Introduction to least squares refinement

Peter Main

Mean and variance

For a set of measurements x_i, $i = 1..n$, the mean is defined as:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

The variance is a measure of the spread of values and is given by:

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

The square root of the variance is the standard deviation, σ.

Weighted mean

If some measurements are considered to be more reliable than others, they may be weighted according to their reliability:

weighted mean: $$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

variance: $$\sigma^2 = \frac{n}{n-1}\frac{\sum_{i=1}^{n} w_i\left(x_i - \bar{x}\right)^2}{\sum_{i=1}^{n} w_i}$$

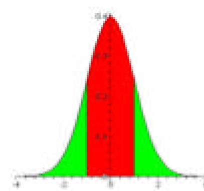
To minimise the variance: $w_i \propto 1/(\text{expected error in } x_i)^2$

Normal (Gaussian) probability distribution

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\left(x-\bar{x}\right)^2}{2\sigma^2}\right)$$

where $\bar{x}$ = mean; σ^2 = variance

This distribution is important because it is produced by a combination of a sufficiently large number of other distributions (Central Limit Theorem).



- The significance of σ is that 68% of observations are expected to be within one standard deviation of the mean.
- 99.7% of observations are expected to be within 3 σ of the mean.
- It is common practice to write a mean of 1.56 and a standard deviation of 0.03 as 1.56(3).

An example calculation

7 measurements of the length of a football pitch:

86.5, 87.0, 86.1, 85.9, 86.2, 86.0, 86.4 m

the mean is: $\bar{x} = \frac{1}{7}\sum_{i=1}^{7} x_i = \frac{604.1}{7} = 86.3$ m

variance: $\sigma^2 = \frac{1}{6}\sum_{i=1}^{7}(x_i - 86.3)^2 = \frac{0.84}{6} = 0.14$ m^2

length = 86.3(4)m

If the first two measurements have twice the expected error of the others, their relative weights will be ¼ that of the others:

weighted mean: $\bar{x} = \frac{\sum w_i x_i}{\sum w_i} = \frac{474.0}{5.5} = 86.2$ m

variance: $\sigma^2 = \frac{7}{6}\frac{\sum w_i \Delta x_i^2}{\sum w_i} = \frac{7}{6}\frac{0.36}{5.5} = 0.076$ m^2

length = 86.2(3)m

Linear regression

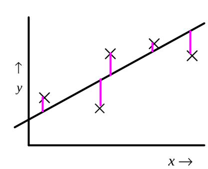
equation of the line: $y = mx + c$

observational equation: $y_i = m x_i + c$

set of observational equations:

$$\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \dots & \dots \\ x_n & 1 \end{pmatrix}\begin{pmatrix} m \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}$$

n equations in 2 unknowns



The residual of each equation is: $\varepsilon_i = y_i - m x_i - c$

which is the vertical distance of each point from the line

Define the best values of m and c as those which minimise $\sum_{i=1}^{n}\varepsilon_i^2$

Write the observational equations as: $A x = b$

A is known as the **design matrix**

Now solve the **normal equations** of least squares: $A^T A x = A^T b$

which is 2 equations in 2 unknowns giving the values of m and c

Use of weights

The observational equations may be given weights w_i according to their reliability, where $w_i \propto 1/(\text{expected value of } e_i)^2$

The least squares process will then minimise $\sum_{i=1}^{n} w_i \varepsilon_i^2$

This is done by multiplying the observational equations by a weight matrix, which is normally a diagonal matrix with the weights on the diagonal: $WAx = Wb$

The normal equations of least squares are: $A^T WAx = A^T Wb$

Expressing the least squares solution of the equations in matrix notation allows the use of as many equations and unknowns as desired.

