5 Polynomials and rational functions

5.1 Polynomials and rational functions

A **polynomial** is a function of the form

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

where the numbers $a_0, a_1, \ldots, a_n$ are constants (called the **coefficients** of the polynomial). *Note: In this class we will only discuss polynomials with real coefficients.* The **degree** of the polynomial is the highest power of x present. So (if $a_n \neq 0$) $P_n(x)$ above has degree n.

Some examples of simple polynomials are

constant	$P_0(x) = a_0$	degree zero
linear	$P_1(x) = a_1\, x + a_0$	degree one
quadratic	$P_2(x) = a_2\, x^2 + a_1 x + a_0$	degree two
cubic	$P_3(x) = a_3\, x^3 + a_2 x^2 + a_1 x + a_0$	degree three
quartic	$P_4(x) = a_4\, x^4 + a_3\, x^3 + a_2 x^2 + a_1 x + a_0$	degree four
quintic	$P_5(x) = a_5\, x^5 + a_4\, x^4 + a_3\, x^3 + a_2 x^2 + a_1 x + a_0$	degree five

For a particular value of x, $x = \alpha$ say, the value of $P_n(x)$ can be evaluated as

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

Examples 5A

Consider two polynomials, $P^1_m(x)$ and $P^2_n(x)$ say, of degree m and n, respectively. If we add, subtract or multiply $P^1_m(x)$ and $P^2_n(x)$, the result is another polynomial. Addition and subtraction are straightforward: we simply add or subtract terms involving the same powers of x. Multiplication of $P^1_m(x)$ and $P^2_n(x)$ gives a polynomial of degree $m + n$. This can be done by using repeated applications of the distributive law (P9) or, more efficiently, in a format similar to multiplication of large numbers.

Dividing one polynomial by another does NOT generally produce another polynomial. Ratios of polynomials, that is, functions of the form

$$\frac{P^1_m(x)}{P^2_n(x)}$$

are called **rational functions**. Note that this expression is only meaningful as long as the denominator is not zero. A rational function is called **proper** if the degree of the polynomial in the numerator is less than the degree of the polynomial in the denominator (i.e. $m < n$ here); otherwise it is called **improper**.

Using long division of polynomials, an improper rational function (i.e. when $m \geq n$) can always be written as

$$\frac{P_m^1(x)}{P_n^2(x)} = Q_{m-n}(x) + \frac{R_k(x)}{P_n^2(x)} \tag{5.1}$$

where $Q_{m-n}(x)$ (called the **quotient**) is a polynomial of degree $m-n$ and $R_k(x)$ (called the **remainder**) is a polynomial of degree k which is strictly less than n. This can be done in a similar format to long division of numbers.

Examples 5B

5.2 Solving polynomial equations

An **equation** is a statement that asserts that two quantities are equal. The quantities are written either side of an **equals sign** $=$. Such equations often involve some unknown quantity, which we must find by **solving** the equation. In this section, we will discuss the solution of **polynomial equations** of the form

$$P(x) = 0 \tag{5.2}$$

where $P(x)$ is a polynomial.

Any value of x which satisfies a polynomial equation (5.2) is called a **solution** or a **root** of the equation, and is a member of the **solution set**. Note that using the term root can be confusing, as we also often use it to denote something different, namely the **root of a real number**. For equation (5.2), a root x is any value which makes the equation true (regardless of the details of the equation). In the second case, we mean something very specific: if x is the n^{th} root of y, then x satisfies the particular equation $y = x^n$.

5.2.1 Linear equations

A **linear equation** is of the form $ax + b = 0$, where a is the coefficient of x and b is the constant term. Such an equation can easily be solved by using the properties (P1)-(P9) to rearrange the equation so that the unknown quantity appears on its own on the left-hand side. Linear equations are also easy to solve graphically.

Important note: do not confuse **linear polynomials** with **linear functions**. We have already seen in §4.6 that every linear function on $\mathbb{R}$ is of the form $f(x) = ax$ for $a \in \mathbb{R}$. A linear polynomial usually also contains a constant term (represented by b here).

Examples 5C

5.2.2 Quadratic equations

A **quadratic equation** is of the form $ax^2+bx+c = 0$ where a, b and c are given coefficients and x is the unknown. Note that b and c can be zero, but a cannot (if it was, we would no longer have a quadratic equation!).

To begin studying the solution of quadratics, we consider the expression $(x + m)(x + n)$. This can be multiplied out to give

$$(x + m)(x + n) = x^2 + (m + n)x + mn,$$

that is, it represents a quadratic polynomial with the coefficient of x given by $m + n$ and a constant term given by mn. So we could replace any quadratic equation of the form

$$x^2 + (m + n)x + mn = 0$$

by the new quadratic equation

$$(x + m)(x + n) = 0.$$

From this we deduce that either $x + m = 0$ or $x + n = 0$, that is, we have reduced the problem of solving the quadratic equation to one of solving two linear equations (which we know how to solve using (P1)-(P9)), giving the solution set $\{-m, -n\}$ here. This suggests a method of solution for quadratic equations, namely trying to write the quadratic polynomial as a product of **linear factors**. The solution is then easily found from the factored form by solving two linear equations.

