17 Appendix 1: Graphs and data analysis

Representing data graphically

There are many quantities that can be measured in a physics experiment. Usually, all but two of them are **controlled** so that they do not change during the course of the experiment. Then one quantity (called the **independent variable**) is deliberately varied or changed and the effect on one other quantity (called the **dependent variable**) investigated.

The best way to analyse the results of such experiments is often to **plot** (draw) a graph. Looking at a graph is a good way to identify a pattern, or trend, in numerical data. Graphs can also provide extra information – **gradients** (slopes), **intercepts** and the **areas** under graphs often have important meanings. Figure 17.1 illustrates the terminology associated with graphs.

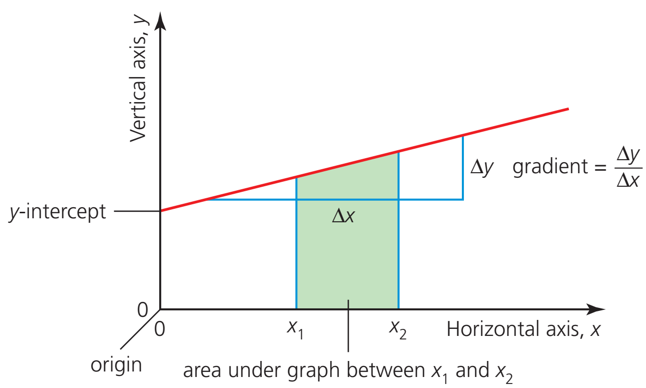


Figure 17.1 Terminology for graphs

Drawing graphs

The drawing of good-quality graphs is a very important skill in physics. The following points should be remembered when drawing a graph.

- The larger a graph is, the more precisely the points can be plotted. A simple rule is that the graph should occupy at least half the available space (in each direction).
- Each **axis** should be labelled with the quantity and the unit used (e.g. force/N, speed/m s^{-1}). Lists of results in tables should be similarly labelled. For example, if you want to record a mass of 5 g as the number five on the axes of a graph, then labelling the axes as mass/g indicates that you have divided 5 g by g to get five.
- The *independent variable* is usually plotted on the horizontal axis and the *dependent variable* on the vertical axis. Sometimes, the choice of what to plot on each axis is made so that the gradient of the graph has a particular meaning. If time is one of the variables, it is nearly always plotted on the horizontal axis.
- The **scales** chosen should make plotting the points and interpreting the graph easy. For example, five divisions might be used to represent 10 or 20, but not 7 or 12.
- Usually, both scales should start at zero, so that the point (0, 0), the **origin**, is included on the graph. This is often important when interpreting a graph. However, this is not always sensible, especially if it would mean that all the readings were restricted to a small part of the graph. Temperature scales in °C do not usually need to start at zero.
- Data **points** should be neat and small. If points are used (rather that crosses), drawing a small circle around them can make sure that they are not overlooked, especially if the line goes through them.
- The more points that can be plotted, the more precisely the line representing the relationship can be drawn. At *least* six points are usually needed, although this may not always be possible.

- When all the points have been plotted, a pattern will usually be clear and a **line of best fit** can be drawn (see Figure 17.2 for two correct examples and one wrong example, in which the points are represented by error bars). These lines are sometimes called *trend lines*. Trend lines may be straight (drawn with a ruler) or a curve. (A straight line is described as being **linear**.) The line should be smooth and thin. Bumpy lines that try to pass through, or near, all the points show that the person drawing the line did not understand that points cannot be perfectly placed, and that there is uncertainty in all measurements. Typically there will be about as many points above a line of best fit as there are below the line. Points on a graph should *never* be joined by a series of straight lines.
- Drawing graphs by hand is a skill that all students should practise. However, knowing how to use a computer program to plot graphs is also a very valuable and time-saving skill (especially for investigation work). A graph generated by a computer (or graphic display calculator) must be judged by the same standards as a hand-drawn graph and sometimes their best-fit lines are not well placed.

