E. STARK°
E. MIKHAIL
Purdue University
Lafayette, Ind. 47907

Least Squares & Non-Linear Functions

Although most literature considers only linear condition equations, most practical problems are non-linear.

INTRODUCTION

THE METHOD of least squares is usually presented to apply to *linear* condition equations. There are a number of techniques for least squares-adjustments which depend on the number and type of condition equations expressing the functional model.

By functional model is meant the totality of relations and assumptions which completely express the physical and geometric elements of the situation or event under study. Of the different techniques possible, we shall concern ourselves

ABSTRACT: *It has been common practice to linearize condition equations at the given observations and at approximations for the parameters, and iterating the solution by updating parameter approximations only. Although this may well be adequate for many classical problems, it is no longer appropriate if modern unified approaches to least-squares adjustment are applied. Therefore, the adjustment technique has been extended to include approximations for the observations, thus allowing for the updating of all variables, parameters and observations alike. Direct nonlinear least squares and the technique of adjustment by "indirect observations" are used to check the correctness of the developed relations which are demonstrated by simple, but realistic, examples.*

in this paper with adjustment with the following three sets of linear condition equations:

Technique 1. Adjustment of observations and parameters combined whose condition equations take the form

$$\underset{c,n}{\mathbf{A}}\ \underset{n,1}{\mathbf{v}} + \underset{c,u}{\mathbf{B}}\ \underset{u,1}{\mathbf{\Delta}} = \underset{c,1}{\mathbf{f}}. \tag{1}$$

Technique 2. Adjustment of observations only, with the condition equations

$$\underset{r,n}{\mathbf{A}}\ \underset{n,1}{\mathbf{v}} = \underset{r,1}{\mathbf{f}}. \tag{2}$$

Technique 3. Adjustment of *indirect observations*, using the condition equations

$$\underset{n,1}{\mathbf{v}} + \underset{n,u}{\mathbf{B}}\ \underset{u,1}{\mathbf{\Delta}} = \underset{n,1}{\mathbf{f}}. \tag{3}$$

In the above three equations, n is the number of observations with residual

° Mr. Stark is with the Institut für Photogrammetry, Stuttgart, West Germany. Prof. Mikhail is with Purdue, and was a visiting professor in Stuttgart when this paper was written.

vector $\mathbf{v}$, r is the number of redundancy or degrees of freedom, u is the number of parameters in the vector $\mathbf{\Delta}$, c is the number of condition equations, $r \leqslant c \leqslant n$ and $c = r + u$, $\mathbf{A}$, $\mathbf{B}$ are coefficient matrices and $\mathbf{f}$ is a column vector.

We note that Equations 2 and 3 are special cases of 1. We also point out that both observations and parameters are considered in the above equations as functionally independent. There are of course other least-squares techniques which deal with instances of adjustment involving functionally dependent parameters. These applications, however, are more involved and are not considered here. The ideas developed in the following sections can nevertheless be extended in a straightforward manner to include these other cases of adjustment.

Non-Linear Condition Equations

In the introductory remarks we gave three sets of equations which were all linear. This is because from a computational standpoint least squares is practical only for such situations. (In rather rare applications one can apply it to non-linear conditions directly; it will be shown later how involved such an approach can be, even for the simplest of problems). In practice, most adjustment problems involve non-linear conditions, whereas directly linear conditions are quite infrequent. Consequently, one has to linearize and iterate the solution on all the variables. However, it has been so far customary in practice to iterate on *the parameters only* for applications involving Techniques 1 and 3, and to have *no iterations* at all in problems applying Technique 2.

The only justification for this practice has been the fact that the linearization is performed at the *given* observations which are *assumed* to be sufficiently close to their final estimate to warrant no iteration. Although such an assumption may apply to classical problems, we need not take such for granted. As a matter of principle, one should treat all variables in the model equally and perform the linearization and iteration on all. Although this may seem to be a purely theoretical consideration, there are situations where linearization and iteration on the observable variables is necessary. For example, in a rather general and unified treatment of least squares (the presentation of which is unfortunately not possible here because of paper length limitation) if the parameters are treated as observations, a lack of proper iteration would cause the method to fail. Consequently, in this paper we shall develop a procedure for iterating the solution, updating approximations for observations as well as parameters. In order to follow the new procedure, we briefly mention current methods.

