7 Looking for relationships: line graphs

> **Key words**: line graph, variable, linear, linear relationship, non-linear, gradient, origin, intercept, proportional, rate, line of best fit, interpolation, extrapolation, outlier.

Many investigations in science are concerned with finding relationships between continuous variables. After collecting a set of data, the data points for two variables can be plotted on a graph, and then a line drawn that best expresses the apparent relationship suggested by the data. This is called a line of best fit.

7.1 Types of relationship and shapes of line graphs

Before looking at drawing lines of best fit, it is useful to consider the possible kinds of relationship between variables that might be found. Figure 7.1 shows a number of different shapes of graphs – this selection is not intended to be comprehensive, but sufficient to illustrate a variety of relationships.

To avoid confusion among pupils, it is important to be aware that the term 'line' has a more precise meaning in mathematics than the way it is often used in science. In mathematics, a *line* (and thus a *line of best fit*) is, by definition, *straight*. In science, however, it is quite common to talk about 'straight lines' and 'curved lines' (which in mathematics would be called 'lines' and 'curves'). Changing habitual ways of talking is hard. A good compromise in science might be to refer to 'straight lines' and 'curves', though it may be hard to avoid using the term 'curved line of best fit'. Even though most of the graphs in Figure 7.1 show curves rather than straight lines, they are still all referred to as ***line graphs***.

Some relationships in science can be described by relatively simple mathematical equations, while others are more complex. This section focuses just on the shapes of the line graphs and on fitting straight lines and curves to data points by eye; the use of mathematical equations is discussed later (see Chapter 9 *Scientific models and mathematical equations* on page 87).

Figure 7.1 A compendium of line graphs
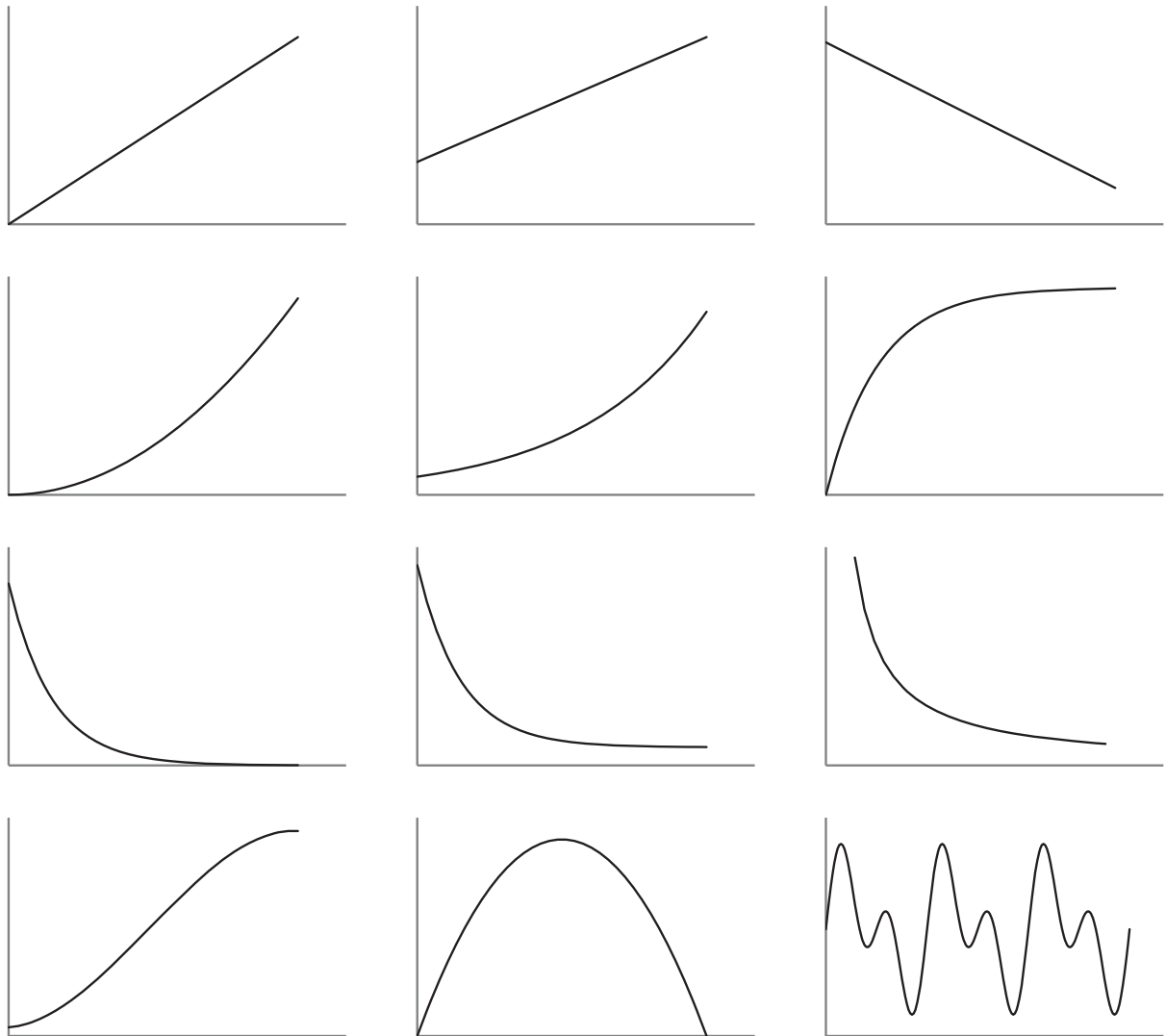
7.2   Developing a descriptive language

It is helpful for pupils to develop ways of identifying and describing the nature of the relationships shown in ***line graphs***. This does not mean simply describing the *superficial shape* of the line, which they might find quite easy to do. What is harder for them is to relate the shape to the *meaning* of the relationship between the real-world ***variables***. This needs building up with practice, using graphs they draw from their own results, as well as completed graphs presented to them. The following discussion includes suggested phrases that can be used as part of such descriptions.

