BRL

REPORT No. 937

A Matrix Treatment Of The General Problem Of Least Squares Considering Correlated Observations

DUANE BROWN

DEPARTMENT OF THE ARMY PROJECT No. 5B0306011
ORDNANCE RESEARCH AND DEVELOPMENT PROJECT No. TB3-0538

BALLISTIC RESEARCH LABORATORIES

ABERDEEN PROVING GROUND, MARYLAND

BALLISTIC RESEARCH LABORATORIES

REPORT NO. 937

MAY 1955

A MATRIX TREATMENT OF THE GENERAL PROBLEM OF LEAST SUARES CONSIDERING CORRELATED OBSERVATIONS

Duane Brown

DEPARTMENT OF THE ARMY PROJECT NO. 5180306011

ORDNANCE RESEARCH AND DEVELOPMENT PROJECT NO. TB3-0538

ABERDEEN PROVING GROUND, MARYLAND

BALLISTIC RESEARCH LABORATORIES

REPORT NO. 937

DBrown/bd
Aberdeen Proving Ground, Md.
May 1955

A MATRIX TREATMENT OF THE GENERAL PROBLEM OF LEAST SQUARES CONSIDERING CORRELATED OBSERVATIONS

ABSTRACT

The most general type problem considered in least squares is formulated and solved with the aid of matrix algebra for the case in which the observations have the general multivariate normal distribution. The criterion for adjustment is the principle of maximum likelihood. Such related topics as the inversion of the normal equations, variance-covariance propagation, direct adjustment of functions of observations, statistical tests of significance, and the geometrical interpretation of the adjustment are considered. It is pointed out that the results of the conventional method of least squares are special cases of the present theory.

1. INTRODUCTION

In recent years several writers have employed matrix analysis to derive special results for the method of least squares. Dwyer (1944) [1] employed matrix algebra in considering the least squares determination of linear regression coefficients. Along the same line, Aitken (1935) [2] and Cohen (1953) [3] derived matrix results for the adjustment of correlated observations using the minimum variance criterion. Perhaps the most extensive work to date is that of the German geodesist Gotthardt (1952) [4] whose matrix presentation covers the basic material treated in such texts on classical least squares as Merriman [5], Weld [6], and Leland [7]. A fairly broad matrix presentation can also be found in a textbook by Arley and Buch (1950) [8].

None of the above writers considered the subject of a matrix treatment of least squares with full generality. What we shall presently regard as the general problem of least squares was first formulated and solved by the geodesist Helmert (1872) [9] [10]. However, Helmert's results apparently remained largely unknown to English speaking mathematicians, and the problem was again formulated and solved by Deming [11], [12], [13], [14] in 1931. Like Helmert, Deming recognized that all problems in least squares could be considered from a single point of view, namely, as problems of constrained minima. Thus, all least squares adjustments consist basically of the minimization of a quadratic form the variables of which are subject to certain equations of constraint, the so called condition equations. Deming's unique contribution was one of interpretation and application. For instance, by fully utilizing this concept he gave a satisfactory solution to the hitherto unsolved general problem of least squares curve fitting with more than one variable in error.

In this paper we shall employ matrix analysis to extend the Helmert-Deming general problem of least squares to the case for which the observations have the general multivariate normal density. However, if desired, the results may also be considered from the minimum variance point of view. Thus, the conventional method of least squares and the principal results of the above references emerge as special cases of the present development.

