A Curve Fitting by the Method of Least Squares

Consider that you have collected a fair number of data points from an experiment you have been running and it appears that, except for some experimental scatter, the relation between two variables of interest, say x and y, is linear. Because of the scatter in the measurements however it is difficult to decide on the position of the "best" straight line through these data points. The method of *Least Squares* is one method that can be used to *justify a linear fit* and place the line through the data points.

If we believe the relationship between x and y is linear then we seek a relation of the form

$$y = Ax + C \tag{1}$$

where A and C are constants to be determined. Suppose that we write this equation,

$$y_i - (Ax_i + C) = S_i^{1/2} \tag{2}$$

where the subscript i refers to the i^{th} set of data points (x_i, y_i). If x_i and y_i fall exactly on a straight line we seek $S_i = 0$; if they do not then $S_i \neq O$. If we square the last equation,

$$S_i = [y_i - (Ax_i + C)]^2 \tag{3}$$

and sum the results for all N data points,

$$S = \sum_{i=1}^{N} [y_i - (Ax_i + C)]^2 \tag{4}$$

Again, if all the data points fall exactly on a line then $S = 0$. Since we expect experimental errors this won't be true. However, we can place a straight line through the data such that we require $S = min$, that is, such that the *sum of the squares of the deviations from the line are a minimum.* To find the minimum we set,

$$\frac{\partial S}{\partial A} = 0 \qquad \text{and} \qquad \frac{\partial S}{\partial C} = 0 \tag{5}$$

When these operations are performed we obtain

$$NC + A\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} y_i \tag{6}$$

$$C\sum_{i=1}^{N} x_i + A\sum_{i=1}^{N} {x_i}^2 = \sum_{i=1}^{N} y_i x_i \tag{7}$$

The two equations may be solved for the coefficients A and C,

$$A = \frac{N\sum x_i y_i - \sum y_i \sum x_i}{(N\sum {x_i}^2) - (\sum x_i)^2} \tag{8}$$

$$C = \frac{(\sum y_i)\sum {x_i}^2 - (\sum x_i y_i)\sum x_i}{(N\sum {x_i}^2) - (\sum x_i)^2} \tag{9}$$

and the line given by

$$y = Ax + C \tag{10}$$

may be drawn. The method of least squares is a powerful tool for fitting equations to data. The method is not limited to linear relations but may be used to fit higher order polynomials to data. It may also be used to fit equations which can be put into linear form. For example, the power law relation

$$y = Ax^n \tag{11}$$

may be written in linear form

$$\log y = n \log x + \log A \tag{12}$$

As always, care must be taken before blind implementation of the method of least squares. It is always proper to ask whether or not the relation of interest *should* be linear. When the data is presented in graphical form it is usually obvious whether a linear fit is appropriate. Also it is *always* a good idea to refer to the equations of interest to see what they may reveal.

References: [16]

B Error Analysis

In experimental work one strives to measure true values and to compute true results from these values. Recognizing that rarely, if ever, does one measure a true value, it is necessary to anticipate the uncertainties which are associated with measured values and hence the error associated with a computed result. In this way one can describe the accuracy (deviation from the true value) of the measured data and the accuracy of the computed results.

There are two fundamental errors associated with measured data; they can be called *fixed* (= *systematic*) *errors*, and *accidental* or *random errors.* The fixed errors are characteristics of the instruments (e.g. a biased voltmeter) and the operator (e.g. parallax in reading a meter) and are made *systematically.* They can be eliminated by calibration. After calibration it is necessary to examine the data for accidental or random errors which themselves can be separated into small and large errors. The latter often are mistakes while the former often constitute errors inherent in the instruments or the operator.

Calibration is achieved by comparing the performance of an instrument with a standard. Standards are defined references and these are classified as primary, secondary, etc. Primary standards are, for example, those of the National Bureau of Standards in Washington, D.C. A secondary standard is one which has been compared to a primary standard. During calibration several readings are taken at each point and averaged to obtain a best estimate. In this way the influence of the random errors associated with the calibration are reduced. Usually one constructs a calibration curve with this information. If an instrument is one whose reading depends only on a well-established physical law, then it is called a primary instrument. A weigh tank and stop watch, or a manometer, are examples of accepted primary instruments.

