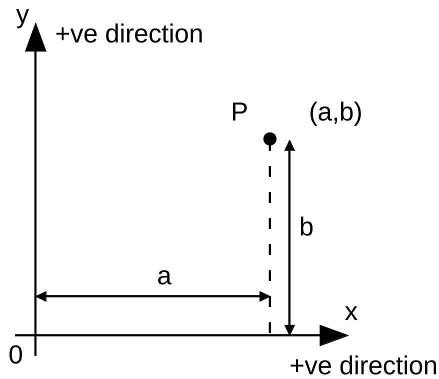
Introduction to Cartesian Coordinate Geometry

Geometry is the branch of mathematics concerned with the properties of lines, curves and surfaces. First we concentrate on two-dimensional geometry, that is, the relationships between points that lie in a plane. Later we touch briefly upon problems in three dimensions. We start by considering what we mean by coordinates.

1 Coordinates in a plane

A rectangular Cartesian coordinate system consists of an origin (O) and mutually perpendicular axes (Ox and Oy) passing through the origin. The *position* of a *point* P in a plane is specified by its *perpendicular distances* from the fixed *perpendicular lines* Ox, Oy. The point P in the diagram has its x-coordinate equal to a, its y-coordinate equal to b. It should be noted that the x-*coordinate* of a point is its perpendicular distance from Oy and its y-coordinate is its perpendicular distance from Ox. Also the distances are directed so that a positive (negative) coordinate is a measure along the positive (negative) direction of the axis.

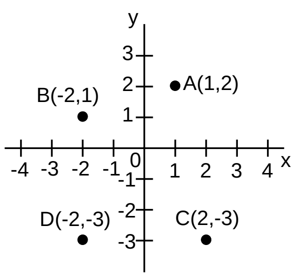


The coordinates are given as an *ordered pair* with the x-coordinate (often called the *abscissa*) first and the y-coordinate (often called the *ordinate*) second. The point P is referred to as (a, b). The coordinates form an ordered pair, in general $P(a, b) \neq P(b, a)$.

In the graphical representation of coordinate axes the scale of the x- and y-axes need not be the same so it is important to annotate the axes.

Example 1

The points $A(1, 2)$, $B(-2, 1)$, $C(2, -3)$, $D(-2, -3)$ are represented as shown.

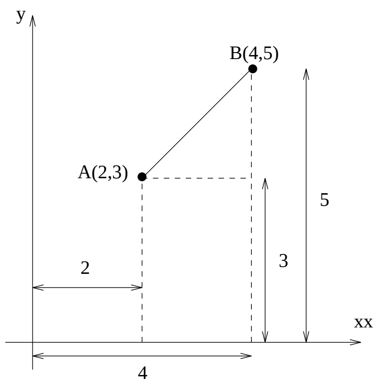


2 The distance between two points in a plane

Given the rectangular Cartesian coordinates of two points we are able to find the distance between them.

Example 2

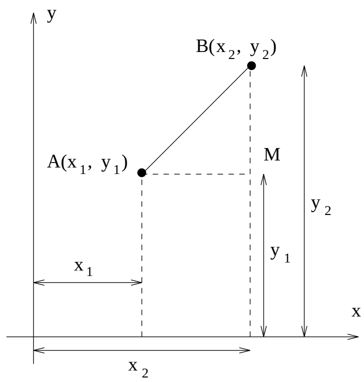
Find the distance between the points $A(2,3)$ and $B(4,5)$.



If we represent the points as shown and draw AM parallel to Ox, then triangle ABM is right angled and by Pythagoras's theorem

$$\begin{aligned} AB^2 &= AM^2 + MB^2 \\ &= (4-2)^2 + (5-3)^2 \\ &= 2^2 + 2^2 = 8 \end{aligned}$$

so $AB = \sqrt{8}$.



In general, if $A(x_1, y_1)$ and $B(x_2, y_2)$ are two points as shown, we have by Pythagoras's theorem

$$\begin{aligned} AB^2 &= AM^2 + MB^2 \\ &= (x_2 - x_1)^2 + (y_2 - y_1)^2 \end{aligned}$$

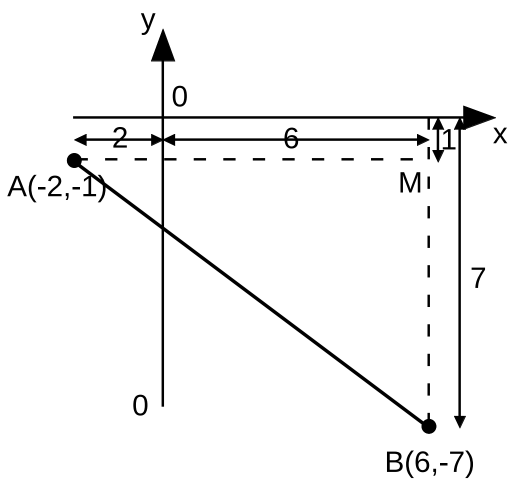
or

$$AB = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

Here the positive square root is used as the distance has no direction (sign) associated with it. In words, we find AB^2 by squaring the difference of the x's, squaring the difference of the y's and adding. This procedure is valid even if some or all of the x's and/or y's are negative.

Example 3

Find the distance between the points $A(-2,-1)$ and $B(6,-7)$



Now

$$AM = 2 + 6 = 8,$$
$$MB = 7 - 1 = 6,$$

so that

$$AB^2 = AM^2 + MB^2 = 64 + 36 = 100$$

and

$$AB = 10.$$

Rule: When due account is taken of the signs of coordinates, the formula

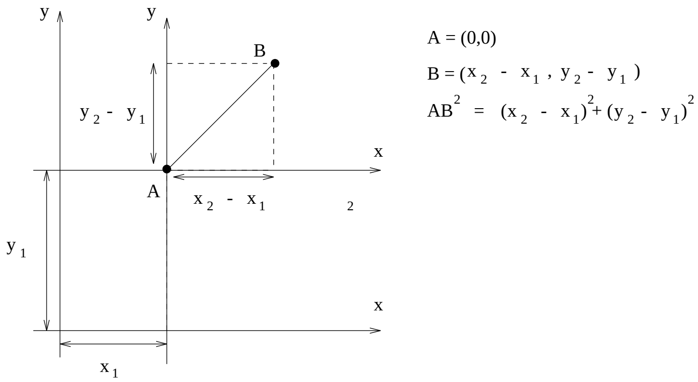
$$AB = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

gives the distance between the points $A(x_1, y_1)$ and $B(x_2, y_2)$.

NB: It does not matter which point you call (x_1, y_1) or (x_2, y_2) - the distance between them is the same. For Example 3 we let $x_1 = -2$, $y_1 = -1$ and $x_2 = 6$, $y_2 = -7$; then the distance between the points A and B is given by

$$\begin{aligned} AB &= \sqrt{(6-(-2))^2 + (-7-(-1))^2} \\ &= \sqrt{8^2 + 6^2} = \sqrt{100} = 10. \end{aligned}$$

The distance between two points is independent of the location of the origin. In the figure below the axes are shifted so that the point A is the new origin. The original coordinates of A and B are (x_1, y_1) and (x_2, y_2) respectively.



3 The midpoint of the straight line joining two given points

We wish to find the coordinates of the midpoint, M, of the line joining the points $A(x_1, y_1)$ and $B(x_2, y_2)$. Let M have coordinates (x_m, y_m).

If[1] *M is the midpoint of AB then*

$$\begin{aligned} AM &= MB, \\ AM^2 &= MB^2. \end{aligned}$$

We

have

$$AM^2 = (x_1 - x_m)^2 + (y_1 - y_m)^2$$

and

$$MB^2 = (x_2 - x_m)^2 + (y_2 - y_m)^2.$$

Therefore

$$x_1^2 - 2x_1x_m + x_m^2 + y_1^2 - 2y_1y_m + y_m^2 = x_2^2 - 2x_2x_m + x_m^2 + y_2^2 - 2y_2y_m + y_m^2.$$

Rearranging this equation

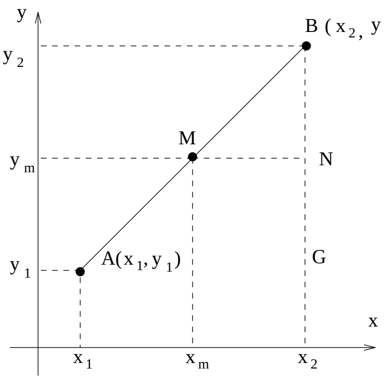
$$\begin{aligned} x_1^2 - x_2^2 - 2x_m(x_1 - x_2) &= y_2^2 - y_1^2 - 2y_m(y_2 - y_1) \\ (x_1 - x_2)(x_1 + x_2 - 2x_m) &= (y_2 - y_1)(y_2 + y_1 - 2y_m). \end{aligned} \tag{3.1}$$

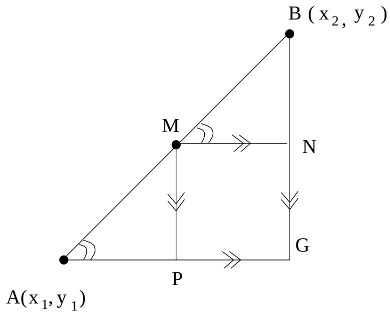
This relationship (equation) is satisfied by all points M that are equidistant from A and B. We need another relationship to determine the coordinates of the midpoint, and we will see later how we can find the point satisfying the above equation that also lies on the line AB. In mathematical terminology we have two unknowns x_m and y_m which can only be determined explicitly if we have two independent equations involving the unknowns.