Examples 5D

Unfortunately, not all quadratic polynomials can be factorised into linear factors involving real numbers in this way. For those which we cannot factorise, we instead seek to rewrite the equation in a form such that the left-hand side is a **perfect square**, that is, it contains only a term of the form X^2 where X depends on x. The equation can then be solved by taking the square root of both sides, with the positive and negative square roots giving the two possible solutions.

The technique used to convert a general quadratic polynomial into an expression involving a perfect square is known as **completing the square**. We write

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

The quadratic equation $ax^2 + bx + c = 0$ can then be solved as follows:

$$
\begin{aligned}
& ax^2 + bx + c = 0 \\
\Leftrightarrow\quad & a\left(x + \frac{b}{2a}\right)^2 + c - \frac{b^2}{4a} = 0 \\
\Leftrightarrow\quad & a\left(x + \frac{b}{2a}\right)^2 = \frac{b^2}{4a} - c \\
\Leftrightarrow\quad & \left(x + \frac{b}{2a}\right)^2 = \frac{1}{a}\left(\frac{b^2}{4a} - c\right) \\
\Leftrightarrow\quad & \left(x + \frac{b}{2a}\right)^2 = \left(\frac{b^2 - 4ac}{4a^2}\right) \\
\Leftrightarrow\quad & x + \frac{b}{2a} = \pm\frac{\sqrt{b^2 - 4ac}}{2a} \\
\Leftrightarrow\quad & x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.
\end{aligned} \tag{5.3}
$$

Expression (5.3) is often referred to as the **quadratic formula**. The number $b^2 - 4ac$ (under the square root sign) is called the **discriminant** of the quadratic. The type of solutions which the quadratic has is characterised by the discriminant:

- if $b^2 - 4ac > 0$, the equation has two real solutions;
- if $b^2 - 4ac = 0$, the equation has two equal real solutions (i.e. one repeated solution);
- if $b^2 - 4ac < 0$, the equation has two complex solutions.

Examples 5E

5.2.3 Solving higher degree polynomial equations

Finding the solutions of a polynomial equation $P_n(x) = 0$ where $n \geq 3$ is much harder than for the linear and quadratic cases. In this section we state some important related results which can help in cases where we have a real polynomial with real roots. Note that the values of x which satisfy $P_n(x) = 0$ are also often called the **zeros** of $P_n(x)$.

Theorem 5.1 ***The Remainder Theorem*** *When a polynomial $P_n(x)$ is divided by the linear polynomial $x - \alpha$, the remainder $R_k(x)$ is a constant function and is equal to $P_n(\alpha)$.*

Proof 5.1 *From equation (5.1) we have*

$$
\frac{P_n(x)}{x - \alpha} = Q_{n-1}(x) + \frac{R_k(x)}{x - \alpha}
$$

where $R_k(x)$ is a constant function (as its degree must be strictly less than the degree of the denominator of the rational function, which is one in this case). In other words, $k = 0$ here so we may write

$$\frac{P_n(x)}{x-\alpha} = Q_{n-1}(x) + \frac{R_0}{x-\alpha}$$

where R_0 is constant. Multiplying both sides by $x - \alpha$ gives

$$P_n(x) = (x-\alpha)Q_{n-1}(x) + R_0. \tag{5.4}$$

Inserting $x = \alpha$ in this equation then gives $R_0 = P_n(\alpha)$, which completes the proof.

Theorem 5.2 ***The Factor Theorem***

$$P_n(\alpha) = 0 \Leftrightarrow x - \alpha \textit{ is a factor of } P_n(x).$$

Proof 5.2 *($\Rightarrow$) By the Remainder Theorem, the remainder on dividing $P_n(x)$ by $x - \alpha$ is $P_n(\alpha)$. This remainder is zero if and only if α is a zero of $P_n(x)$. If this is the case, R_0 in (5.4) is zero, so $P_n(x) = (x-\alpha)Q_{n-1}(x)$ for some polynomial $Q_{n-1}(x)$ of degree $n-1$, that is, $x - \alpha$ is a factor of $P_n(x)$.*

($\Leftarrow$) If $x-\alpha$ is a factor of $P_n(x)$ then $P_n(x) = (x-\alpha)Q_{n-1}(x)$ where $Q_{n-1}(x)$ is a polynomial of degree $n-1$ so $P_n(\alpha) = 0$.

These two theorems show that solving the polynomial equation $P_n(x) = 0$ is intimately related to finding linear factors of $P_n(x)$ (exactly as we have already seen for quadratics). Specifically, the theorems tell us that if we can factorise the polynomial

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

into linear factors as

$$P_n(x) = a_n(x-\alpha_1)(x-\alpha_2)\ldots(x-\alpha_n), \tag{5.5}$$

then $\alpha_1,\ \alpha_2,\ \ldots,\ \alpha_n$ are the zeros of $P_n(x)$, that is, $P_n(\alpha_1) = P_n(\alpha_2) = \ldots = P_n(\alpha_n) = 0$.

Examples 5F

Although we now have a connection between solving $P_n(x) = 0$ and finding the linear factors of $P_n(x)$, it is still the case that factorising general polynomials is difficult. In addition, as we have already seen with quadratic equations, some polynomial equations do not have any (real) solutions, and therefore will not have a factorisation of the form (5.5). For example, the graph of the polynomial $p(x) = x^2 + 1$ never crosses the $x-$axis so there are no real values of x such that $p(x) = 0$. We say that such polynomials are **irreducible**. In general, finding linear factorisations of polynomials is only feasible for a limited range of low degree polynomials (such as some quadratics and cubics).