The large random errors (mistakes) can be eliminated by careful examination of the data. Intuition tells us that the small random errors will be reduced if many readings are taken and averaged. However, when one takes data, an error is always associated with the results and it often is important to include an error estimate when reporting data.

Before proceeding it is necessary to define several terms which are used to discuss errors.

Accuracy: - the ability of an instrument to determine the true value of a quantity.

Precision: - the ability of an instrument to repeat a measurement.

The distribution curves shown in Figure 5 illustrates these definitions for high precision and low precision instruments which are equally inaccurate. A statement about error and limits is necessary to describe precision.

Error: $\pm\Delta$, i.e., ± 2 deg, $\pm 2\ ft.$, $\pm 2\ psia$, etc.

Limits: It is most difficult to put limits on error estimates because total familiarization with the instrumentation and the experiment is required. Even when one reaches the point where it is possible to attach limits to errors, this does not enable one to immediately understand the propagation of errors. Since engineering results usually consist of several measurements it is necessary to have some laws by which the individual errors in measurement propagate to form an error in the computed result.

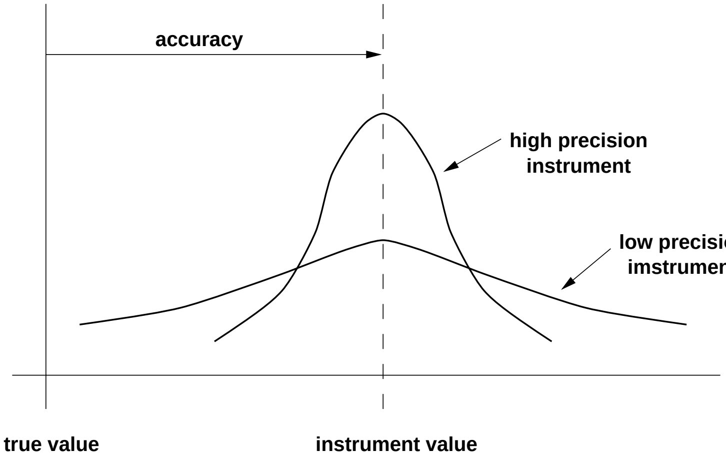

Figure 5: Distribution Curves for Equally Inaccurate Instruments.

Uncertainty: In this study two kinds of errors will be considered. Suppose that the value of a parameter, R, depends on the independent variables x_i as

$$R = R(x_1, x_2, x_3 \ldots x_i) \tag{13}$$

For example, the volume of a parallelepiped depends on width, depth, and height. The maximum possible error in R, which we call $(\Delta R)_{max}$, is obtained if no distinction is made as to whether the deviations from the mean value of the measured independent variables are positive or negative. In this case there is no compensation for cancelling errors. It can be shown that if equation (13) applies, then

$$(\Delta R)_{max} = \pm \sum_i \left| \frac{\partial R}{\partial x_i} \Delta x_i \right| \tag{14}$$

where Δx_i is the uncertainty in that independent variable [16].

If the uncertainties in the independent variables are all given with the same odds, and if allowance is made for the fact that random errors tend to cancel out (since they are both ±) over a large number of measurements, then the 'uncertainty' in a computed result is given by

$$(\Delta R) = \pm \left(\sum_i \left(\frac{\partial R}{\partial x_i} \Delta x_i \right)^2 \right)^{1/2} \tag{15}$$

with the same odds as the independent variables [16].

Suppose equation (13) is of the form

$$R = K\,(x_1)^{a_1}\,(x_2)^{a_2}\,(x_3)^{a_3} \ldots (x_n)^{a_n} \tag{16}$$

where K is a constant coefficient and the a_n are constant exponents on the independent variables. For such a relationship it may be shown that

$$\left(\frac{\Delta R}{R} \right)_{max} = \pm \sum_i \left| a_i \frac{\Delta x_i}{x_i} \right| \tag{17}$$

and

$$\frac{\Delta R}{R} = \pm \left(\sum_i \left(a_i \frac{\Delta x_i}{x_i} \right)^2 \right)^{1/2} \tag{18}$$

It is important to remember that equations (17) and (18) are valid only if R has the form of equation (16).