In fact we can determine the coordinates of the midpoint of AB by simple trigonometry.

Draw MN parallel to the x-axis. Then angle BMN is equal to angle BAG as AB intercepts the parallel lines AG and MN at the same angle. So the triangles ABG and MBN are

[1]This section, in italics, is for those who are more familiar with the basic concepts.





similar *(three equal angles); we have $BM = \frac{1}{2}AB$, therefore $BN = \frac{1}{2}BG$. By similar reasoning $AP = \frac{1}{2}AG$. So*

$$\begin{aligned} x_m &= x_1 + \frac{1}{2}(x_2 - x_1) = \frac{1}{2}(x_1 + x_2), \\ y_m &= y_1 + \frac{1}{2}(y_2 - y_1) = \frac{1}{2}(y_1 + y_2). \end{aligned}$$

As one would expect these values for x_m, y_m satisfy the relationship (3.1).

So why bother using a rather long-winded algebraic approach to find x_m and y_m? Firstly you will find that it is much simpler to program algebraic procedures - computers are not good at trigonometry which is essentially a visual tool. In this case you would of course use the known result but as a general rule one should think algebraically not trigonometrically.

Rule: The midpoint of the line joining the points (x_1, y_1) and (x_2, y_2) has coordinates

$$\left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}\right).$$

This result holds when some or all of the coordinates are negative.

Example 4

Find the coordinates of the *midpoint* of the line AB, where A and B are the points (2,3) and (4,7).

Let M be the midpoint of AB. Then, writing 'x of M' to mean the x coordinate of M,

$$\begin{aligned} x \text{ of } M &= \frac{1}{2}(x \text{ of } A + x \text{ of } B) = \frac{1}{2}(2 + 4) = 3, \\ y \text{ of } M &= \frac{1}{2}(y \text{ of } A + y \text{ of } B) = \frac{1}{2}(3 + 7) = 5, \end{aligned}$$

Thus M is the point (3,5).

4 Gradient of a line

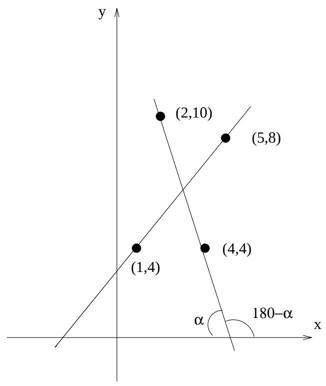
The gradient (slope) of a line is given by the increase in the y-coordinate divided by the increase in the x-coordinate as you move from one point on the line to another. For example, the gradient of the line passing through the points (1,4) and (5,8) is

$$\frac{8-4}{5-1} = \frac{4}{4} = 1.$$

Similarly the gradient of the line joining (2,10) and (4,4) is

$$\frac{4-10}{4-2} = \frac{-6}{2} = -3.$$

Here the gradient is negative because y decreases as x increases.



It is not important which two points on a line are used to calculate the gradient of the line as the gradient is the same throughout the line. We say that a line has a constant gradient.

Rule: The gradient of a line passing through the two points (x_1, y_1) and (x_2, y_2) is given by

$$\text{gradient of line} = \frac{y_2 - y_1}{x_2 - x_1}.$$

When $x_2 - x_1 = 0$ the gradient is infinite and the line is parallel to the y-axis. When $y_2 - y_1 = 0$ the line is parallel to the x-axis.

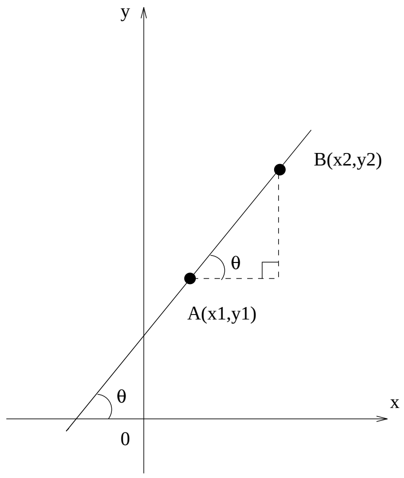
The gradient of a line is related to the angle that the line makes with the x-axis. Consider the line passing through the points A and B in the diagram below which makes an angle, θ, with the x-axis.

Now

$$\tan\theta = \frac{\text{opposite}}{\text{adjacent}} = \frac{y_2 - y_1}{x_2 - x_1} = \text{gradient}.$$

The line passing through the points (2,10) and (4,4) makes an angle, α, with the x-axis. Here

$$\tan\alpha = \frac{y_2 - y_1}{x_1 - x_2} = -\text{gradient}.$$



But we know that $\tan(180-\alpha) = -\tan\alpha$, therefore $\tan(180-\alpha) =$ gradient.

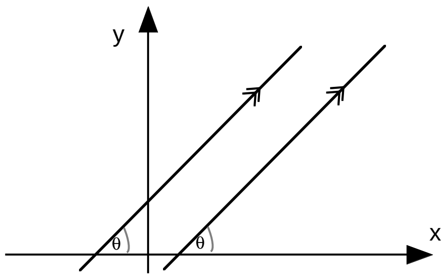
We define the angle θ as follows: θ is the angle measured from the positive x-axis to the line in an anti-clockwise direction. With this definition for θ we always have $\tan\theta =$gradient.

Example 5

Find the gradient of the line passing through $(-2,3)$ and $(3,5)$.

$$\text{gradient} \;=\; \frac{5-3}{3-(-2)} = \frac{2}{5}.$$

Parallel lines make the same angle, θ, with the x-axis. Since the gradient of the line is $\tan\theta$ in each case it follows that *parallel lines have equal gradients.*



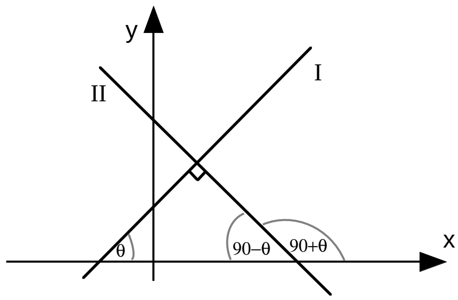
Example 6

Show that the line passing through $(3,7)$ and $(5,15)$ is parallel to the line through $(-6,-8)$ and $(2,24)$.

$$\begin{aligned} \text{gradient of first line} &= \frac{7-15}{3-5} = \frac{-8}{-2} = 4 \\ \text{gradient of second line} &= \frac{24-(-8)}{2-(-6)} = \frac{32}{8} = 4 \end{aligned}$$

The gradients of the two lines are equal, therefore the lines are parallel. Note that in this example we again use algebra rather than trigonometry to obtain the result.

Lines I and II in the diagram are at right angles to one another. We can find a relationship between the gradients of perpendicular lines.



$$\begin{aligned}
\text{gradient of line I} &= \tan\theta \\
\text{gradient of line II} &= \tan(90+\theta) \\
&= -\tan(90-\theta) \\
&= -\frac{1}{\tan\theta}.
\end{aligned}$$

The product of the gradients of perpendicular lines is $\tan\theta\left(-\dfrac{1}{\tan\theta}\right) = -1$.

Example 7

Show that the line joining $A\,(1,3)$ and $B\,(3{,}6)$ is perpendicular to the line passing through $C(8{,}6)$ and $D(5{,}8)$.

$$\begin{aligned}
\text{gradient } AB &= \frac{6-3}{3-1} = \frac{3}{2}. \\
\text{gradient } CD &= \frac{6-8}{8-5} = -\frac{2}{3}. \\
\text{Product of the gradients} &= -\frac{3}{2} \times \frac{2}{3} = -1.
\end{aligned}$$

Therefore the lines are perpendicular.

5 Equation of a line

The equation of a line is a relationship between the coordinates of any point on the line. So if, for example, we want to know the y-coordinate of a point on a line for a given value of the x-coordinate we can use the equation of the line to find it.

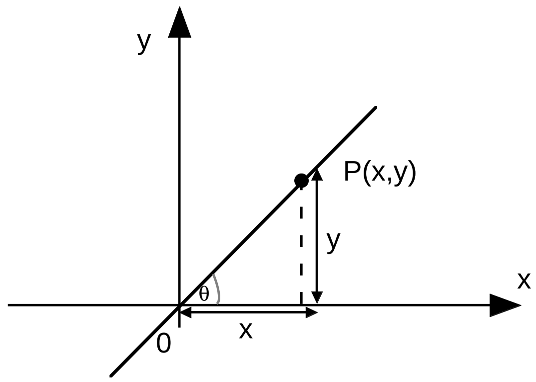
Example 8

A line has equation $y = 2x - 1$. What is the y-coordinate of the point on the line with x-coordinate equal to 3? What point with a y-coordinate of -5 lies on the line?

When $x = 3$ we have $y = 2 \times 3 - 1 = 5$. So the y-coordinate is 5.

When $y = -5$ we have $-5 = 2x - 1$, so $2x = -4$ and hence $x = -2$. The point is $(-2, -5)$.