For Technique 1, the general non-linear conditions are $F(l,\mathbf{x}) = \mathbf{0}$, where l and $\mathbf{x}$ are observations and parameters, respectively. With

$$\mathbf{A} = \left.\frac{\partial F}{\partial l}\right|_{\substack{l=l \\ x=x^\circ}}$$

$$\mathbf{B} = \left.\frac{\partial F}{\partial x}\right|_{\substack{l=l \\ x=x^\circ}}$$

$$\mathbf{f} = -F(l,\mathbf{x}^\circ)$$

the linearized equations become

$$\mathbf{A}\mathbf{v} + \mathbf{B}\,\mathbf{\Delta} = \mathbf{f}.$$

Note that l is the vector of *given values* for the observations, while $\mathbf{x}^\circ$ is a vector of *approximate values* for the parameters. The vector of corrections $\mathbf{\Delta}$ is computed by least squares and added to $\mathbf{x}^\circ$ and the process iterated until the last $\mathbf{\Delta}$ is insignificantly different from $\mathbf{0}$. At this point, the vector $\mathbf{v}$ is computed only once. The same applies to Technique 3 with the exception of $\mathbf{A}$ being the identity matrix.

In contrast to the above practice, a simple extension can be effected where the iterative process applies to both $\mathbf{v}$ and $\mathbf{\Delta}$. For simplicity in presentation we treat Technique 2 first, then the more general technique. It should be easy to recognize that Technique 3 needs no extension because the *conditions are originally linear in the observations.* This is why in the numerical examples the results from Technique 3 are used to check those from the other two.

LINEARIZATION FOR ADJUSTMENT OF OBSERVATIONS ONLY (Technique 2)

The original non-linear condition equations may be written as

$$F(l) = \mathbf{0}. \tag{4}$$

Now we distinguish between the actual numerical values of the observations l and an approximate vector l°. The vector l remains unaltered, while l° changes in value from one iteration to the next. This is different from current practice where l is always used for the linearization and no iterating is performed. At any iteration we have

$$l + \mathbf{v} = l^\circ + \mathbf{\Delta}l. \tag{5}$$

Linearizing Equation 4 we get

$$\left.\frac{\partial F}{\partial l}\right|_{l=l^\circ} \cdot \mathbf{\Delta}l = -F(l^\circ)$$

or, with obvious correspondence in terms,

$$\mathbf{A} \cdot \mathbf{\Delta}l = -F(l^\circ)$$

$$\mathbf{A}(l + \mathbf{v} - l^\circ) = -F(l^\circ)$$

or

$$\mathbf{A}\mathbf{v} = -F(l^\circ) - \mathbf{A}(l - l^\circ). \tag{6}$$

Then

$$\mathbf{A}\mathbf{v} = \mathbf{f}^\circ. \tag{6a}$$

Conveniently, the first iteration begins by an $l_1^\circ = l$ which makes $\mathbf{f}_1^\circ$ in Equation 6 equal to $\mathbf{f}$ in Equation 2 and the solution, $\mathbf{v}_1$, will be identical to that from the conventional practice. The second iteration, however, applies $l_2^\circ = l + \mathbf{v}_1$ and both $\mathbf{A}$ and $\mathbf{f}^\circ$ take new values.

The solution $\mathbf{v}_2$ from the second iteration can then be used to compute an updated vector $l_3^\circ = l + \mathbf{v}_2$. Note that in every iteration a *total* $\mathbf{v}$ is computed and not an incremental one. This is the reason why the residual vector is always added to the *given and fixed* observation vector l. The iterative process terminates when two successive vectors l°, or equivalently, when two successive vectors $\mathbf{v}$ differ by an insignificantly small value.

To ascertain the correctness of the above technique, the method of adjustment of *indirect observations* will be used as a check. Here, the non-linear conditions are of the form $l + F(x) = \mathbf{0}$ which are linearized to the form $\mathbf{v} + \mathbf{B}\ \mathbf{\Delta} = \mathbf{f}$. The solution by least squares gives $\mathbf{\Delta}$ which is added to an approximate parameter vector $\mathbf{x}^\circ$ and the scheme repeated until the last $\mathbf{\Delta}$ is tolerably small. The vector of residuals $\mathbf{v}$ computed at the end of iterations should be the proper least-squares estimate. The following examples will demonstrate the developed method and the check.