[*Historical note:* We have seen that, independent of whether or not the quadratic polynomial $p(x) = ax^2 + bx + c$ can be factorised, it the solution set

$$\left\{ \frac{-b + \sqrt{b^2 - 4ac}}{2a}, \frac{-b - \sqrt{b^2 - 4ac}}{2a} \right\}.$$

Similar formulae (i.e. expressions combining only the polynomial coefficients) can be constructed for solutions of cubic and quartic equations, although they are much more complicated. However, in 1830, the French mathematician Evariste Galois proved that there cannot be a formula for the solutions of a general polynomial equation of degree ≥ 5.]

We next look at a method of factorising a polynomial P_n of degree n which satisfies the following two particular conditions:

(i) a_0, a_1,...,$a_{n-1} \in \mathbb{Z}$ (i.e. all of the coefficients are integers);
(ii) $a_n = 1$ (i.e. the coefficient of the x^n term is one).

In this special case, a method of factorising is suggested by a corollary to the following theorem.

Theorem 5.3 *A polynomial*

$$P_n(x) = a_0 + a_1\, x + a_2\, x^2 + \ldots + a_{n-1}\, x^{n-1} + x^n$$

where a_0, a_1,...,$a_{n-1} \in \mathbb{Z}$ can have only integer or irrational roots.

Proof 5.3 *Assume that P_n has a rational zero $\alpha = \dfrac{p}{q}$ where p and q have no common factors and q is positive. Then*

$$\begin{aligned} a_0 + a_1\,\alpha + a_2\,\alpha^2 + \ldots + a_{n-1}\,\alpha^{n-1} + \alpha^n &= 0 \\ \Leftrightarrow a_0 + a_1\,\frac{p}{q} + a_2\left(\frac{p}{q}\right)^2 + \ldots + a_{n-1}\left(\frac{p}{q}\right)^{n-1} + \left(\frac{p}{q}\right)^n &= 0 \\ \Leftrightarrow a_0 + a_1\,\frac{p}{q} + a_2\left(\frac{p}{q}\right)^2 + \ldots + a_{n-1}\left(\frac{p}{q}\right)^{n-1} &= -\left(\frac{p}{q}\right)^n \\ \Leftrightarrow a_0q^{n-1} + a_1\,pq^{n-2} + a_2\,p^2q^{n-3} + \ldots + a_{n-1}\,p^{n-1} &= -\frac{p^n}{q}. \end{aligned}$$

The LHS of this equation is an integer (as the coefficients are integers) and the RHS is a fraction (as p and q have no common factors) so the only way that this equation can make sense is if $q = 1$. So the zero of P_n is $\alpha = p$ (which is an integer).

Corollary 5.4 *Any integer zero of P_n in Theorem 5.3 must be a divisor of a_0.*

Proof 5.4 *If α is a zero of P_n, then*

$$\begin{aligned} a_0 + a_1\,\alpha + a_2\,\alpha^2 + \ldots + a_{n-1}\,\alpha^{n-1} + \alpha^n &= 0 \\ \Leftrightarrow \alpha(a_1 + a_2\,\alpha + \ldots + a_{n-1}\,\alpha^{n-2} + \alpha^{n-1}) &= -a_0 \end{aligned}$$

so α must be a divisor of a_0.

This corollary suggests the following method for factorising polynomials of this type.

1. Find a value for one of the zeros, α say, by trial and error. Thanks to the above corollary, this is not as hard as it sounds, as α must be a divisor of the constant term in the polynomial (with a plus or minus sign).
2. Divide the polynomial by the associated linear factor $x - \alpha$, which will reduce the degree of the polynomial by one.
3. Repeat this procedure can until a complete factorisation is obtained, or until what remains is irreducible.

Note: Although we could in principle find all of the zeros by trial and error, there may sometimes be too many divisors to check!

The following are some standard factorisation results which can be verified by expansion and/or long division.

$$\begin{aligned} x^2 - y^2 &= (x+y)(x-y) \\ x^2 + y^2 & \quad \text{has no real linear factors} \\ x^3 - y^3 &= (x-y)(x^2 + xy + y^2) \\ x^3 + y^3 &= (x+y)(x^2 - xy + y^2) \\ x^4 - y^4 &= (x^2+y^2)(x^2-y^2) = (x^2+y^2)(x+y)(x-y) \\ x^n - y^n &= (x-y)(x^{n-1} + x^{n-2}y + x^{n-3}y^2 + \ldots + xy^{n-2} + y^{n-1}), \quad \text{for any integer } n \geq 3. \end{aligned}$$

Examples 5G

We finish this section by proving one more useful theorem about general polynomial functions. First we need a preliminary result to help in the proof.

Lemma 5.5 *Let*

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

be a polynomial of degree $\leq n$. *If there are more than* n *different values of* x *such that* $P_n(x) = 0$, *then* $P_n(x) = 0$ *for all values of* x. *In other words, all of the coefficients of* P_n *must be zero.*

Proof 5.5 *Suppose that* $P_n(x) = 0$ *for the* n *different values* $x = \alpha_1, \alpha_2, \ldots, \alpha_n$ *so that we can write*

$$P_n(x) = a_n(x - \alpha_1)(x - \alpha_2) \ldots (x - \alpha_n). \tag{5.6}$$

Now suppose that $x = c$ *is also a zero of* P_n, *where* c *is different from any* α_r. *We know that* $P_n(c) = 0$, *so*

$$a_n\, (c - \alpha_1)(c - \alpha_2) \ldots (c - \alpha_n) = 0.$$

Each term in brackets on the left-hand side is non-zero, hence we must have $a_n = 0$. *So from (5.6), we see that as* $a_n = 0$, *we must have* $P_n(x) = 0$ *for any value of* x.