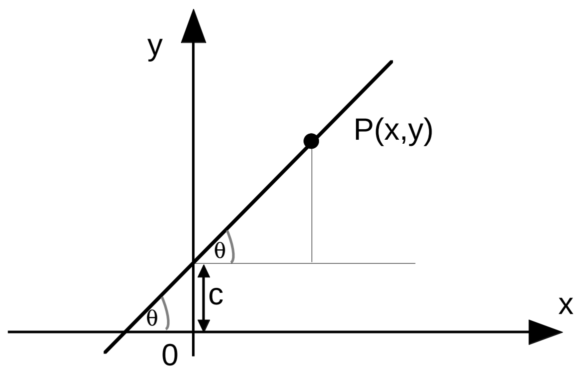
So how do we determine the equation of a line?



Consider first a line of given gradient, m say, passing through the origin. Any point P on the line is given by the coordinates (x, y) and the line makes an angle θ with the x-axis. We know from basic trigonometry that $\tan\theta = \frac{y}{x}$, so $y = x\tan\theta$. The gradient of the line, m, is also equal to $\tan\theta$ so $y = mx$.

Rule: The equation of a line with gradient m which passes through the origin is $y = mx$.

A line of given gradient, m, intersects the y-axis a distance c from the origin. Again $P(x, y)$ represents any point on the line and θ is the angle the line makes with the x-axis.

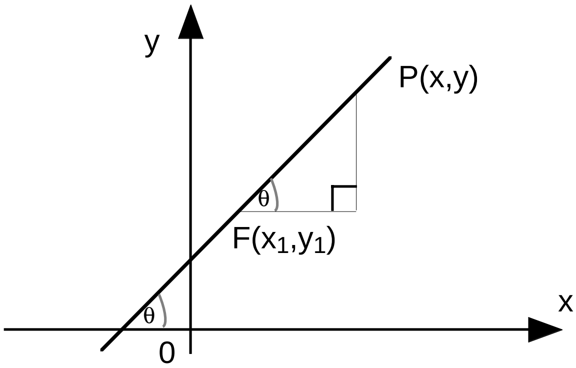


Now

$$m = \tan\theta = \frac{y - c}{x},$$

so $mx = y - c$ or $y = mx + c$.The distance c is often referred to as the *intercept* on the y-axis.

Rule: The equation of a line with gradient m which intersects the y-axis a distance c from the origin is $y = mx + c$.



A line of given slope, m, passing through a fixed point $F(x_1, y_1)$ has equation

$$m = \tan\theta = \frac{y - y_1}{x - x_1}.$$

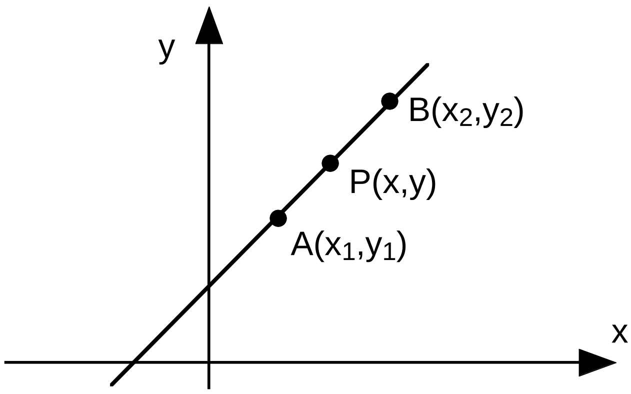
Equivalently

$$m(x - x_1) = y - y_1,$$

or

$$y = mx - mx_1 + y_1.$$

In relation to the last case this is a line with gradient m and intercept $c = -mx_1 + y_1$.



The equation of a line passing through two given points $A(x_1, y_1)$ and $B(x_2, y_2)$ can be determined as follows. Again $P(x, y)$ is any point on the line, we have

$$\text{gradient of line} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{y - y_1}{x - x_1}.$$

So

$$y - y_1 = \frac{(y_2 - y_1)}{(x_2 - x_1)}(x - x_1) = m(x - x_1),$$

or

$$y = mx - mx_1 + y_1 \qquad \text{as before.}$$

In general the equation of a line is given by $y = mx + c$ where m is the gradient of the line and c is the intercept on the y-axis (i.e. the y-coordinate of the point where the line crosses the y-axis). The equation can be determined if two attributes of the line are known, for example if it passes through two given points or it passes through a given point with given gradient, or it has given gradient and intercept.

Example 9

Find the equation of the line with gradient -2 which passes through the point (2,1).

Let (x, y) be any point on the line. Then

$$\text{gradient of line} = \frac{y-1}{x-2} = -2,$$

equivalently

$$y - 1 = -2(x - 2)$$

or

$$y = -2x + 5.$$

Alternatively, let the equation be $y = mx + c$. We are given $m = -2$, so $y = -2x + c$. But $(2, 1)$ lies on the line, so $x = 2$ when $y = 1$. Substituting these values into the equation gives

$$1 = -4 + c, \text{ so } \quad c = 5$$

and as before

$$y = -2x + 5.$$

Example 10

Find the equation of the line passing through the points $(2, -1)$ and $(-1, 6)$.

First we find the gradient of the line:

$$\text{gradient of line} = \frac{6-(-1)}{-1-2} = -\frac{7}{3}.$$

Then, as in the previous example, we equate this to the gradient obtained using a general point (x, y) and one of the given points:

$$\begin{aligned} -\frac{7}{3} &= \frac{y+1}{x-2} \\ -7(x-2) &= 3(y+1) \\ -7x + 14 - 3 &= 3y. \end{aligned}$$

Hence the equation is

$$y = \frac{1}{3}(-7x + 11).$$

Alternatively, let the equation of the line be $y = mx + c$ where m and c are constants to be determined. Now when $x = 2$, $y = -1$ and when $x = -1$, $y = 6$, so

$$-1 = 2m + c \tag{5.1}$$

$$6 = -m + c \tag{5.2}$$

and we have two simultaneous equations for the two unknowns m and c. Subtracting (5.2) from (5.1) we have

$$-7 = 3m, \quad m = -\frac{7}{3}.$$

Substituting this in (5.2) we have

$$6 = \frac{7}{3} + c; \quad c = 6 - \frac{7}{3} = \frac{18-7}{3} = \frac{11}{3}.$$

Therefore $y = -\frac{7}{3}x + \frac{11}{3} = \frac{1}{3}\left(-7x + 11\right)$.

Now[2] *we return to the question we left unresolved. Let $M(x_m, y_m)$ be the mid-point of the line segment AB. Find the coordinates of M in terms of the coordinates of A and B, namely, x_1, y_1, x_2, y_2. We have*

$$(x_1 - x_2)(x_1 + x_2 - 2x_m) = (y_2 - y_1)(y_1 + y_2 - 2y_m) \tag{5.3}$$

if $AM = MB$. We also require that M lies on the line passing through A and B. The gradient of this line is

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{y_2 - y_m}{x_2 - x_m}$$

If $x_1 = x_2$ the line is parallel to the y-axis and obviously the midpoint has coordinates $(x_1, \frac{1}{2}(y_2 - y_1) + y_1) = (x_1, \frac{1}{2}(y_1 + y_2))$. If $y_1 = y_2$ the line is parallel to the x-axis and clearly the mid-point has coordinates $\left(\frac{1}{2}(x_1 + x_2), y_1\right)$. Otherwise $(y_2 - y_1)(x_2 - x_m) = (y_2 - y_m)(x_2 - x_1)$ is a second relationship which, together with (5.3), enables us to determine x_m and y_m. We could substitute

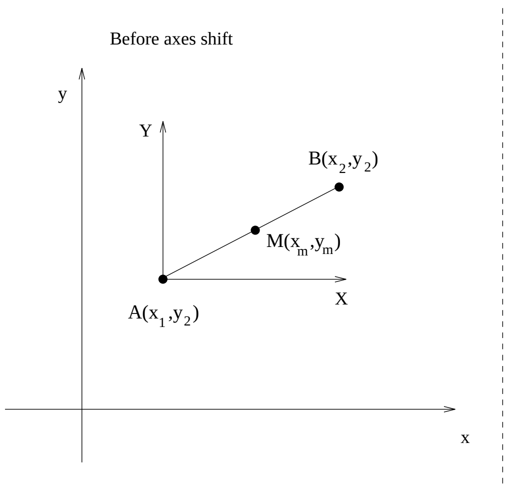
$$x_m = \frac{-(y_2 - y_m)(x_2 - x_1)}{y_2 - y_1} + x_2, \tag{5.4}$$

provided that $y_2 - y_1 \neq 0$, into (5.3) to give an equation for y_m in terms of x_1, y_1, x_2 and y_2. Then y_m could be replaced in (5.4) to give x_m in terms of x_1, y_1, x_2 and y_2. However, the algebra is rather tedious and we will not pursue it here, rather we will shift the axes to simplify things somewhat. We shift the axes parallel to the current axes until A becomes the origin. Label the new axes X, Y and the points $M(X_m, Y_m)$, $B(X_2, Y_2)$ and of course $A(0,0)$.

The x-axis has been shifted right by an amount x_1, the y-axis shifted up a distance y_1. So we have the following relationships between the old and new coordinates:

$$\begin{aligned} X_2 &= x_2 - x_1, \quad Y_2 = y_2 - y_1, \\ X_m &= x_m - x_1, \quad Y_m = y_m - y_1. \end{aligned}$$

[2] This section, in italics, is for those who are more familiar with the basic concepts.



