Dr. Edward M. Mikhail*
*Purdue University*
*Lafayette, Ind. 47907*

# Parameter Constraints in Least Squares

**For example, points on a lake shore in a photogrammetric model may be constrained to have the same elevation.**

(*Abstract on next page*)

## Introduction

Almost all of those interested in computational photogrammetry are well aware of the increasing use of the method of least squares in estimating variables in overdetermined models. Furthermore, as most of the analytical methods used rely heavily on digital computers, and are best expressed in matrix algebra, it follows that modern treatments of least squares are in matrix notation. Some noted photogrammetrists realized the importance of this fact and authored papers[1,2] introducing the fundamental models of least squares in matrices. Other works[3,4] made valuable contributions to more generalized concepts which are currently used in modern analytical systems. Although lacking any derivations, Madkour[5] includes a good overall coverage of the method of least squares with its varied cases of application.

All of the writings cited above may be considered as general in nature covering a variety of situations. By contrast, this paper is intended to address, rather in detail, one type of adjustment problem. Case[6], to this writer's knowledge, was the first to discuss the subject of constraints encountered in analytical photogrammetry in general. In a recent presentation[7], this author enumerated a multitude of constraints which may be met in extraterrestrial applications. Neither of these two works, which were concerned mostly with photogrammetric aspects, devoted sufficient attention to adjustment aspects. Consequently, it was deemed timely and useful to discuss in this paper the method of least squares in the presence of constraints.

Lest the reader may be unfamiliar with what is meant by constraints, we should first give a few examples. Points appearing on a lake shore in a photogrammetric model may be *constrained* to have the same elevation. Points identified along a straight stretch of highway or railway may be constrained to *lie on a straight line*. In extraterrestrial work, the total base between two camera positions may be constrained to a given value, if such a distance is measured.

In this study we shall consider constraints from two viewpoints. The first encompasses situations where *functional constraint equations* involving parameters only, and which must be rigorously satisfied, augment the basic condition equations arising from the physical or geometric considerations of the problem. The second point of view addresses the occasions where the parame-

Dr. Edward M. Mikhail

* Presented at the 1970 Symposium on Computational Photogrammetry, January 7–9, Alexandria Virginia.

ters involved in the mathematical model are also treated as *observations*. Numerical examples are included to help in the understanding of the various concepts discussed.

Before commencing the detailed treatment, a short general remark regarding least squares might be helpful. It must be emphasized that the method of least squares applies only to linear mathematical models. The fact that we normally use some series expansion, such as Taylor's, to linearize non-linear models should not cause us to consider such linearization as a part of least squares. This is particularly so when an iterative procedure is used to compensate for the neglected higher-order terms. In such situations the basic conditions of the problem may change from one iteration to another and may cause undue misinterpretation of the results. This point is elaborated upon in detail in the latter part of this study.

ABSTRACT: *Modern applications of computational photogrammetry, such as extraterrestrial problems, encounter and make use of a variety of constraints. Least squares is invariably used in estimating the unknown variables in the usually overdetermined photogrammetric models. Procedures of least-squares solutions are involved in the presence of two types of constraints. Three methods of solution can be applied for the first type, namely, functional parameter constraints which must be rigorously satisfied. Constraints are also treated from a second point of view which includes two methods: one where the unknown variables are simultaneously considered as parameters and observations, and the other which regards them simply as observed data. Both methods are equivalent under a condition which is totally practical.*

## FUNCTIONAL CONSTRAINTS OF PARAMETERS ONLY

### DERIVATION OF GENERAL CASE

We shall begin our discussion by treating the most general problem then point out the special applications. Let the linear mathematical model take the general form

$$\boldsymbol{AV} + \boldsymbol{B\Delta} = \boldsymbol{F}^0 \tag{1}$$

where $\boldsymbol{V}$ is the vector of observational residuals, $\boldsymbol{\Delta}$ is the vector of parameters (or corrections thereto if the model is a linearized one), $\boldsymbol{A}$ and $\boldsymbol{B}$ are coefficient matrices, and $\boldsymbol{F}^0$ is a constant term vector.

In addition to the condition equations given in Equation 1, a set of equations may exist that relate some or all of the elements of $\boldsymbol{\Delta}$ as well as, to be general, the elements of another vector $\boldsymbol{\Delta}'$. Such a vector, $\boldsymbol{\Delta}'$, represents a set of *extraneous* parameters that might be necessary for the formulation of the functional constraints. Consequently, let these constraints be of the form

$$\boldsymbol{C}_1\boldsymbol{\Delta} + \boldsymbol{C}_2\boldsymbol{\Delta}' = \boldsymbol{G}^0 \tag{2}$$

in which $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ are coefficient matrices and $\boldsymbol{G}^0$ is a corresponding constant-term vector. Applying least squares to the model of Equation 1 augmented by the constraints of Equation 2, we minimize the scalar

$$\phi'' = \boldsymbol{V}^t\boldsymbol{W}_0\boldsymbol{V} - 2\boldsymbol{K}_0{}^t(\boldsymbol{AV} + \boldsymbol{B\Delta} - \boldsymbol{F}^0) - 2\boldsymbol{K}_c{}^t(\boldsymbol{C}_1\boldsymbol{\Delta} + \boldsymbol{C}_2\boldsymbol{\Delta}' - \boldsymbol{G}^0) \tag{3}$$

where $\boldsymbol{W}_0$ is the weight matrix of the observations, which is normally taken as the inverse of the covariance matrix, and $\boldsymbol{K}_0$ and $\boldsymbol{K}_c$ are two vectors, of appropriate dimensions, of as yet unknown Lagrange multipliers. If $\phi''$ is differentiated with respect to the free variables $\boldsymbol{V}$, $\boldsymbol{\Delta}$ and $\boldsymbol{\Delta}'$, the corresponding partial differentials are equated to zero, and the result is combined with Equations 1 and 2 and rearranged, we get the following set of five vector equations:

$$\begin{aligned} -W_0V + A^tK_0 \qquad\qquad\qquad &= 0 \\ A\ V \qquad + B\,\Delta \qquad\qquad &= F^0 \\ B^tK_0 \qquad + C_1{}^tK_c \qquad &= 0 \\ C_1\Delta \qquad + C_2\Delta' &= G^\circ \\ C_2{}^tK_c \qquad &= 0. \end{aligned} \tag{4}$$