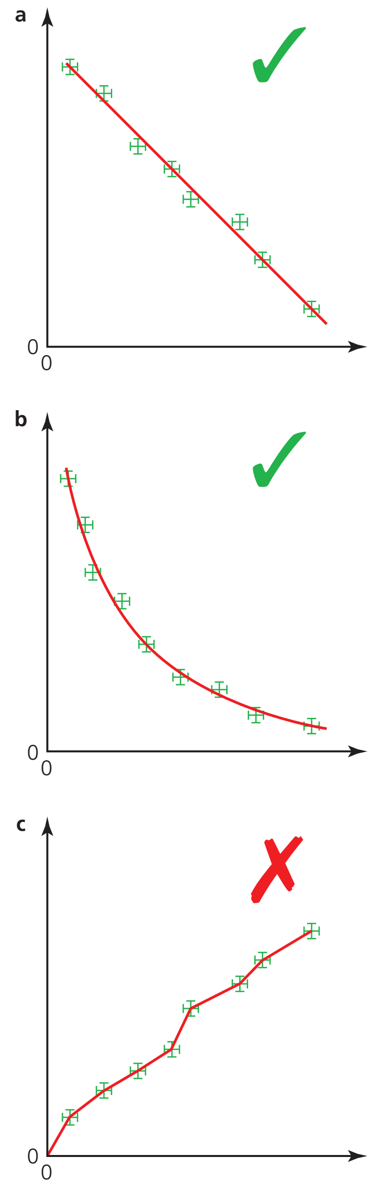


Figure 17.2 Right and wrong ways to draw best-fit lines

Extrapolating and interpolating

A line of best fit is usually drawn to cover a specific *range* of measurements recorded in an experiment, as shown in Figure 17.3. If we want to predict other values *within* that range, we can do that with confidence. The diagram indicates how a value for y can be determined for a chosen value of x. This is called **interpolation**.

If we want to predict what would happen *outside* the range of measurements (**extrapolation**) we need to *extend* the line of best fit. Lines are often extrapolated to see if they pass through the origin, or to find an intercept, as shown in Figure 17.3.

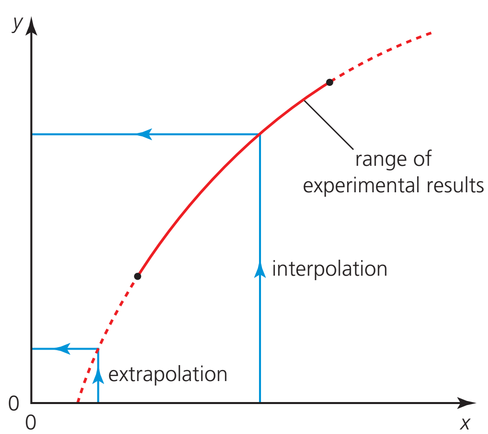


Figure 17.3 Interpolating and extrapolating to find the intercept on the *y*-axis

Predictions made by extrapolation should be treated with care, because it may be wrong to assume that the behaviour seen within the range of measurements also applies outside that range.

Proportionality

The simplest possible relationship between two variables is that they are **proportional** to each other (sometimes called *directly* proportional). This means that if one variable, say x, doubles, then the other variable, y, also doubles; if y is divided by five, then x is divided by five; if x is multiplied by 17, then y is multiplied by 17 etc. In other words, the ratio of the two variables (x/y or y/x) is constant. Proportionality is shown as:

$$y \propto x$$

Many basic experiments are aimed at investigating if there is a proportional relationship between two variables, and this is usually best checked by drawing a graph.

If two variables are (directly) proportional, then their graph will be a **straight line passing through the origin** (Figure 17.4). It is important to stress that a linear graph that does not pass through the origin *cannot* represent proportionality (Figure 17.5).