Working with the new axes

$$\begin{aligned} AM^2 &= X_m^2 + Y_m^2 \\ MB^2 &= (X_2 - X_m)^2 + (Y_2 - Y_m)^2 \\ &= X_2^2 - 2X_2X_m + X_m^2 + Y_2^2 - 2Y_2Y_m + Y_m^2. \end{aligned}$$

As $AM^2 = MB^2$ *we can equate these expressions to give*

$$X_2^2 - 2X_2X_m + Y_2^2 - 2Y_2Y_m = 0,$$

or equivalently

$$X_2(X_2 - 2X_m) + Y_2(Y_2 - 2Y_m) = 0. \qquad (5.5)$$

The gradient of line AB is

$$\frac{Y_m}{X_m} = \frac{Y_2}{X_2}$$

provided X_2, X_m *are non-zero. If* $X_2 = X_m = 0$ *the mid-point is* $\left(0, \frac{Y_2}{2}\right)$*.Otherwise*

$$Y_m = \frac{Y_2X_m}{X_2}. \qquad (5.6)$$

Substituting in (5.5) we have

$$X_2(X_2 - 2X_m) + Y_2(Y_2 - \frac{2Y_2X_m}{X_2}) = 0$$

$$X_2^2 + Y_2^2 = 2X_mX_2 + 2Y_2^2\frac{X_m}{X_2} = 2X_m\left(\frac{X_2^2 + Y_2^2}{X_2}\right).$$

Therefore

$$\frac{X_2}{2} = X_m.$$

Substituting this expression for X_m *in (5.6) gives*

$$Y_m = \frac{Y_2X_2}{2X_2} = \frac{Y_2}{2}.$$

In terms of the original axes we have

$$X_m = x_m - x_1 = \frac{x_2 - x_1}{2},$$

therefore

$$x_m = \frac{1}{2}(x_1 + x_2).$$

Similarly

$$Y_m = y_m - y_1 = \frac{y_2 - y_1}{2}$$

therefore

$$y_m = \frac{1}{2}(y_1 + y_2).$$

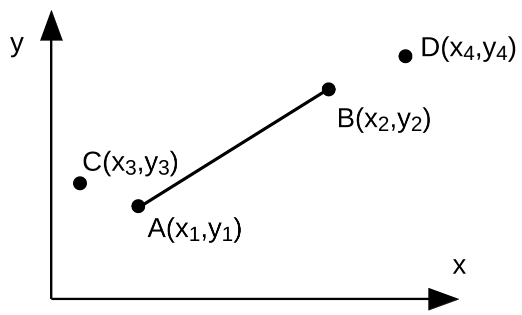
Important note: Shifting and rotating coordinate axes can often simplify the algebra required to solve a problem. Routines to shift and rotate axes are commonly provided in graphical packages.

Example*[3] *11: Problem of the wall and line of sight

A wall of fixed length obscures the view of people that are on either side of it. Design an algorithm to establish whether or not two people are in each other's line of sight.

We will pose the problem as follows. Let the wall extend between the two points A and B. Let the people be positioned at points C and D. The people are not *in each other's line of sight when the line segment joining C to D intersects the wall AB. This is the case when:*

(1) C and D are on opposite sides of the wall,
(2) the lines passing through the points A and B, and C and D intersect
and (3) the point of intersection is between A and B on the line AB.



Choose coordinate axes; let A be the point (x_1, y_1) and B be the point (x_2, y_2). We can determine the equation *of the* line *passing through* A *and* B. *If $y_1 = y_2$ then the line is parallel to the x-axis and has equation $y = y_1$. If $x_1 = x_2$ then the line is parallel to the y-axis and has equation $x = x_1$. We shall consider these two cases later. First we assume that the wall is not parallel to either of the axes and thus the equation of the line passing through A and B is $y = ax + b$, where $a = (y_2 - y_1)/(x_2 - x_1)$ and $b = -ax_1 + y_1$.*

[3]This section, in italics, is for those who are more familiar with the basic concepts

First we establish whether or not $C(x_3, y_3)$ and $D(x_4, y_4)$ are on opposite sides of this line. A point (x, y) is on the line if $y = ax + b$, above the line if $y > ax + b$ and below the line if $y < ax + b$. For point C compare y_3 with $ax_3 + b$ and for point D compare y_4 with $ax_4 + b$. Points C and D are on opposite sides of the line passing through A and B if

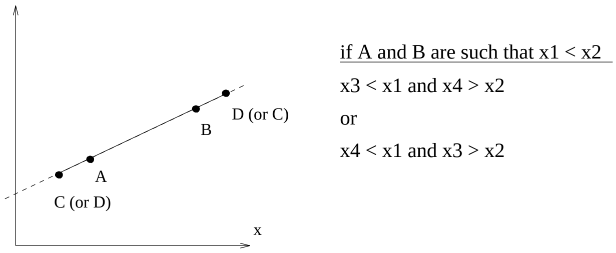
$$y_3 > ax_3 + b \text{ and } y_4 < ax_4 + b$$

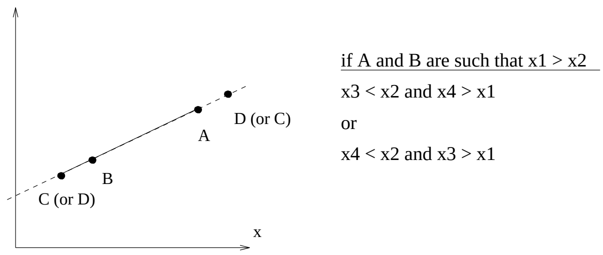
or

$$y_3 < ax_3 + b \text{ and } y_4 > ax_4 + b.$$

What if $y_3 = ax_3 + b$ or $y_4 = ax_4 + b$? That is C or D (or both) lie on the line passing through A and B. The context of the problem would tell us if it is feasible for C or D to be on the line between A and B, that is on the wall here. Points C or D are on the wall if ($x_1 < x_3 < x_2$ or $x_1 < x_4 < x_2$) or if ($x_2 < x_3 < x_1$ or $x_2 < x_4 < x_1$) and the coordinates of the point satisfy the equation of the line through A and B. We proceed assuming that C, D are not on the wall.

If C (or D) lies on the line (but not on the wall) but D (or C) does not, then (3) cannot be true and C can see D. If both C and *D lie on the line but not between A and B we need to consider their relative positions on the line.*

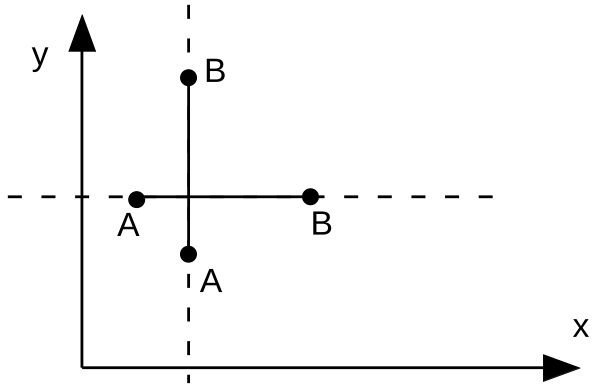




If A, B are such that $x_1 < x_2$ then C, D are on opposite sides if ($x_3 < x_1$ and $x_4 > x_2$) or ($x_4 < x_1$ and $x_3 > x_2$). On the other hand, if A, B are such that $x_1 > x_2$ then C, D are on opposite sides if ($x_3 < x_2$ and $x_4 > x_1$) or ($x_4 < x_2$ and $x_3 > x_1$).

Another possibility we must consider is that of the wall being parallel to one of the axes. If the wall lies parallel to the x-axis, the equation of the line passing through A and B is $y = y_1$ and we compare this constant value of y with y_3 and y_4. If $y_3 < y_1$ and $y_4 > y_1$, or $y_3 > y_1$ and $y_4 < y_1$, then points lie on opposite sides of the line through AB. If y_3 (or y_4) $= y_1$ then one or both of the points lie on the line. If one point, C or D, is on the line but not on the wall, then point (3) cannot be true. If ($x_1 < x_3 < x_2$ or $x_1 < x_4 < x_2$) or ($x_2 < x_3 < x_1$ or $x_2 < x_4 < x_1$) then C or D are on the wall. If both points are on the line, but not between A and B, they are on opposite sides of the wall if $x_1 < x_2$ and (($x_3 < x_1$ and $x_4 > x_2$) or

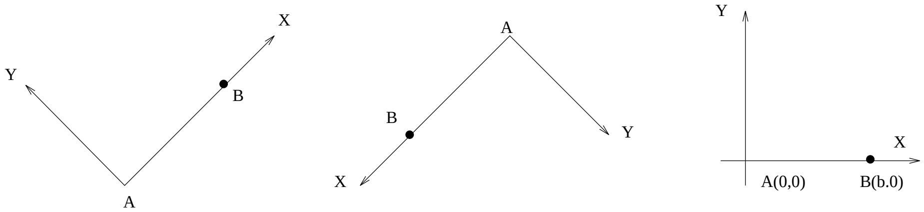
$(x_4 < x_1$ *and* $x_3 > x_2))$ *or if* $x_2 < x_1$ *and* $((x_3 < x_2$ *and* $x_4 > x_1)$ *or* $(x_4 < x_2$ *and* $x_3 > x_1))$
.