Note: If $P_n(x) = 0$ for all values of x, then $P_n(x)$ is sometimes called the **null polynomial**.

Theorem 5.6 *If two polynomials*

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

and

$$Q_n(x) = b_0 \ + b_1\, x + b_2\, x^2 + \ldots + b_{n-1}\, x^{n-1} + b_n\, x^n$$

of degree $\leq n$ *are equal in value for more than* n *different values of* x, *then they must be equal for all values of* x *and the corresponding coefficients must be equal (so* $a_j = b_j$ *for all* $j = 0, 1, \ldots, n$*).*

Proof 5.6 *Consider the polynomial* $d(x) = P_n(x) - Q_n(x)$. *From the assumptions in the theorem we know that* $d(x)$ *has degree* $\leq n$ *and vanishes for more than* n *different values of* x. *Hence, from Lemma 5.6,* $d(x)$ *must be zero for all values of* x, *so* $P_n(x) = Q_n(x)$ *for all values of* x.

One important application of this theorem is that it shows that if we have $n + 1$ points in space, we can uniquely fit them with a polynomial of degree n. In other words,

- two points uniquely define a linear polynomial (straight line);

- three points uniquely define a quadratic polynomial;
- four points uniquely define a cubic polynomial;

etc.

Examples 5H

5.3 Inequalities

We have now seen some techniques for solving polynomial equations. In this section, we will focus on solving polynomial **inequalities**, that is, expressions involving $<$, $\leq$, $>$ or $\geq$ (instead of $=$). The concept of solving an inequality means finding the **range of values** of the unknown variable for which the inequality holds. Note that expressions involving $>$ and $<$ are sometimes called **strict** inequalities.

The following key facts about inequalities can be deduced using properties (P1)-(P12):

1. If $a < b$ then $b > a$.
2. If $a < b$ then $a + c < b + c$.
3. If $a < b$ then $a - c < b - c$.
4. If $c > 0$ then $a < b \iff ac < bc$.
5. If $c < 0$ then $a < b \iff ac > bc$.
6. If $a > 0$ then $\dfrac{1}{a} > 0$.
7. If $0 < a < b$ then $\dfrac{1}{b} < \dfrac{1}{a}$.

Note in particular rule 5: **multiplying by a negative quantity reverses the inequality**.

Like linear equations, linear inequalities can be solved by rearranging so that the unknown quantity appears on its own on the left-hand side.

Examples 5I

For inequalities involving higher degree polynomials, we can factorise and use the number line to look at the sign of each linear factor. In general, this involves a three-step procedure:

1. Rearrange the inequality to get zero on the right-hand side: this reduces the problem to one of deciding whether the rearranged expression on the left-hand side is positive (> 0) or negative (< 0).
2. Simplify/factorise the expression on the left-hand side.
3. Examine the signs of the linear factors and work out the sign of the complete expression.

Examples 5J

5.4 Equations and inequalities involving the modulus function

Consider the equation

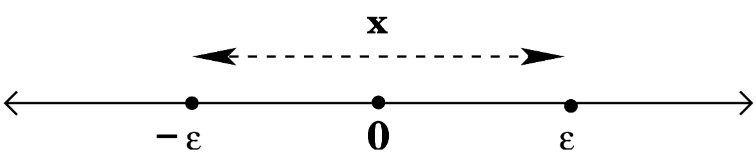
$$|x| = \epsilon \tag{5.7}$$

for some given number ϵ. There are three possibilities:

- if $\epsilon < 0$, (5.7) has no solutions (recall that the modulus of a number must be positive);
- if $\epsilon = 0$, (5.7) has solution set $\{0\}$;
- if $\epsilon > 0$, (5.7) has solution set $\{\epsilon, -\epsilon\}$ (that is, the two points that lie a distance ϵ from the origin on the real line).

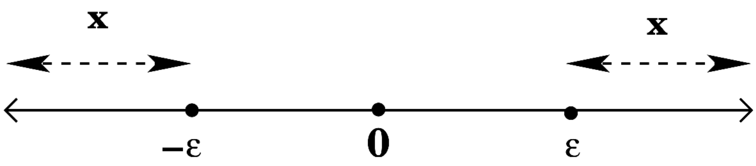
For similar expressions involving inequalities, it is often useful to interpret the absolute value as a distance. For example, assuming $\epsilon > 0$, the inequality $|x| < \epsilon$ means that the distance from x to 0 is less than ϵ, so x must lie between $-\epsilon$ and ϵ. That is,

$$|x| < \epsilon \quad \Leftrightarrow \quad x > -\epsilon \qquad \textbf{and} \qquad x < \epsilon.$$