EXAMPLE 1

Figure 1 shows a much simplified problem of an object point P which is photographed by three terrestrial cameras. All three camera stations S_1, S_2, S_3 are assumed to lie on the X_1-axis of the object space coordinate system. The X_2-axis coincides with the optical axis of the first camera position, S_1. The three camera axes are parallel and lie in the same horizontal plane. All interior orientation ele-

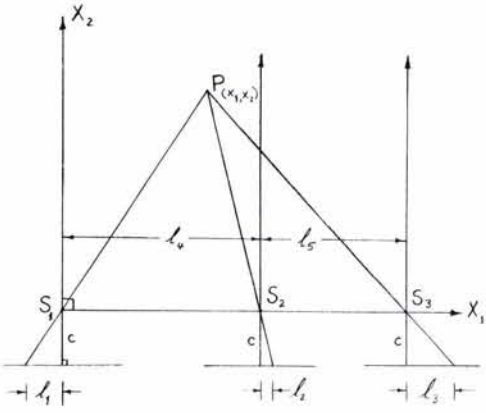

FIG. 1. A simplified three-photo problem.

ments are known to be zero except for the principal distance $c = 100$ *mm* which is to be considered as constant. The observations l are the five denoted by that symbol on the figure; their values and standard deviations, with no correlation, are given in Table 1. All observations are assumed, for simplicity, to lie in the x_1-x_2 plane.

Examining Figure 1 one would agree that the mathematical model is a rather simple one concerning geometry in a plane. The functional model is essentially specified by four points, S_1, S_2 and S_3 lying on the X_1-axis, and point P. Point S_1 is the origin of the coordinate system, S_2 and S_3 are located by one coordinate each, and P by two coordinates. This makes a total of four elements necessary and sufficient for a unique determination, and with five observations we end up with one degree of freedom, or $r = 1$. The one condition equation among the observations may be written directly from the figure as

$$-l_1 l_5 - l_2 l_4 - l_2 l_5 + l_3 l_4 = 0.$$

With regard to the stochastic model, the *a priori* cofactor matrix (covariance matrix) is given by

$$\mathbf{Q}_l = \text{diag. } \{0.01 \text{ mm}^2, 0.01 \text{ mm}^2, 0.01 \text{ mm}^2, 0.0025 \text{ m}^2, 0.0025 \text{ m}^2\}.$$

TABLE 1. DATA FOR EXAMPLE 1.

Observation	*Value*	*Standard deviation*
1	16.5 mm	0.10 mm
2	3.8 mm	0.10 mm
3	20.4 mm	0.10 mm
4	10.0 m	0.05 m
5	8.0 m	0.05 m

TABLE 2. RESIDUAL VECTORS FOR EXAMPLE 1.

$$\mathbf{v}_1 = \begin{bmatrix} 0.043,642,149,5\ (\text{mm}) \\ 0.098,194,836,5\ (\text{mm}) \\ -0.054,552,686,9\ (\text{mm}) \\ -0.022,639,365,0\ (\text{m}) \\ 0.027,685,488,6\ (\text{m}) \end{bmatrix} \quad \mathbf{v}_f = \begin{bmatrix} 0.043,775,467,4\ (\text{mm}) \\ 0.098,182,590,1\ (\text{mm}) \\ -0.054,407,122,6\ (\text{mm}) \\ -0.022,421,679,9\ (\text{m}) \\ 0.027,867,185,9\ (\text{m}) \end{bmatrix}$$

With first approximation vector $l^\circ = l$, the linearized condition equation according to Equation 6 becomes

$$[-8.0(\text{m}) \quad -18.0(\text{m}) \quad 10.0(\text{m}) \quad 16.6(\text{mm}) \quad -20.3(\text{mm})]\mathbf{v} = -3.6(\text{mm m}).$$

Solving by the method of least squares the residual vectors after the first iteration, $\mathbf{v}_1$, and the last (fourth) iteration, $\mathbf{v}_f$, are shown in Table 2. It should be emphasized that we carried 10 digits only for the purpose of checking the exactness of the method rather than for practical considerations.