If the first two equations are solved for $V$ and $K_o$, respectively, and the result substituted into the third equation, Equation 4 reduces to:

$$\begin{aligned} -N\,\Delta + C_1{}^tK_c \qquad &= -T \\ C_1\Delta \qquad + C_2\Delta' &= G^0 \\ C_2{}^tK_c \qquad &= 0 \end{aligned} \tag{5}$$

in which

$$\begin{aligned} N &= B^t(AW_0{}^{-1}A^t)^{-1}B \\ T &= B^t(AW_0{}^{-1}A^t)^{-1}F^0 \end{aligned} \tag{6}$$

The first equation of Equation 5 may be solved for $\Delta$:

$$\Delta = \Delta^0 + N^{-1}C_1{}^tK_c \tag{7}$$

where

$$\Delta^0 = N^{-1}T \tag{8}$$

and the result substituted into the remaining two which, when solved simultaneously, yield

$$K_c = (C_1N^{-1}C_1{}^t)^{-1}(G^0 - C_1\Delta^0 - C_2\Delta') \tag{9}$$

$$\Delta' = [C_2{}^t(C_1N^{-1}C_1{}^t)^{-1}C_2]^{-1}[C_2{}^t(C_1N^{-1}C_1{}^t)^{-1}(G^0 - C_1\Delta^0)]. \tag{10}$$

Finally, if Equation 9 is used in Equation 7, the vector $\Delta$ is obtained as:

$$\Delta = \Delta^0 + N^{-1}C_1{}^t(C_1N^{-1}C_1{}^t)^{-1}(G^0 - C_1\Delta^0 - C_2\Delta'). \tag{11}$$

Equations 10 and 11 represent the least-squares solution for the general problem as expressed by the model consisting of Equations 1 and 2 combined. Next we discuss briefly some simpler cases.

### SPECIAL CASES

*a*. In many situations, particularly in analytical photogrammetry, the set of condition equations comprising the mathematical model takes a form simpler than that of Equation 1 where $A$ is an identity matrix, $I$. In this instance, the solution will still be given by Equations 10 and 11 but the matrix $N$ and the vector $T$ would be computed from the following simpler forms than those given in Equation 6:

$$\begin{aligned} N &= B^tW_0B \\ T &= B^tW_0F^0. \end{aligned} \tag{12}$$

*b*. Even though we have so far considered cases with additional parameters, $\boldsymbol{\Delta}'$, situations may arise where no new parameters are added when functional constraints are used. In such cases, $\boldsymbol{\Delta}'$, and correspondingly Equation 10, will be absent and the solution will be given by the following equation which replaces Equation 11:

$$\boldsymbol{\Delta} = \boldsymbol{\Delta}^0 + \boldsymbol{N}^{-1}\boldsymbol{C}_1{}^t(\boldsymbol{C}_1\boldsymbol{N}^{-1}\boldsymbol{C}_1{}^t)^{-1}(\boldsymbol{G}^0 - \boldsymbol{C}_1\boldsymbol{\Delta}^0). \tag{13}$$

Of course $\boldsymbol{N}$ and $\boldsymbol{T}$ may be computed from either Equation 6 or 12.

## A TREATMENT BASED ON DIRECT ELIMINATION

It must be agreed that the solutions given above are relatively complex and require a large number of matrix manipulations. Consequently, wherever these manipulations can be reduced in number, savings in the computational effort may be realized. In this section, we show that such savings are possible particularly where the model and constraints are both originally linear. This will be achieved by altogether avoiding the direct use of the constraint equations in the least-squares solution. Instead, these equations are first solved for as many parameters as the number of constraint equations. This will lead not only to the cancellation of the constraints from the least squares, but also the reduction of the number of unknown parameters to be determined directly from the least-squares solution.

Starting with the general case where additional parameters exist, let the system of Equations 2 be partitioned to

$$\begin{bmatrix} \boldsymbol{C}_{11} \\ \boldsymbol{C}_{21} \end{bmatrix} \boldsymbol{\Delta} + \begin{bmatrix} \boldsymbol{C}_{12} \\ \boldsymbol{C}_{22} \end{bmatrix} \boldsymbol{\Delta}' = \begin{bmatrix} \boldsymbol{G}_1{}^0 \\ \boldsymbol{G}_2{}^0 \end{bmatrix} \tag{14}$$

such that $\boldsymbol{C}_{12}$ is a square matrix. The system of Equation 14 represents, in fact, two matrix equations from which $\boldsymbol{\Delta}'$ can be eliminated. This reduces the constraints to the form

$$\boldsymbol{D}\boldsymbol{\Delta} = \boldsymbol{H}^0 \tag{15}$$

where

$$\begin{aligned} \boldsymbol{D} &= \boldsymbol{C}_{21} - \boldsymbol{C}_{22}\boldsymbol{C}_{12}{}^{-1}\boldsymbol{C}_{11} \\ \boldsymbol{H}^0 &= \boldsymbol{G}_2{}^0 - \boldsymbol{C}_{22}\boldsymbol{C}_{12}{}^{-1}\boldsymbol{G}_1{}^0. \end{aligned} \tag{16}$$

The matrix $\boldsymbol{C}_{12}$ will always be non-singular as long as the constraint equations are independent, which should obviously be the case. At this point, Equation 15 represents a case where the functional constraints involve the parameters of the mathematical model only. Consequently, the following treatment applies equally to that case as well as to the one with the vector of added parameters, $\boldsymbol{\Delta}'$, after it has been eliminated.