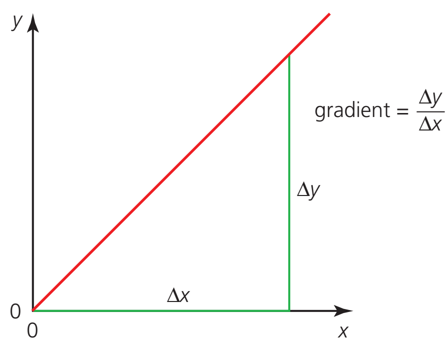


Figure 17.4 A proportional relationship

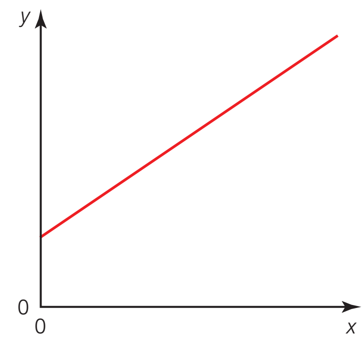


Figure 17.5 A linear relationship that is not proportional does not pass through the origin

Gradients of lines

The gradient of a line is given the symbol m and it is calculated by dividing a change in y, Δy, by the corresponding change in x, Δx, as shown in Figure 17.4. (A delta sign, Δ, is used to represent a change of a quantity.)

$$m = \frac{\Delta y}{\Delta x}$$

It is important to note that a *large* triangle should be used when determining the gradient of a line, because the percentage uncertainty will be less when using larger values.

The gradients of many lines have a physical meaning – for example the gradient of a mass – volume graph equals the density of the material.

The gradient of a curved line, such as that shown in Figure 17.6, is constantly changing. The gradient of the curve at any point, P, can be determined from a **tangent** drawn to the curve at that point. (In mathematics, if the equation of a line is known, then the gradient at any point can be determined by a process called *differentiation*.)

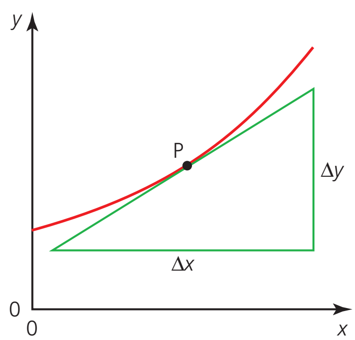


Figure 17.6 Finding the gradient of a curve at point P with a tangent

Worked examples

1 Figure 17.7 shows a best-fit line produced from an experiment in which the masses and volumes of different pieces of the same metal alloy were measured.

a Calculate a value for the density of the alloy, which is equal to the gradient of the line.

b Suggest why the graph does not pass through the origin.

c Explain why using the gradient to find the density is a much better method than just calculating a value from one pair of readings of mass and volume.

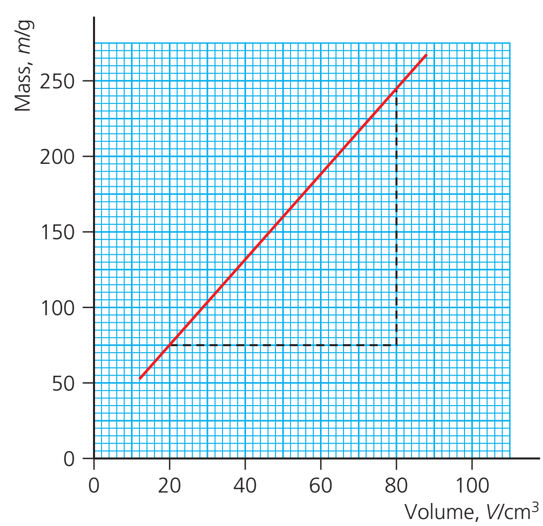


Figure 17.7 Graph of mass m against volume V of different pieces of alloy

a Using the triangle shown on the graph:

$$\text{gradient} = \text{density} = \frac{\Delta m}{\Delta V} = \frac{245 - 75}{80 - 20} = 2.8\,\text{g cm}^{-3}$$

b The instrument used to measure mass had a zero offset error (of about +20 g).

c Individual readings may be inaccurate. The best-fit line reduces the effect of random errors and the zero offset error does not affect the result of the calculation.