Example: A rotameter was calibrated using a stop watch and a graduated cylinder. If each can be read accurately to ± 1 scale division, what is the error in the mass flow rate reading (= rotameter reading) after the calibration?

The flow rate is given by the equation

$$Q = \frac{3600}{434} \gamma \rho_{H_2O} \frac{y}{t} \tag{19}$$

with the variables

ρ_{H_2O} : density of water in gms. per cc
γ : specific gravity of the fluid, dimensionless
y : volume, in cc
t : time of measurement, in seconds
Q : flow rate in lbm per hour

The following information is obtained from an examination of a hypothetical experimental setup.

Smallest scale divisions

time : = 1 sec
graduated cylinder : = 5 cc
rotameter : = 0.4 lbm/hr

Other data are

ρ_{H_2O} : = 1 gm/cc
γ : = 0.850 ± 0.005
graduated cylinder capacity : = 300 cc

The relationship has the form of equation (16) with

$a_1 = a_2 = a_3 = -a_4 = 1$
$\Delta x_1 \;=\; 0.005$
$\Delta x_2 \;=\; 0$ (ρ_{H_2O} is not measured)
$\Delta x_3 \;=\; 5$
$\Delta x_4 \;=\; 1$

The maximum error in the "flow rate" is obtained from equation (17)

$$\begin{aligned} \pm\left(\frac{\Delta Q}{Q}\right)_{max} &= \left|a_1\frac{\Delta x_1}{x_1}\right| + \left|a_2\frac{\Delta x_2}{x_2}\right| + \left|a_3\frac{\Delta x_3}{x_3}\right| + \left|a_4\frac{\Delta x_4}{x_4}\right| \\ &= \left|1\cdot\frac{0.005}{0.850}\right| + \left|1\cdot\frac{0}{62.4}\right| + \left|1\cdot\frac{5}{y}\right| + \left|-1\cdot\frac{1}{t}\right| \end{aligned} \tag{20}$$

At the lowest flowrate, we measure $y = 300$ cc and $t = 80.8$ sec, which from equation (19) gives a mass flowrate of $Q = 25$ lbm/hr. The maximum error for this computed flowrate is $\pm(\Delta Q/Q)_{max}$ = 0.0349. Since there also is an error in reading the flow meter later (= after the calibration is finished), the computations are still incomplete. The maximum error in the *observed* flow rate (rotameter reading after calibration) is

$$\pm\left(\frac{\Delta R}{R}\right)_{max} = \left(\frac{\Delta Q}{Q}\right)_{max} + \left(\frac{\Delta r}{r}\right) \tag{21}$$

where r is the meter reading and Δr is the error associated with it. At 25 lbm/hr:

$$\pm\left(\frac{\Delta R}{R}\right)_{max} = 0.0349 + \frac{0.4}{25} = 0.0509\ . \tag{22}$$

Similarly, at the highest flowrate, we measure $y = 300$ cc and $t = 40.4$ sec, which from equation (19) gives a flowrate of $Q = 50$ lbm/hr. For this flowrate we get:

$$\pm\left(\frac{\Delta R}{R}\right)_{max} = 0.0473 + \frac{0.4}{50} = 0.0553 \tag{23}$$

Thus the maximum error associated with the use of the rotameter varies from 5.1 to 5.5 percent over the flow range of interest.

Now considering 'uncertainty' rather than 'maximum error', and using equation (18) for the uncertainty $\frac{\Delta Q}{Q}$, then the uncertainty in reading the rotameter after calibration $\frac{\Delta R}{R}$ is calculated from

$$\pm\left(\frac{\Delta R}{R}\right)=\frac{\Delta Q}{Q}+\frac{\Delta r}{r} \tag{24}$$

to give

$$\pm\left(\frac{\Delta R}{R}\right) = 3.75\% \text{ at } 25 \text{ lbm/hr} \tag{25}$$

$$\pm\left(\frac{\Delta R}{R}\right) = 3.84\% \text{ at } 50 \text{ lbm/hr} . \tag{26}$$