We now proceed to eliminate the constraint Equations 15 and as many parameters, $\boldsymbol{\Delta}_1$, from the total parameter vector $\boldsymbol{\Delta}$. This is accomplished by partitioning $\boldsymbol{\Delta}$ such that Equations 1 and 15 take the forms

$$\begin{aligned} \boldsymbol{A}\boldsymbol{V} + \boldsymbol{B}_1\boldsymbol{\Delta}_1 + \boldsymbol{B}_2\boldsymbol{\Delta}_2 &= \boldsymbol{F}^0 \\ \boldsymbol{D}_1\boldsymbol{\Delta}_1 + \boldsymbol{D}_2\boldsymbol{\Delta}_2 &= \boldsymbol{H}^0 \end{aligned} \tag{17}$$

where $\boldsymbol{D}_1$ is a square matrix, and hence $\boldsymbol{\Delta}_1$ is of the same order. Now solving the second of Equation 17 for $\boldsymbol{\Delta}_1$ we get

$$\boldsymbol{\Delta}_1 = \boldsymbol{D}_1{}^{-1}(\boldsymbol{H}^0 - \boldsymbol{D}_2\boldsymbol{\Delta}_2) \tag{18}$$

which, when substituted into the first equation, yields

$$AV + \bar{B}\Delta_2 = \bar{F}^0 \tag{19}$$

where

$$\begin{aligned} \bar{B} &= B_2 - B_1 D_1^{-1} D_2 \\ \bar{F}^0 &= F^0 - B_1 D_1^{-1} H^0. \end{aligned} \tag{20}$$

Again $D_1^{-1}$ will always exist unless the mathematical model inadequately describes the problem and includes dependent equations.

Equation 19 can now be solved by least squares in the usual manner:

$$\Delta_2 = \bar{N}^{-1}\bar{T} \tag{21}$$

where

$$\begin{aligned} \bar{N} &= \bar{B}^t (A W_0^{-1} A^t)^{-1} \bar{B} \\ \bar{T} &= \bar{B}^t (A W_0^{-1} A^t)^{-1} \bar{F}^0. \end{aligned} \tag{22}$$

With the value of $\Delta_2$ computed from Equation 21, $\Delta_1$ can now be computed from Equation 18 and the final answer vector will simply be

$$\Delta = \begin{bmatrix} \Delta_1 \\ \Delta_2 \end{bmatrix}. \tag{23}$$

The application of this treatment to the case where $A$ is the identity matrix is sufficiently evident that it needs no further comment.

Before leaving this section we should clarify a point that may have crossed the reader's mind. The treatment given above appears sufficiently straightforward that one might wonder why then do we bother with the apparently more complex general case given earlier. The answer to this is to say simply that it all depends on the problem at hand. Certainly numerous situations occur where it would be illogical not to use the method of elimination explained in this section. However, there may be other instances where this approach may become impractical. For such a case the reader is referred to this writer's paper (Mikhail[7]). There, the direct elimination of the added parameters gave rise to such complicated condition equations that it was deemed more realistic to solve the problem directly using equations comparable to Equations 10 and 11.

## A PROCEDURE TREATING CONSTRAINTS AS OBSERVATION EQUATIONS WITH RELATIVELY LARGE WEIGHTS

Schmid and Schmid[8] presented a most general procedure for least squares in which all elements of the mathematical model may be considered as observations, placing the burden of classification on the weight matrix. This concept can be applied here by considering the constraint equations with an observational residual vector $V_c$ and an associated weight matrix $W_c$. This changes them from the form of Equation 2 to the following form:

$$V_c + C_1\Delta + C_2\Delta' = G^0. \tag{24}$$

Equation 24 can be combined with Equation 1 and gives

$$\tilde{A}\tilde{V} + \tilde{B}\tilde{\Delta} = \tilde{F}^0 \tag{25}$$

where

$$\tilde{A} = \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix}, \quad \tilde{V} = \begin{bmatrix} V \\ V_c \end{bmatrix}$$
$$\tilde{B} = \begin{bmatrix} B & 0 \\ C_1 & C_2 \end{bmatrix}, \quad \tilde{\Delta} = \begin{bmatrix} \Delta \\ \Delta' \end{bmatrix}, \quad \tilde{F}^0 = \begin{bmatrix} F^0 \\ G^0 \end{bmatrix}. \tag{26}$$

The solution of Equation 25 can be readily given by

$$\tilde{\Delta} = [\tilde{B}^t(\tilde{A}\tilde{W}^{-1}\tilde{A}^t)^{-1}\tilde{B}]^{-1}[\tilde{B}^t(\tilde{A}\tilde{W}^{-1}\tilde{A}^t)^{-1}\tilde{F}^0] \tag{27}$$

in which

$$\tilde{W} = \begin{bmatrix} W_0 & 0 \\ 0 & W_c \end{bmatrix}. \tag{28}$$

It is clear from Equation 28 that no correlation is assumed between the two systems of equations. The magnitudes of the elements of $W_c$ are considerably larger than those of $W_0$, which leads to very small values for $V_c$ thus bringing Equation 24 close to its true form of Equation 2.

Before we move on to the next section, we shall first give a numerical example to illustrate some of the concepts covered so far.

*Example 1*

Suppose that the mathematical model of a certain problem is of the form

$$V + B\Delta = F^0$$

where

$$B = \begin{bmatrix} 2 & -3 \\ -1 & 2 \\ 0 & 1 \end{bmatrix}, \quad \Delta = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad F^0 = \begin{bmatrix} -1.1 \\ 1.2 \\ 1.0 \end{bmatrix}.$$

We shall assume here that the weight matrix of the observations $W_0$ is the identity matrix $I$. Next, we consider the set of constraints

$$C_1\Delta + C_2\Delta' = G^0$$

where

$$C_1 = \begin{bmatrix} 1 & -1 \\ 2 & -1 \end{bmatrix}, \quad C_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \quad \Delta' = x_3, \quad G^0 = \begin{bmatrix} -1 \\ 3 \end{bmatrix},$$

that is to say, two equations will be added, with one extra parameter $x_3$.