To check the above method we solved the problem by the method of adjustment of *indirect observations*. The two coordinates of point P and the distances S_1S_2 and S_2S_3 were used as four parameters. The solution was iterated until the last values of the parameter corrections were all less than 10^{-10}. The vector of five residuals was identical to $\mathbf{v}_f$ above to all 10 decimal places.

An interesting alternative is to check the solution by Equation 6 of the problem in Example 1. It is possible to treat the example directly as a non-linear least-squares adjustment problem. This means that we here minimize $\mathbf{v}^t\mathbf{P}\mathbf{v}$ subject to the *non-linear* condition, or

$$\varphi = \mathbf{v}^t\mathbf{P}\mathbf{v} - 2k_n(-l_1l_5 - l_2l_4 - l_2l_5 + l_3l_4) \longrightarrow \min.$$

where k_n is the one Lagrange multiplier for the non-linear condition. Replacing l_i by $(l_i + v_i)$ and realizing that $\mathbf{P}$ is a 5×5 diagonal matrix, φ becomes

$$\begin{aligned}\varphi = {} & p_1v_1^2 + p_2v_2^2 + p_3v_3^2 + p_4v_4^2 + p_5v_5^2 + 2k_n[(l_1 + v_1)(l_5 + v_5) \\ & + (l_2 + v_2)(l_4 + v_4) + (l_2 + v_2)(l_5 + v_5) - (l_3 + v_3)(l_4 + v_4)] \longrightarrow \min.\end{aligned}$$

Differentiating φ with respect to each v_i and equating to zero we get

$$\begin{aligned} p_1v_1 + k_n(l_5 + v_5) &= 0 \\ p_2v_2 + k_n[(l_4 + v_4) + (l_5 + v_5)] &= 0 \\ p_3v_3 - k_n(l_4 + v_4) &= 0 \\ p_4v_4 + k_n[(l_2 + v_2) - (l_3 + v_3)] &= 0 \\ p_5v_5 + k_n[(l_1 + v_1) + (l_2 + v_2)] &= 0 \end{aligned}$$

These five equations together with the condition equation constitute six *non-linear* equations in six unknowns v_1 to v_5 and k_n. It is possible, with some numerical scheme, to solve directly for these six unknowns. However, we can check the correctness of $\mathbf{v}_f$ by a much simpler way: by computing five values for k_n from the above five equations using the known values of $\mathbf{v}$. Applying this technique, using $\mathbf{v}_f$ from example 1, we get the following values for k_n:

$$\begin{aligned} k_n &= -0.545,293,867,41 \\ &= -0.545,293,867,44 \\ &= -0.545,293,867,41 \\ &= -0.545,293,867,43 \\ &= -0.545,293,867,44 \end{aligned}$$

which agree with each other to the 10th decimal place. This is a good indication that the procedure is correct. (It is important to note here how a complex, non-

TABLE 3. DATA FOR EXAMPLE 2.

i	x_{1i}	x_{2i}	y_{1i}	y_{2i}
1	0.0	1.0	−2.1	1.1
2	1.0	0.0	1.0	2.0
3	1.0	1.0	−0.9	2.8

linear, least-squares adjustment can be even for a problem with one condition and a diagonal weight matrix!).

EXAMPLE 2

The two-parameter transformation equations

$$y_{1i} = ax_{1i} - bx_{2i}$$

$$y_{2i} = bx_{1i} + ax_{2i}$$

relate the x-coordinates of any point i to its y-coordinates by the two parameters a and b. One point of known coordinates in both systems would be sufficient for determining a and b. Instead we have three such points as indicated in Table 3. $\mathbf{Q}_x = \mathbf{S}_x = 0.01\mathbf{I}$ for each point with no correlation between them. Considering the y-coordinates as constants, and the x-coordinates as observations with the given covariance matrix we are interested in computing a and b.

