

Data fitting and error analysis

During this first lab session you will meet your lab-mates and lab demonstrators, and to learn about the `physicalab` software that you will use to collect, analyze, and plot data in all subsequent labs. We will also review essential data analysis concepts, and introduce you to error-analysis calculations.

The Cartesian coordinate system for plotting data

The Cartesian coordinate system specifies each point in a two-dimensional grid by a unique pair of coordinates, x and y . The horizontal x -axis and the vertical y -axis cross at the point $(0, 0)$, known as the origin of the coordinate system.

By marking the x and y axes at regular intervals determined by the data to be plotted, the plane is divided into a scaled array of intersecting lines, as shown in Figure 1. To plot a point at coordinates (x, y) ,

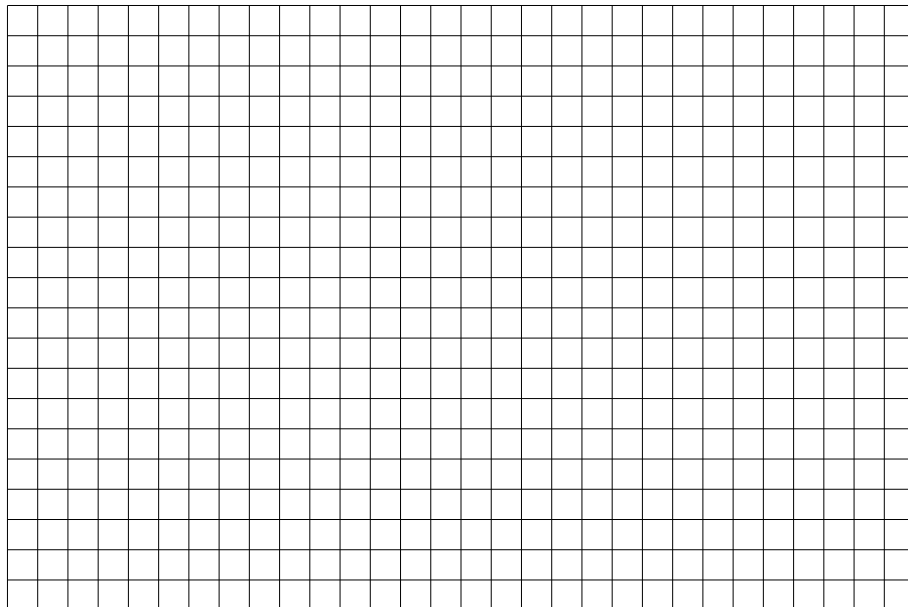


Figure 1: Cartesian coordinate grid

move horizontally along the x -axis to position x and then vertically parallel to the y -axis to position y .

What is curve fitting?

Suppose that you perform an experiment and collect some data. How would you analyse this set of measurements, or extract useful conclusions from it? For example, you might expect that the measured variables are related in some specific way; how would you decide whether the data supports the hypothesized relationship?

One can review the data and make a qualitative judgement regarding the trend that the data follows. Curve fitting attempts to quantify the behaviour of a set of data points by determining a function that represents this behaviour.

The fitting process uses a trial function that includes one or more variable parameters, called fitting parameters. The fitting parameters are adjusted by the fitting algorithm to yield a curve of best fit. The curve of best fit does not necessarily pass through any of the data points, although it might do so. Instead, the curve of best fit is placed so that a certain kind of average distance from the curve to the points is minimized. (You might like to think about what kind of average might be helpful.)

The failure of a fitted curve to pass through the data points is typically due to uncertainties associated with the values of these data points; e.g., due to the limited precision of a measuring instrument, or the presence of noise.

A mis-fit may also occur if the fit function is inappropriate for this particular data. For example, trying to fit the equation of a straight line to a periodic data set of pendulum oscillations would yield meaningless results.

Once the data is properly fitted, the fitting formula can be used to evaluate various properties inherent in the data. For example, fitting a sine function to a series of points that describe the periodic change in distance with time of a pendulum bob allows the fitting algorithm to precisely determine the period and amplitude of the pendulum oscillation.

Graphing and fitting to a linear data set

A linear data set consists of a series of coordinate points (x, y) where the relationship between the x and y coordinates is always such that a change Δy in the variable y is directly proportional to a change Δx in the variable x .

This relationship can be represented by a function $y = m * x + b$. This is the equation of a straight line with constant slope $m = \Delta y / \Delta x$. The y -intercept b is the value of y when $x = 0$.