When interpreting a graph, the first thing always is to pay attention to *what the graph is about*, i.e. the variables involved, as indicated by the labels on the axes, and any other information. It is also essential to be aware of the *range for each axis*, since this affects the visual appearance of the graph (see Section 4.3 *Choosing the range of each axis* on page 37). This is particularly important when comparing two or more similar graphs.

Note that since the graphs shown below are intended to be 'abstract' they do not have the axes labelled with particular variables. The discussion will therefore be in terms of changes to '*x*' and '*y*'. In a real context, the names of the variables would be used to give the descriptions a real-world meaning.

When thinking about the meaning of the line on a graph, a starting point is to identify whether the line goes up or down. For the graphs in Figure 7.2a, '*as x increases, y increases*'; for those in Figure 7.2b, '*as x increases, y decreases*'. This distinction is a very basic aspect of a relationship, and indeed does not really need a graph to be able to identify – it would be obvious from the table of results. Note that some graphs do not simply go either up or down, and these will be discussed later.

**Figure 7.2** Increasing and decreasing

(a)
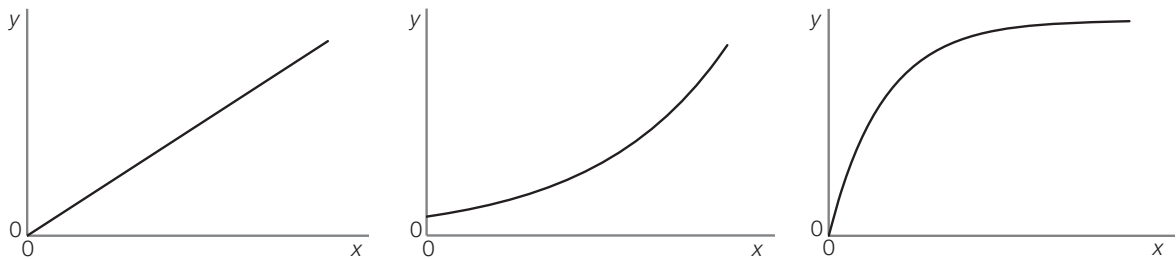
(b)
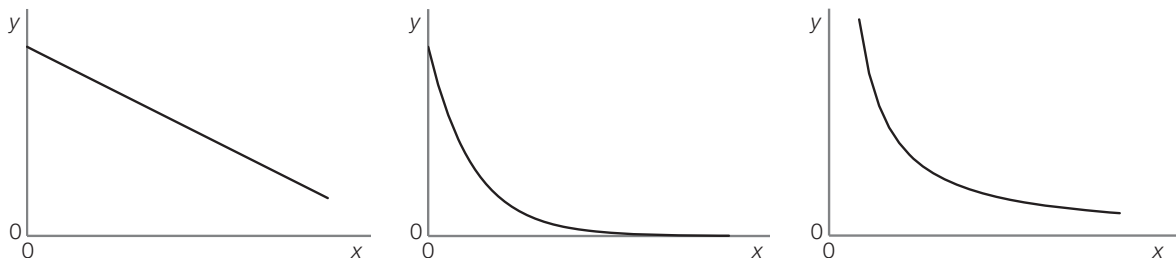
What we can see from a line graph, but would not be so obvious from the table of results, is whether the line is straight or curved. So, we could talk of a '*straight line graph*' for which '*as x increases, y steadily increases*' (Figure 7.3a) or for which '*as x increases, y steadily decreases*' (Figure 7.3b). The term 'steadily' is a rather informal term but it gives a good sense of what is happening.

Figure 7.3 Graphs with straight lines
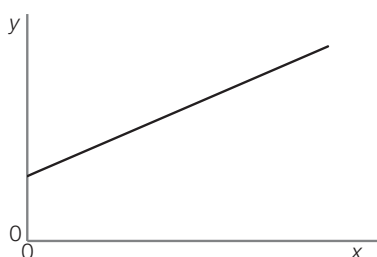
**(a)**   **(b)**
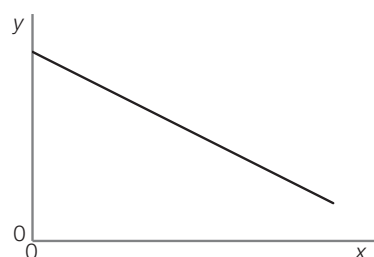

The relationships shown in '*curved graphs*' are more complex to describe. One possibility is that '*as x increases, y increases slowly at first and then more rapidly*' (Figure 7.4a) or that '*as x increases, y increases rapidly at first and then more slowly*' (Figure 7.4b). Similar descriptions can be used for the curved graphs that show a '*decrease of y with x*'.