*Method a*: We apply the procedure derived earlier for the General Case and for the Special Case. From Equation 6,

$$N = \begin{bmatrix} 5 & -8 \\ -8 & 14 \end{bmatrix}, \quad T = \begin{bmatrix} -3.4 \\ 6.7 \end{bmatrix},$$

and

$$N^{-1} = \tfrac{1}{6}\begin{bmatrix} 14 & 8 \\ 8 & 5 \end{bmatrix}.$$

Hence, the solution vector $\Delta^0$ in the absence of constraints is

$$\Delta^0 = N^{-1}T = \begin{bmatrix} 1.00 \\ 1.05 \end{bmatrix}. \tag{1.1}$$

Now, applying Equation 10 we compute $\boldsymbol{\Delta}'$ as

$$\Delta' = x_3 = -0.97922. \tag{1.2}$$

Next, using Equation 11 and the values of $\Delta^0$ and $\Delta'$ thus computed we get

$$\Delta = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1.00 \\ 1.05 \end{bmatrix} + \begin{bmatrix} 0.06233 \\ 0.03311 \end{bmatrix} = \begin{bmatrix} 1.06233 \\ 1.08311 \end{bmatrix}. \tag{1.3}$$

This leads to the total answer vector as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1.06233 \\ 1.08311 \\ -0.97922 \end{bmatrix}. \tag{1.4}$$

*Method b*: Based on Direct Elimination. The matrices $\boldsymbol{C}_1$, $\boldsymbol{C}_2$ and $\boldsymbol{G}^0$ will first be partitioned as

$$\boldsymbol{C}_{11} = [1 \quad -1] \qquad \boldsymbol{C}_{12} = 1 \qquad \boldsymbol{G}_1{}^0 = -1$$
$$\boldsymbol{C}_{21} = [2 \quad -1] \qquad \boldsymbol{C}_{22} = -2 \qquad \boldsymbol{G}_2{}^0 = 3.$$

Then from Equation 16, we compute $\boldsymbol{D} = [4 \quad -3]$ which may be partitioned to $\boldsymbol{D}_1 = 4$ $\boldsymbol{D}_2 = -3$, and $\boldsymbol{H}^0 = 1$. The $\boldsymbol{B}$ matrix can be partitioned accordingly:

$$\boldsymbol{B}_1 = [2 \quad -1 \quad 0]^t \quad \text{and} \quad \boldsymbol{B}_2 = [-3 \quad 2 \quad 1]^t.$$

Using Equation 20 we get

$$\bar{\boldsymbol{B}} = [-1.5 \quad 1.25 \quad 1.0]^t \quad \text{and} \quad \bar{\boldsymbol{F}}^0 = [-1.6 \quad 1.45 \quad 1.0]^t$$

which lead to

$$\boldsymbol{\Delta}_2 = x_2 = (\bar{\boldsymbol{B}}^t\bar{\boldsymbol{B}})^{-1}(\bar{\boldsymbol{B}}\bar{\boldsymbol{F}}^0)$$

or $x_2 = 1.08311$. Equation 18 can now be used to compute $\Delta_1 = x_1$, hence $x_1 = 1.06233$. Finally, if desired, $\Delta' = x_3$ may be computed from

$$\Delta' = \boldsymbol{C}_{12}{}^{-1}(\boldsymbol{G}_1{}^0 - \boldsymbol{C}_{11}\boldsymbol{\Delta}) \tag{1.5}$$

or $x_3 = -0.97922$. The total vector will then be

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1.06233 \\ 1.08311 \\ -0.97922 \end{bmatrix}$$

which is identical to that given by 1.4.

*Method c*: Based on Observation Equations. We start by choosing $\tilde{\boldsymbol{W}} = \boldsymbol{I}$; the $\boldsymbol{B}$ and $\boldsymbol{F}^0$ matrices are

$$\boldsymbol{B} = \left[\begin{array}{rrr|rr} 2 & -1 & 0 & 1 & 2 \\ -3 & 2 & 1 & -1 & -1 \\ \hline 0 & 0 & 0 & 1 & -2 \end{array}\right]^t, \qquad \boldsymbol{F}^0 = [-1.1 \quad 1.2 \quad 1.0 \;\vdots\; -1 \quad 3]^t.$$

The solution vector is obtained from

$$\tilde{\boldsymbol{\Delta}} = (\tilde{\boldsymbol{B}}^t\tilde{\boldsymbol{W}}\tilde{\boldsymbol{B}})^{-1}(\tilde{\boldsymbol{B}}^t\tilde{\boldsymbol{W}}\tilde{\boldsymbol{F}}^0)$$

or

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1.04486 \\ 1.07383 \\ -0.98785 \end{bmatrix}. \tag{1.7}$$

It is obvious that this vector is not the same as the answer given before. This is because by giving the constraints the same weight as the condition equations, they were not strictly enforced. Consequently, we next construct the $\tilde{\boldsymbol{W}}$ matrix as

$$\tilde{\boldsymbol{W}} = \begin{bmatrix} W_0 & 0 \\ 0 & \boldsymbol{W}_c \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & 0 \\ 0 & 10\boldsymbol{I} \end{bmatrix}.$$

This leads to the answer vector

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1.06000 \\ 1.08188 \\ -0.98038 \end{bmatrix} \tag{1.8}$$

which is closer to the answer than that of 1.7. Therefore, we now increase the weight of the constraints to $\boldsymbol{W}_c = 100\boldsymbol{I}$ and get

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1.06210 \\ 1.08299 \\ -0.97934 \end{bmatrix}. \tag{1.9}$$

This answer is essentially the same as the exact one given in 1.6. If a closer vector than that of 1.9 is desired the value of $\boldsymbol{W}_c$ relative to $\boldsymbol{W}_0$ can be increased to, say, 1000 $\boldsymbol{WI}$.