2 Figure 17.8 shows a distance – time graph for an accelerating car. Determine the speed of the car after 4 s (equal to the gradient of the line at that time).

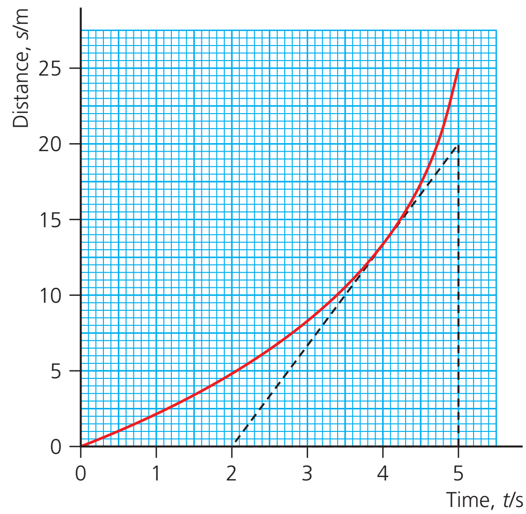


Figure 17.8 Graph of distance *s* against time *t* for an accelerating car

The sloping dashed line is a tangent to the curve at $t = 4$ s.

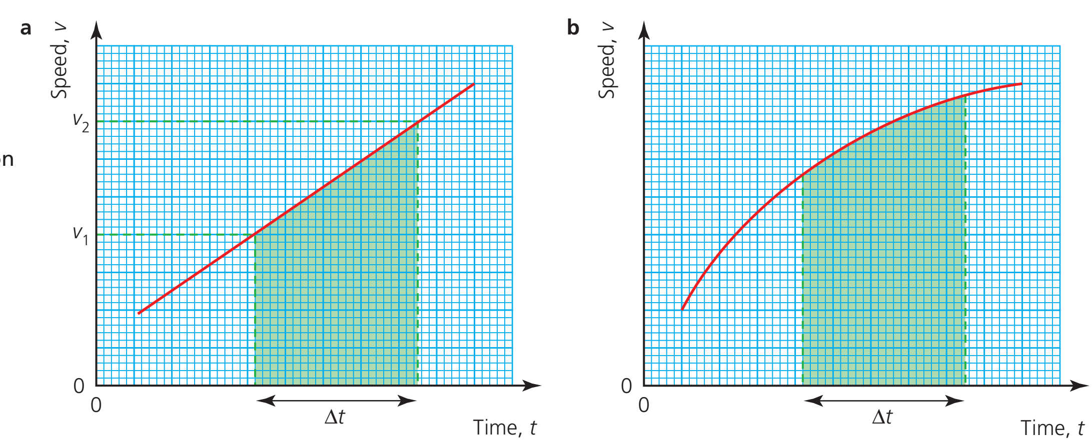
$$\text{gradient} = \text{speed} = \frac{\Delta s}{\Delta t} = \frac{20 - 0}{5.0 - 2.0} = \frac{20}{3.0} = 6.7\,\text{m s}^{-1}$$

Areas under graphs

The area under many graphs has a physical meaning. As an example, consider Figure 17.9a, which shows part of a speed – time graph for a vehicle moving with constant acceleration. The area under the graph (the shaded area) can be calculated from the average speed, given by $\frac{(v_1 + v_2)}{2}$, multiplied by the time, Δt. The area under the graph is therefore equal to the distance travelled in time Δt.

In Figure 17.9b a vehicle is moving with a changing (decreasing) acceleration, so that the graph is curved, but the same rule applies – the area under the graph (shaded) represents the distance travelled in time Δt.