The formal term to describe a straight line graph is ***linear***, whether or not it goes through the origin, and the relationship between the two variables is called a ***linear relationship***. Similarly, the relationship shown by a curved graph is called ***non-linear***.

Figure 7.4 Graphs with curves

(a)

(b)
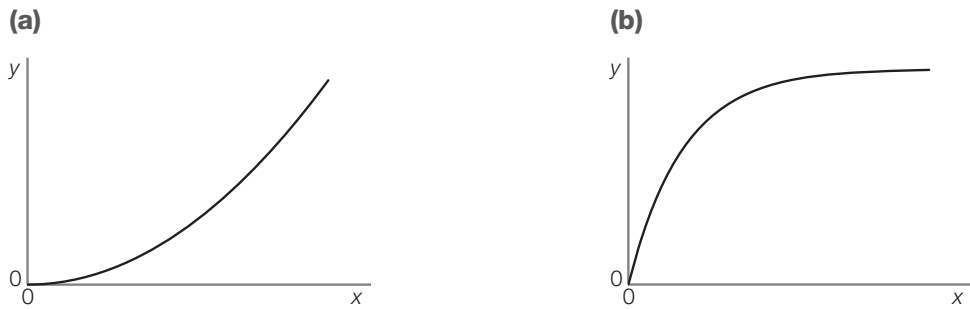
When we talk of a variable changing 'slowly' or 'rapidly', we are using these terms in a relative sense to describe how the **_gradient_** (or *slope*) of a line changes. For a linear relationship, the gradient at any point along the line is the same. For a curve, the gradient varies at different points along the curve.

An important feature of a relationship is whether the line goes through the **_origin_** (the point at which the values of *x* and *y* are zero). Figures 7.5a and 7.5b are both linear relationships. However, while the first shows '*a straight line that goes through the origin*', the second shows '*a straight line with an intercept on the y-axis*'. The point at which it meets the *y*-axis is called the **_intercept_**. Figure 7.5a shows a **_proportional_** relationship, i.e. doubling the value of *x* doubles the value of *y*. So '*as x increases, y increases, and y is proportional to x*'. However, although Figure 7.5b represents a *linear* relationship, it is not a *proportional* relationship, since the line does not go through the origin. (See Section 9.11 *Mathematical equations and relationships in science* on page 99.)

**Figure 7.5** Straight line graphs and the origin

(a)

(b)
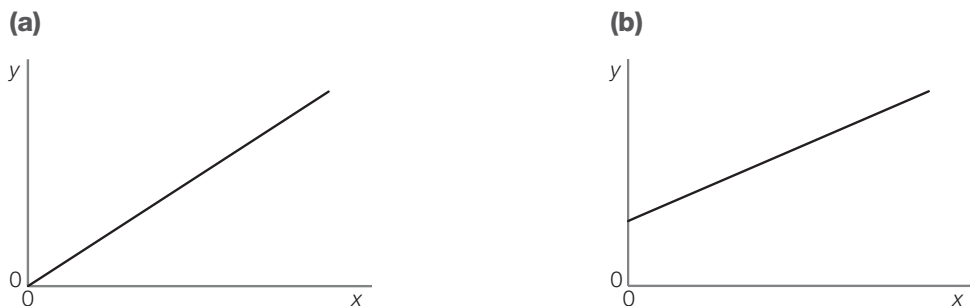
Finally, while some curves may appear to increase indefinitely (Figure 7.6a), others may '*level out towards a maximum*' (Figure 7.6b). Similarly, other curves showing decreasing values may '*level out towards a minimum*' (Figure 7.6c).

Figure 7.6 Some curves tend towards a maximum or a minim level

(a)

(b)

(c)
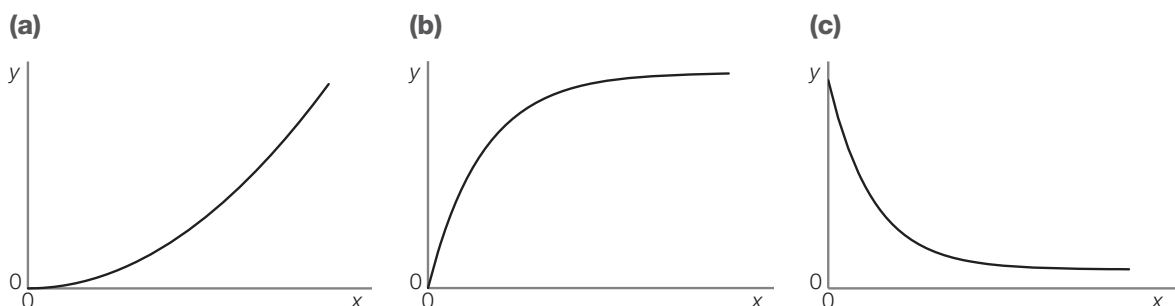
The graph shown in Figure 7.7 provides an opportunity to put all of these ideas together. Relevant phrases might be: '*curved graph*', '*intercept on y-axis*', '*as x increases, y increases*', '*y increases slowly at first, then more rapidly, then slows down again*' and '*reaches a maximum level*'.