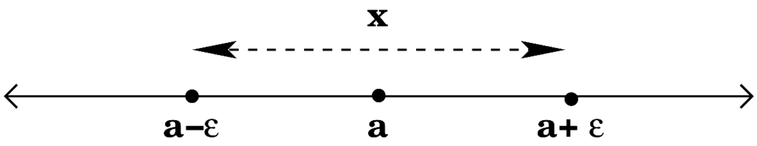
Similarly, $|x| > \epsilon$ means that the distance from x to 0 is greater than ϵ, so either x is less than $-\epsilon$ or x is greater than ϵ. That is,

$$|x| > \epsilon \quad \Leftrightarrow \quad x < -\epsilon \qquad \textbf{or} \qquad x > \epsilon.$$



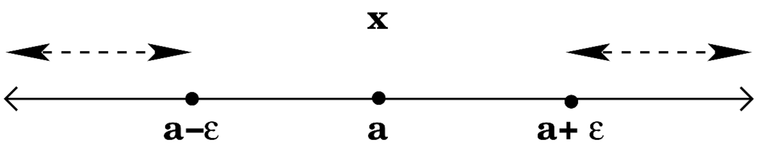
These results generalise to the inequalities $|x-a| < \epsilon$, $\quad |x-a| > \epsilon$ in an obvious way. Specifically, $|x-a| < \epsilon$ means that the distance from x to a is less than ϵ, so x must lie between $a-\epsilon$ and $a+\epsilon$. That is,

$$|x-a| < \epsilon \quad \Leftrightarrow \quad a-\epsilon < x \qquad \textbf{and} \qquad x < a+\epsilon.$$



Similarly, $|x-a| > \epsilon$ means that the distance from x to a is greater than ϵ so

$$|x-a| > \epsilon \quad \Leftrightarrow \quad x < a-\epsilon \qquad \textbf{or} \qquad x > a+\epsilon.$$



To solve linear inequalities of this type, we use exactly this approach and rewrite the version with the modulus sign as two simpler inequalities:

$$\begin{aligned} |x-a| < \epsilon &\quad \Leftrightarrow \quad x > a-\epsilon \textbf{ and } x < a+\epsilon; \\ |x-a| > \epsilon &\quad \Leftrightarrow \quad x < a-\epsilon \textbf{ or } x > a+\epsilon. \end{aligned}$$

An equivalent technique works for expressions involving $\geq$ and $\leq$.

Examples 5K

5.5 Simultaneous equations in two variables

Many mathematical applications involve solving more than one equation at a time. For n linear equations in n unknowns, a systematic (but tedious!) approach can always be used. For example, for the two linear equations in two unknowns x and y

$$\begin{aligned} ax+by &= p \\ cx+dy &= q, \end{aligned}$$

(where a, b, c, d, p and q are known constants) we can add (or subtract) multiples of the two equations to eliminate one of the variables. This method can be extended to larger systems of linear equations in an efficient way, as you will see next year.

Unfortunately, there is no systematic way to solve systems of **nonlinear** equations. Indeed, solving even one nonlinear equation in one unknown can be tricky (e.g. $e^x = x$). For the two nonlinear equations in two unknowns x and y

$$\begin{aligned} f(x, y) &= 0 \\ g(x, y) &= 0, \end{aligned}$$

techniques to try are

- solve one equation for y in terms of x (or vice-versa) and substitute into the other;
- factorise $f(x, y)$ and/or $g(x, y)$ and set the factors equal to zero.

Note that in general there is no way of telling how many solutions to expect, so you must be careful to write down all possible solutions to any equation (e.g. the equation $x^2 = 4$ has solution set $\{2, -2\}$).

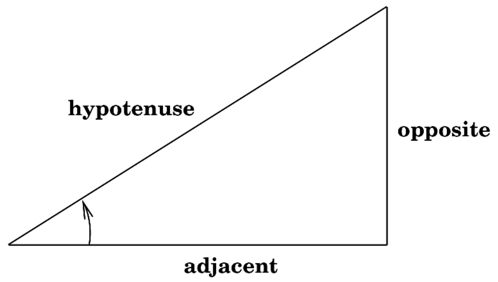
Examples 5L

6 Trigonometric functions

6.1 The sine and cosine functions

Defining the functions **sin** and **cos** is a more subtle procedure than might be expected. We will deal with some basic points here, and revisit these functions in the second half of the course.

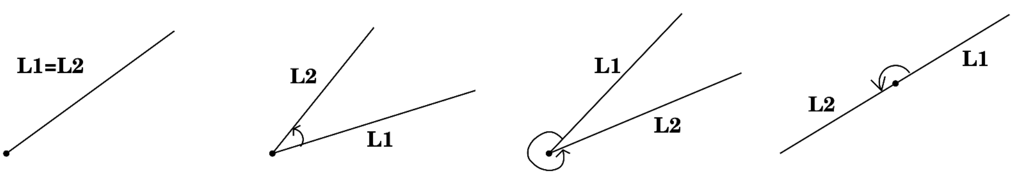
For an acute angle, trigonometric functions are commonly defined as ratios of two sides of a right-angled triangle containing the angle.



