How to Prepare and Interpret Graphs

James W. Phillips

1. Why prepare graphs?

A graph is perhaps the most effective tool that an engineer has, not only for understanding the relationship between two or more variables, but also for conveying relational information to others. Indeed, the graph is to the engineer what a painting is to an artist: a complete story can be told; the reader learns something not only about the way y depends on x, but also about the way the graph's creator *thinks*.

Like technical writing, graph preparation can be fun. There really are only a few rules for making effective and attractive graphs. Axes should be drawn in a fine line, each axis being labeled with the variable name, symbol, and units; numbers should appear at major tic marks or grid lines, and these numbers should be multiples of 1, 2, or 5 (with preceding or trailing zeroes as necessary). Curves representing theoretical (or computational) results are drawn in heavy lines without points; these curves are generally solid unless they represent extrapolated or comparative results, in which case they may be dashed. Raw experimental data are usually plotted as points, with or without error bars, depending on the nature of the data reduction.

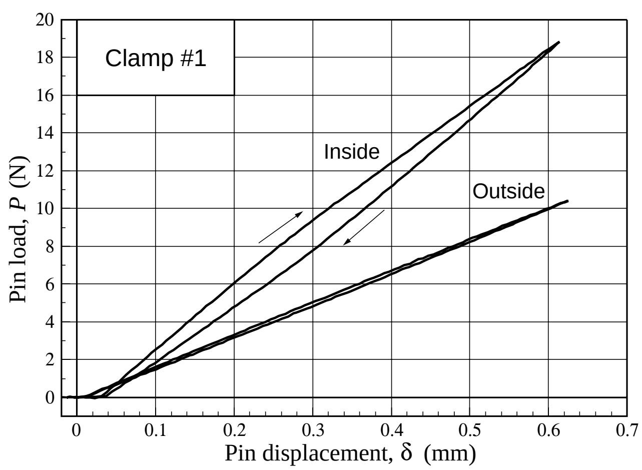

Fig. 1. Example of a technical graph.

There are cases where theoretical or computational results are not available except at a few points; then points should be used tell this story. Likewise, experimental data may be obtained as continuous curves, as in the case of an XY recorder plot of load versus deflection; then a solid curve for the experiment is appropriate.

The first step in preparing a graph is to select which variable to plot on which axis. The rule is to plot the independent variable on the horizontal axis and the dependent variable on the vertical axis. It usually requires some thought to decide which variable is independent and which is dependent. An example is shoe size and age: shoe size depends on age, not the other way around, so age would be plotted on the horizontal axis and shoe size on the vertical axis. If it is not clear which variable is independent and which is dependent, ask your laboratory supervisor for advice.

More than one dependent variable may be plotted on the same graph, but each must then be clearly identified, either by labeling each curve or by using different symbols or line types.

Generally, a "key to symbols" should be avoided; individual curves should be labeled by placing words or symbols near the curves, preferably without the use of arrows. An example of a graph that follows these rules is presented in Fig. 1.

Axes require a scale large enough to plot all points and small enough that the points are spread out across the entire graph. It is essential to give the beginning and ending scale values and to use others as necessary to indicate the scale interval. Axes must be labeled with the name of the parameter and the scale unit.

Often, a graph is prepared with experimental data, and a curve fit of some type is sought. The following discussion pertains to the proper construction and interpretation of straight lines drawn on various combinations of linear and logarithmic axes.

2. Linear

If the data appear to be represented by a straight line on linear–linear graph paper, the following form is appropriate:

$$y = mx + b, \tag{1}$$

where m is the slope and b is the intercept along the line $x = 0$. In general, two points are chosen from the line drawn through the data, as shown in Fig. 2.