Variance and covariance

If a quantity x is derived from two measured parameters a and b as:

$$x = \alpha a + \beta b$$

the variance of x is: $\sigma_x^2 = \alpha^2 \sigma_a^2 + \beta^2 \sigma_b^2 + 2\alpha\beta\sigma_a\sigma_b\,\mu_{ab}$

where $\sigma_a\sigma_b\mu_{ab}$ = covariance of a and b

μ_{ab} = correlation coefficient $\quad -1 \le \mu_{ab} \le +1$

- measures how strongly errors in a and b are related

The variance-covariance matrix

observational equations: $Ax = b$

normal equations: $A^T WAx = A^T Wb$

parameters x will minimise $\sum_{i=1}^{n} w_i \varepsilon_i^2$

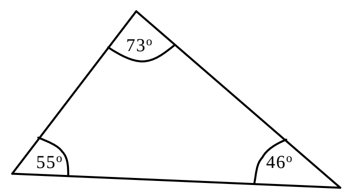
the variance-covariance matrix is: $M = \dfrac{n}{n-p}\dfrac{\sum w_i \varepsilon_i^2}{\sum w_i}\left(A^T WA\right)^{-1}$

n = number of observational equations
p = number of parameters
$n - p$ = number of degrees of freedom

M contains variances on the diagonal, covariances off-diagonal

e.g. $M = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\mu_{12} \\ \sigma_1\sigma_2\mu_{12} & \sigma_2^2 \end{pmatrix}$

Darkie's field



An inept surveyor measures the angles of a triangular field

Restraints

Observational equations: $a = 73^\circ$, $b = 46^\circ$, $g = 55^\circ$

Restraint equation: $a + b + g = 180^\circ$

Restraints are only approximately satisfied.

They are added to the observational equations:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} a \\ b \\ g \end{pmatrix} = \begin{pmatrix} 73 \\ 46 \\ 55 \\ 180 \end{pmatrix}$$

The normal equations are formed by:

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} a \\ b \\ g \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} 73 \\ 46 \\ 55 \\ 180 \end{pmatrix}$$

which gives:

$$\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}\begin{pmatrix} a \\ b \\ g \end{pmatrix} = \begin{pmatrix} 253 \\ 226 \\ 235 \end{pmatrix}$$

and the solution is: $a = 74.5^\circ \quad b = 47.5^\circ \quad g = 56.5^\circ$

The previous sum of angles was 174°

The new sum of angles is 178.5°

Use of weights

Assume the expected error in a is ½ the expected errors in ß and ?.

To minimise $\sum_{i=1}^{n} w_i e_i^2 \qquad w_i \propto 1/(\text{expected error})^2$

In addition, we will apply the constraint more strongly.

The weighted observational and constraint equations will therefore be:

$$\begin{pmatrix}1&0&0&1\\0&1&0&1\\0&0&1&1\end{pmatrix}\begin{pmatrix}4&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&4\end{pmatrix}\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\\1&1&1\end{pmatrix}\begin{pmatrix}a\\ ß\\ ?\end{pmatrix}=\begin{pmatrix}1&0&0&1\\0&1&0&1\\0&0&1&1\end{pmatrix}\begin{pmatrix}4&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&4\end{pmatrix}\begin{pmatrix}73\\46\\55\\180\end{pmatrix}$$

Multiplying the matrices gives the normal equations of least squares:

$$\begin{pmatrix}8&4&4\\4&5&4\\4&4&5\end{pmatrix}\begin{pmatrix}a\\b\\g\end{pmatrix}=\begin{pmatrix}1012\\766\\775\end{pmatrix}$$

and their solution is $a = 73.6^o,\ b = 48.4^o,\ g = 57.4^o$

- Notice that a has moved away from its measured value by a smaller amount than either ß or ? because of its assumed smaller error.
- Notice also that the sum of angles is now 179.4^o - greater than its previous value of 178.5^o because of the stronger restraint.

Further observations

It appears that any desired result can be obtained by adjusting the weights to some appropriate value.

If you are to trust the numbers you get from the calculation, you must put genuine information into it.