Four condition equations exist between the six observations. It can be shown that these equations are:

$$f_{1i} = (x_{11}y_{11} + x_{21}y_{21})x_{1i} - (x_{11}y_{21} - x_{21}y_{11})x_{2i} - y_{1i}(x_{11}^2 + x_{21}^2) = 0$$

$$f_{2i} = (x_{11}y_{11} + x_{21}y_{21})x_{2i} + (x_{11}y_{21} - x_{21}y_{11})x_{1i} - y_{2i}(x_{11}^2 + x_{21}^2) = 0$$

with $i = 2$ and 3 where (x_{11},x_{21}) and (y_{11},y_{21}) are the coordinates of the first point. The linearized form is $\mathbf{Av} = \mathbf{f}^\circ$ with a $\mathbf{Q}_l = 0.01\mathbf{I}$. For $l^\circ = l$ we get the first value, $\mathbf{v}_1$, and then updating l° we iterate until the final residual vector $\mathbf{v}_f$ as shown in Table 4. The problem was also programmed in the form $\mathbf{v} + B\,\boldsymbol{\Delta} = \mathbf{f}$ and the final residual vector $\mathbf{v}_f$ agrees with the one above to even more digits than those shown, which are carried to 10 places here *only for reasons of checking.*

Consider the possibility of the conventional solution of $\mathbf{Av} = \mathbf{f}$ (Equation 2) to obtain a vector of estimated observations $\hat{l}$ and a corresponding cofactor matrix $\mathbf{Q}_{\hat{l}}$. One may consider that a relinearization with these new values can be performed and another adjustment solution carried out. Although this may seem to be correct, it will not work because $\mathbf{Q}_{\hat{l}}$ is always a singular matrix. To demonstrate this point, let us consider the data of Example 2. In the first iteration we have:

$\mathbf{A}_1$ is a 4×6 matrix with a rank of 4,
$\mathbf{Q}_l$ is a 6×6 matrix with a rank of 6,
$\mathbf{N}_1 = \mathbf{A}_1\mathbf{Q}_l\mathbf{A}_1^t$ is a 4×4 normal-equations coefficient matrix of a rank of 4.

TABLE 4. RESIDUAL VECTORS FOR EXAMPLE 2.

$$\mathbf{v}_1 = \begin{bmatrix} 0.009,296,191,9 \\ 0.080,495,537,0 \\ 0.014,389,283,7 \\ 0.011,362,012,3 \\ -0.061,013,161,5 \\ -0.063,078,981,9 \end{bmatrix} \qquad \mathbf{v}_f = \begin{bmatrix} 0.009,340,944,5 \\ 0.078,360,145,3 \\ 0.017,125,064,9 \\ 0.010,378,827,2 \\ -0.053,450,960,1 \\ -0.054,488,842,8 \end{bmatrix}$$

After the first iteration, the computed $\hat{l}$ will be functionally dependent with the consequence that $\mathbf{Q}_{\hat{l}}$ will have a rank of two. This leads to a set of four normal equations whose coefficient matrix $\mathbf{N}_2$ has a rank of two or less, and one cannot therefore iterate the solution with this scheme.

Linearization for Adjustment of Observations and Parameters Combined (Technique 1)

In a manner similar to the derivation given in the foregoing section we begin with the non-linear condition equations

$$F(l,\mathbf{x}) = \mathbf{0}.$$

Letting l and l° denote the actual observations and their approximations for the purpose of linearization, $\mathbf{x}^\circ$ denote approximations for the parameters, Δl corrections to l°, and $\boldsymbol{\Delta}$ corrections to $\mathbf{x}^\circ$, the condition equations may be rewritten as

$$F(l^\circ + \Delta l,\ \mathbf{x}^\circ + \boldsymbol{\Delta}) = \mathbf{0}. \tag{7}$$

Linearization of Equation 7 gives

$$F(l^\circ,\mathbf{x}^\circ) + \left.\frac{\partial F}{\partial l}\right|_{l^\circ,\mathbf{x}^\circ} \cdot \Delta l + \left.\frac{\partial F}{\partial \mathbf{x}}\right|_{l^\circ,\mathbf{x}^\circ} \cdot \boldsymbol{\Delta} = \mathbf{0}$$

which from Equation 5, and using $\mathbf{A}$ for $\frac{\partial F}{\partial l}$ and $\mathbf{B}$ for $\frac{\partial F}{\partial \mathbf{x}}$, becomes

$$F(l^\circ,\mathbf{x}^\circ) + \mathbf{A}(\mathbf{v} + l - l^\circ) + \mathbf{B}\,\boldsymbol{\Delta} = \mathbf{0}$$

or

$$\mathbf{A}\mathbf{v} + \mathbf{B}\,\boldsymbol{\Delta} = -F(l^\circ,\mathbf{x}^\circ) - \mathbf{A}(l - l^\circ) \tag{8}$$

and finally,

$$\mathbf{A}\mathbf{v} + \mathbf{B}\,\boldsymbol{\Delta} = \mathbf{f}^\circ. \tag{8a}$$