The slope may be calculated by writing Eqn. (1) twice, once for each (x, y) pair, and subtracting:

$$m = \frac{y_2 - y_1}{x_2 - x_1}. \tag{2}$$

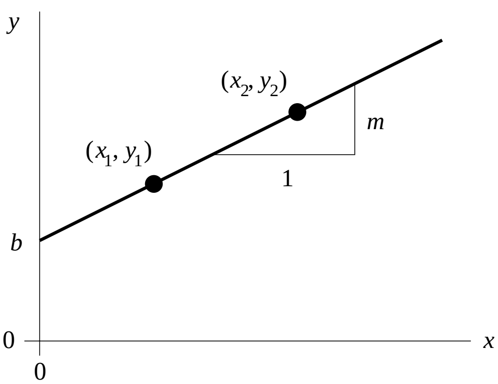

Fig. 2. Linear–linear slope and intercept.

For best accuracy, the two points should be separated widely along the line.

The constant b is the value of y when $x = 0$; if the graph is drawn with the y axis passing through $x = 0$ then b is just the intercept shown in Fig. 2. However,

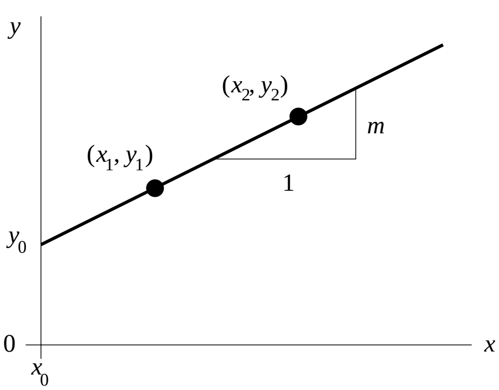

Fig. 3. Linear–linear slope and intercept at nonzero value of *x*.

if the graph is drawn with the y axis passing through some other value of x, say x_0 (as shown in Fig. 3), then the value of b is not y_0 but rather

$$b = y_0 - m x_0, \tag{3}$$

which follows from Eqn. (1).

Even in this simple case of linear–linear plots, one must be careful about obtaining the value of m by a strictly graphical procedure. Only in the rare case when the *scale* for x is the same as that for y will m be found correctly by taking $\Delta x = 1$ as shown in Figs. 2 or 3, and determining the slope *graphically*.

For example, if y denotes velocity, which is plotted at 5 km/s per cm, and x denotes time, which is plotted at 0.1 s per cm, and the data fall on a straight line with slope equal to 2.1 (cm/cm) on the graph paper, then the acceleration a is not 2.1 $\mathrm{km/s^2}$ but rather (2.1)(5 km/s)/(0.1 s/cm) or 105 $\mathrm{km/s^2}$.

One can obviously take the scales into account if they are not the same, but it is recommended to use Eqns. (1–3) directly to avoid the problem altogether.

3. Log–log

If the data appear to be linear on log–log paper,[1] as shown in Fig. 4, then the power-law relation

$$y = A x^m, \tag{4}$$

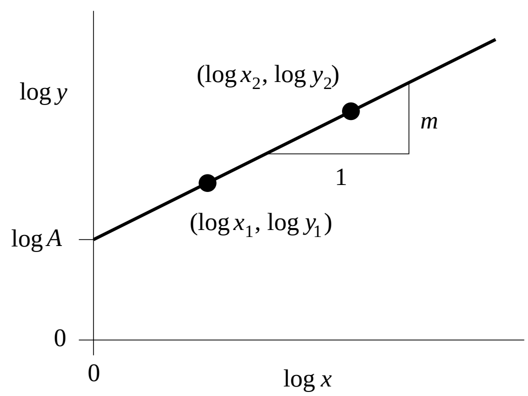

Fig. 4. Log–log slope and intercept.

where A and m are constants, may be an appropriate representation.[2] Taking the base-10 logarithm[3] of both sides of Eqn. (4) gives

[1] The term 'log–log paper' refers to graph paper in which the *logarithms* of the indicated values along both axes are already scaled linearly. However, the axis labels and the indicated values placed at tic marks are the values of x and y, not their logarithms. It is unnecessary (and incorrect) to take the logarithms of numbers *before* plotting on log–log paper.