Similarly, if the wall is parallel to the y*-axis then the equation of the line passing through* A *and* B *is* $x = x_1$*. Points* C *and* D *are on opposite sides of the line if* $(x_3 < x_1$ *and* $x_4 > x_1)$ *or if* $(x_3 > x_1$ *and* $x_4 < x_1)$*. Again we must consider the possibility that* C *and/or* D *lie on the line* $x = x_1$*. If* x_3 *(or* x_4*)* $= x_1 = x_2$ *then* C *(or* D*) lies on the line. If only one of the points is on the line, but not on the wall, then* C *and* D *must be in sight of one another. If* $(y_1 < y_3 < y_2$ *or* $y_1 < y_4 < y_2)$ *or if* $(y_2 < y_3 < y_1$ *or* $y_2 < y_4 < y_1)$ *then* C *or* D *lie on the wall. If both* C *and* D *are on the line, but not on the wall, they are on opposite sides of the wall if* $y_1 < y_2$ *and* $((y_3 < y_1$ *and* $y_4 > y_2)$ *or* $(y_4 < y_1$ *and* $y_3 > y_2))$ *or if* $y_2 < y_1$ *and* $((y_3 < y_2$ *and* $y_4 > y_1)$ *or* $(y_4 < y_2$ *and* $y_3 > y_1))$.

We now know under what conditions C *and* D *are on opposite sides of the line through* A, B*. We continue in these cases to determine whether the wall obscures their view of each other. We establish whether or not the lines passing through* A, B *and* C, D *intersect. As we have established that* C *and* D *are on opposite sides of the line passing through* A *and* B *(and not on it) then the lines cannot be parallel and must intersect at some point. We determine the* equation *of the* line *passing through* C *and* D*,* $y = cx + d$ *say. Next we solve the equations* $y = ax + b$ *and* $y = cx + d$ *simultaneously to find the point of intersection,* (x_p, y_p) *say. Finally we must show that* (x_p, y_p) *lies between* $A(x_1, y_1)$ *and* $B(x_2, y_2)$*. If* $x_1 < x_2$ *then if* $x_1 < x_p < x_2$ *the point of intersection is between* A *and* B*, as it is if* $x_2 < x_1$*, and* $x_2 < x_p < x_1$*. Alternatively we could compare the* y*-coordinates in a similar manner.*

The detail is quite tedious and we shall see how a more judicious choice of axes can simplify the procedure considerably. We translate and rotate the axes so that A *is the new origin and the positive* x*-axis lies along* AB*.*



(i) A to left of B

(ii) A to right of B

(iii) New axes system
in either case

Let the shifted (translated) axes be labelled u, v. A point (x, y) in the original axes system has coordinates $u = x - x_1$, $v = y - y_1$ in the u, v system. The u, v-axes are rotated through an angle θ to produce X, Y-axes. The positive X-axis must run from A to B, so the angle of rotation is such that $\tan(\theta) = (y_2 - y_1)/(x_2 - x_1)$. The new coordinates are related to the original coordinates (x, y) by the equations

$$\begin{aligned} X &= u\cos\theta + v\sin\theta \\ &= (x - x_1)\cos\theta + (y - y_1)\sin\theta \end{aligned}$$

and

$$\begin{aligned} Y &= v\cos\theta - u\sin\theta \\ &= (y - y_1)\cos\theta - (x - x_1)\sin\theta. \end{aligned}$$

Let B be the point $(b, 0)$ in the X, Y-axes system. Clearly using the above equations $b = (x_2 - x_1)\cos\theta + (y_2 - y_1)\sin\theta$ and it can be confirmed that the Y-coordinate is zero. The equation of the line passing through A and B is $Y = 0$. In terms of the new axes we have $C(X_3, Y_3)$ and $D(X_4, Y_4)$, where the coordinates are obtained from the above equations with $x = x_3$, $y = y_3$ and $x = x_4$, $y = y_4$ respectively. If the signs of Y_3 and Y_4 are the same then both C and D are on the same side of the wall; the two people can see each other. If Y_3 (or Y_4) $= 0$ then either C or D, or both, lie on the x-axis. In this case, if $0 < X_3 < b$ or $0 < X_4 < b$ then C or D lie on the wall and presumably they can see each other. This is also true if C or D, but not both, lie on the x-axis but not on the wall. When both C and D lie on the x-axis, but not on the wall, then they are not in each other's line of sight if ($X_3 < 0$ and $X_4 > b$) or if ($X_4 < 0$ and $X_3 > b$).

NB: *Checking the sign of a real number is much more efficient computationally than comparing the values of two real numbers.*

When one of Y_3 and Y_4 is positive, and the other is negative, then C and D are on opposite sides of the wall. We must establish whether or not the line joining C, D intersects the line segment AB. We find the equation of the line passing through C and D using $\dfrac{Y - Y_3}{X - X_3} = \dfrac{Y_3 - Y_4}{X_3 - X_4}$, provided that $X_3 \neq X_4$. If $X_3 = X_4$ then CD is parallel to the Y-axis and intersects AB if $0 \leq X_3 \leq b$. Otherwise we have $Y = Y_3 + m(X - X_3)$, where $m = \dfrac{Y_3 - Y_4}{X_3 - X_4}$. This line intersects the X-axis when $Y = 0$, that is

$$0 = Y_3 + m(X - X_3),$$

equivalently

$$-Y_3 + mX_3 = mX.$$

Now $m \neq 0$, because Y_3, Y_4 have opposite signs, so we can write

$$X = \frac{-Y_3 + mX_3}{m}.$$

The point of intersection is between A and B if

$$0 \leq \frac{mX_3 - Y_3}{m} = \frac{X_3 Y_4 - Y_3 X_4}{Y_4 - Y_3} \leq b.$$

We note that when $X_3 = X_4$ then

$$\frac{X_3Y_4 - Y_3X_4}{Y_4 - Y_3} = \frac{X_3(Y_4 - Y_3)}{Y_4 - Y_3} = X_3.$$

So the line CD intersects the line segment AB if

$$0 \leq \frac{X_3Y_4 - Y_3X_4}{Y_4 - X_3} \leq b,$$

whether $X_3 = X_4$ or otherwise.

We can summarise the algorithm as follows. People at points C and D are unsighted

$$\begin{aligned} \textit{if } Y_3Y_4 &< 0 \textit{ and } 0 \leq \frac{X_3Y_4 - Y_3X_4}{Y_4 - Y_3} \leq b \\ \textit{or if } Y_3 &= Y_4 = 0 \textit{ and } ((\textit{if } X_3 < 0 \textit{ and } X_4 > b) \textit{ or } (\textit{if } X_4 < 0 \textit{ and } X_3 > b)). \end{aligned}$$

We note that there is no need to convert back to the original coordinates. The answer we require is independent of the chosen axes.

6 Equations of Curves

We have determined the equation of a line in various ways:

(i) Given the gradient, m, and the intercept on the y-axis, c, we have

$$y = mx + c.$$

(ii) Given the gradient, m, and the point (x_1, y_1) through which the line passes, we have

$$y = mx - mx_1 + y_1.$$

(iii) Given two points (x_1, y_1) and (x_2, y_2) through which the line passes, we have

$$\frac{y_2 - y_1}{x_2 - x_1} = \frac{y - y_1}{x - x_1}.$$

All of these can be written in the form

$$\alpha x + \beta y + \gamma = 0,$$

where α, β, γ are constants.

For (i) $mx - y + c = 0$ $(\alpha = m,\ \beta = -1,\ \gamma = c)$,

for (ii) $mx - y + y_1 - mx_1 = 0$ $(\alpha = m,\ \beta = -1,\ \gamma = y_1 - mx_1)$

and for (iii) $(y_2 - y_1)x + (x_1 - x_2)y + x_2y_1 - x_1y_2 = 0$ $(\alpha = y_2 - y_1,\ \beta = x_1 - x_2,\ \gamma = x_2y_1 - x_1y_2)$.

In fact when the equation relating x and y is of the form $\alpha x + \beta y + \gamma = 0$ the geometric object is always a line. Some of the constants may be zero. For example, $x = 0$, $y - 1/2 = 0$ and $y + 2x = 0$ are equations describing lines, as is the equation $2x + 3y - 8 = 0$. For a point to lie on the line its coordinates must satisfy the equation of the line.

Example 12

Do the points $(3, 5)$ and $(6, 10)$ line on the line $y - 2x + 1 = 0$?

When $x = 3$, $y = 5$ we have $y - 2x + 1 = 5 - 6 + 1 = 0$, so $(3, 5)$ is on the line.

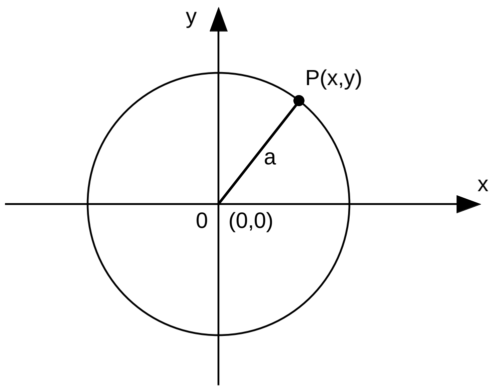
When $x = 6$, $y = 10$ we have $10-12+1 = -1 \neq 0$, so $(6, 10)$ is not on the line $y-2x+1 = 0$.