Figure 7.7 A more complex shape
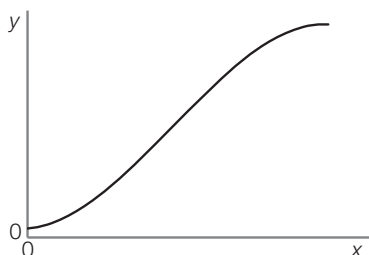
So far, all of the graphs discussed have had a line that either always goes up or always goes down (the technical term for this is *monotonic*). Other relationships are more complex, with lines than can go up and down (called *non-monotonic*).

The last two examples are included to illustrate this. Figure 7.8a could represent the height of a ball as it is thrown in the air and then falls to the ground. Figure 7.8b could represent the amplitude of a loudspeaker producing a musical note. Since the pattern repeats, it can be called periodic (another example would be a sine wave).

Figure 7.8 Some curves are neither 'up' nor 'down'

(a)
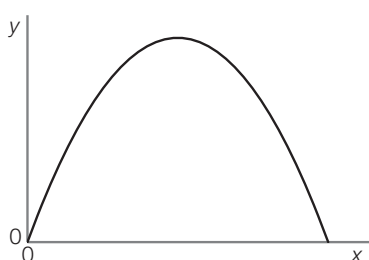
(b)
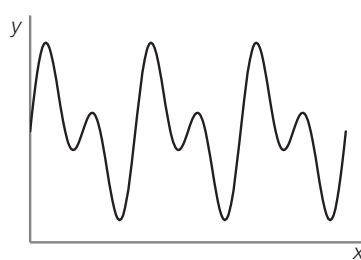
## 7.3 Gradients and rates of change

Many graphs have 'time' as the variable on the horizontal axis, and indeed the language that we use reflects this. For example, Figure 7.9a shows a graph that represents the progress of a chemical reaction between a carbonate and an acid to produce carbon dioxide. At first, the increase in the volume of carbon dioxide is quite fast but then it slows down. Notice how a graph reflects the way we use language to describe how the sizes of things can go *up and down* (represented on the *vertical* axis) and happen *one thing after another* in time (represented from left to right on the *horizontal* axis).

On a graph that shows a change over time, the steepness of the line represents how fast the change is happening. In other words, the **_gradient_** of the line represents a **_rate_** of change. Since, this is the rate of change at a *particular instant in time*, it is called an *instantaneous* rate of change.

Figure 7.9b shows the way that the current through a filament bulb varies with the potential difference across it. The curve has a shape that has some similarities to the previous graph (note that it is not linear and does not follow Ohm's Law). Even though it does not represent a change over time, we may still use the same language to describe it – rising rapidly at first

and then more slowly. It is also possible to use the term 'rate' to describe the gradient here as well, i.e. as 'the rate at which current changes with potential difference'). When talking about relationships between variables in 11–16 science, however, *rate* is usually only used to refer to a *rate of change over time*.

Figure 7.9 Some changes are against time, others are not

(a)

(b)
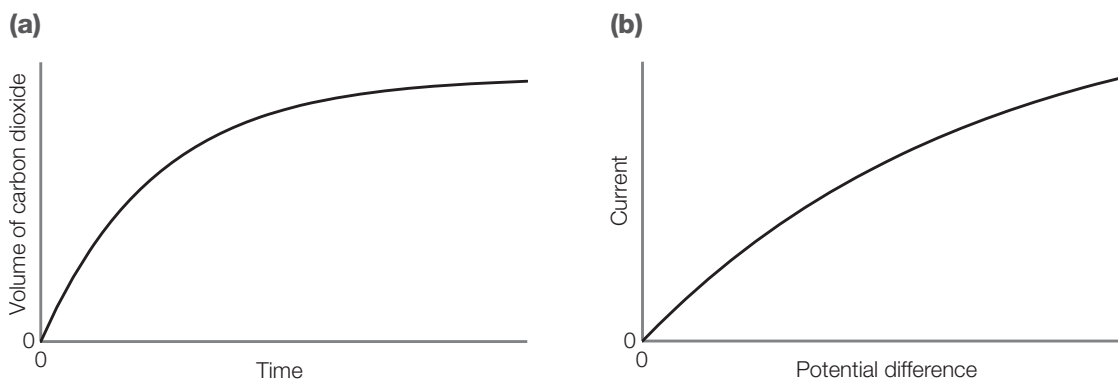
7.4 Lines of best fit: linear relationships

Sometimes it happens that when the data points from an experiment are plotted, a straight line can be drawn which appears to pass *exactly* through all of the points. More usually, even if the underlying relationship is linear, the data points do not lie exactly on a straight line because of measurement uncertainty. In such a case, a ***line of best fit*** may be found that passes as close as possible to the points.

It is important to be clear about the meaning of a fitted line *for data of this kind*. It has a very different meaning to the line segments on a graph where each pair of data points are joined; it also has a very different meaning from a line of best fit on a scatter graph (see Section 3.6 *Line graphs and scatter graphs: two related quantities* on page 29).

These differences arise because of the differences in the nature of the data. For 'line graph' type data, the data points may not all lie on the fitted line because of *measurement uncertainty*. For 'scatter graph' type data, the differences from a fitted line are due to *differences between individuals in a population*. The distinction between these two different kinds of variability is discussed in Section 6.1 *Where does variability come from?* on page 50.