## "Parameter/Observation" Consideration

One of the most powerful tools used in modern analytical photogrammetric systems has been the ability of introducing the unknown parameters also as *observations* with known *a priori* covariance matrices. Brown[4] should probably be credited as being the first to introduce this technique, although it is also embodied within the paper by Schmid and Schmid[3]. This technique allows for the writing, in addition to the original mathematical model as in Equation 1, of a set of *simple linear* equations reflecting the fact that the parameters can also and simultaneously be considered as *observations*. Consequently, if a parameter $x$ is denoted by $x^0$ as an observation with a residual $v_p$, and its approximate value by $x^{00}$ with a correction $\delta$, then obviously

$$x = x^0 + v_p = x^{00} + \delta \tag{29}$$

which leads to

$$v_p - \delta = c^0 \tag{30}$$

where

$$c^0 = \boldsymbol{x}^{00} - \boldsymbol{x}^0. \tag{31}$$

Equation 30 may be written for all the parameters, thus

$$\boldsymbol{V}_p - \boldsymbol{\Delta} = \boldsymbol{C}^0. \tag{32}$$

Equation 32, combined with equation 1, yields

$$\begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} V \\ V_p \end{bmatrix} + \begin{bmatrix} B \\ -I \end{bmatrix} \Delta = \begin{bmatrix} F^0 \\ C^0 \end{bmatrix} \tag{33}$$

which may be considered as being of the general form of Equation 1. Least squares can then be applied directly to Equation 33 and, after a few straightforward manipulations, gives

$$[B^t(AW_0^{-1}A^t)^{-1}B + W_p]\Delta = [B^t(AW_0^{-1}A^t)^{-1}F^0 - W_pC^0] \tag{34}$$

where, the only term so far undefined, $W_p$ is the *a priori* weight matrix of the parameters which is usually the inverse of the *a priori* covariance matrix. In view of Equation 6, Equation 34 may be written as

$$[N + W_p]\Delta = [T - W_pC^0] \tag{35}$$

whose solution is

$$\Delta = [N + W_p]^{-1}[T - W_pC^0]. \tag{36}$$

This writer has pondered over the reasons for the dual role of $\Delta$ as parameters and as observations, and also over the consequent necessity of writing the additional set of Equations 32. This led to the speculation that perhaps only one set of equations will be sufficient. Therefore, we will attempt next to consider the parameters strictly as observations* with the *a priori* weight matrix $W_p$, and write the mathematical model as

$$[A \quad B] \begin{bmatrix} V \\ V_\Delta \end{bmatrix} = F^0 \tag{37}$$

where $V_\Delta$ simply replaces the vector $\Delta$. Equation 37 is in the form of the so-called "Adjustment by Conditions" model whose least-squares solution is given by Madkour[5] and Mikhail[8].

$$\begin{bmatrix} V \\ V_\Delta \end{bmatrix} = \begin{bmatrix} W_0^{-1} & 0 \\ 0 & W_p^{-1} \end{bmatrix} \begin{bmatrix} A^t \\ B^t \end{bmatrix} \left[ [A \quad B] \begin{bmatrix} W_0^{-1} & 0 \\ 0 & W_p^{-1} \end{bmatrix} \begin{bmatrix} A^t \\ B^t \end{bmatrix} \right]^{-1} F^0. \tag{38}$$

We will now show whether, and under what conditions, $V_\Delta$ is equal to $\Delta$. From Equation 38 we get for $V_\Delta$

$$V_\Delta = W_p^{-1}B^t(AW_0^{-1}A^t + BW_p^{-1}B^t)^{-1}F^0. \tag{39}$$

It can be shown from matrix algebra that

$$(AW_0^{-1}A^t + BW_p^{-1}B^t)^{-1} = (AW_0^{-1}A^t)^{-1}\{I - B[B^t(AW_0^{-1}A^t)^{-1}B + W_p]^{-1}B^t(AW_0^{-1}A^t)^{-1}\}. \tag{40}$$

Substituting Equation 40 into 39 and making use of Equation 6 leads to

$$V_\Delta = W_p^{-1}T - W_p^{-1}N(N + W_p)^{-1}T. \tag{41}$$

From Equation 35 we obtain

$$W_p^{-1}T = C^0 + W_p^{-1}(N + W_p)\Delta$$

and

$$(N + W_p)^{-1}T = \Delta + (N + W_p)^{-1}W_pC^0$$

* A treatment similar to the one given here is included in reference (10) in which a portion of the parameter vector was considered as observed data.

which when substituted into Equation 41 yields

$$V_\Delta = \Delta + [C^0 - W_p^{-1}N(N + W_p)^{-1}W_pC^0]. \tag{42}$$

Equation 42 shows that in order for $V_\Delta$ to be equal to $\Delta$, the vector $C^0$ must be equal to zero. This means that the approximate values must be chosen equal to the *observed* values of the parameters, which is usually the case. To recapitulate, then, when *a priori* estimates of the unknown parameters in the model (as well as the corresponding *a priori* covariance matrix) are known, two ways are possible for a least-squares solution: one in which they are considered as both parameters and observations, and the other regards them simply as observed data. The fundamental condition for this possibility is the vanishing of the vector $C^0$. Whenever the value of this vector becomes different from zero, as for example in the case of iterative solutions due to non-linearity of the model, this possibility does not hold which will be amply demonstrated in the following examples.

*Example 2*

Let us take a very simple case of a one-loop level net with point $A$ as the reference point with an elevation of 5.000 m (Figure 1).

The observed differences in elevation are given in Table 1 and their weight matrix

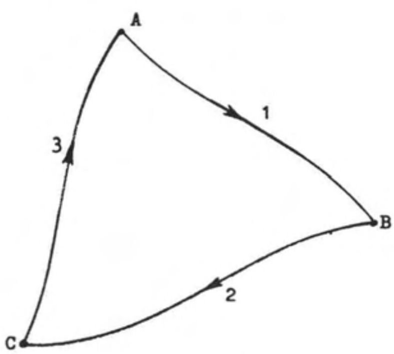


FIGURE 1

TABLE 1

| *Line* | *Difference in Elevation (meters)* |
|---|---|
| 1 | −0.793 |
| 2 | −2.310 |
| 3 | +3.106 |

is taken as $W_0 = I$. Let the *a priori* estimates of the elevations of points $B$ and $C$ be 4.205 m and 1.893 m, respectively, and their weight matrix $W_p = 0.01\ I$. It is required to compute the least squares estimates of these two elevations based on the given data.

