A new way of resolving the linear equations that occur in the least squares method

C. G. J. Jacobi

Astronomische Nachrichten number 523, 1845

This is a partial translation of Jacobi's paper.

The difficulty of strictly resolving a large number of linear equations, like in many cases for the method of least squares, has led to the use of approximation methods. This is the case if in the different equations there is a variable multiplied by a large coefficient. The equations are:

$$\begin{aligned}(00)x + (01)x_1 + (02)x_2 + \cdots &= (0m),\\ (10)x + (11)x_1 + (12)x_2 + \cdots &= (1m),\\ (20)x + (21)x_1 + (22)x_2 + \cdots &= (2m),\end{aligned}$$

and so on, and all coefficients (ik) are very small compared to those in the diagonal (ii). You get an approximation of the unknowns x, x_1, x_2, etc. from the equations:

$$(00)x = (0m), \quad (11)x_1 = (1m), \quad (22)x_2 = (2m), \ldots$$

If you denote these values with a, a_1, a_2 etc., you get their first corrections, which I denote by Δ, Δ_1, Δ_2 etc., from the equations:

$$\begin{aligned}(00)\Delta &= -\left\{(01)a_1 + (02)a_2 + \cdots\right\},\\ (11)\Delta_1 &= -\left\{(10)a + (12)a_2 + \cdots\right\},\end{aligned}$$

and so on, and, in general,

$$\begin{aligned}x &= a + \Delta + \Delta^2 + \Delta^3 + \cdots,\\ x_1 &= a_1 + \Delta_1 + \Delta_1^2 + \Delta_1^3 + \cdots,\\ x_2 &= a_2 + \Delta_2 + \Delta_2^2 + \Delta_2^3 + \cdots\end{aligned}$$

and so on.

If the upper indices mean the successive decreasing corrections, Δ^{i+1} will be obtained from Δ^i by the equations

$$\begin{aligned}(00)\Delta^{i+1} &= -\left\{(01)\Delta_1^i + (02)\Delta_2^i + \cdots\right\},\\ (11)\Delta_1^{i+1} &= -\left\{(10)\Delta^i + (12)\Delta_2^i + \cdots\right\},\end{aligned}$$

and so on.

In the equations to which the least squares method leads, the coefficients in the diagonal as a whole are predominant because they are aggregates of squares, while the other coefficients are created by adding positive and negative numbers, which cancelled in each case. However, several of the coefficients located outside the diagonal could assume such significant values such that the success of the approximation method just mentioned will be hindered.

However, I will show in the following that one can transform the equations iteratively by repeating an easy calculation, in which the mentioned disadvantage arises less and less, so that the equations finally get a form that verifies the application of the above approximation method.

I assume, as is always the case in the equations to which the least squares method leads, that two coefficients outside the diagonal, (ik) and (ki) are equal, and I assume that the coefficient (01) has a significant value, the influence of which prevents the application of the approximation method. To annihilate this coefficient, I state

$$\begin{aligned} x &= \cos\alpha\cdot\eta+\sin\alpha\cdot\eta_1, \\ x_1 &= \sin\alpha\cdot\eta-\cos\alpha\cdot\eta_1. \end{aligned}$$

whereby

$$\begin{aligned} (00)x+(01)x_1 &= \{(00)\cos\alpha+(01)\sin\alpha\}\,\eta+\{(00)\sin\alpha-(01)\cos\alpha\}\,\eta_1, \\ (10)x+(11)x_1 &= \{(10)\cos\alpha+(11)\sin\alpha\}\,\eta+\{(10)\sin\alpha-(11)\cos\alpha\}\,\eta_1, \end{aligned}$$

and set the two equations:

$$\begin{aligned} u &= (00)x+(01)x_1+(02)x_2+\cdots-(0m)=0, \\ u_1 &= (10)x+(11)x_1+(12)x_2+\cdots-(1m)=0, \end{aligned}$$

through the other two

$$\begin{aligned} \nu &= \cos\alpha\cdot u+\sin\alpha\cdot u_1=0, \\ \nu_1 &= \sin\alpha\cdot u-\cos\alpha\cdot u_1=0. \end{aligned}$$

Note from the translator. This amounts to multiply

$$\begin{pmatrix}\cos\alpha & \sin\alpha\\ \sin\alpha & -\cos\alpha\end{pmatrix}\begin{pmatrix}a_{0,0} & a_{0,1}\\ a_{0,1} & a_{1,1}\end{pmatrix}\begin{pmatrix}\cos\alpha & \sin\alpha\\ \sin\alpha & -\cos\alpha\end{pmatrix}$$

and to set the $(1,0)$ coefficient to zero

$$(a_{0,0}-a_{1,1})\sin\alpha\,\cos\alpha+a_{1,0}(\sin^2\alpha-\cos^2\alpha)=0.$$