Figure 7.10a shows a set of eight data points plotted on a graph (for convenience, the discussions in this chapter on drawing lines of best fit use 'abstract' graphs with no variables specified). Inspecting these points by eye suggests that this may be a ***linear relationship***, so a straight line could be fitted. A useful technique with points plotted on graph paper is to hold the paper almost at right angles to your face and then to rotate the paper to look down along the direction of the data points. This is a quick way of seeing how close the points would be to a straight line.

To draw a line on graph paper, it is better to have a *transparent* ruler so that all of the data points can be seen. Figure 7.10b shows a line of fit drawn by eye. It seems to be quite a good fit as it passes close to all of the points. But what is meant by a line of *best* fit?
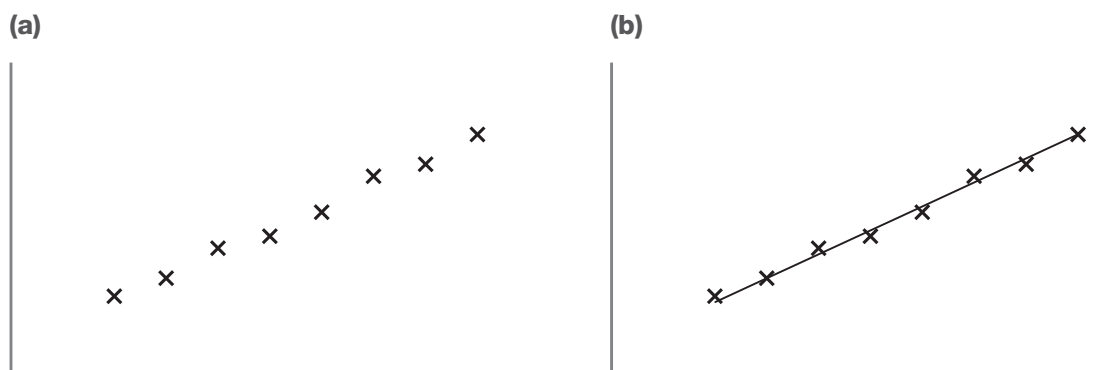
Figure 7.10 A good line of fit

(a)

(b)

Figure 7.11 shows two lines that are most definitely not good lines of fit. The 'badness' of fit has been deliberately exaggerated in each case to illustrate the criteria for fitting a good line. In Figure 7.11a, the gradient of the line matches the gradient of the data points but the line is too high. In the good line of fit, there were data points on *both sides* of the line, but here they are all *below* the line. By contrast, in Figure 7.11b, there are similar numbers of points above and below the line but the gradient of the line is wrong. All the points below the line are on the left and all the points above the line are on the right.

Figure 7.11 What makes a bad line of fit?

(a) Too high
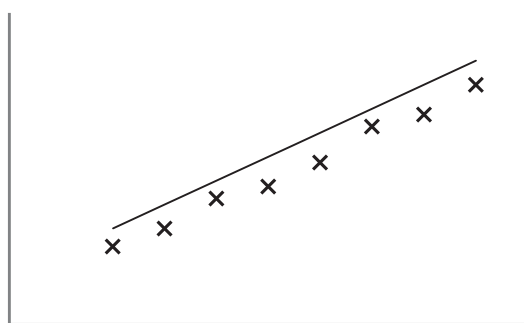
(b) Wrong gradient
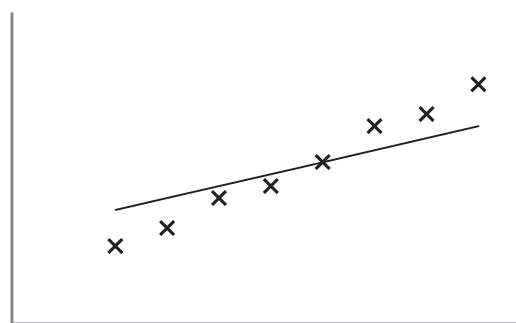
So, when drawing a line by hand on graph paper, there are two things to think about: getting the height of the ruler right and getting its slope right, so that the line is as close to all the points as possible. Sometimes a line may pass though some of the points but this is not essential – it is possible to have a line of best fit that does not actually pass through any of the points.

There are no hard-and-fast rules for producing a 'best fit' by eye; it is a matter of judgement to find the one that looks best. When a line of best fit is done by a computer, various rules are used to decide exactly what is meant by 'as close to all the points as possible'. However, even with a computer, there is no single method to produce a 'unique' best line, and different rules may produce different 'best' lines.
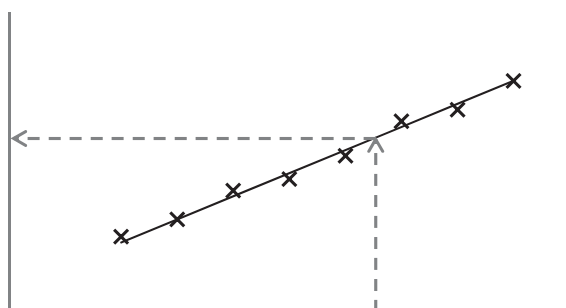
7.5 Interpolation and extrapolation on a line graph

The graph in Figure 7.10a had eight data points, each representing a value for *y* corresponding to a particular value of *x*. Once a line of best fit has been drawn, it is possible to use it to estimate a value for *y* corresponding to *any* value for *x*. Figure 7.12a illustrates how a value for *y* can be 'read off' the graph for a value of *x* that is in between the original data points. This technique is called ***interpolation***. It can be used, for example, in calibrating instruments

such as thermometers – a number of measurements are made to create a calibration curve (or straight line), and this is used to infer all of the other values.