[2] If x has dimensions (such as meters or pascals) and m is not a rational number, then there is an inherent problem with identifying the dimensions of the constant A. For this

$$\log y = \log A + m \log x, \tag{5}$$

which is basically the same form as that of Eqn. (1). The slope m is now given by

$$m = \frac{\log y_2 - \log y_1}{\log x_2 - \log x_1}. \tag{6}$$

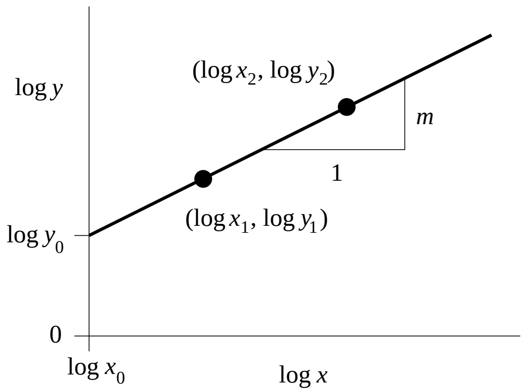

Fig. 5. Log–log slope and intercept at nonzero value of log*x*.

The intercept along the $\log y$ axis is easily interpreted as $\log A$ as long as the $\log y$ axis passes through the origin of the $\log x$ axis (i.e. at $x = 1$). For example, when plotting fatigue data in the form of *S-N* diagrams or plots of strain amplitude ($\Delta\varepsilon/2$) vs. reversals (2*N*), the logarithm of a stress or strain variable is often plotted as a function of the logarithm of the number of cycles *N* to failure. Calculations for straight-line curve fitting are simplified if the log*S* or $\log(\Delta\varepsilon/2)$ axis passes through *zero* on the log*N* or log(2*N*) axis.

Sometimes, however, it is impractical to include $\log x = 0$ in the plotted range of $\log x$, and then the $\log y$ axis will pass through a nonzero value of $\log x$, say $\log x_0$. As shown in Fig. 5, the intercept of the straight line with the $\log y$ axis occurs at $\log y_0$, not at $\log A$ as before. It follows from Eqn. (5) that

$$\log A = \log y_0 - m \log x_0, \tag{7}$$

that is,

$$A = \frac{y_0}{x_0^m}. \tag{8}$$

As in the case of linear–linear plots, one must be careful about obtaining the value of m in a log–log plot by a strictly graphical procedure. If log–log paper is used, and the *scale* for $\log x$ is the same as that for $\log y$, i.e. decade values of x and y are plotted with the same spatial increment, then m will be found correctly by taking the slope graphically.

However, if *unequal* scales are used for $\log x$ and $\log y$, i.e. if decade values of x and y are plotted with unequal spatial increments, then it is easy to make a mistake in calculating m graphically. Again, one can take the scales into account if they are not the same, but it is recommended that Eqns. (4–8) be used directly to avoid the problems of graphical interpretation altogether.

As an example of a log–log plot drawn with unequal log scales, consider Fig. 6, which is taken from the laboratory on fatigue. It is required, among other things, to find the slope c of the "plastic strain" line, which is given empirically as

$$\frac{\Delta\varepsilon_p}{2} = \varepsilon'_f (2N)^c, \tag{9}$$

where ε'_f is the "fatigue ductility coefficient" and c is the "fatigue ductility exponent." Observe that the "plastic strain" line intersects the strain-amplitude axis at $\Delta\varepsilon/2 = 4.0 \times 10^{-2}$, and at this point, $(2N) = 1.0 \times 10^0$, i.e. $(2N) = 1$. Let this point be point number 1. A second point can be identified by

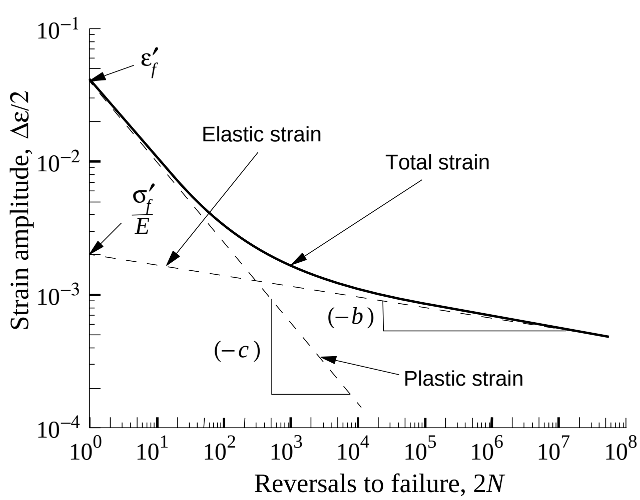

Fig. 6. Example of a log–log plot: the total-strain life curve.