Figure 17.9 Area under a speed–time graph for **a** constant acceleration and **b** changing acceleration



The area in Figure 17.9b can be estimated in a number of different ways, for example by counting small squares, or by drawing a rectangle that appears (as judged by eye) to have the same area. (If the equation of the line is known, it can be calculated using the process of *integration*.)

Worked example

3 Figure 17.10 represents the motion of a train that travels at a constant speed for 30 s and then decelerates for 60 s. Calculate the distance travelled in 90 s (equal to the area under the graph).

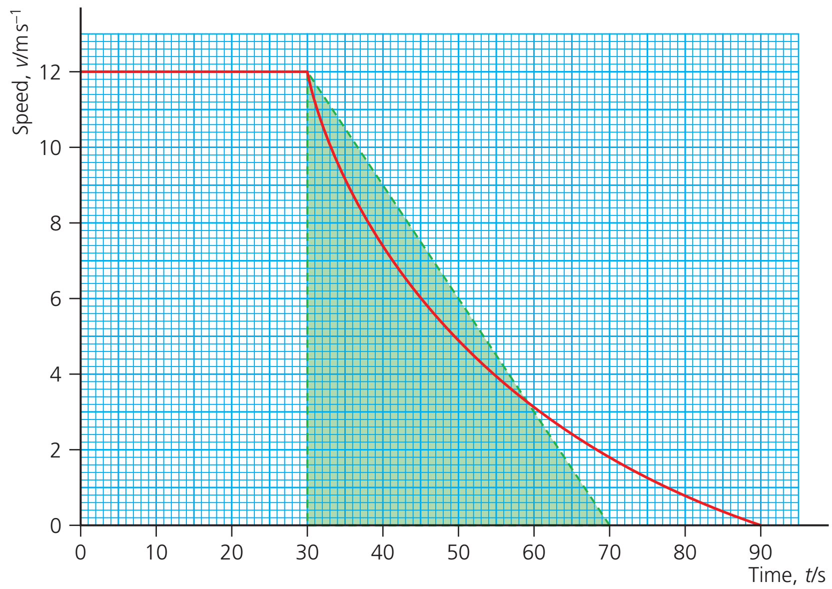


Figure 17.10 Graph of velocity, *v*, against time, *t*, for a train

The area under the graph up to a time of 30 s = 12 × 30 = 360 m.

The area under the graph between 30 s and 90 s can be estimated from the shaded triangle, which has been drawn so that its area appears to be the same as the area under the curved line:

area = $\frac{1}{2}$ × 12 × (70 − 30) = 240 m

total distance (area) = 360 + 240 = 600 m

The usefulness of straight-line graphs

Straight lines are much easier to understand and analyse than curved lines, but when directly measured experimental data are plotted against each other (x and y, for example), the lines are often curved rather than linear.

Data that give an $x - y$ curve can be used to draw other graphs to check different possible relationships. For example:

- A graph of y against x^2 could be drawn to see if a straight line through the origin is obtained, which would confirm that y was proportional to x^2.
- A graph of y against $\frac{1}{x}$ that passed through the origin would confirm that y was proportional to $\frac{1}{x}$ (in which case x and y are said to be **inversely proportional** to each other).
- A graph of y against $\frac{1}{x^2}$ passing through the origin would represent an **inverse square relationship**.

Figure 17.11 shows graphs of the most common relationships.