Plotting and fitting a linear data set by hand

Here are some (x, y) coordinate points that follow a linear trend:

$$(0.2, 2.4) , (1.1, 3.6) , (2.2, 4.4) , (3.1, 5.7) , (4.1, 6.6) , (4.9, 7.3) , (6.0, 8.4)$$

Proceed as follows to plot these points on a Cartesian grid, then sketch an estimated line of best fit by hand, and finally determine the slope and y -intercept of your estimated line of best fit.

1. Review the data set and determine the x -coordinate minimum and maximum values. Subtract the minimum value from the maximum value, then divide this difference by the number of horizontal grid divisions. Round up the result, if necessary, to get a nice step increment that you can apply to the horizontal axis. (i.e., $x_{max} - x_{min} = 6.0 - 0.2 = 5.8$, $5.8/30 = 0.193$, so use 0.2 for the step size)

Label the bottom of the grid with these incremental step values, starting from the left side, so that all the x values fit on the grid. (i.e., $x_{min} = 0.2$ so start with 0.2, then 0.4, 0.6, etc.)

- Repeat the above procedure using the y -coordinates, dividing by the number of vertical grid divisions and labelling the left edge of the grid. Data graphs need not include the origin $(0, 0)$.
- With the grid now scaled, plot each of the data points onto the grid. Carefully estimate the position of points that do not lie exactly on the intersection of two grid lines.

You will note that the points appear to follow a linear behaviour; they appear to lie on a straight line. This observation gives you the hint that if you were going to try and fit some function to this data set, then the a straight line might be a good choice.

- Now use a straight edge to draw a straight line that you think best estimates your data points, extending it past the left and right edges of the grid.

? How did you choose where to place your *line of best fit*?

- Estimate the coordinates of these two edge points (x_1, y_1) on the left edge and (x_2, y_2) on the right edge, then determine the slope and y -intercept of your line of best fit, as described above. This choice of points will typically yield the least relative error.

$$\begin{aligned}
 \text{slope : } m &= \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1} = \dots\dots\dots = \dots\dots\dots \\
 y - \text{intercept : } b &= y_1 - mx_1 \equiv y_2 - mx_2 = \dots\dots\dots = \dots\dots\dots
 \end{aligned}$$

Plotting and fitting a linear data set using Physicalab

Now that you have estimated a line of best fit by hand, use the **Physicalab** software to determine a more precise line of best fit for the same set of data.

- In **Physicalab**, click File, then **Load linear data** to place some points in the data window.
- Select **fit to: y=** and enter **A*x+B** in the fitting equation box then click Draw to generate a graph of the data with a straight line fitted to it.

The slope **A** and y -intercept **B** are the *fitting parameters* adjusted by the fitting algorithm to yield the equation of the line of best fit that is drawn over your data points. The resulting values of the fitting parameters appear below the graph.

Note that the points do not lie exactly on the line. This total difference between the y -values of the points and the y -values predicted by the fitted line is used by the fitting program to determine the uncertainty in the values of the fit parameters.

The uncertainty, or error, in **A** is denoted by $\delta\mathbf{A}$ and the error in **B** is denoted by $\delta\mathbf{B}$. These results are typically displayed in the form **A** $\pm\delta\mathbf{A}$, **B** $\pm\delta\mathbf{B}$. The smaller the error values, the better the fit of the equation to the data set; if all the points were to lie exactly on the line, then these uncertainties would be zero.

? Record below the resulting values from the **Physicalab** fit of the linear data set. How do these results compare with your fit estimates for the slope and the y -intercept?

$$\begin{aligned}
 \text{slope : } \mathbf{A} &= \dots\dots\dots \pm \dots\dots\dots \\
 y - \text{intercept : } \mathbf{B} &= \dots\dots\dots \pm \dots\dots\dots
 \end{aligned}$$

Graphing and fitting to a periodic data set

A data set that is periodic has y values that repeat regularly after an interval in x . Many such periodic data sets can be fitted to a sine function, or sine wave. The basic sine function has formula $y = \sin(x)$, and a general sine function has formula $y = A \sin(Bx + C) + D$. Do you recall from high school how changing the values of A, B, C , and D influence the graph?