reason, some investigators prefer to *normalize* the variable x with respect to some reference value, say $\bar{x}$, and write $y = A(x/\bar{x})^m$; then at least the dimensions of A are the same as those of y. Otherwise, it is difficult to give a physical significance to the constant A.

[3]Here, 'log' is used to denote the base-10 logarithm ($\log_{10}$) but the natural logarithm (ln) could also be used. Since in this lab emphasis is placed only on base-10 operations, the subscript '10' will be assumed.

extending the "plastic strain" line until it intersects the reversals-to-failure axis. This intersection occurs at $(2N) = 2.0 \times 10^4$, at which point $\Delta\varepsilon/2 = 10^{-4}$. The slope c is therefore given by

$$c = \frac{\log(\Delta\varepsilon/2)_2 - \log(\Delta\varepsilon/2)_1}{\log(2N)_2 - \log(2N)_1}$$
$$= \frac{\log(10^{-4}) - \log(4.0 \times 10^{-2})}{\log(2.0 \times 10^4) - \log(10^0)}$$
$$= \frac{(-4) - (-1.40)}{(+4.30) - (0)} = -0.605.$$

Now it may appear that a mistake in the calculations must have been made because it is obvious from Fig. 6 that the slope of the "plastic strain" line is more negative than -1. That is, the way Fig. 6 is drawn, it appears that $(-c) \cong 1.2$ or $c \cong -1.2$. However, the strain-amplitude logarithmic scale is stretched by a factor of about 2, compared with the reversals-to-failure logarithmic scale, so that the correct value of c is about –1.2/2, or –0.6, as was obtained analytically.

Also note that, by passing the (log of the) strain-amplitude axis through the zero value of the (log of the) reversals-to-failure axis, one can "pick off" the value of ε_f' easily; its value is 4.0×10^{-2} or 0.040.

4. Log–linear (semi-log)

Some data, when plotted on semi-log paper, appear to fall along a straight line. If so, the representation

$$y = A e^{mx} \tag{10}$$

may be appropriate,[4] since by taking logarithms of both sides, one obtains

$$\log y = \log A + (m \log e)x. \tag{11}$$

In this case, one would choose the logarithmic axis for y and the linear axis for x, as shown in Figs. 7 and 8. As before, calculations are simplified if the $\log y$ axis passes through the x axis at $x = 0$. Either way, the slope of the line is $m \log e$, and hence

$$m = \frac{1}{\log e} \frac{\log y_2 - \log y_1}{x_2 - x_1}. \tag{12}$$

If the $\log y$ axis intersects the x axis at $x = 0$ (Fig. 7), then the intercept is simply $\log A$, i.e. the value of A can be read off directly from the plot. If the $\log y$ axis intersects the x axis at some other value of x, say x_0 (Fig. 8), then the intercept is not $\log A$ but rather $\log y_0$, and (from Eqn. (10))

$$y_0 = A e^{mx_0}$$

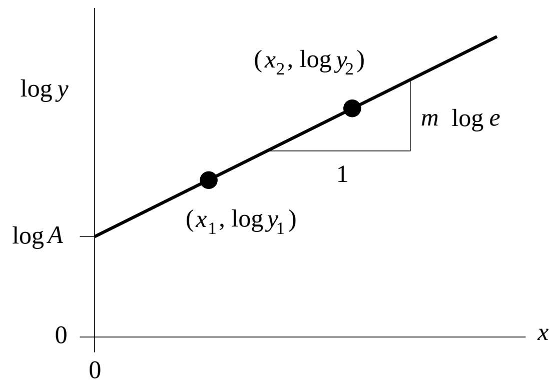

Fig. 7. Log–linear slope and intercept.