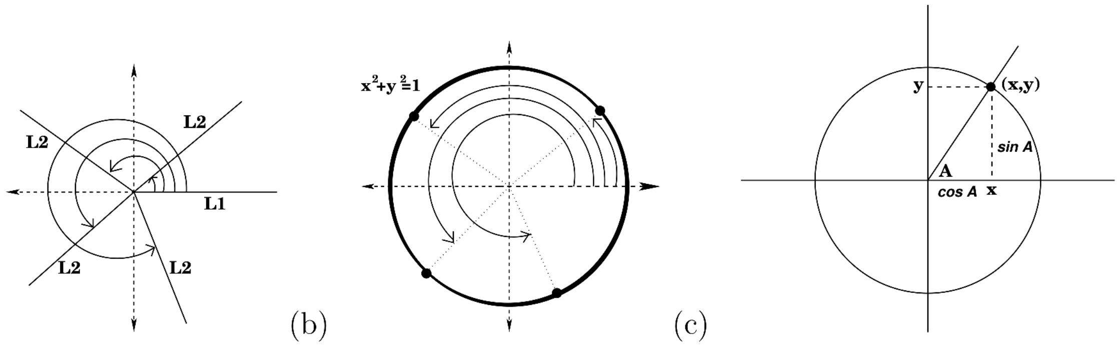
$$\text{sine of angle} \equiv \frac{\text{length of opposite side}}{\text{length of hypotenuse}}, \qquad \text{cosine of angle} \equiv \frac{\text{length of adjacent side}}{\text{length of hypotenuse}}.$$

For more general angles, we can give a definition in terms of points on the **unit circle** (the circle $x^2 + y^2 = 1$ with centre at the origin and radius 1). Note that trigonometric functions are often also called **circular functions**.

We begin with the idea of a **directed angles** between a pair of lines, $L1$ and $L2$ emanating from the same initial point.



If we choose $L1$ to always lie along the positive x-axis, a directed angle is completely defined by the second line, $L2$ (see diagram (a) below). In fact, we can describe a directed angle even more simply by using the point (x, y) on the unit circle $x^2 + y^2 = 1$ where each line $L2$ crosses it (diagram (b) below). The sine and cosine of a directed angle are then defined in terms of the co-ordinates of this point.



(a) (b) (c)

For example, consider the angle A in diagram (c). It can be seen that

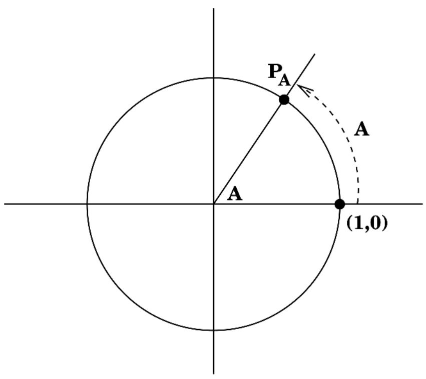
$$\text{sine of angle } A \equiv y, \qquad \text{cosine of angle } A \equiv x.$$

What we have defined so far are the sine and cosine of a directed angle; in calculus, what we need is to define $\sin(x)$ and $\cos(x)$ for any **real number** x. The usual way of doing this involves associating every number with an angle, traditionally by measuring an angle in **degrees**, that is,

$$\sin^{\circ}(x) = \text{sine of the angle of } x \text{ degrees.} \tag{6.1}$$

Here, the angle going all the way round the circle is associated with 360° and other angles are measured against this. However, this choice of 360 is somewhat arbitrary and is unlikely to lead to nice mathematical theory. In mathematics, we therefore use instead **radians** to measure angles.

In radian measure, given any number A, we choose a point P_A on the unit circle such that A is the length of the arc of the circle starting from (1,0) and going counterclockwise to P_A. The directed angle determined by P_A is called the angle of A radians.



As the length of the whole circle is 2π (circumference $= 2\pi\times$ radius), the angle of x radians and the angle of $2\pi + x$ radians will be identical. We define

$$\sin^{r}(x) = \text{sine of the angle of } x \text{ radians.} \tag{6.2}$$

Note the connection between definitions (6.1) and (6.2). As we want $\sin^\circ(360) = \sin^r(2\pi)$, we can define

$$\sin^\circ(x) = \sin^r\left(\frac{2\pi x}{360}\right) = \sin^r\left(\frac{\pi x}{180}\right).$$

From now on, we will always measure angles in radians (unless otherwise stated). That is, we will drop the superscript r and adopt (6.2) as our definition for the function $\sin(x)$.

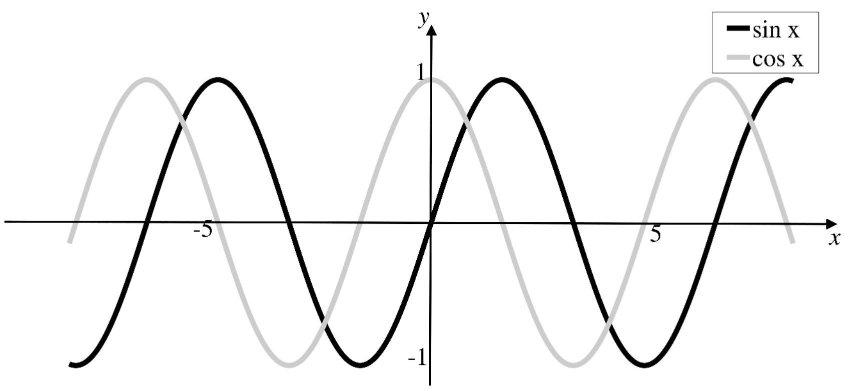
Some points on notation: the brackets surrounding the argument of sine and cosine are commonly omitted unless there is ambiguity. That is, we often write

$$\boxed{\sin x \quad (= \sin(x)); \qquad \text{and} \qquad \cos y \quad (= \cos(y))}$$

as the meaning is clear, but we always need the brackets in an expression like $\sin(x+y)$ (otherwise it could be read as $\sin(x) + y$, which is of course different!). In addition, powers of trigonometric functions usually have the superscript before the argument (e.g. we usually write $\sin^2 x$, $\cos^4 y$ instead of $\sin(x)^2$, $\cos(y)^4$, which is easier to misinterpret).