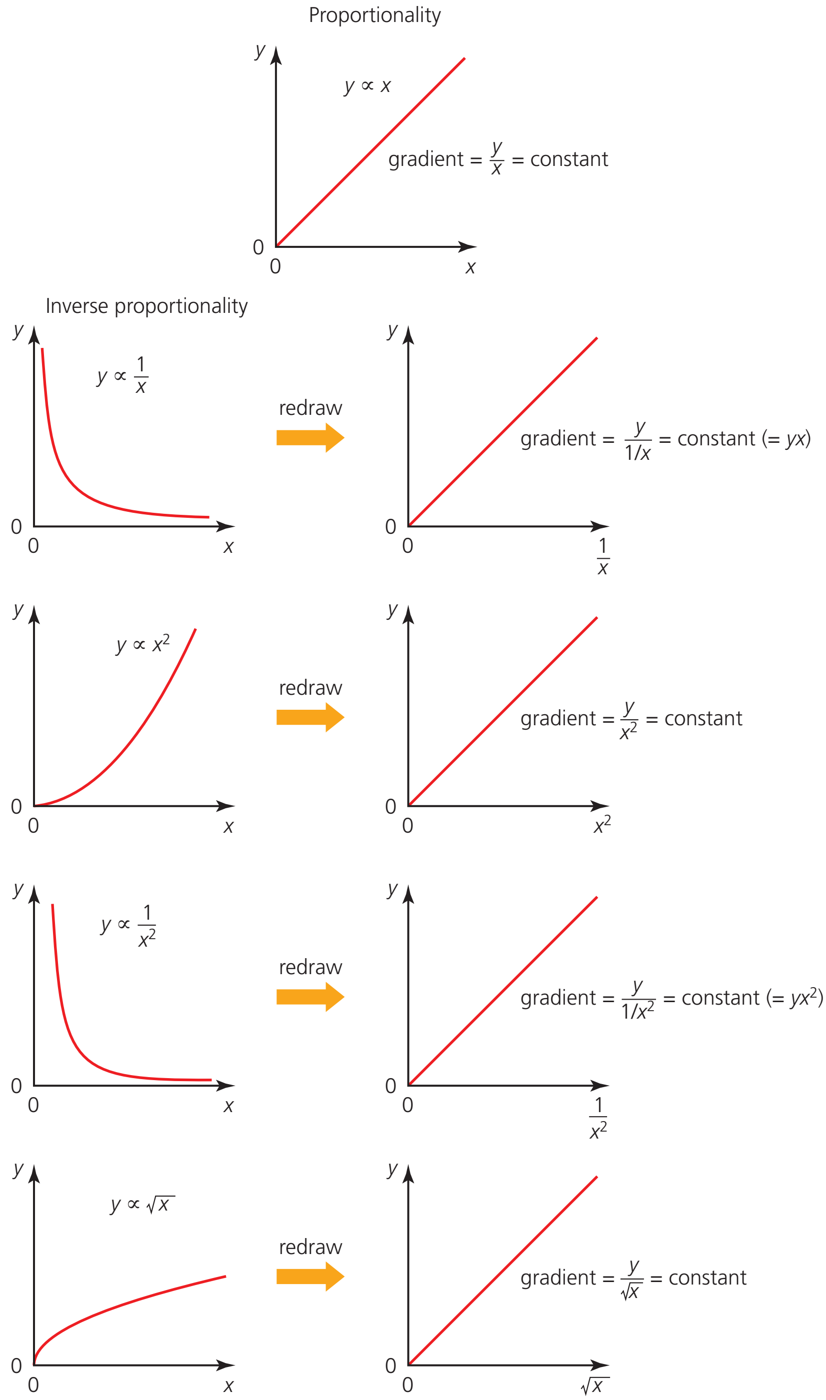


Figure 17.11 Some common graphical relationships, showing how curves can be replotted to produce straight lines

Worked example

4 In an internal assessment students were asked to investigate one factor that affected the deflection of a wooden beam fixed only at one end (a cantilever), and which had a mass (load) hanging from somewhere on the part that extended from the bench top (Figure 17.12).

A student listed the following variables: (i) type of wood, (ii) thickness of wood, (iii) width of wood, (iv) length of wood from where it was fixed, L, (v) position of load, (vi) mass of load. He decided to investigate how the deflection, D, depended on the length, L, keeping all the other variables constant. His results are shown in Table 17.1 (for simplicity, uncertainties have not been included).