2. STATEMENT OF THE GENERAL PROBLEM

Let $x_1^o, x_2^o, \cdots, x_n^0$ be the elements of a set of observations. Assume temporarily that all the observations are ideal, i.e., error free. Then the set is overdetermined if any subset is sufficient to determine the whole set. The degrees of freedom r of the set is equal to the number of observations in excess of the minimum number n_o required to determine the whole set ($r = n - n_o$). The number of independent condition equations existing between the observations is equal to the degrees of freedom of the set. In many cases, curve fitting for instance, it is convenient to introduce p unknown parameters in setting up the relations between the observations. The total number of independent condition equations existing between the observations and parameters is then $m = r + p$. Let this set of m condition equations be denoted by

(2.1) $$f_i(x_1, x_2, \cdots, x_n, \alpha_1, \alpha_2, \cdots, \alpha_p) = 0 \qquad i = 1, 2, \cdots, m \quad .$$

where $\alpha_1, \alpha_2, \cdots, \alpha_p$ are the unknown parameters and the x_i are adjusted values of the original observations: that is

(2.2) $$x_i = x_i^o + v_i = \text{ observed value + residual} \, ,$$

Since the condition equations must be independent, it is necessary that the relations $m > p,\ n > m - p$ (or more compactly $n + p > m > p$) hold. If the p parameters are not mutually independent, the relations existing between them must be included among the condition equations. Hence $s < p$ of the m condition equations may involve parameters only. Introducing known approximation values α_i^o for the parameters with

(2.3) $$\alpha_i = \alpha_i^o + \delta_i \quad ,$$

the condition equations can be written

(2.4) $$f_i(x_1^o + v_1, x_2^o + v_2, \cdots, x_n^o + v_n, \alpha_1^0 + \delta_1, \alpha_2^0 + \delta_2, \cdots, \alpha_p^0 + \delta_p) = 0$$
$$i = 1, 2, \cdots, m$$

Assuming the $v's$ and $\delta's$ to be sufficiently small, (2.4) can be approximated by the zero and first order terms of its Taylor expansion. The linearized condition equations are thus

(2.5)
$$\sum_{j=1}^{n} f_{ij} v_j + \sum_{j=1}^{p} f_{i\alpha_j} + f_{io} = 0, \qquad i = 1, 2, \cdots, m$$

in which

(2.6)
$$f_{ij} = \frac{\partial f_{io}}{\partial x_j^o}$$

(2.7)
$$f_{i\alpha_j} = \frac{\partial f_{io}}{\partial \alpha_j^o}$$

(2.8)
$$f_{io} = f_i(x_1^o, x_2^o, \cdots, x_n^o, \alpha_1^o, \alpha_2^o, \cdots, \alpha_p^o) = 0 \quad .$$

Let σ_{ii} denote the variance of the observation x_i^o. Then the general problem of least squares as considered by Helmert and Deming is to determine the set of residuals and parameter corrections which minimizes the sum

(2.9)
$$S = \sum_{i=1}^{n} \left(\frac{\sigma_{oo}}{\sigma_{ii}} \right) \cdot v_i^2 \quad ,$$

while satisfying the condition equations (2.4) or equivalently (2.5). The quantity $\sigma_{oo} / \sigma_{ii}$ is the weight of the i^{th} observation with σ_{oo} an arbitrary constant termed the unit variance or variance of unit weight. If the observational errors have the normal distribution, the residuals so obtained are the most probable values, and the least squares and maximum likelihood adjustments are equivalent.
Suppose, however, that the errors have the general maltivariate normal distribution. Temporarily considering the residuals as the actual errors, the distribution is

(2.10)
$$h(v_1, v_2, \cdots, v_n) = \sqrt{\frac{|\sigma^{-1}|}{2\pi}} \cdot e^{-\frac{1}{2} V^T \sigma^{-1} V}$$

where σ is the covariance matrix of the observations, $\left|\sigma^{-1}\right|$ the determinant of σ^{-1} and ν the vector

(2.11) $$\nu^T = (\nu_1, \nu_2, \cdots, \nu_n)$$

the superscript T denoting transposition. For this case the most probable set of residuals is clearly that which minimizes the quadratic form

(2.12) $$S = \boldsymbol{\nu}^T \boldsymbol{\sigma}^{-1} \boldsymbol{\nu},$$

or a multiple thereof, while satisfying the specified condition equations. The classical method of least squares corresponds to the special case for which σ is diagonal, i.e., the observations are mutually independent and are normally distributed. Henceforth we shall use the terminology 'method of least squares' in a broad sense to denote the maximum likelihood adjustment of observations having the general multivariate normal distribution.