*Method (i)*. It is easily seen that the mathematical model relating the observations and parameters is composed of linear equations. Despite this fact, one may linearize the model using the given estimates as approximate values for the parameters (hence $C^0 = 0$) or $z_b^{00} = z_b^0 = 4.204\ m$ and $z_c^{00} = z_c^0 = 1.893\ m$. Hence, the *linearized* model will be

$$V + \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \delta_b \\ \delta_c \end{bmatrix} = \begin{bmatrix} 4.205 + 0.793 - 5.000 \\ 2.310 + 1.893 - 4.205 \\ 5.000 - 3.106 - 1.893 \end{bmatrix} = \begin{bmatrix} -0.002 \\ -0.002 \\ +0.001 \end{bmatrix}. \tag{2.1}$$

Noting that $A = W_0 = I$, Equation 34 applied to the data gives

$$\begin{bmatrix} 2.01 & -1 \\ -1 & 2.01 \end{bmatrix} \begin{bmatrix} \delta_b \\ \delta_c \end{bmatrix} = \begin{bmatrix} 0 \\ +0.003 \end{bmatrix}$$

or

$$\begin{bmatrix} \delta_b \\ \delta_c \end{bmatrix} = \frac{1}{3.0401} \begin{bmatrix} 2.01 & 1 \\ 1 & 2.01 \end{bmatrix} \begin{bmatrix} 0 \\ 0.003 \end{bmatrix} \tag{2.2}$$

or

$$\begin{aligned} \delta_b &= \frac{0.003}{3.0401} \simeq 0.001\ m \\ \delta_c &= \frac{0.00603}{3.0401} \simeq 0.002\ m. \end{aligned} \tag{2.3}$$

These lead to the final elevations

$$b = 4.206\ m \qquad c = 1.895\ m. \tag{2.4}$$

(Note that the corrections in (2.3) are sufficiently small that no further iteration is necessary.)

*Method (ii)*. In this method we shall simply consider the given elevations of points $B$ and $C$ as *observations* whose residuals $v_b$, $v_c$ we now seek. In this instance the mathematical model corresponding to (2.1) will be

$$\begin{bmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_b \\ v_c \end{bmatrix} = \begin{bmatrix} -.002 \\ -.002 \\ +.001 \end{bmatrix} \text{ with } \boldsymbol{W} = \begin{bmatrix} \boldsymbol{I} & 0 \\ 0 & .01\boldsymbol{I} \end{bmatrix} \tag{2.5}$$

which is in the same form as Equation 37. Using Equation 30 we get

$$\boldsymbol{V}_\Delta = 100\boldsymbol{I} \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \frac{1}{30401} \begin{bmatrix} 10301 & 10100 & 10000 \\ 10100 & 10201 & 10100 \\ 10000 & 10100 & 10301 \end{bmatrix} \begin{bmatrix} -.002 \\ -.002 \\ +.001 \end{bmatrix} \tag{2.6}$$

or

$$\begin{aligned} v_b &= \frac{0.3}{304.01} \simeq 0.001\ m \\ v_c &= \frac{0.603}{304.01} \simeq 0.002\ m. \end{aligned} \tag{2.7}$$

These two values are *exactly* the same as those given by (2.3) and will lead to the same elevations shown in (2.4). If one computes the *observational* residuals, he gets all three values approximately equal to $-0.001$. As the corrections in (2.7) are of the same order of magnitude as the residuals, there is no need for any iterations. This situation will be contrasted in the following example where the solution must be repeated several times in an iterative manner.

*Example 3*

Suppose that it is required to fit a straight line with the equation $y - mx = 0$ through the two points (1.1, 2.1) and (2.1, 4.0) whose covariance matrices are the identity matrix. The *a priori* estimate for $m$ is $m^0 = 1.0$ with a variance $\sigma_m{}^2 = 100$.

*Method (i)*. Denoting the observed coordinates by $o$, and the approximate value for $m$ by $m^{00}$, the linearized form of the equation for one point is $\boldsymbol{AV}+\boldsymbol{B}\Delta=\boldsymbol{F}^0$ in which:

$$\boldsymbol{A} = [-m^{00} \quad 1], \qquad \boldsymbol{B} = -x^0, \qquad \boldsymbol{F}^0 = (m^{00}x^0 - y^0).$$

Using the data given we get

$$\boldsymbol{A} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \qquad \boldsymbol{B} = \begin{bmatrix} -1.1 \\ -2.1 \end{bmatrix}, \qquad \boldsymbol{F}^0 = \begin{bmatrix} -1.0 \\ -1.9 \end{bmatrix},$$

$w_0=\boldsymbol{I}$, $\boldsymbol{W}_p=0.01$ and we choose $\boldsymbol{C}^0=0$ (i.e., $m^{00}=m^0$) Hence, using Equation 34 we get

$$\begin{aligned} 2.82\ \Delta_1 &= 2.545 \\ \Delta_1 &= 0.902482. \end{aligned} \tag{3.1}$$

Obviously the value of $\Delta_1$ is of the same order of magnitude as $m^0$ and therefore we must iterate the solution to overcome the effects of the neglected higher-order terms. In this case $m^{00}$ will become 1.902482, $(m^0+\Delta_1)$, and $c^0$ will no longer be zero but will take the value 0.902482. Hence,

$$(1.216598 + 0.01)\ \Delta_2 = -\ 0.005118 \tag{3.2}$$

$$\Delta_2 = -\ 0.004172. \tag{3.3}$$

If we iterate a third time we will get

$$\Delta_3 = 0.000025 \tag{3.4}$$

which will lead to the final value of $m$ as

$$m = 1.898335 \tag{3.5}$$

assuming that $\Delta_3$ is sufficiently small to terminate the iterative process.