Examples 6A

The graphs of $\sin(x)$ and $\cos(x)$ are shown below.



We now list some of the functions' important properties, which can be deduced from the fact that they are the coordinates of a point (x, y) on the unit circle.

1. The range of sine and cosine is the same:

$$-1 \leq \sin(x) \leq 1, \qquad -1 \leq \cos(x) \leq 1.$$

2. For every real number x,

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

3. Both functions are **periodic** with period 2π:

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

4. Sine is an **odd** function and cosine is an **even** function:

$$\sin(-x) = -\sin(x), \qquad \cos(-x) = \cos(x).$$

We say that two angles are **complementary** if their sum is $\pi/2$, and **supplementary** if their sum is π. This gives rise to the following two sets of identities involving sine and cosine:

5. Complementary angle identities:

$$\sin\left(\frac{\pi}{2} - x\right) = \cos(x), \qquad \cos\left(\frac{\pi}{2} - x\right) = \sin(x).$$

6. Supplementary angle identities:

$$\sin(\pi - x) = \sin(x), \qquad \cos(\pi - x) = -\cos(x).$$

The following **addition formulas** are helpful for finding the sine and cosine of a sum of difference of two numbers. The proofs of these results follow from the definitions above but are omitted here.

$$\begin{aligned} \sin(x \pm y) &= \sin(x)\,\cos(y) \pm \cos(x)\,\sin(y), \\ \cos(x \pm y) &= \cos(x)\,\cos(y) \mp \sin(x)\,\sin(y). \end{aligned}$$

Putting $x = y$ into these gives the so-called **double-angle formulas**

$$\begin{aligned} \sin(2x) &= 2\sin(x)\,\cos(x), \\ \cos(2x) &= \cos^2(x) - \sin^2(x) = 2\cos^2(x) - 1 = 1 - 2\sin^2(x). \end{aligned}$$

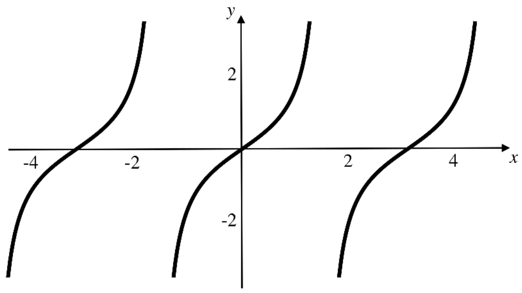
Examples 6B

6.2 The tangent function

Another very common trigonometric function is the **tangent function** $\tan(x)$ which is defined as the ratio of the sine and cosine:

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

The graph of $\tan(x)$ is shown below.



Note that $\tan(x)$ is undefined when $\cos(x) = 0$, that is, when $x = k\pi + \frac{\pi}{2}$, $k \in \mathbb{Z}$. The tangent is an odd function with period π:

$$\tan(x+\pi) = \frac{\sin(x+\pi)}{\cos(x+\pi)} = \frac{\sin(x)\cos(\pi) + \cos(x)\sin(\pi)}{\cos(x)\cos(\pi) - \sin(x)\sin(\pi)} = \frac{-\sin(x)}{-\cos(x)} = \tan(x).$$

Addition formulas involving the tangent can be derived from those for sine and cosine.

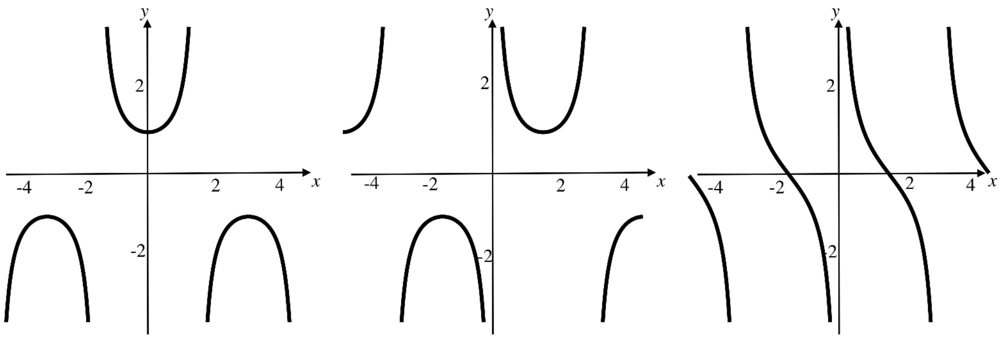
$$\tan(x \pm y) = \frac{\tan(x) \pm \tan(y)}{1 \mp \tan(x)\tan(y)}.$$

6.3 Other trigonometric functions

Three other useful trigonometric functions which can be defined in terms of sin and cos.

secant	$\sec(x) = \dfrac{1}{\cos(x)}$;
cosecant	$\csc(x) = \dfrac{1}{\sin(x)}$;
cotangent	$\cot(x) = \dfrac{1}{\tan(x)}$.