The set of linearized condition equations (2.5) can be expressed in matrix notation by

(2.13) $$F_{X_o} \nu + F_{A_o} \Delta + F_o = O$$

Where in addition to ν defined above we have equations (2.14) below

$$F_{x_o} = \begin{pmatrix} f_{11} & f_{12} & \cdot & \cdot & f_{1n} \\ f_{21} & f_{22} & \cdot & \cdot & f_{2n} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ f_{m1} & f_{m2} & \cdot & \cdot & f_{mn} \end{pmatrix}, F_{A_o} = \begin{pmatrix} f_{1\alpha_1} & f_{1\alpha_2} & \cdot & \cdot & f_{1\alpha_n} \\ f_{2\alpha_1} & f_{2\alpha_2} & \cdot & \cdot & f_{2\alpha_n} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ f_{m\alpha_1} & f_{m\alpha_2} & \cdot & \cdot & f_{m\alpha_n} \end{pmatrix}, \Delta = \begin{pmatrix} \delta_1 \\ \delta_2 \\ \cdot \\ \cdot \\ \delta_p \end{pmatrix}, F_o = \begin{pmatrix} f_{1o} \\ f_{2o} \\ \cdot \\ \cdot \\ f_m \end{pmatrix}$$

The problem to be considered is to determine, of all possible vectors ν and Δ which satisfy (2.13), those which result in the minimization of (2.12). References [2] and [3] treat basically the case in which F_{x_o} is a square, diagonal matrix with σ nondiagonal(correlated observations). In [1], [4], and [8], F_{x_o} is also square and diagonal, but the observations are uncorrelated(σ diagonal). In addition, references [4] and [8] also consider the case in which several observations appear in each of the condition equations but no parameters are involved(F_{x_o} rectangular and filled and F_{A_o} nonexistent).

3. THE NORMAL EQUATIONS

Problems in constrained minima are most conveniently solved by the method of Lagrange multipliers, also called the method of correlates. Accordingly let

(3.1) $$\Lambda^T = (\lambda_1, \lambda_2, \cdots, \lambda_m)$$

be a vector of m undetermined constant multipliers, one for each equation of constraint(condition equation). We must then minimize the expression

(3.2) $$S = \boldsymbol{\nu}^T \boldsymbol{\sigma}^{-1} \boldsymbol{\nu} - 2\Lambda^T \left(F_{x_o} \boldsymbol{\nu} + F_{A_o} \Delta + F_o \right) \ .$$

Differentiating this with respect to the free variables ν and Δ gives

(3.3) $$dS = 2(\nu^T \sigma^{-1} - \Lambda^T F_{x_o}) d\nu - 2\Lambda^T F_{A_o} d\Delta$$

At the minimum of S , dS must equal zero for all possible variations of $d\nu$ and $d\Delta$. This requires that

(3.4) $$\nu^T \sigma^{-1} - \Lambda^T F_{x_o} = 0 \ ,$$

(3.5) $$F_{A_o}^T \Delta = 0 \ .$$

Solving (3.4) for ν gives

(3.6) $$\nu = \sigma F_{x_o}^T \Lambda \ ,$$

and since ν must satisfy (2.13), we have

(3.7) $$(F_{x_o} \sigma F_{x_o}^T)\Lambda + F_{A_o} \Delta + F_o = 0 \ .$$

The vectors Λ and Δ can be obtained by solving (3.5) and (3.7) simultaneously, and then ν can be determined from (3.6). Combining (3.5) and (3.7) into the single matrix equation

(3.8)

$$\begin{bmatrix} F_{x_o}\sigma F_{x_o}^T & F_{A_o} \\ F_{A_o}^T & 0 \end{bmatrix} \cdot \begin{bmatrix} \Lambda \\ \Delta \end{bmatrix} + \begin{bmatrix} F_o \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

gives the general system of normal equations for the adjustment. Except for the recognition that σ is not necessarily diagonal, equations (3.6) and (3.8) are the matrix equivalents of the results obtained by Helmert and Deming.