The weights must therefore reflect the actual reliability of the observations.

If you put rubbish into your calculation, you will get rubbish out.

Constraints

The sum of the angles of a triangle is exactly 180 degrees so it is sensible to impose this restriction rigorously.

This is done using constraints which are satisfied exactly, whereas restraints are satisfied only approximately.

To solve the equations $\boldsymbol{Ax} = \boldsymbol{b}$

subject to the constraints $\boldsymbol{Gx} = \boldsymbol{f}$

introduce a new set of variables **?** such that

$$\boxed{\begin{aligned}Ax + G^T l &= b\\ Gx &= f\end{aligned}}$$

These equations are solved for ***x*** and **?**.

The constraints will be satisfied exactly and the observational equations will be satisfied approximately.

Constraints example

Observational equations:

$$\left.\begin{aligned}a &= 73^o\\ b &= 46^o\\ g &= 55^o\end{aligned}\right\}$$

Constraint equation: $a + b + g = 180^o$

Now solve:

$$\begin{pmatrix}1&0&0&1\\0&1&0&1\\0&0&1&1\\1&1&1&0\end{pmatrix}\begin{pmatrix}a\\b\\g\\l\end{pmatrix}=\begin{pmatrix}73\\46\\55\\180\end{pmatrix}$$

The solution is: $a = 75^o,\ b = 48^o,\ g = 57^o,\ l = -2^o$

The sum of angles is 180^o exactly.

Weighted least squares with constraints

The equations in the previous example can be expressed as:

$$\begin{pmatrix}A & \vdots & G^T\\ \cdots & & \cdots\\ G & \vdots & 0\end{pmatrix}\begin{pmatrix}x\\ \cdots\\ l\end{pmatrix}=\begin{pmatrix}b\\ \cdots\\ f\end{pmatrix}$$

For a weighted least squares solution, use:

$$\begin{pmatrix}A^T W A & \vdots & G^T\\ \cdots & & \cdots\\ G & \vdots & 0\end{pmatrix}\begin{pmatrix}x\\ \cdots\\ l\end{pmatrix}=\begin{pmatrix}A^T W b\\ \cdots\\ f\end{pmatrix}$$

Reduce the number of parameters

Crystallographic least squares programs will typically use the constraints to eliminate some of the parameters to be determined.

The constraint a + ß + ? = 180

can be rearranged to give ? = 180 – a – ß

Now use a reduced set of unknowns u, v such that

$$\begin{pmatrix} a \\ b \\ g \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 180 \end{pmatrix}$$

The observational equations will then become:

$$\begin{pmatrix} a \\ b \\ g \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 180 \end{pmatrix} = \begin{pmatrix} 73 \\ 46 \\ 55 \end{pmatrix}$$

so that

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 73 \\ 46 \\ -125 \end{pmatrix}$$

Premultiply by the lhs matrix to give the normal equations: $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 198 \\ 171 \end{pmatrix}$

The solution is $u = 75$, $v = 48$

so that $a = 75^{o}$, $b = 48^{o}$, $g = 57^{o}$

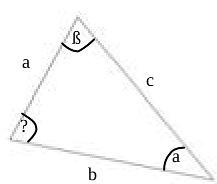
Non-linear least squares

If the sides of the triangle are known, they can be added to the observational equations: $a = 21$ m, $b = 16$ m, $c = 19$ m.

The sides are related to the angles through the sine and cosine rules,

e.g. $\dfrac{a}{\sin(a)} = \dfrac{b}{\sin(b)} = \dfrac{c}{\sin(g)}$

which can be used as restraints or constraints.



Non-linear least squares contd.

Observational equations:

$$\left.\begin{array}{r} a = 73 \\ b = 46 \\ g = 55 \\ a = 21 \\ b = 16 \\ c = 19 \end{array}\right\}$$

Restraint equations:

$$\left.\begin{array}{r} a\sin(b) - b\sin(a) = 0 \\ b\sin(g) - c\sin(b) = 0 \\ a + b + g = 180 \end{array}\right\}$$

The equations are now non-linear and no general method of solving non-linear simultaneous equations exists.