The equation of a line is a *first degree* curve, that is it only involves x to the power one and y to the power one. When the equation of a curve is not of first degree, in x and y, the graph will not be a straight line. We will now look at some curves of higher degree.

7 Circle

A *circle* is a curve on which points are all equidistant from a fixed point, that is its centre.

Consider a circle of radius a, centre at the origin.



Let $P(x, y)$ be any point on the circle. The distance between the origin and P is the radius, a. Therefore $a^2 = OP^2 = (x - 0)^2 + (y - 0)^2$ and the equation of the circle is

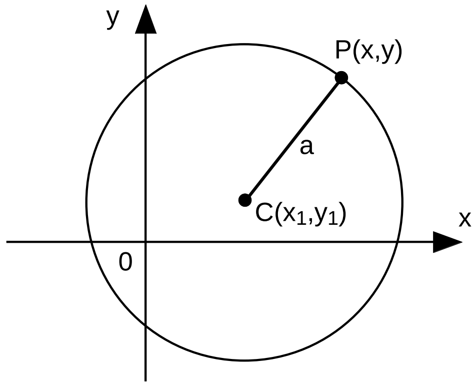
$$a^2 = x^2 + y^2.$$

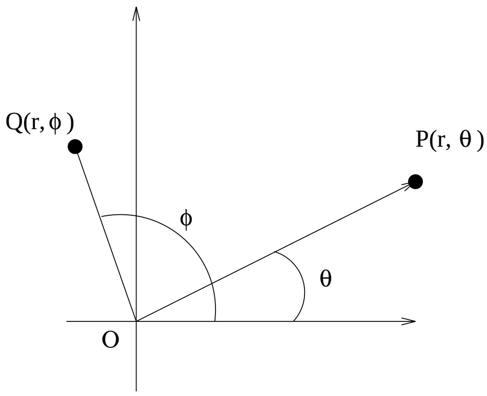
This is the equation of the circle, it is of degree two. We can determine the values of y, for points on the circle, for any given value of x and vice versa. For any value of x(or y) there are at most two corresponding y-coordinates (or x-coordinates).

If the centre is at the point (x_1, y_1), rather than at the origin, we have $a^2 = CP^2 = (x - x_1)^2 + (y - y_1)^2$ and the equation of the circle is

$$a^2 = (x - x_1)^2 + (y - y_1)^2.$$

In fact there is a simpler equation for a circle if we use *polar coordinates* rather than Cartesian coordinates. In polar coordinates, the position of a point is given as (r, θ), where r is the





distance from the origin (which is always positive) and θ is the angle between the horizontal axis and the line joining the origin to the point (θ is measured anticlockwise).

The different angle distinguishes the points $P(r,\theta)$ and $Q(r,\phi)$ which have the same radial distance, r. The relationship between polar and Cartesian coordinates is

$$x = r\cos\theta, \quad y = r\sin\theta.$$

So for the equation of a circle with centre at the origin, $x^2 + y^2 = a^2$, we have

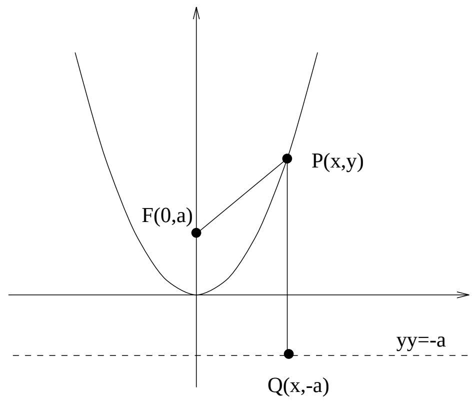
$$\begin{aligned} r^2\cos^2\theta + r^2\sin^2\theta &= a^2 \\ r^2(\cos^2\theta + \sin^2\theta) &= a^2 \\ r^2 &= a^2 \\ r &= a. \end{aligned}$$

In polar coordinates the equation of a circle is $r = a$, for all values of θ.

When working with circles it is often simpler to use polar coordinates. They also have an obvious use in navigation where position is given by distance and direction, e.g. 200 miles NW of Shetland.

8 Parabola

A *parabola* is a curve on which points are equidistant from a given point (its *focus*) and a line (its *directrix*). Consider a parabola which passes through the origin, with its focus on the y-axis and its directrix parallel to the x-axis.



By definition, we require $FP = PQ$. Now

$$FP^2 = (x-0)^2 + (y-a)^2$$

and

$$PQ^2 = (x-x)^2 + (y+a)^2.$$

So

$$FP^2 = PQ^2$$

if

$$x^2 + y^2 - 2ay + a^2 = y^2 + 2ay + a^2$$

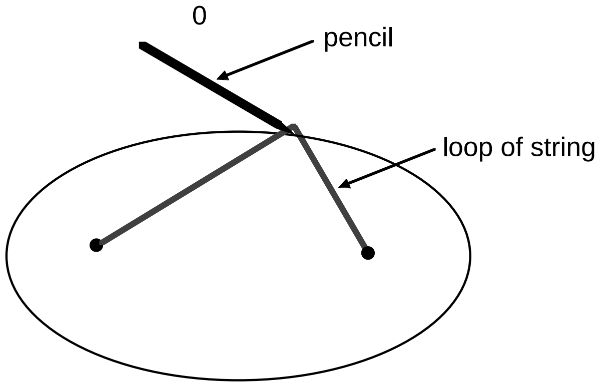
that is

$$x^2 = 4ay.$$

This is the equation of a parabola which passes through the origin and has its focus at $(0, a)$. It is symmetric about the y-axis; for any given value of y there are at most two points on the parabola, one with x-coordinate x_1, say, the other with x-coordinate $-x_1$. On the other hand, for a given value of x there is only one corresponding value for y. Like the equation of a circle it is also a second degree curve.

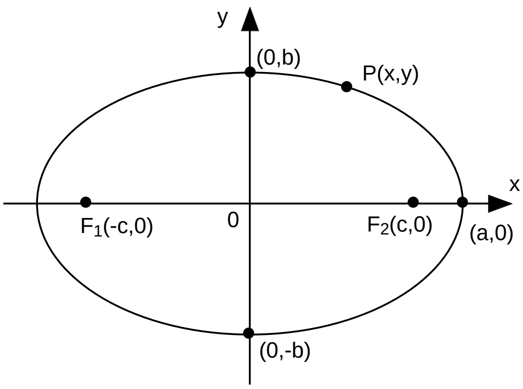
9 Ellipse

An *ellipse* is a curve on which points whose distances from two fixed points (the *foci*) have a constant sum. Think of a piece of string fixed at its ends. Place a pencil on the loop of the string and pull taut. As the pencil is moved it traces out an ellipse.



The points where the string is fixed are the foci. The length of the string is the sum of the distances from the foci.

We now determine the equation of an ellipse. Choose axes with the origin mid-way between the foci and the x-axis passing through them. Let the point where the ellipse cuts the positive x-axis be $(a, 0)$.



We have that, for an ellipse, $PF_1 + PF_2 = a$ constant. When P is the point $(a, 0)$, $PF_1 = a+c$ and $PF_2 = a - c$. Therefore $PF_1 + PF_2 = 2a$ and hence the constant is equal to $2a$. For the general point $P(x, y)$ we have

$$\begin{aligned} PF_1 + PF_2 &= 2a \\ \sqrt{(x+c)^2 + y^2} + \sqrt{(x-c)^2 + y^2} &= 2a \\ \sqrt{(x+c)^2 + y^2} &= 2a - \sqrt{(x-c)^2 + y^2}. \end{aligned}$$

Squaring

$$(x+c)^2 + y^2 = 4a^2 + (x-c)^2 + y^2 - 4a\sqrt{(x-c)^2 + y^2}$$

and simplifying

$$cx - a^2 = -a\sqrt{(x-c)^2 + y^2}.$$

Squaring again

$$c^2x^2 + a^4 - 2a^2cx = a^2((x-c)^2 + y^2).$$

Divide through by a^2, then

$$\begin{aligned}
\frac{c^2}{a^2}x^2 + a^2 - 2cx &= x^2 - 2cx + c^2 + y^2 \\
x^2\left(\frac{c^2}{a^2} - 1\right) - y^2 &= c^2 - a^2 \\
x^2\frac{(c^2-a^2)}{a^2} - y^2 &= c^2 - a^2 \\
\frac{x^2}{a^2} - \frac{y^2}{c^2-a^2} &= 1 \\
\frac{x^2}{a^2} + \frac{y^2}{a^2-c^2} &= 1.
\end{aligned}$$

As $a > c$ then $a^2 > c^2$, so $a^2 - c^2 > 0$. We can replace $a^2 - c^2$ by b^2, which is always positive.

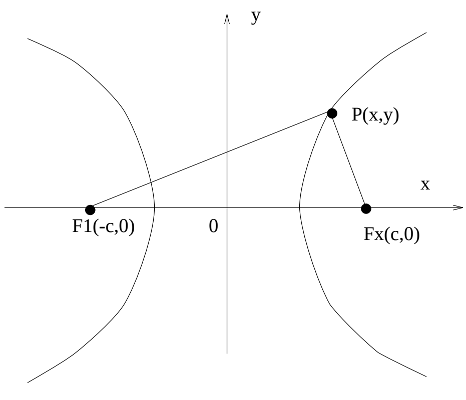
The equation of an ellipse is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

We note that when $x = 0$, $y = \pm b$, so $(0, b)$ is the point where the ellipse cuts the positive y-axis and $(0, -b)$ where it cuts the negative y-axis. The ellipse is symmetric about both axes.