4. INVERSION OF NORMAL EQUATION COEFFICIENT MATRIX

If any of the condition equations involve parameters alone, F_{x_o} will be of rank less than m and consequently the $m_x m$ matrix $F_{x_o}\sigma F_{x_o}^T$ will be singular. To allow for this possibility in the solution of the normal equations we assume that the last $s < p$ of the m condition equations involve the parameters only. F_{x_o} will then have the form

(4.1)

$$F_{x_o} = \left[\begin{array}{ccccc} f_{11} & f_{12} & \cdot & \cdot & f_{1n} \\ f_{21} & f_{22} & \cdot & \cdot & f_{2n} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ f_{m-s,1} & f_{m-s,2} & \cdot & \cdot & f_{m-s,n} \\ \hline 0 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 0 \end{array}\right] = \left[\begin{array}{c} F_{x_o}^1 \\ \hline 0 \end{array}\right],$$

where the last s rows are composed of zeros. The broken line indicates the partitioning employed. Partitioning the matrices F_{A_o}, Λ and F_o in a corresponding manner, we have

(4.2)

$$F_{A_o} = \left[\begin{array}{ccccc} f_{1\alpha_1} & f_{1\alpha_2} & \cdot & \cdot & f_{1\alpha_p} \\ f_{2\alpha_1} & f_{2\alpha_2} & \cdot & \cdot & f_{2\alpha_p} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ f_{m-s,\alpha_1} & f_{m-s,\alpha_2} & \cdot & \cdot & f_{m-s,\alpha_p} \\ \hline f_{m-s+1,\alpha_1} & f_{m-s+1,\alpha_2} & \cdot & \cdot & f_{m-s+1,\alpha_p} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ f_{m,\alpha_1} & f_{m,\alpha_2} & \cdot & \cdot & f_{m,\alpha_p} \end{array}\right] = \left[\begin{array}{c} F_{A_o}^1 \\ \hline F_{A_o}^2 \end{array}\right],$$

(4.3)

$$\Lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ . \\ . \\ \lambda_{m-s} \\ \hline \lambda_{m-s+1} \\ . \\ \lambda_m \end{bmatrix} = \begin{bmatrix} \Lambda^1 \\ \hline \Lambda^2 \end{bmatrix}, \qquad F_o = \begin{bmatrix} f_{1o} \\ f_{2o} \\ . \\ . \\ f_{(m-s)o} \\ \hline f_{(m-s+1)o} \\ . \\ f_{mo} \end{bmatrix} = \begin{bmatrix} F_o^1 \\ \hline F_o^2 \end{bmatrix}$$

From (4.1) it follows that

(4.4)

$$F_{x_o} \sigma F_{x_o}^T = \begin{bmatrix} F_{x_o}^1 \sigma F_{x_o}^{1T} & 0 \\ 0 & 0 \end{bmatrix},$$

and this substituted into the general normal equations along with (4.2) and (4.3) gives

(4.5)

$$\begin{bmatrix} F_{x_o}^1 \sigma F_{x_o}^{1T} & 0 & F_{A_o}^1 \\ 0 & 0 & F_{A_o}^2 \\ F_{A_o}^{1T} & F_{A_o}^{2T} & 0 \end{bmatrix} \cdot \begin{bmatrix} \Lambda^1 \\ \Lambda^2 \\ \Delta \end{bmatrix} + \begin{bmatrix} F_o^1 \\ F_o^2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

or in a more convenient arrangement

(4.6)

$$\begin{bmatrix} F_{x_o}^1 \sigma F_{x_o}^{1T} & F_{A_o}^1 & 0 \\ F_{A_o}^{1T} & 0 & F_{A_o}^{2T} \\ 0 & F_{A_o}^2 & 0 \end{bmatrix} \cdot \begin{bmatrix} \Lambda^1 \\ \Delta \\ \Lambda^2 \end{bmatrix} + \begin{bmatrix} F_o^1 \\ 0 \\ F_o^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

This arrangement of the normal equations is generally suitable for elimination solutions. If such a solution should break down at any point, i.e., division by zero occurs, it is merely necessary to delete the row and column where the difficulty occurs and remove them to the end of the matrix, taking care to remove the corresponding unknown also. Such a breakdown would happen, for instance, if $F_{x_o}^1 \sigma F_{x_o}^{1T}$

were singular, or for that matter, if any square submatrix containing an entire left hand section of the diagonal were singular.