Because we may now iterate the solution with updating both l and $\mathbf{x}$, two computational procedures can be effected. The first is simply to update both sets of variables in every cycle. The second is to use *nested looping* by iterating on $\mathbf{x}$ for every given value of l, and repeating the process for other updated values for l until the solution converges. Both methods were tried in the following examples and the first procedure required slightly less computational effort.

EXAMPLE 3

The problem of Example 1 was reworked carrying the coordinates of point P as two parameters, giving the following three condition equations:

$$\begin{aligned} l_1x_2 - cx_1 &= 0 \\ l_2x_2 - c(l_4 - x_1) &= 0 \\ l_3x_2 - c(l_4 + l_5 - x_1) &= 0 \end{aligned}$$

in which c is the principle distance (see Figure 1). The solution was iterated (five iterations) until the last $\boldsymbol{\Delta}$ was negligibly small. The final residual vector was the same as $\mathbf{v}_f$ in Example 1 to one unit in the tenth decimal place.

EXAMPLE 4

In a similar manner, Example 2 was recomputed carrying the parameters a and b in the adjustment. In this instance, *six* condition equations were written:

$$f_{1i} = ax_{1i} - bx_{2i} - y_{1i}$$
$$f_{2i} = bx_{1i} + ax_{2i} - y_{2i}$$

with $i = 1, 2, 3$. The approximations are $a^\circ = 1.0$, $b^\circ = 2.0$ and the observations are those given in Table 3. The linearized equations were solved for corrections to approximations for both parameters and observations. After five iterations, the vector of residuals was computed and found to be identical to $\mathbf{v}_f$ computed in Example 2 (Table 4).

Conclusions

Although most of the presentations on least squares limit consideration only to linear condition equations, the overwhelming majority of practical problems are non-linear. The common practice of solution by iteration on linearized equations has been extended here to include observations in addition to parameters which are customarily considered alone. In doing so, distinction was made between observational values, and approximations for the variables representing observations. The given values for the observations, as well as their covariance (or cofactor) matrix, remain unaltered all through the adjustment.

All derived relationships were programmed and tested with simple, but realistic, examples which demonstrated the correctness of the derivations. It is believed that the account given here extends least-squares adjustment to a level which is applicable to non-linear conditions in a general way. The same concept has also been utilized for a considerably more unified approach yielding quite a complete treatment of least squares. The latter effort is rather lengthy and is therefore considered beyond the scope of this paper.

Acknowledgment

The authors wish to thank Prof. F. Ackermann of the Photogrammetric Institute in Stuttgart, Germany, for his valuable suggestions and contributions to this paper.

Literature*

Madkour, "Non-Linear Least Squares Algorithm," *Journal of the Surveying & Mapping Division*, American Society of Civil Engineers, Vol. 98, SN2, Nov. 1972.

Pope, Allen J., "Some Pitfalls to be Avoided in the Iterative Adjustment of Nonlinear Problems, *Proceedings*, 38th Annual Meeting, American Society of Photogrammetry, Washington, D. C., March 1972.

Articles for Next Month

R. J. Anderle, Geodetic control with doppler.
M. E. Lafferty, Accuracy/costs with analytics.
B. Payandeh et al., Bias from photo intensity.
R. J. Reimold et al., Remote sensing of tidal marsh.
F. L. Scarpace and *P. R. Wold*, Atmospheric refraction.
J. S. Seher and *P. T. Tueller*, Color aerial photos for marshland.
J. D. Turinetti and *O. W. Mintzer*, Computer analysis of imagery.

* Added by Review Board of this Journal inasmuch as this article was received before these references were published.—*Editor.*