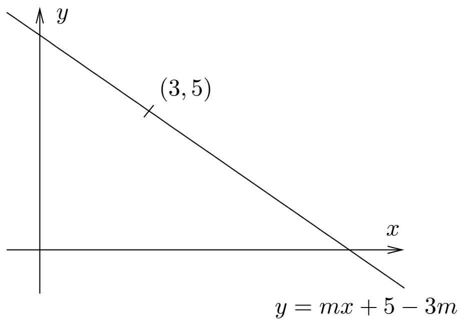
$$D''(x) = 12x^2 + 2 > 0$$

for all x, so that we have a local minimum when $x = 1$. Thus the point on $y = x^2$ that is closest to $(3, 0)$ is when $x = 1$ and so $y = 1^2 = 1$ as well. Thus the required point is $(1, 1)$.

Exercise 4.17. Find the area of the largest rectangle that can be inscribed in a semicircle of radius r, with one side of the rectangle on the straight side of the semicircle.

Example 4.18. Find the equation of the line through the point $(3, 5)$ that cuts off the least area from the first quadrant.

Solution. Let the gradient of the line through $(3,5)$ be m, so the set up looks something like:



The line has equation

$$y - 5 = m(x-3) \;\Rightarrow\; y = mx + 5 - 3m.$$

When $x = 0$ we have $y = 5 - 3m$ and when $y = 0$ we have $x = (3m-5)/m$. Note that we must have $m < 0$ for the line to actually cut off a triangle, whose area is thus

$$A(m) = \frac{1}{2} \times (5-3m) \times \frac{3m-5}{m} = \frac{1}{2}\left(30 - 9m - \frac{25}{m}\right).$$

Differentiating this function of m we get

$$A'(m) = \frac{1}{2}\left(-9 + \frac{25}{m^2}\right)$$

which is equal to zero when

$$m^2 = \frac{25}{9} \;\Rightarrow\; m = \pm\frac{5}{3}.$$

But recall that we need $m < 0$, and so the critical point of interest occurs when $m = -5/3$. Note also that

$$A''(m) = -\frac{25}{m^3}$$

which is greater than zero for all $m < 0$. Thus we do have a minimum when $m = -5/3$. Hence the equation of the line that cuts off the least area is

$$y = -\frac{5}{3}x + 5 + 5 \;\Leftrightarrow\; 3y = 30 - 5x.$$