The graphs of these functions are shown below.



Each of these functions is undefined at points where the function in the denominator of the definition has value 0. The cosecant and cotangent are odd functions and secant is an even function.

We can derive further useful identities involving these functions from the identities involving $\sin(x)$ and $\cos(x)$ quoted above. For example, dividing the identity $\sin^2(x) + \cos^2(x) = 1$ by $\cos^2(x)$ and $\sin^2(x)$ gives the identities

$$\tan^2(x) + 1 = \sec^2(x)$$

and

$$1 + \cot^2(x) = \csc^2(x),$$

respectively.

Examples 6C

6.4 Inverse trigonometric functions

Recall from §4.7 that for a function to have an inverse, it must be one-to-one. As this is NOT the case for trigonometric functions, in order to define their inverses, we must first restrict them to suitable intervals. The intervals usually chosen are

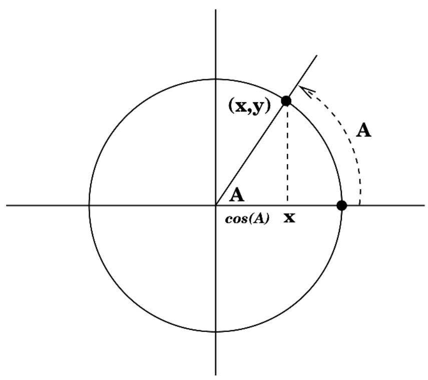
$$\begin{aligned} x \in [0, \pi] &\quad \text{for} \quad \cos(x) \\ x \in [-\pi/2, \pi/2] &\quad \text{for} \quad \sin(x) \\ x \in (-\pi/2, \pi/2) &\quad \text{for} \quad \tan(x). \end{aligned}$$

(The inverses of the other trigonometric functions are rarely used, so we will not consider them here).

The inverse of the function

$$f(x) = \cos(x), \qquad 0 \le x \le \pi$$

is denoted by **arccos**. This notation is used because, as we saw earlier, if $x = \cos(A)$ for an angle A, then $A = \arccos(x)$ is the length of the **arc** of the unit circle measured anticlockwise from $(1, 0)$ (see diagram below).



Similarly, the inverse of the function

$$f(x) = \sin(x), \qquad -\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$$

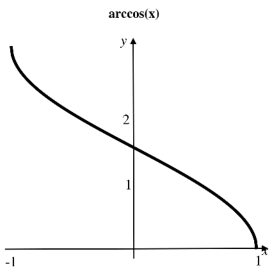
is denoted by **arcsin**.

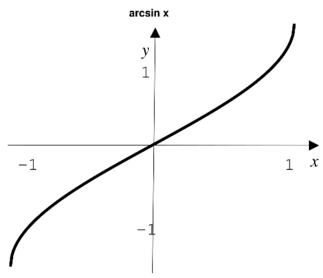
Finally, the inverse of the function

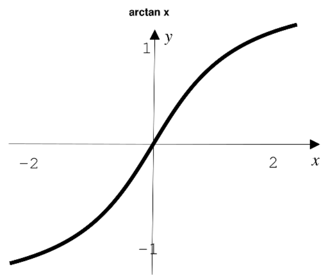
$$f(x) = \tan(x), \qquad -\frac{\pi}{2} < x < \frac{\pi}{2}$$

is denoted by **arctan**.

These three functions are plotted below.







From the plots we see that the slope of arccos is negative everywhere. We also see that $\mathrm{Dom}(\arccos) = [-1, 1]$ and $\mathrm{Im}(\arccos) = [0, \pi]$ with $\arccos(0) = \frac{\pi}{2}$, $\arccos(1) = 0$ and $\arccos(-1) = \pi$.

Similarly, from the plot of arcsin we see that the slope is positive everywhere with $\mathrm{Dom}(\arcsin) = [-1, 1]$, $\mathrm{Im}(\arcsin) = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Also, $\arcsin(0) = 0$, $\arcsin(1) = \frac{\pi}{2}$ and $\arcsin(-1) = -\frac{\pi}{2}$.

Finally, from the plot of arctan we see that $\mathrm{Dom}(\arctan) = (-\infty, \infty)$, $\mathrm{Im}(\arctan) = \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. Also, $\arctan(0) = 0$, $\arctan(1) = \frac{\pi}{4}$, $\arctan(-1) = -\frac{\pi}{4}$.

Important note on notation: In many textbooks, the notation $\sin^{-1}(x)$, $\cos^{-1}(x)$ and $\tan^{-1}(x)$ is used for these inverse functions. However, this can be confusing for two reasons. Firstly, it implies that these functions are the inverses of $\cos(x)$, $\sin(x)$ and $\tan(x)$ which is not the case (they are the inverses of different functions, namely the trigonometric functions with restricted domains). Also, it offers the prospect of confusion with the reciprocals of $\cos(x)$, $\sin(x)$ and $\tan(x)$, that is,

$$1/\cos(x) = \sec(x), \qquad 1/\sin(x) = \csc(x), \qquad 1/\tan(x) = \cot(x).$$

For these reasons, we will use the notation $\arccos(x)$, $\arcsin(x)$ and $\arctan(x)$ on this course.

Examples 6D