or

$$A = \frac{y_0}{e^{mx_0}}. \tag{13}$$

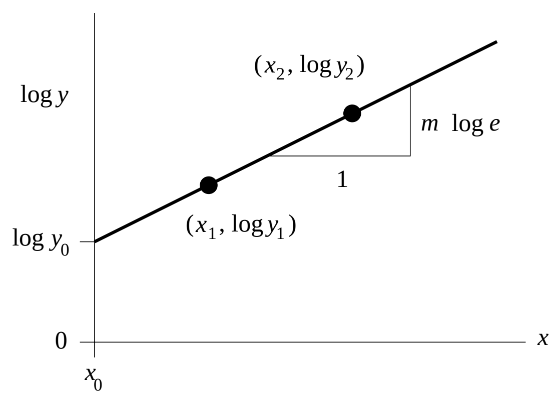

Fig. 8. Log–linear slope and intercept at nonzero value of x.

5. Special

Besides the linear, log–log, and semi-log plots described above, many other types of nonlinear plot are used for special purposes. There are polar plots, probability plots, Weibull distribution plots, and others. Often the purpose of these plots is to suggest that a particular mathematical representation is reasonable if the data fall along a straight line when plotted on special nonlinear axes.

[4]The base need not be e; it could be any constant, such as 10. However, many rate laws are written with base e.

One such case arises in the laboratory on asphalt cements. It is found that the viscosity of bitumens is an extremely sensitive function of temperature, and it is convenient to plot the *logarithm of the logarithm* of the viscosity as a function of the *logarithm* of the absolute temperature. The generic situation is shown in Fig. 9, where a straight-line fit to data has been indicated.

A simple expression that provides a straight-line plot on these axes is

$$y = 10^{Ax^m}, \tag{14}$$

where A and m are constants. Taking the logarithm of both sides gives

$$\log y = A\,x^m \tag{15}$$

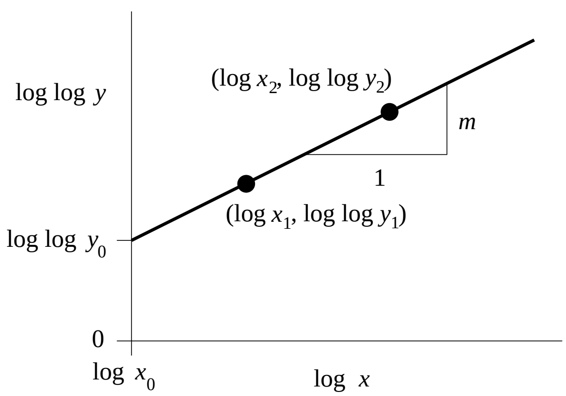

Fig. 9. Log(log)–log slope and intercept at nonzero value of log*x*.

and taking the logarithm again gives

$$\log(\log y) = \log A + m\log x. \tag{16}$$

As before, the slope m can be found by writing Eqn. (16) twice, once for each (x,y) pair, and subtracting:

$$m = \frac{\log(\log y_2) - \log(\log y_1)}{\log x_2 - \log x_1}. \tag{17}$$

The intersection of the $\log(\log y)$ axis with the $\log x$ axis may or may not occur at $\log x_0 = 0$; in general, as illustrated in Fig. 9, it will not; suppose this intersection occurs at $\log x = \log x_0$ as shown. Then the intercept of the straight line with the $\log(\log y)$ axis will occur at $\log(\log y_0)$ and, from Eqn. (15),

$$A = \frac{\log y_0}{x_0^m}. \tag{18}$$

Observe that if the $\log(\log y)$ axis does intersect the $\log x$ axis at $\log x_0 = 0$ (i.e. $x_0 = 1$), then A is simply given by $\log y_0$.