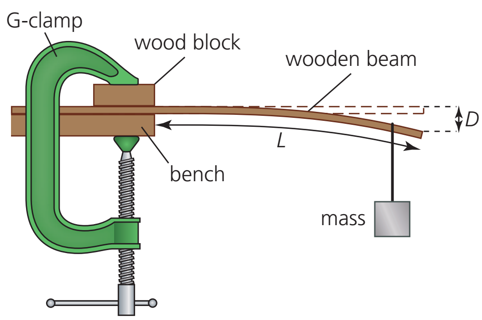


Figure 17.12

Table 17.1

L/cm	D/cm
30.0	0.8
40.0	2.0
50.0	3.7
60.0	6.5
70.0	10.8
80.0	15.8
90.0	22.2
100.0	31.2

a A graph of the raw data produces a curved line, so the student thought that maybe the deflection was proportional to the length squared or the length cubed. Perform numerical checks on the data to see if either of these possibilities is correct.

b Plot a suitable graph to confirm the correct relationship.

a If the deflection, D, is proportional to the length, L squared, $\frac{L^2}{D}$ (or $\frac{D}{L^2}$) will be constant, within the limits of experimental uncertainties. Calculations produce the following results (all × 10^2 cm): 11.0, 8.0, 6.8, 5.5, 4.5, 4.1, 3.6, 3.2. These values are getting smaller for longer lengths, and are clearly *not* constant.

If the deflection, D, is proportional to the length, L cubed, $\frac{L^3}{D}$ (or $\frac{D}{L^3}$) will be constant, within the limits of experimental uncertainties. Calculations produce the following results (all × 10^4 cm^2): 3.4, 3.2, 3.4, 3.3, 3.2, 3.2, 3.3, 3.2. These values are all very similar (within 3% of their average), confirming that the deflection is proportional to the length cubed.

b See Figure 17.13. A graph of D against L^3 produces a straight line through the origin. Note that it would have been better if the student had used lengths such that the points were spread out evenly along the line.

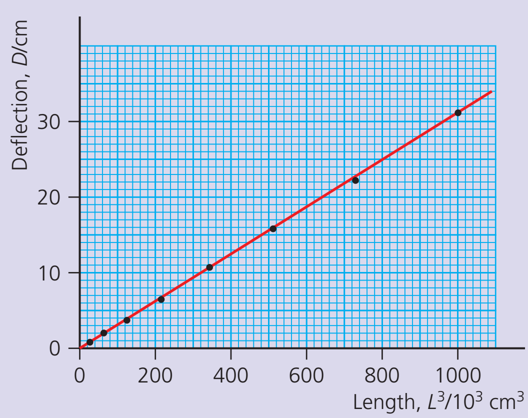


Figure 17.13

Equation of a straight line

All linear graphs can be represented by an equation of the form:

$$y = mx + c$$

where m is the gradient and c is the value of y when $x = 0$, known as the ***y*-intercept** (Figure 17.14).

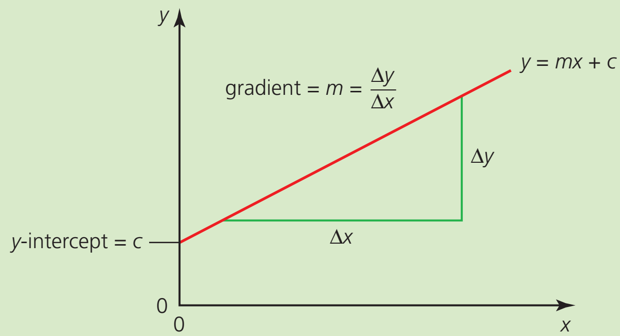


Figure 17.14 Graph of $y = mx + c$

Once a linear graph has been drawn, values for the gradient and the y-intercept can be determined and the results used to produce a mathematical equation to describe the relationship.