10 Hyperbola

An *hyperbola* is a curve on which every point is such that its distances from two fixed points (the foci) have a constant difference.



Again we choose axes with the foci at $(c, 0)$ and $(-c, 0)$ with the x-axis passing through them. Let $(a, 0)$ be the point where the hyperbola cuts the positive x-axis. We require

$PF_1 - PF_2 = constant$. As $P(a,0)$ is on the hyperbola we have

$$PF_1 - PF_2 = a + c - (c - a) = 2a.$$

There is another branch of the hyperbola when $PF_2 - PF_1 = 2a$. Hyperbolas have two symmetric branches. Let $P(x,y)$ be any point on the hyperbola. Then

$$\begin{aligned} PF_1 - PF_2 &= \pm 2a \\ \sqrt{(x+c)^2 + y^2} - \sqrt{(x-c)^2 + y^2} &= \pm 2a. \end{aligned}$$

Simplifying this, as in the equation of the ellipse, gives

$$\frac{x^2}{a^2} + \frac{y^2}{a^2 - c^2} = 1.$$

This is the same equation as that of the ellipse but *here $a < c$ so $a^2 - c^2 < 0$ and we let $b^2 = -(a^2 - c^2) > 0$.*

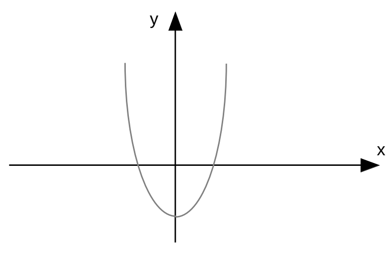
The equation of an hyperbola is

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$

The hyperbola is symmetric about both axes.

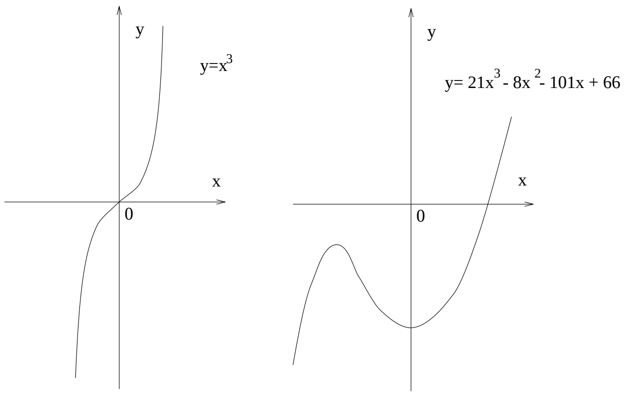
11 Other curves

We have already seen an example of a degree two curve in which y is of degree one and x is of degree two: a parabola. We can write $y = Ax^2 + Bx + C$ and often refer to y as a 'quadratic in x.' This can be thought of as a graph of y against x.



It has one turning point at most, that is one change of direction. In the diagram as x increases y decreases until $x = 0$, then y increases as x increases. The sign of the coefficient A determines the orientation; if A is positive the curve forms a 'bowl', if A is negative the 'bowl' is upside down. For a given value of y there are at most two points on the curve; that is two values of x satisfy the equation when y is given. At the turning point these two values of x coincide.

The higher the degree of the curve, the more turning points (changes of direction) it has. In general a curve of degree n has $n-1$ turning points. Some turning points may coincide; this is known as a point of inflexion. The following example shows two cubic (degree 3) curves.



12 Intersection of curves

A point at which two curves meet is called a point of intersection. At this point the x and y-coordinates must satisfy the equations describing both curves. The higher the degree of the curve, the more points of intersection it can have with another curve. For example, two lines can have at most one point of intersection; a line and a degree three curve can have up to three points of intersection. Some curves do not intersect at all.

Example 14

Consider the intersection of the line $y = 3x$ (line is a curve of degree one) and the second degree curve $y = x^2 + 8x + 6$. We solve the equations simultaneously to find values for x and y that satisfy both equations

$$y = 3x \tag{12.1}$$
$$y = x^2 + 8x + 6. \tag{12.2}$$

Substitute $y = 3x$ from equation (12.1) into equation (12.2)

$$\begin{aligned} 3x &= x^2 + 8x + 6 \\ 0 &= x^2 + 5x + 6 \\ 0 &= (x+3)(x+2) \end{aligned}$$

This is satisfied when $x = -3$ or $x = -2$. When $x = -3$, $y = -9$, from (12.1) and when $x = -2$, $y = -6$. Therefore $(-3, -9)$ and $(-2, -6)$ are points of intersection of the two curves.

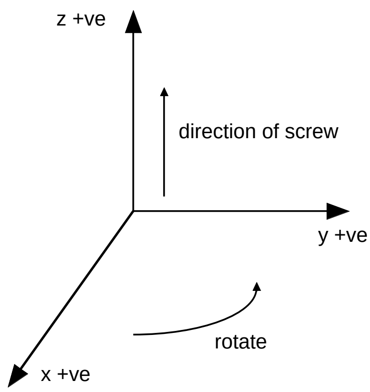
There cannot be more than two points where a line intersects a curve of degree two. Of course curves of higher degree can have more points of intersection with other curves. The technique for finding the points of intersection is to solve the two equations of the curves simultaneously, whatever their degree.

13 Three-dimensional coordinate geometry

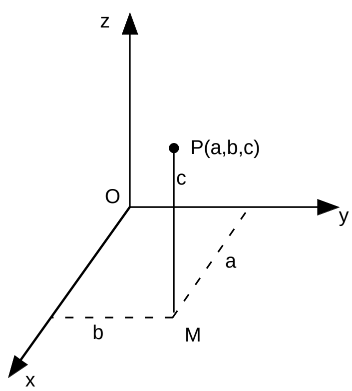
So far we have considered two-dimensional geometry, that is geometry relating to points in a plane. We are now going to consider three-dimensional geometry, that is geometry of points not all of which are in the same plane. A three-dimensional Cartesian frame of reference consists of a fixed point O (the origin) and three mutually perpendicular axes Ox, Oy, Oz.

In three-dimensional geometry we fix the position of a point by projecting it onto the x-y plane, thus giving coordinates x, y as in two-dimensional axes and for the third dimension we also require its height, z, above this plane. It is possible to take the positive z-axis up or down, both directions being perpendicular to Ox and Oy. We use the convention that gives a right-handed set of axes.

A simple way to find the direction of the z-axis: Imagine turning a screwdriver from the positive x-axis towards the positive y-axis, the direction of motion of the screw being along the positive z-axis.



A point has coordinates (x, y, z), the order in which the coordinates are specified is important.



In the diagram above, point M is the foot of the perpendicular from the point P to the x-y plane, that is the projection of P in the x-y plane. So point M has coordinates $(a, b, 0)$. Point P is at a height c above the x-y plane, therefore $P = (a, b, c)$.

14 Distance between two points in three-dimensional space

The distance from the origin to the point P is given by

$$\begin{aligned} OP^2 &= OM^2 + MP^2 \\ &= a^2 + b^2 + c^2. \end{aligned}$$

We use this relationship to deduce an expression for the distance between two points $A(x_1, y_1, z_1)$ and $B(x_2, y_2, z_2)$. Assume that the axes have been translated so that A is the new origin. The distance AB is the same whatever axes are used. Point $A(x_1, y_1, z_1)$ becomes $A'(0,0,0)$ with reference to the new axes and $B(x_2, y_2, z_2)$ becomes $B'(x_2 - x_1, y_2 - y_1, z_2 - z_1)$. So the length AB is given by the distance from the origin (A') to the point B', therefore

$$AB^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2.$$

As in two dimensions this formula represents the distance between two points (x_1, y_1, z_1) and (x_2, y_2, z_2) so long as due account is taken of the signs of the coordinates.

Example 15

Find the distance from the origin, O, to the point $P(4, -1, 2)$.

$$OP^2 = 16 + 1 + 4, \quad OP = \sqrt{21}.$$

Example 16

What is the distance between $A(2, 1, 3)$ and $B(1, 2, 3)$?

$$AB^2 = (1-2)^2 + (2-1)^2 + (3-3)^2 = 2; \quad AB = \sqrt{2}.$$

Example 17

Find the length AB if A is the point $(2, 1, 1)$ and B is $(-3, -2, -1)$.

$$\begin{aligned} AB^2 &= (-3-2)^2 + (-2-1)^2 + (-1-1) \\ &= 25 + 9 + 4 = 38; \quad AB = \sqrt{38}. \end{aligned}$$

15 Midpoint of a line joining two points in three-dimensional space

Rule: The midpoint of a line joining the two points $A = (x_1, y_1, z_1)$ and $B = (x_2, y_2, z_2)$ is

$$M = (\tfrac{1}{2}(x_1 + x_2), \tfrac{1}{2}(y_1 + y_2), \tfrac{1}{2}(z_1 + z_2)).$$

Example 18

The midpoint of a line joining $A(-3, 1, 4)$ and $B(-2, 4, 3)$ is

$$M = \left(\frac{-5}{2}, \frac{5}{2}, \frac{7}{2}\right).$$

Example 19

If $M = (-1, 4, 2)$ is the midpoint of AB and A has coordinates $(1, 2, 3)$ find the coordinates of B.