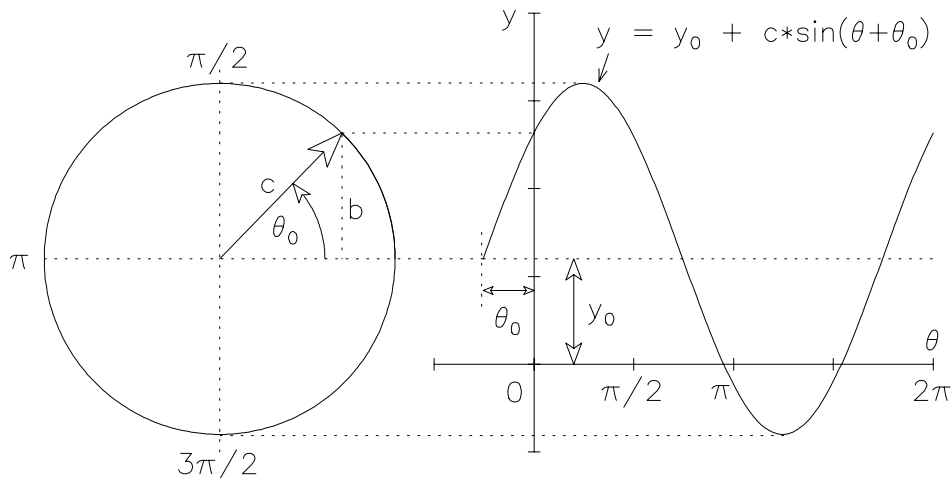


Figure 2: Projection of a circular motion to an xy -plane to generate a sine wave

Consider the radius vector of magnitude c that describes the circumference of a circle, as shown in Figure 2. Recall that a circle of radius r has a circumference of $2\pi r$. If we sweep the radius vector along this circumference by an arc distance equal to the circle radius, then the angle θ is increased by one radian. If we increase θ at a constant rate ω from 0 to 2π radians and plot the magnitude of the vertical line segment $b = c \sin \theta$ as a function of θ , a sine wave of amplitude c and period of 2π radians is generated.

You can see this in Figure 2 if you imagine a spotlight shining on the radius vector from the left so that the radius vector casts a shadow on a screen located towards the right (say, on the y -axis). As the radius vector rotates in a circle, the shadow of its tip just oscillates up and down on the screen. The location of the shadow of the tip of the radius vector is plotted on the sine curve at the right of the figure as time passes.

The variable $\omega = \Delta\theta/\Delta t$, represents the angular velocity of the radius vector as it rotates in a circle on the left diagram of Figure 2. This same quantity determines how quickly the sine wave evolves in time. For motions that are periodic, but are not circular, the quantity ω is called the angular frequency. An example of such a motion is the back-and-forth motion of the shadow of the tip of the radius vector.

If the sine wave advances by one cycle then the angle θ changes by $\Delta\theta = 2\pi$ radians. The time needed for the sine wave to advance by one cycle is called the period of the motion. Thus, the period T is related to the angular frequency ω by

$$\omega = \frac{2\pi}{T} \quad \text{which is equivalent to} \quad T = \frac{2\pi}{\omega}$$

Relative to some arbitrary coordinate system, in this case the xy -axes shown, the sine wave is shifted by an *offset distance* of y_0 from the horizontal axis and by a *phase angle* of θ_0 from the vertical axis. In this (θ, y) coordinate system, the sine wave is represented by the formula $y = c \sin(\theta + \theta_0) + y_0$. Expressing the formula using ω we get

$$y = c \sin(\omega t + \theta_0) + y_0 \tag{1}$$

Physicalab uses the same format, but different symbols, as follows:

$$y = A \sin(Bx + C) + D$$

By referencing Figure 2 you will be able to provide graphical interpretations for the Physicalab fitting parameters A , B , C , and D .

Fitting to a sine wave using Physicalab

- Click **File**, then **Load sinusoidal data**. This is some sample data obtained from the pendulum experiment (Experiment 1) that you will perform later in the course.
- Click **Draw** to generate a graph of the data. Always check that the data points follow the general shape of the function that you are going to fit to the data. In this case, the set of points should resemble a nice smooth sine wave, without spikes, stray points or flat spots.
- Select **fit to: y=** and enter **A*sin(B*x+C)+D** in the fitting equation box. Comparing this equation with Equation 1, we see that \mathbf{x} represents the independent variable (typically in units of time), \mathbf{A} is the amplitude of the sine wave, \mathbf{C} is the initial phase angle (in radians) of the wave when $\mathbf{x} = 0$, and \mathbf{D} is the offset distance of the wave measured from the x -axis, so that when $\mathbf{A} = 0$, $y = \mathbf{D}$.

The fit parameter \mathbf{B} (in radians/s) is the rate of change in angle with time, so that $\mathbf{B}*\mathbf{x}$ is an angle in radians. If this angle is advanced by 2π radians, so that $\mathbf{B}*\mathbf{x} = 2\pi$, then the time changes by an amount $\mathbf{x} = T$, where T represents the period of the sine wave in seconds. Then $\mathbf{B} = 2\pi/T$.