Getting good estimates from interpolation assumes that the fitted line is a good representation of what happens in between the measured data points, and that there are no unexpected variations. The more measurements that are made, the greater the chance that interpolation will give good estimates. In fact, since the fitted line may compensate for measurement uncertainties in the data, it can actually give better estimates of the *y* values for the original data points than the actual data values themselves. This is the reason that a fitted line is used to find the gradient on such a graph, and not just the two extreme values.

Figure 7.12 Using the fitted line to estimate values
(a) Interpolation   **(b)** Extrapolation
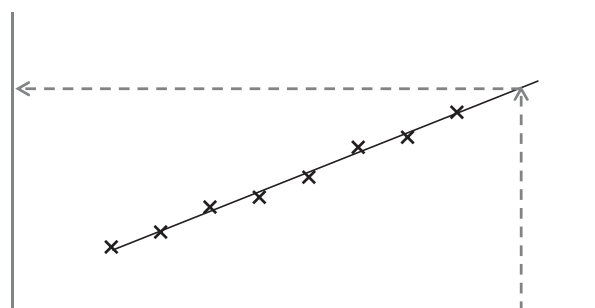
The fitted line may also be extended in order to make estimates of values beyond the range of the original data. Figure 7.12b shows the line being extended to higher values, and a value of *y* being 'read off' the graph for a value of *x* that is greater than the original range. This process is called ***extrapolation***. For example, a graph showing the extension of a spring against applied force could be extended to find the extension of the spring for a greater force. Care needs to be taken with extrapolation, however, since the linear relationship may not apply outside the data range. In the example of the spring, a point is reached when it becomes 'overstretched' and the extension is no longer proportional to the force.

Extrapolation can also be done by extending the line towards lower values. In such a case, it may be of interest to find out whether it passes through the origin, or, if it does not, to find the value of the intercept on the *x*-axis or *y*-axis.

7.6   Origin and intercepts: the meaning of where a fitted line starts

If a straight line on a graph goes through the ***origin*** then this represents a ***proportional*** relationship (see Section 5.2 *Proportionality and visual representation* on page 41). Deciding whether the first data point is at the origin, i.e. $(0, 0)$, is important.

Figure 7.13a shows a proportional relationship – it is a graph of current against potential difference for a resistance that follows Ohm's Law. Here, it is apparent that the line *must* go through the origin – if there is no potential difference then there is no current. This could also easily be confirmed by reading these values from the voltmeter and ammeter. For other relationships, we may know that the line must start at the origin, even though in practice there might not be any measurements to show this. Care needs to be taken, however, about making any assumptions about the values at the intercepts, and a line only drawn through the origin when *in principle* it could not be otherwise.

Figure 7.13b shows how the length of a spring varies with force. It is a linear relationship (it obeys Hooke's Law) but it does not pass through the origin, so it is not a proportional relationship. Usually in such a case, the ***intercept*** on the *y*-axis has a real-world meaning – here, it represents the length of the spring when the value on the *x*-axis (the force on the spring) is zero. In other words, it is its 'normal' length when no force is acting on it to stretch it.

If the original length of the spring is subtracted from all of the data values, this gives the extension of the spring: plotting this against force would be a straight line through the origin, and this a proportional relationship.

In this example, the intercept on the *y*-axis can easily be found by measurement – it is simply the length of the spring with no force. However, in other situations, the intercept cannot be measured directly though it may be found by extrapolation.

Figure 7.13 Lines can meet the axes in different ways
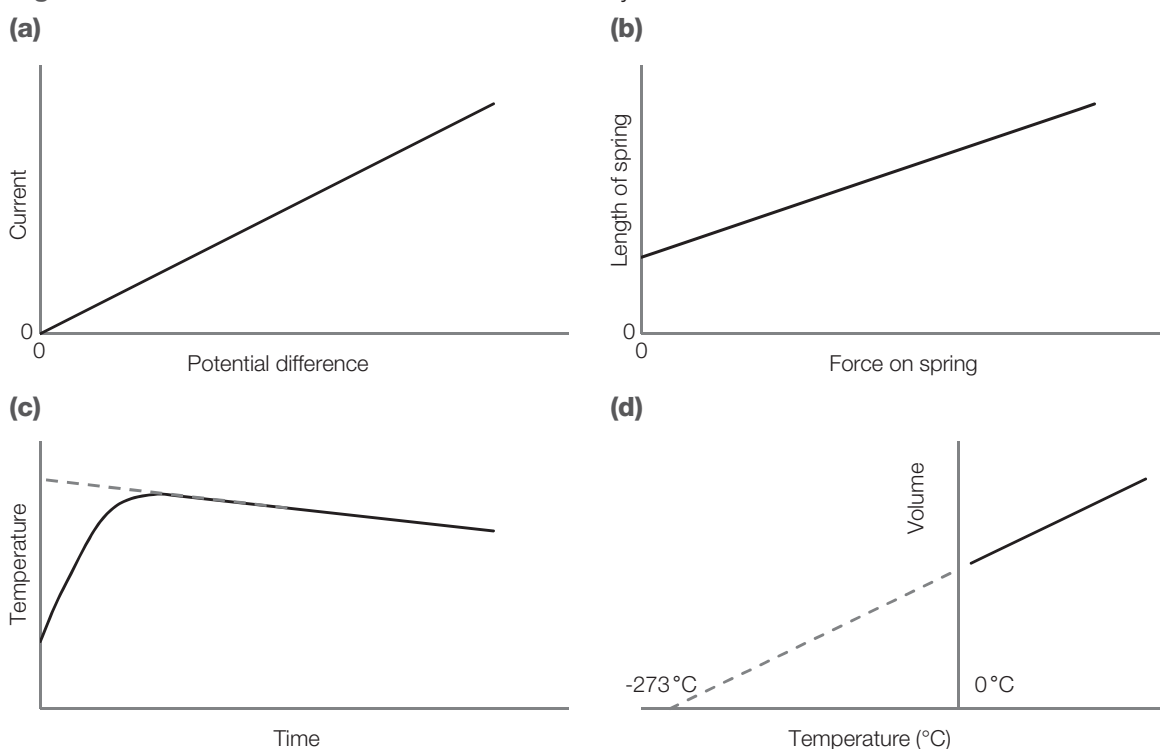