*Method (ii)*. The linearized form is in this case $\boldsymbol{AV}=\boldsymbol{F}^0$ where

$$\boldsymbol{A} = \begin{bmatrix} -m^0 & 1 & 0 & 0 & -x_1 \\ 0 & 0 & -m^0 & 1 & -x_2 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 & 0 & -1.1 \\ 0 & 0 & -1 & 1 & -2.1 \end{bmatrix}$$

$$\boldsymbol{F}^0 = \begin{bmatrix} -1.0 \\ -1.9 \end{bmatrix}, \qquad \boldsymbol{W}_0 = \boldsymbol{I} \quad \text{and} \quad \boldsymbol{W}_p = 0.01.$$

Applying the equivalent of Equation 39 we get

$$V_{m_1} = V_{\Delta_1} = 0.902482 \tag{3.6}$$

which is exactly identical to the value of $\Delta_1$ given in (3.1) where $\boldsymbol{C}^0$ was equal to zero at that cycle. Now, if we add $V_{m_1}$ to $m^0$ and recompute another value we get

$$V_{m_2} = 0.003187 \tag{3.7}$$

which is obviously quite different from the value of $\Delta_2$ given in (3.3). The reason for this should be clear by now; the value of $\boldsymbol{C}^0$ at that iteration was 0.902482 which is different from zero. As a form of a check let us compute the second term on the right-hand side of Equation 42. Referring to (3.2) and realizing that it is equivalent to (35), we have

$$\begin{aligned} \boldsymbol{C}^0 - \boldsymbol{W}_p^{-1}\boldsymbol{N}(\boldsymbol{N} + \boldsymbol{W}_p)^{-1}\boldsymbol{W}_p\boldsymbol{C}^0 &= [\boldsymbol{I} - \boldsymbol{W}_p^{-1}\boldsymbol{N}(\boldsymbol{N} + \boldsymbol{W}_p)^{-1}\boldsymbol{W}_p]C^0 \\ &= \left[1 - 100\frac{1.216598}{1.226598} \times 0.01\right] 0.902482 \\ &= 0.007358. \end{aligned}$$

TABLE 2. SUMMARY

| *Method* | *a priori estimate* | *Correction No.* 1 | 2 | 3 | *Final value of m* | $V^tWV$ *(weighted sum of squares)* |
|---|---|---|---|---|---|---|
| (i) | 1.0 | .902482 | −.004172 | .000025 | 1.898335 | 0.008140 |
| (ii) | 1.0 | .902482 | .003187 | .000025 | 1.905694 | 0.008207 |
| (iii) | 1.0 | .902482 | −.001562 | .000065 | 1.900985 | 0.008156 |

If this value is now added to $\Delta_2$ we get 0.003186 which is equal to $V_{m_2}$ to a one in the sixth decimal place. This demonstrates the correctness of Equation 42.

If a third iteration is applied we get

$$V_{m_3} = 0.000025, \tag{3.8}$$

and the final value is then

$$m = 1.905694. \tag{3.9}$$

If the idea occurs to the reader, a third method (*iii*) was attempted in which the total residual vector was computed and all the *observations* were updated each iteration. The results of this attempt are included in the summary Table 2.

If the final value of $m$ is used to plot the straight line with respect to the given two points we will get Figure 2. This figure may lead to a misinterpretation of the results; nevertheless it is included to point out certain subtle factors which are worthy of discussion. The first glance at the figure may lead one to believe that method (*ii*) gives the *best*, or most logical, solution followed by method (*iii*) and then by method (*i*). This is because our experience tells us that a well-fitted straight line should pass between the two points. Unfortunately, this will be, in this case, an erroneous interpretation of both the problem and the results. For a given problem there is only one rigorous least-squares solution based on the minimization of the quadratic form ($V^tWV$). In this instance it is that given by method (*i*) as evidenced by the smallest value of the sum of squares in the table. The other two methods, therefore, must be only approximate. This can be ascertained by the fact that, after the first iteration, the conditions of the problem changed by the addition of the residuals to the *observations*—be it the one representing the parameter (Method *ii*) or all of them (Method *iii*). Consequently, we are in effect solving a different problem each iteration. This is not the case for the first method as the values of the observations remain unaltered all through the solution. This discussion also applies to the weight matrix of the observations.

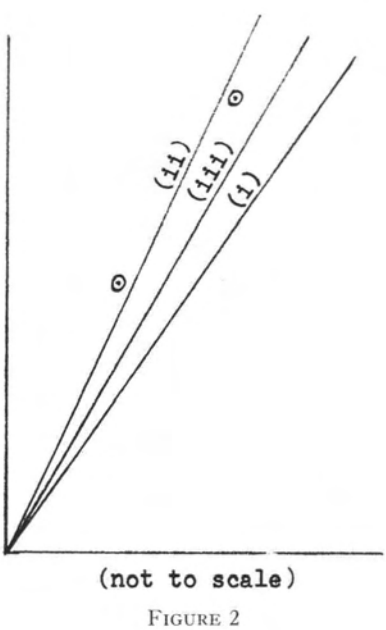


FIGURE 2

In closing this discussion we must emphasize that all three methods are *identical* at the end of the first iteration as was shown by the corrections computed. This indicates that the process of iteration, which is necessary because of linearization, is not really a part of least squares which is based completely on a *linear* model.