- Look at your graph and enter some reasonable *approximate* values for the fitting parameters. You can get an initial guess for \mathbf{B} by estimating the time \mathbf{x} between two adjacent minima, or one period, of the sine wave and approximate the value of $2\pi = 6.2831\dots$ with the number 6.

amplitude : $\mathbf{A} \approx \dots\dots\dots$
angular velocity : $\mathbf{B} \approx \dots\dots\dots$
initial phase angle : $\mathbf{C} \approx \dots\dots\dots$
offset : $\mathbf{D} \approx \dots\dots\dots$

- Click **Draw**. If you get an error message the initial guesses for the fitting parameters may be too distant from the required values for the fitting program to properly converge. Try again.

Introduction to the analysis of uncertainty (error analysis)

Beware! In general conversation, the term **error** usually refers to some sort of **mistake**. In sciences the term **error** refers specifically to the **uncertainty** δX in the **magnitude** of a measured quantity X . A measured quantity X without an associated error δX is meaningless, as there is no way to establish how reliable the value of X is. A proper result is expressed as a pair of numbers $X \pm \delta X$.

A sample data set

Table 1 contains a sample data set obtained from the Pendulum experiment. In the experiment, the length s of a string supporting a ball of diameter d that acts as the pendulum bob is adjusted and the length s is

Run, i	mass	m (kg)	d (m)	s (m)	L (m)	B (rad/s)	g_i (m/s ²)
1	m_1	0.0225	0.02540	0.300	0.3127	5.59641±0.00267	9.79370
2	m_1	0.0225	0.02540	0.450	0.4627	4.60396±0.00224	9.80760
3	m_1	0.0225	0.02540	0.600	0.6127	4.00688±0.00163	9.83695
4	m_1	0.0225	0.02540	0.750	0.7627	3.58703±0.00182	9.81350
5	m_1	0.0225	0.02540	0.900	0.9127	3.27880±0.00229	9.81201
1	m_2	0.0095	0.01904	0.500	0.5095	4.38192±0.00240	9.78344

Table 1: Table of experimental results

measured and recorded. The pendulum length L is the string length plus the radius of the ball: $L = s + \frac{1}{2}d$. (Note that the radius of the ball is half of its diameter.)

The pendulum ball is set swinging and the motion is recorded and fitted to a sine function. The fit parameter \mathbf{B} is obtained from the fit and the acceleration due to gravity g is obtained from $g = \mathbf{B}^2 L$.

There were five trials done using a ball of mass m_1 and a single trial using a ball of mass m_2 . The balls have a different diameter d . We will use this data to determine by different methods values for g and their associated errors δg .

Determining the uncertainty in a single measurement

Because there was only one trial using the ball of mass m_2 , the only option is to use error propagation rules to determine error estimates for L and g .

First, we need to determine the magnitude of the uncertainties in the measured values of the string length s and the diameter of the ball d . The **measurement errors** in s and d , represented by δs and δd respectively, are determined from the **precision** of the scales of the measuring instruments. This error is expressed as \pm one-half of the smallest increment, or **resolution**, of the scale used to make the measurement. The scale used to set the string length s had a resolution of 0.001 m, while the micrometer used to measure the ball diameter d had a scale increment, of 0.00001 m. The errors are:

$$\delta s = \pm \dots \dots \dots \quad \delta d = \pm \dots \dots \dots$$

The proper way to show a calculation is in three steps: first display the relevant equation, then replace the variables by their unrounded values and finally show the numerical result. Do not include units. The error equation for L is obtained from the error rules in the Appendix. Then for m_2 ,

$$L = s + \frac{1}{2}d = \dots \dots \dots = \dots \dots \dots$$

$$\delta L = \sqrt{(\delta s)^2 + (\frac{1}{2}\delta d)^2} = \dots \dots \dots = \dots \dots \dots$$

Once these two values are obtained, the final result for $L \pm \delta L$, is obtained as follows:

1. round the error term, in this case δL , to one significant digit. Round up this digit by adding one to it if the digit that followed it was greater than four, i.e. 0.0149 \rightarrow 0.01 and 0.0150 \rightarrow 0.02;

2. round the result, in this case L , to the same precision (or decimal place) as the error term, rounding up this digit if the digit that followed it was greater than four.