Let $B = (x, y, z)$. Then the coordinates of the midpoint are as follows:

x-coordinate: $-1 = \frac{1}{2}(1 + x)$ which implies $x = -3$;

y-coordinate: $4 = \frac{1}{2}(2 + y)$ which implies $y = 6$;

z-coordinate: $2 = \frac{1}{2}(3 + z)$ which implies $z = 1$.

Hence the coordinates of the point B are $(-3, 6, 1)$.

16 Equations in three dimensions

In two-dimensional space equations represent curves, whereas in three-dimensional space they define surfaces. No single equation can represent a curve in three dimensions; two equations are required, the curve being the intersection of two surfaces.

A first degree equation in two dimensions such as $\alpha x + \beta y + \gamma = 0$, where α, β, γ are constants, represents a line. In three dimensions the equation $Ax + By + Cz + D = 0$, where A, B, C, D are constants, is that of a plane. If two planes intersect, they do so in a line. However, no single equation defines this line in three-dimensional space.

Example 20

The equations $2x + 3y - z + 6 = 0$ and $x - 2y + z - 4 = 0$ represent planes. We can use one of these equations to eliminate x (or y or z) from the other equation. The resulting equation is that of a line in the y-z (or x-z or x-y) plane, which is the projection of the line of intersection in that plane. So, for example, eliminating y we have $7x + z = 0$. This is the equation of a line in the $x - z$ plane; it is not the equation of the line of intersection of the two planes.

Rule: Any point which lies in two planes must lie on the line of intersection of the two planes.

Example 21

Show that $A(1, -5, -7)$ and $B(-1, 1, 7)$ lie on the line of intersection of the planes $2x + 3y - z + 6 = 0$ and $x - 2y + z - 4 = 0$.

We need to show that A and B lie in both planes. For the point A, $x = 1$, $y = -5$ and $z = -7$. Substituting these values into the two equations we have

$$2 - 15 + 7 + 6 = 0$$

and

$$1 + 10 - 7 - 4 = 0.$$

So point A lies on the line of intersection. Similarly we can show that point B lies on the line of intersection but what about the point $C(0, 1, 6)$? Here $x = 0$, $y = 1$ and $z = 6$, which when substituted into the equations of the planes gives

$$0 + 3 - 6 + 6 = 3 \neq 0$$

and

$$0 - 2 + 6 - 4 = 0.$$

Therefore point C lies in the plane with equation $x - 2y + z - 4 = 0$ but not in the plane with equation $2x + 3y - z + 6 = 0$. It does not lie on the line of intersection of the two planes.

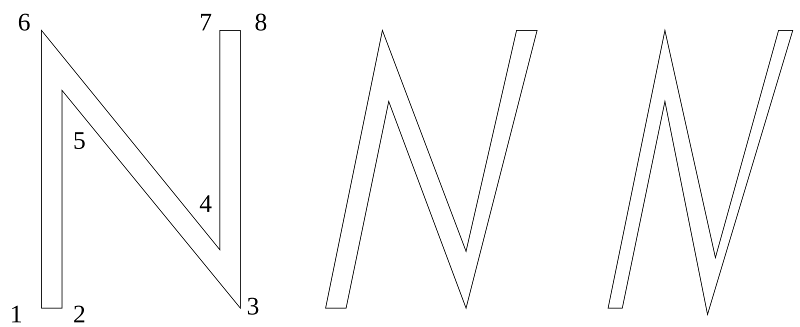
Equations of higher degree in three dimensions also represent surfaces. In general the higher the degree the more convoluted the surface. Two such equations are necessary to define a curve in three dimensions. For example, the second degree equation $x^2 + y^2 + z^2 = a^2$ represents the surface of a sphere of radius 'a' with centre at the origin. When a plane intersects this sphere, the curve of intersection is a circle in three dimensional space – the two equations representing the sphere and the plane define the circle. Similarly $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$ is the equation of the surface of an ellipsoid. The sphere and the ellipsoid are examples of closed surfaces; their counterparts in two dimensions, the circle and the ellipse, are closed curves. Of course there are many equations representing surfaces that are not closed.

17 Computer graphics

Computer graphics have widespread applications in business and industry as well as their pervasive use in computer games. As a computer screen is two-dimensional all graphical images to be displayed must be confined to a plane. Three-dimensional objects are conveyed by perspective images, that is a two-dimensional image that gives the impression of three dimensions. Coordinate geometry is fundamental in creating the image to be displayed.

Graphical images can be constructed by specifying a number of points, giving details of how points are to be connected and by providing information on colour (and shading) for any enclosed regions so formed. Often curves are approximated by line segments. The shorter the line segments the better the approximation to the actual curve. We now consider an example in which points and lines are used to describe a character which can then be transformed algebraically using shearing and scaling operations.

Example 22



Regular N　　Slanted N　　Composite transformation of N

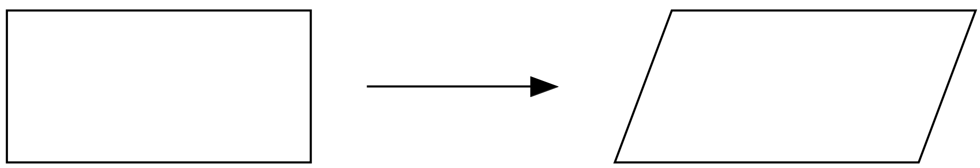
The capital letter N in the picture is described by eight points (or vertices) connected by straight lines. In order to draw this figure graphically we need to store the coordinates of the vertices and to specify which ones are joined by a line. The image is defined by the points

$$
\begin{aligned}
&P_1(x_1, y_1)\\
&P_2(x_2, y_2)\\
&\vdots\\
&P_8(x_8, y_8)
\end{aligned}
$$

where P_1 is the point labelled 1 in the figure, etc. and the vertices are joined in the sequence 1,2,5,3,8,7,4,6,1. The image is drawn by plotting the points and connecting P_1 to P_2, P_2 to $P_5, ..., P_6$ to P_1, using lines. For convenience choose axes with origin at P_1, x-axis passing through P_1, P_2 (and of course P_3) and y-axis passing through P_1, P_6. Then the following points could represent the letter N:

$$
\begin{aligned}
&P_1(0, 0)\\
&P_2(0.5, 0)\\
&P_3(6, 0)\\
&P_4(5.5, 1.58)\\
&P_5(0.5, 6.42)\\
&P_6(0, 8)\\
&P_7(5.5, 8)\\
&P_8(6, 8)
\end{aligned}
$$

We may wish to make the N more like an italic N by performing a *shear* transformation on the vertices. In a shear transformation the shape is deformed in the direction of the shear, as illustrated below.



The order for joining the points remains unchanged. We want to keep the same vertical height for our image so we perform a shear on the x-coordinates only. We also want the x-coordinate to move further to the right for vertices with larger y-coordinates. We choose a shear of $0.25y$ in the positive x-direction. So if the y-coordinate was zero for the original vertex the point will not be moved at all. So points P_1, P_2, P_3 will remain fixed when the image of N undergoes the given shear transformation. That is, the base of the character remains fixed. This is usually what is required when manipulating characters. However, a point like P_6 which is on the y-axis at $(0, 8)$ will move to the point $(2, 8)$. The x-coordinate is increased by one quarter of the y-coordinate value.

The sheared figure is the second of the three N's pictured earlier, the 'slanted N'. The new coordinates (X, Y) are related to the original coordinates (x, y) by the equations:

$$
\begin{aligned}
X &= x + 0.25y\\
Y &= y.
\end{aligned}
$$

and the transformed vertices are:

$$
\begin{aligned}
&P_1'(0,0)\\
&P_2'(0.5,0)\\
&P_3'(6,0)\\
&P_4'(5.895,1.58)\\
&P_5'(2.105,6.42)\\
&P_6'(2,8)\\
&P_7'(7.5,8)\\
&P_8'(8,8)
\end{aligned}
$$

Note that all y-coordinates remain unchanged. Once the new vertices have been calculated the image can be re-drawn by connecting appropriate vertices using straight lines. Perhaps the second N is too wide for an italic N so we wish to shrink the width. This can be done by performing a *scale* transformation. We only want to scale the x-coordinates leaving the y-coordinates unchanged. For example we could scale x by a factor of 0.75 such that the point $P_3'(6,0)$ becomes $P_3''(4.5,0)$. Similarly for other points, then the new vertices are:

$$
\begin{aligned}
&P_1''(0,0)\\
&P_2''(0.35,0)\\
&P_3''(4.5,0)\\
&P_4''(4.42125,1.58)\\
&P_5''(1.57875,6.42)\\
&P_6''(1.5,8)\\
&P_7''(5.625,8)\\
&P_8''(6,8)
\end{aligned}
$$

The third N shows the result of the composite transformation, a shear followed by scaling. Mathematically these can be combined into one operation. The governing equations are

$$
\begin{aligned}
X &= 0.75(x+0.25y),\\
Y &= y,
\end{aligned}
$$

where (x,y) are the original coordinates and (X,Y) are the sheared and scaled coordinates.

Note that in this example P_8 has returned to its start position and in fact we have retained the image within a box defined by the vertices P_1, P_3, P_6, P_8. This is precisely the type of manipulation that is required when dealing with text.