Worked example

5 Experimental data connecting two variables, x and y, are represented by the graph in Figure 17.15. Take measurements from the graph to enable you to write an equation to represent the relationship.

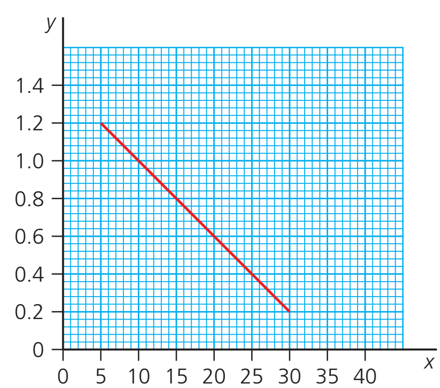


Figure 17.15

The gradient of the line, m, is $\frac{0.2 - 1.2}{30 - 5} = -0.04$ and the y-intercept, c, is 1.4.

Substituting into $y = mx + c$ we get:

$$y = -0.04x + 1.4$$

(which could be rewritten as $25y = 35 - x$).

Power laws and logarithmic graphs

Sometimes there is no 'simple' relationship between two variables, or we may have no idea what the relationship may be. So, in general, we can write that the variables x and y are connected by a relationship of the form:

$$y = kx^p$$

where k and p are constants. That is, y is proportional to x to the power p.

Taking logarithms of this equation we get:

$$\log y = p \log x + \log k$$

Compare this to the equation for a straight line, $y = mx + c$.

If a graph is drawn of $\log y$ against $\log x$, it will have a gradient p and an intercept of $\log k$. Using this information, a mathematical equation can be written to describe the relationship. Note that logarithms to the base 10 have been used in the above equation, but **natural logarithms (ln)** could be used instead (Higher Level only).

Worked examples

6 The relationship between two variables, x and y, is shown in Figure 17.16. Take measurements from the graph so that you can write an equation to represent the relationship.

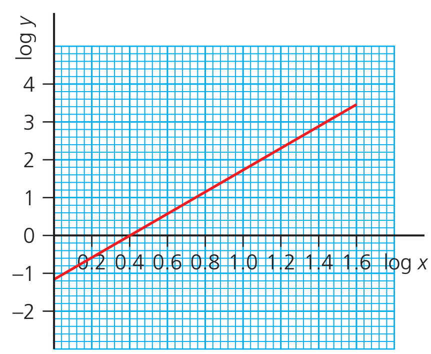


Figure 17.16

The gradient of the line, p, is 2.9 and the intercept on the $\log y$ axis, $\log k$, is –1.2. Substituting into $\log y = p \log x + \log k$ we get:

$$\log y = 2.9 \times \log x - 1.2$$

So,

$$y = 0.063x^{2.9}$$

7 Refer back to Worked example 4. Use the data to draw a log graph that verifies that the relationship is described by the equation $D = kL^m$. Determine values for m and k from the graph.

Taking logs of the equation:

$$\log D = m \log L + \log k$$

Comparing this to $y = mx + c$, we know that a graph of $\log D$ against $\log L$ will have a gradient of m, and k can be determined from an intercept (Figure 17.17).

$$m = 3.1 \text{ and } k = 2.2 \times 10^{-5}$$

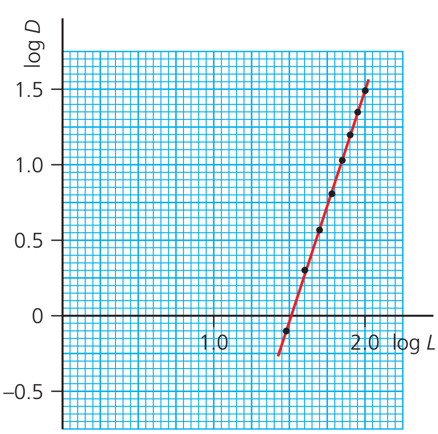


Figure 17.17