For example, if $\delta X = 0.014$ and $X = 4.354$ then $X \pm \delta X = 4.35 \pm 0.01$. If $\delta X = 0.015$ and $X = 4.355$ then $X \pm \delta X = 4.36 \pm 0.02$.

Finally, use the format below to show the properly rounded pair of values, with the appropriate units:

$$L = \dots\dots\dots \pm \dots\dots\dots$$

To get a final result for g using the ball of mass m_2 , the error equation for $g = \mathbf{B}^2 L$ is needed. This time it is a good idea to perform a change of variables to determine δg .

Note that the error equation is a product of two terms, \mathbf{B}^2 and L . Let $A = \mathbf{B}^2$ then $g = AL$. Using the power rule and the product rule respectively:

$$F = X^2 \rightarrow \frac{\delta F}{F} = 2 \frac{\delta X}{X}, \quad F = XY \rightarrow \frac{\delta F}{F} = \sqrt{\left(\frac{\delta X}{X}\right)^2 + \left(\frac{\delta Y}{Y}\right)^2} \quad (2)$$

we replace the variables and substitute the $\delta A/A$ term in the g error equation:

$$A = \mathbf{B}^2 \rightarrow \frac{\delta A}{A} = 2 \frac{\delta \mathbf{B}}{\mathbf{B}}, \quad g = AL \rightarrow \frac{\delta g}{g} = \sqrt{\left(\frac{\delta A}{A}\right)^2 + \left(\frac{\delta L}{L}\right)^2} = \sqrt{\left(2 \frac{\delta \mathbf{B}}{\mathbf{B}}\right)^2 + \left(\frac{\delta L}{L}\right)^2} \quad (3)$$

$$g = \mathbf{B}^2 L = \dots\dots\dots = \dots\dots\dots$$

$$\delta g = g \sqrt{\left(2 \frac{\delta \mathbf{B}}{\mathbf{B}}\right)^2 + \left(\frac{\delta L}{L}\right)^2} = \dots\dots\dots = \dots\dots\dots$$

Finally,

$$g = \dots\dots\dots \pm \dots\dots\dots$$

Determining the uncertainty from a series of measurements

Note: the symbol Δ is traditionally used to represent a difference of two values, as with $\Delta x = x_2 - x_1$, whereas the symbols δ and σ represent the uncertainty associated with a result.

In Table 1, there are five trials ($i = 1, \dots, N = 5$) using the large ball of mass m_1 . These five results for g are *expected to have the same value*. In this case, you can invoke the theory of statistics to evaluate a sample average $\langle g \rangle$, or mean value, of the five trials as well as the standard deviation of the sample $\sigma(g)$ that gives a measure of the scattering of the trials around $\langle g \rangle$.

The sample average of N values g_i is given by the sum of the samples divided by the number of samples:

$$\langle g \rangle = \frac{1}{N} \sum_1^N g_i = \frac{g_1 + g_2 + \dots + g_N}{N} \quad (4)$$

To get a feel for what is involved, let's perform a manual standard deviation calculation.

- To begin, we use Equation 4 to calculate $\langle g \rangle$.
- Then for each g_i we calculate the difference Δg_i and $(\Delta g_i)^2$.

i	g_i	$\Delta g_i = g_i - \langle g \rangle$	$(\Delta g_i)^2$
1			
2			
3			
4			
5			
$\langle g \rangle =$		$variance =$	
		$\sigma(g) = \sqrt{variance} =$	

Table 2: Calculation template for g and $\sigma(g)$.

- The variance of the sample is a sum of the $(\Delta g_i)^2$ terms, this time divided by $N - 1$:

$$variance = \sigma(g)^2 = \frac{1}{N - 1} \sum_1^N (\Delta g_i)^2 = \frac{(\Delta g_1)^2 + (\Delta g_2)^2 + \dots + (\Delta g_N)^2}{N - 1}.$$

- Finally, we determine the standard deviation $\sigma(g)$ from the variance in the sample of g values:

$$g = \dots \pm \dots$$

- You can also enter the five g values as a column in **Physicalab**, then from the **Edit** menu select **Insert X Index** to add a column of index values. Check **bellcurve** to view your data as a distribution.

Compare the mean and standard deviation values from the graph with your results from Table 2. They should be the same.

.....

- To view a distribution of g values with a larger number of samples, click **File** and **Load distribution data**. Check the **bargraph** box to display the number of samples in each bin as a series of vertical bars. See how the graph changes as you partition the data into a different number of bins.

$$g = \dots \pm \dots$$

How does this result for a larger data sample compare with the result for a sample of five g values?

.....