(a)

(b)

(c)

(d)

Figure 7.13c represents a graph of the temperature of the reaction mixture during an exothermic chemical change, such as the reaction of zinc with acid in an insulated container. It is not easy to find the 'true' value of the temperature rise, because the reaction takes a little time to complete and during this time energy escapes from the warm reaction mixture. After the initial temperature rise and the completion of the reaction, the mixture starts to cool. Since the container is insulated, the cooling is relatively slow, and approximates to a linear decrease over a small time period. By extrapolating this line backwards, the 'true' temperature rise (i.e. if there had been no cooling) can be estimated from the intercept on the *y*-axis.

An interesting historical use of extrapolation was to estimate the temperature at absolute zero (and this is still a valuable practical activity in post-16 physics). The solid line in Figure 7.13d represents the relationship between the volume of a fixed mass of gas and its temperature. As the gas is cooled, its volume decreases. The theoretical interpretation of this is that the decrease is due to the molecules of the gas moving more slowly. If the molecules stop moving

at absolute zero then the volume would approach zero (assuming the volume of the molecules themselves is negligible). Extrapolating the line back to zero volume (here, this means to the point where it meets the *x-axis*), gives a temperature of about −273 °C, which is close to the accepted value.

7.7 When a straight line does not fit all the points

Not every graph has data points that clearly all lie close to a straight line. Two possibilities which may arise are:

- the underlying relationship is linear but there are **_outliers_**, perhaps due to mistakes in measurement – they may need to be ignored or rechecked
- the underlying relationship is not linear – the line of best fit is a *curve*, not a straight line.

The more data that are collected, the clearer the nature of the relationship becomes. Figure 7.14a shows a graph with just four data points. It is not obvious what the line of best fit might be. Perhaps it is best to draw a straight line close to all of the points (Figure 7.14b), or perhaps treat the final point as an outlier and draw a straight line through the other three (Figure 7.14c), or perhaps it is best to draw a curve close to the points (Figure 7.14d). With only four points, you cannot really decide.

Figure 7.14 It is hard to identify a relationship with a small number of data points

(a)
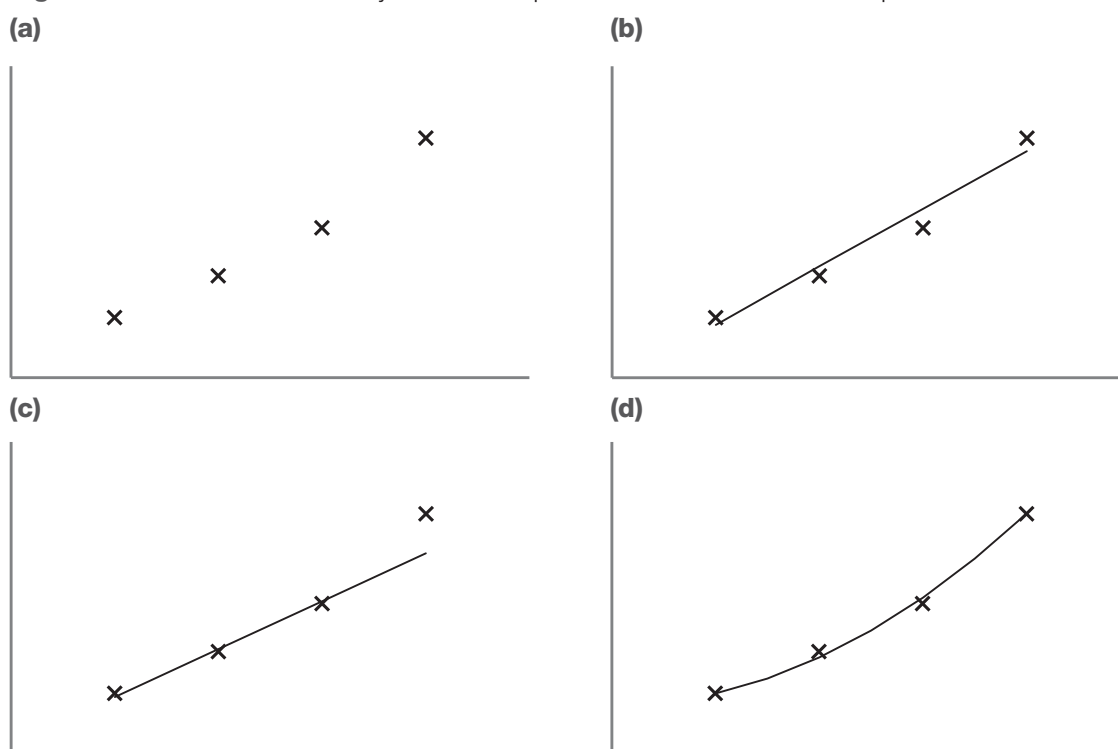
(b)