As an example, consider the relation between kinematic viscosity ν and temperature T given in Fig. 10 for an unknown grade of asphalt cement. (See Lab 4 for details.) At T_1 = 60°C the viscosity ν_1 is found to be 100,000 cSt, whereas at T_2 = 135°C the viscosity ν_2 is only 500 cSt. If, as shown in the figure, the data for intermediate temperatures are linear when $\log(\log \nu)$ is plotted as a function of $\log T$, then the viscosity–temperature relation must be of the form

$$\nu(T) = 10^{AT^m}. \tag{19}$$

Note that although the temperature axis is marked in degrees Celsius, the locations of the tic marks are determined by taking the logarithms of temperature in

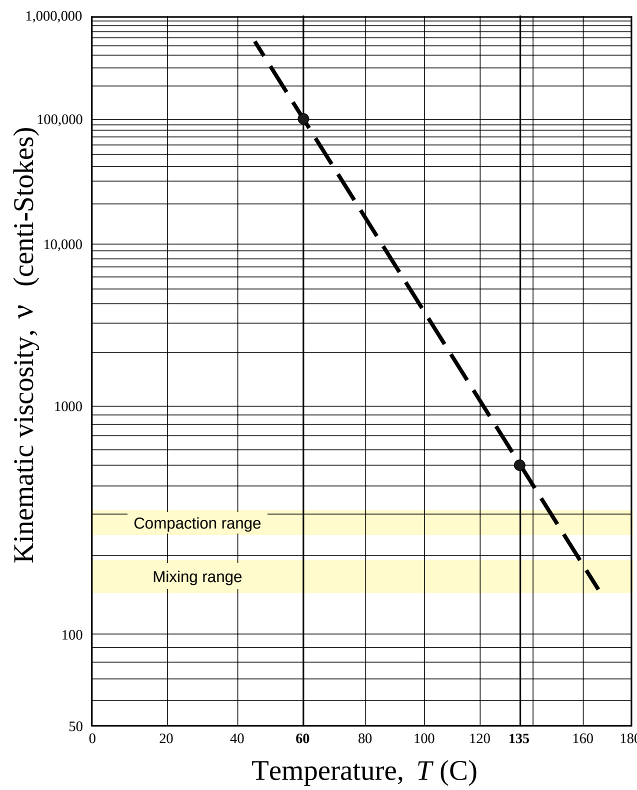

Fig. 10. Viscosity–temperature relation for a certain grade of asphalt cement.

degrees Kelvin (i.e. absolute temperature). From Eqn. (17), the dimensionless slope m is given by

$$m = \frac{\log(\log(500)) - \log(\log(100{,}000))}{\log(273+135) - \log(273+60)}$$
$$= \frac{0.431 - 0.699}{2.6107 - 2.5224} = \frac{-0.268}{0.0882}$$
$$= -3.04 \,.$$

Then, from Eqn. (15), using the data at $T = T_1$, one finds that

$$A = \frac{\log(100{,}000)}{(273+60)^{-3.04}} = 227 \times 10^6 .$$

Unlike the exponent m, the coefficient A has units, but it is difficult to ascertain what they are, given the empirical nature of the assumed relation (14). In a circumstance like this, one must specify exactly the units used for all variables in the equations. Note that even m, which is dimensionless, has a value that depends on the units chosen to express viscosity and temperature.

Once the values of A and m are determined, one can interpolate or extrapolate to find y for any x. In the present example, suppose the extrapolated value of asphalt viscosity at the freezing point of water (T = 0°C) is sought. Then, from Eqn. (19),

$$\nu(0) = 10^{(227\times 10^6)(273+0)^{-3.04}}$$
$$= 1.38 \times 10^9 \text{ cSt} ,$$

which is more than 3 orders of magnitude off the top of the graph (Fig. 10). An extrapolation to this extent may not be reasonable; but it is fair to estimate that the viscosity of this asphalt cement will be 3 to 4 orders of magnitude greater at 0°C than it is at 60°C.

6. Summary

Several plotting schemes have been presented, and there are others that have not been presented. With whatever scheme is chosen, one must exercise care in determining the constants that emanate from the straight-line fits that are sought. A good way to check results for the determined coefficients is to substitute known values of x, such as x_1 and x_2, into the appropriate expressions for y, and to make sure that the known values of y are recovered.

Printed 7/17/03