It is necessary to start with approximate values for the parameters and calculate shifts that will reduce the value of $\sum_i w_i e_i^2$

Care must be taken with the weights as the angles and lengths are completely different quantities.

Calculation of parameter shifts

Restraint equation: $a\sin(ß) - b\sin(a) = 0$

With approximate parameter values, the equation is not satisfied:

$$f(a, ß, a, b) = a\sin(ß) - b\sin(a) = e$$

Adjust the values of a, ß, a, b to reduce e to zero. This means the shifts to the parameters should satisfy:

$$\frac{\partial f}{\partial a}\Delta a + \frac{\partial f}{\partial b}\Delta b + \frac{\partial f}{\partial a}\Delta a + \frac{\partial f}{\partial b}\Delta b = -e$$

i.e. $-b\cos(a)\Delta a + a\cos(b)\Delta b + \sin(b)\Delta a - \sin(a)\Delta b = -e$

which is a linear equation in the shifts ? a, ? ß, ? a and ? b

Parameter shifts contd.

Repeating this for all observational and restraint equations gives a system of linear simultaneous equations:

$$A ? x = -e$$

matrix of derivatives | vector of shifts | vector of residuals

A weighted least squares solution of these equations is obtained from

$$A^T W A\Delta x = -A^T W e$$

Then the parameters are updated: $x_{new} = x_{old} + \Delta x$

and the new values are used to repeat the calculation to obtain further shifts until the shifts become negligibly small.

Ill - conditioning

Consider the equations:

$$23.3\,x + 37.7\,y = 14.4$$
$$8.9\,x + 14.4\,y = 5.5$$

The exact solution is: $x = -1, \quad y = +1$

Now change the rhs of the first equation to 14.39 and solve again:

$$23.3\,x + 37.7\,y = 14.39$$
$$8.9\,x + 14.4\,y = 5.5$$

This gives: $x = 13.4, \quad y = -7.9$

The tiny change in one of the equations has made the solution unrecognisable.

Ill – conditioning contd.

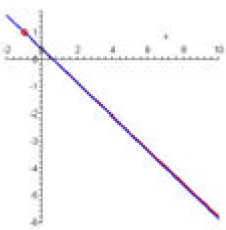
The reason for the ill-conditioning is that the second equation is an almost exact copy of the first, multiplied by a scale factor.

Putting them on the same scale gives:

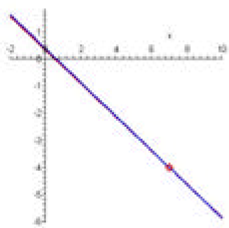
$$23.300\,x + 37.700\,y = 14.400$$
$$23.300\,x + 37.699\,y = 14.399$$

The second equation adds almost no new information to the first.

Graphical representation



These two lines intersect at (-1,+1).



Move the blue line upwards by a small amount and they now intersect at (7,-4).

Symptoms of ill-conditioning

1. A small change to any of the equations results in a large change in the solution.

2. The determinant of the lhs matrix has a small value:

$$\begin{vmatrix} 23.3 & 37.7 \\ 8.9 & 14.4 \end{vmatrix} = -0.01$$

3. The elements of the inverse matrix are large:

$$\begin{pmatrix} 23.3 & 37.7 \\ 8.9 & 14.4 \end{pmatrix}^{-1} = \begin{pmatrix} -1440 & 3770 \\ 890 & -2330 \end{pmatrix}$$

4. The estimated standard deviations of the parameters obtained from a least squares solution are large.

$$M = \frac{n}{n-p} \frac{\sum w_i \varepsilon_i^2}{\sum w_i} \left(A^T W A\right)^{-1}$$