(c)

(d)

Figure 7.15a shows the original four points but now includes an additional four points to give eight in total. It is now much clearer to see a pattern – it looks like this is a linear relationship but that one of the measurements is an outlier. Figure 7.15b shows a straight line as a line of best fit, using seven of the points and ignoring the outlier.

If a computer is used to find a fitted line, for example on a spreadsheet graph, it may use all eight points including the outlier. The fitted line would then slope more steeply, and by eye would not look like a good fit.

Figure 7.15 Ignoring an outlier when drawing a line of best fit

(a)   **(b)**
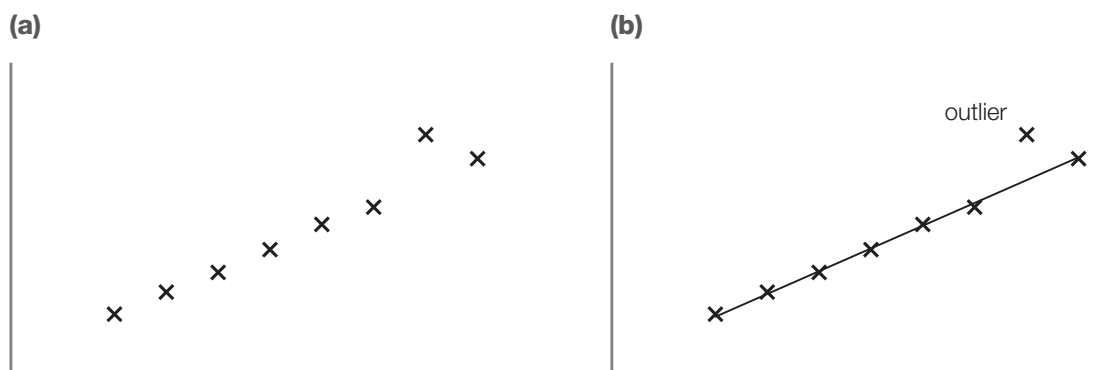
These additional four points were necessary to identify the relationship, but if they had been different then the relationship might have looked very different. Figure 7.16a also shows the original four points but with a different set of four points added. Now, the pattern of data points suggests that the line of best fit should be curved. As before, if the data points are plotted on graph paper, holding it up and looking along the points by eye is a good way of getting a sense of the shape. Drawing good curves by hand needs practice – it can be done using a sweeping movement of the hand with the wrist or elbow as a pivot 'inside' the curve, or by using an instrument known as a 'flexible curve'.

Figure 7.16 A curved line of best fit

(a)   **(b)**
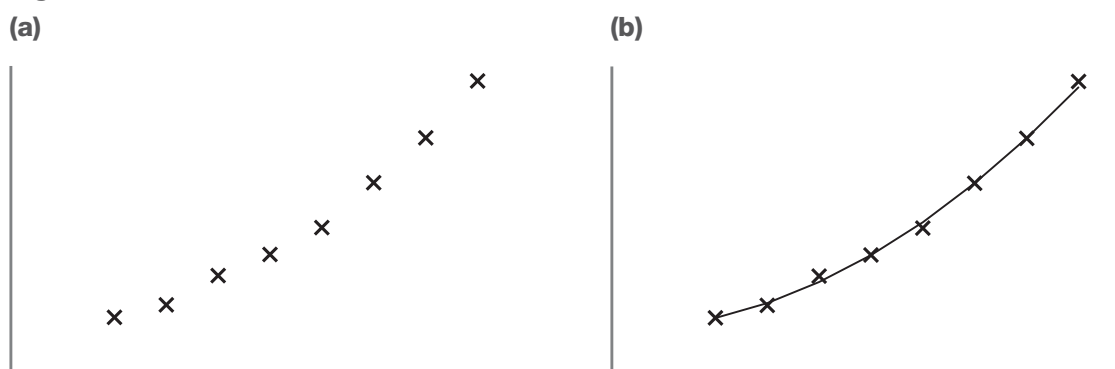
If, instead of a curve, a straight line had been drawn that passed as close to the points as possible then the points in the middle would have been below the line and the ones at each end above. This is a sign that a curve would be a better fit.

Drawing curves is much easier by computer. For example, a formula for a curve could be entered into a spreadsheet but the difficulty may be in finding the correct formula that produces a good fit. In 11–16 science, straight line graphs are more common, though students studying post-16 science come across graphs with curved lines more frequently: in particular, in physics, students need to be able to determine the equations of some non-linear relationships by using logarithmic scales.