In order to derive a useful general expression for the inverse of the coefficient matrix in (4;6), it is necessary that certain matrices to be defined presently be nonsingular. Since this can always be achieved by proper formulation or manipulation of the condition equations prior to the adjustment, the ensuing development may be considered general.

Denoting the coefficient matrix in (4.6) by N , it is a straightforward matter to show by the method of submatrices that

(4.7)
$$N^{-1} = \begin{bmatrix} G^{-1} - H^T Q H & H^T Q & -H^T K^T L^{-1} \\ QH & -Q & K^T L^{-1} \\ -L^{-1} K H & L^{-1} K & L^{-1} \end{bmatrix}$$

in which

(4.8)
$$D = F^1_{x_o} \sigma$$

(4.9)
$$G = D F^{1T}_{x_o}$$

(4.10)
$$H = F^{1T}_{A_o} G^{-1}$$

(4.11)
$$J = H F^1_{A_o}$$

(4.12)
$$K = F^2_{A_o} J^{-1}$$

(4.13)
$$L = K F^{2T}_{A_o}$$

(4.14)
$$Q = J^{-1} - K^T L^{-1} K$$

From (4.6) and (4.7) the roots of the normal equations are

(4.15) $$\Lambda^1 = -(G^{-1} - H^T Q H)F_o^1 + (H^T K^T L^{-1})F_o^2$$

(4.16) $$\Delta = -(QH)F_o^1 - (K^T L^{-1})F_o^2$$

(4.17) $$\Lambda^2 = (L^{-1}KH)F_o^1 - (L^{-1})F_o^2$$

Unless appropriate measures are taken beforehand, it is possible for G or J to be singular, since the ranks of the factors $F_{x_o}^1$ and $F_{A_o}^1$ can be less respectively than the orders of G and J. On the other hand, it is not necessary to assume that L is nonsingular, for this is assured by the independence of the condition equations.

Some relations among the auxiliary matrices (4.9)-(4.14) are

(4.18) $$HGH^T = J$$

(4.19) $$KJK^T = L$$

(4.20) $$QJQ = Q$$

Note that since for the general case $Q \neq J^{-1}$, equation (4.20) implies that Q is singular. It will be shown later that Q is the covariance matrix of the parameters. Hence for the general case considered in which the parameters are not necessarily independent, it is to be expected that Q is singular and of rank $p - s$.

5. COVARIANCE MATRICES RELATED TO THE ADJUSTMENT

Let Y and Z be arbitrary vectors of variates with

(5.1) $$Y^T = (y_1, y_2, \cdots, y_k)$$

(5.2) $$Z^T = (z_1, z_2, \cdots, z_l)$$

Then clearly

(5.3) $$YZ^T = \begin{bmatrix} y_1z_1 & y_1z_2 & \cdot & \cdot & y_1z_l \\ y_2z_1 & y_2z_2 & \cdot & \cdot & y_2z_l \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ y_kz_1 & y_kz_2 & \cdot & \cdot & y_kz_l \end{bmatrix}$$

and we define

(5.4) $$\sigma_{YZ^T} = \begin{bmatrix} \sigma_{y_1z_1} & \sigma_{y_1z_2} & \cdot & \cdot & \sigma_{y_1z_l} \\ \sigma_{y_2z_1} & \sigma_{y_2z_2} & \cdot & \cdot & \sigma_{y_2z_l} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \sigma_{y_kz_1} & \sigma_{y_kz_2} & \cdot & \cdot & \sigma_{y_kz_l} \end{bmatrix}$$

as the covariance matrix of the vectors Y, Z. It readily follows that

(5.5) $$\sigma_{YZ^T} = (\sigma_{Z^TY})^T$$

If $Y = Z$ we shall simply refer to (5.4) as the covariance matrix of Y.