Now I determine the angle α so that

$$\{(00)-(11)\}\cos\alpha\cdot\sin\alpha=(01)\left\{\cos^2\alpha-\sin^2\alpha\right\},$$

or

$$\frac{1}{2}\tan 2\alpha=\frac{(01)}{(00)-(11)},$$

so the two new equations are

$$\begin{aligned}\{(00)\cos^2\alpha &+ 2(01)\cos\alpha\cdot\sin\alpha+(11)\sin^2\alpha\}\eta+\{(02)\cos\alpha+(12)\sin\alpha\}x_2+\cdots\\ &= (0m)\cos\alpha+(1m)\sin\alpha,\end{aligned}$$

$$\begin{aligned}\{(00)\sin^2\alpha &- 2(01)\cos\alpha\cdot\sin\alpha+(11)\cos^2\alpha\}\eta_1+\{(02)\sin\alpha-(12)\cos\alpha\}x_2+\cdots\\ &= (0m)\sin\alpha-(1m)\cos\alpha.\end{aligned}$$

The coefficients of x_2, x_3 etc. are easily calculated trigonometrically by auxiliary angles whose tangents are the same

$$\frac{(12)}{(02)},\quad\frac{(13)}{(03)},$$

with special attention to the correctness of the signs of the coefficients.

A useful control in this respect is obtained by making the assumption

$$\nu=u\cos\alpha+u_1\sin\alpha,\quad \nu_1=u\sin\alpha-u_1\cos\alpha,$$

in u and u_1, ν and ν_1, and by checking the equality of the values

$$x=\cos\alpha+\sin\alpha,\ x_1=\sin\alpha-\cos\alpha,\ \eta=\eta_1=x_2=x_3=\cdots=1.$$

The coefficients of η and η_1 can also be represented as follows:

$$\frac{(00)+(11)}{2}+\sqrt{R},\quad\frac{(00)+(11)}{2}-\sqrt{R},$$

where

$$R=\left\{\frac{(00)-(11)}{2}\right\}^2+(01)^2.$$

and the character of $\sqrt{R}$ depends on the quadrant in which 2α is taken, by means of the double formula

$$\sqrt{R}=\frac{(00)-(11)}{2\cos 2\alpha}=\frac{(01)}{\sin 2\alpha},$$

which also offers a control. Each of the other equations, like

$$(20)x+(21)x_1+(22)x_2+\cdots=(2m),$$

is transformed by introducing η and η_1 for x and x_1 into the following:

$$\{(20)\cos\alpha+(21)\sin\alpha\}\eta+\{(20)\sin\alpha-(21)\cos\alpha\}\eta_1+(22)x_2+(23)x_3+\cdots=(2m).$$

Since here the coefficients of η and η_1 are the same as the coefficients of x_2 in the first two transformed equations, one sees that the transformed equations keep the symmetry with respect to the diagonal, and that one therefore only has to calculate the coefficients of x_2, x_3 etc. in the two transformed equations in order to have also the coefficients of η and η_1 in the remaining equations. in the two transformed equations in order to have also the coefficients of η and η_1 in the remaining equations, in which in addition the coefficients of x_2, x_3 etc., as well as the constant member remain unchanged.

The coefficient corresponding to (01) is $=0$ in the transformed equations; the sum of the coefficients in the diagonal remains the same that is, $(00)+(11)$; on the other hand, the sum of their squares increases by $2(01)^2$; from which it follows that these coefficients diverge further, the larger becomes larger, the smaller becomes smaller.

However, the smaller one can never disappear if the coefficients of the defined equations are composed as is the case with the applications of the least squares method. The product of both coefficients becomes

$$\left\{\frac{(00)+(11)}{2}\right\}^2 - R = (00)(11)-(01)^2,$$

if one sets

$$\begin{aligned}(00) &= \alpha\alpha+\beta\beta+\gamma\gamma+\cdots,\\ (11) &= \alpha_1\alpha_1+\beta_1\beta_1+\gamma_1\gamma_1+\cdots,\\ (01) &= \alpha\alpha_1+\beta\beta_1+\gamma\gamma_1+\cdots\end{aligned}$$

always a positive quantity

$$(00)(11)-(01)^2=\sum(\alpha\beta_1-\beta\alpha_1)^2,$$

where the sum comprises all squares formed by combining two squares each from the elements $\alpha,\beta,\gamma,\delta$ etc. can never vanish if not all sizes $\alpha,\beta,\gamma,\delta$ etc. are proportional to the sizes $\alpha_1,\beta_1,\gamma_1,\delta_1$ etc.

The sums of the squares of the coefficients of x_2, of x_3, etc., also remain unchanged in the two transformed equations, $(02)^2+(12)^2$, $(03)^2+(13)^2$, etc. Likewise, in each of the remaining equations, the sum of the squares of the coefficients of η and η_1 become the same as of x and x_1, in the original system. The sum of the squares of the coefficients located outside the diagonal decreases by $2\,(01)^2$, which is the same amount by which the sum of the squares of the two coefficients in the diagonal has increased, so that the sum of the squares of all coefficients of the equations remains unchanged, which also applies to the sum of the squares of the constant members. From this it follows that if the transformed system is transformed again in a similar way and thus the given transformation is applied several times in succession, always taking away the most influential of the coefficients located on the diagonal, in the last obtained system equations

1) the sum of the coefficients in the diagonal, the sum of the squares of all coefficients and the sum of the squares of the completely constant members is the same as in the original system;

2) The sum of the squares of the coefficients located in the diagonal is increased, the sum of the squares of the coefficients located outside the diagonal is decreased by the same amount, namely by twice the sum of the squares of the coefficients zeroed in the individual transformations.

[...]

Berlin, Nov. 17, 1844