## A Note on Error Propagation

It is often of interest to compute the *a posteriori* estimate of the cofactor (or relative covariance) matrices of the estimated parameters. This can, in all cases, be obtained by applying the law of propagation of covariance[8] to the relationships from which the parameters are computed. Consequently, it can be shown, for example, that from Equation 10, the cofactor matrix $\boldsymbol{Q}$ of the added parameters is

$$\boldsymbol{Q}_{\Delta'} = [\boldsymbol{C}_2{}^t(\boldsymbol{C}_1\boldsymbol{N}^{-1}\boldsymbol{C}_1{}^t)^{-1}\boldsymbol{C}_2]^{-1} \tag{43}$$

from Equation 11, the cofactor matrix of the original parameters is

$$\boldsymbol{Q}_{\Delta} = \boldsymbol{N}^{-1}[\boldsymbol{I} - \boldsymbol{C}_1{}^t\boldsymbol{M}^{-1}\boldsymbol{C}_1\boldsymbol{N}^{-1} + \boldsymbol{C}_1{}^t\boldsymbol{M}^{-1}\boldsymbol{C}_2\boldsymbol{R}^{-1}\boldsymbol{C}_2{}^t\boldsymbol{M}^{-1}\boldsymbol{C}_1\boldsymbol{N}^{-1}] \tag{44}$$

where

$$\begin{aligned} \boldsymbol{M} &= \boldsymbol{C}_1\boldsymbol{N}^{-1}\boldsymbol{C}_1{}^t \\ \boldsymbol{R} &= \boldsymbol{C}_2{}^t\boldsymbol{M}^{-1}\boldsymbol{C}_2. \end{aligned} \tag{45}$$

The remaining relationships can be treated similarly and appropriate expressions for the cofactor matrices accordingly derived.

Perhaps one point which should be discussed is whether the propagated cofactors of the parameters will be equal if computed by each of the two methods given previously in the section on "Parameter/Observation." Referring to Example 2 the cofactor matrix of the parameters from Equation 2.2 (Method *ii*) is obviously

$$\boldsymbol{Q}_p = \frac{1}{3.0401}\begin{bmatrix} 2.01 & 1 \\ 1 & 2.01 \end{bmatrix}.$$

To compute the corresponding matrix from Method *ii*, it can easily be shown[5,8] that for a model of the form $\boldsymbol{AV} = \boldsymbol{F}^0$, we have

$$\boldsymbol{Q}_V = \boldsymbol{W}^{-1}\boldsymbol{A}^t(\boldsymbol{A}\boldsymbol{W}^{-1}\boldsymbol{A}^t)^{-1}\boldsymbol{A}\boldsymbol{W}^{-1}. \tag{46}$$

Equation 46 may now be applied to the data of Problem 2 and we get

$$\boldsymbol{Q}_V = \frac{1}{3.0401}\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} 10301 & 10100 & 10000 \\ 10100 & 10201 & 10100 \\ 10000 & 10100 & 10301 \end{bmatrix}\begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}$$

$$\boldsymbol{Q}_V = \frac{1}{3.0401}\begin{bmatrix} 302 & -1 \\ -1 & 302 \end{bmatrix}.$$

It can be shown further[8] that the *a posteriori* cofactor matrix is equal to the *a priori* matrix minus $\boldsymbol{Q}_V$, or

$$\boldsymbol{Q}_P = 100\boldsymbol{I} - \frac{1}{3.0401}\begin{bmatrix} 302 & -1 \\ -1 & 302 \end{bmatrix} = \frac{1}{3.0401}\begin{bmatrix} 2.01 & 1 \\ 1 & 2.01 \end{bmatrix}$$

which agrees exactly with the one computed above.

We may also consider the same thing with regard to Problem 3. From 3.1 it is obvious that

$$\boldsymbol{Q}_{\Delta_1} = \frac{1}{2.82} = 0.354610.$$

As for the second (or third) method,

$$Q_V = 100[-1.1 \quad -2.1] \frac{1}{1128} \begin{bmatrix} 443 & -231 \\ -231 & 123 \end{bmatrix} \begin{bmatrix} -1.1 \\ -2.1 \end{bmatrix} 100$$

$$= \frac{11.24}{0.1128}$$

and

$$Q_{\Delta_1} = 100 - \frac{11.24}{0.1128} = \frac{0.04}{0.1128} = 0.354610$$

which is the exact value obtained above. We must note, however, that in this instance of iterative solution the above comparison will hold for the values from the first iteration only.

## Conclusion

An extensive treatment of the use of constraints in least squares is presented. A balance has been attempted between both excessive abstraction and brevity, and too much detail. Furthermore, despite their simplicity, the worked examples are included to demonstrate the underlying concepts. It is hoped, therefore, that this paper will be of value to students and practicing photogrammetrists alike, particularly in tackling problems with terse explicit control where constraints play a useful role.

## Acknowledgments

The author wishes to thank Dr. Atef Elassal for many valuable remarks made during informal discussions on the subject and for bringing Reference 10 to his attention. Thanks are also due Mr. Raymond J. Helmering for reading the draft.

## References

1. Brown, D., "A Matrix Treatment of the General Problem of Least Squares Considering Correlated Observations", *Report No. 937*, Ballistic Research Laboratories, Aberdeen, Md., 1955.
2. Thompson, E. H., "The Theory of the Method of Least Squares", *Photogrammetric Record*, Vol. IV, No. 19, April 1962.
3. Schmid, H. H. and Schmid, E., "A Generalized Least Squares Solution for Hybrid Measuring Systems", *The Canadian Surveyor*, Vol. IX, No 1, March 1965.
4. Brown, D. C., et al, "Research in Mathematical Targeting, The Practical and Rigorous Adjustment of Large Photogrammetric Nets", Report by D. Brown Associates, Inc.
5. Madkour, M. F., "Precision of Adjusted Variables by Least Squares", *Journal of the Surveying and Mapping Division*, ASCE, September 1968, pp. 119–136.
6. Case, J. B., "The Utilization of Constraints in Analytical Photogrammetry", *Photogrammetric Engineering*, Vol. XXVII, No. 5, December 1961.
7. Mikhail, E. M., "Use of Relative Control Information in Analytical Extraterrestrial Photogrammetry", *Photogrammetric Engineering*, Vol. XXXVI, No. 4, April 1970, pp. 381–389.
8. Mikhail, E. M., "Advanced Geometronics Data Adjustment", printed class notes (unpublished).
9. Bodewig, E., *Matrix Calculus*, Interscience Publishers, Inc., New York, 1959, pp. 217–218.
10. Anon., "Relative Mapping Triangulation (RELMAP) Program", Vol. II. *Mathematical Analysis*, Appendix, Report No. 69–3295 by Autometric/Raytheon for U. S. Army Engineer Topographic Laboratory, June 1969.