Now let

(5.6) $$u_i^o = u_i(x_1^o, x_2^o, \cdots, x_n^o), \qquad i = 1, 2, \cdots, q$$

be arbitrary functions of the observations x_i^o . Differentiation gives

(5.7) $$du_i^o = \sum_{j=1}^{n} u_{ij}^o dx_j^o$$

where u_{ij}^o is the partial derivative of u_i^o with respect to x_j^o. Regarding du_i^o as

the error in u_i^o resulting from errors dx_j^o in the observations, the covariance of du_g^o, du_h^o; and hence that of u_g^o, u_h^o is

(5.8) $$\sigma_{u_g^o u_g^o} = \sum_{i=1}^{n}\sum_{j=1}^{n} u_{gi}^o \sigma_{ij} u_{hj}^o$$

where σ_{ij} denotes the covariance of x_i^o, x_j^o. Letting

(5.9) $$u_o = \begin{bmatrix} u_1(x_1^o, x_2^o, \cdots, x_n^o) \\ u_2(x_1^o, x_2^o, \cdots, x_n^o) \\ \cdot \\ \cdot \\ u_q(x_1^o, x_2^o, \cdots, x_n^o) \end{bmatrix}, x_o = \begin{bmatrix} x_1^o \\ x_2^o \\ \cdot \\ \cdot \\ x_n^o \end{bmatrix}, \frac{\partial}{\partial x_o} = \begin{bmatrix} \frac{\partial}{\partial x_1^o} \\ \frac{\partial}{\partial x_2^o} \\ \cdot \\ \cdot \\ \frac{\partial}{\partial x_n^o} \end{bmatrix}$$

we can express (5.7) as

(5.10) $$du_o = \left(\frac{\partial}{\partial x_o} u_o^T \right)^T dx_o = u_{x_o} dx_o$$

and it follows readily from (5.8) that

(5.11) $$\sigma_{u_o u_0^T} = u_{x_o} \sigma_{x_o x_o^T} u_{x_o}^T = u_{x_o} \sigma u_{x_o}^T$$

is the covariance matrix of u_o. By means of this result the covariance matrix of any vector of functions of the observations can be determined.

Of particular interest is the covariance matrix of the unknowns in the normal equations. Letting

(5.12) $$W_o^T = \begin{pmatrix} \Lambda^{1T} & \Delta^T & \Lambda^{2T} \end{pmatrix}$$

(5.13) $$C_o^T = \begin{pmatrix} F_o^{1T} & 0 & F_o^{2T} \end{pmatrix}$$

the solution of the normal equations can be written

(5.14) $$W_o = -N^{-1} C_o$$

Since N is essentially unaffected by the observational errors and hence may be considered constant, we have from (5.14)

(5.15) $$dW_o = -N^{-1} dC_o$$

with

(5.16) $$dC_o = \left(\frac{\partial}{\partial x_o} C_o^T \right)^T dX_o = \left(\frac{\partial}{\partial x_o} F_o^{1T} \quad 0 \quad \frac{\partial}{\partial x_o} F_o^{2T} \right)^T dx_o$$

But by definition

(5.17) $$\frac{\partial}{\partial x_o} F_o^{1T} = F_{x_o}^{1T}$$

and since F_o^2 involves parameters only,

(5.18) $$\frac{\partial}{\partial x_o} F_o^{2T} = 0 \quad .$$

Hence in (5.16)

(5.19) $$dC_o = \begin{pmatrix} F_{x_o}^{1T} & 0 & 0 \end{pmatrix}^T dx_o \quad .$$

From (4.7), (5.15) and (5.19) it follows that

(5.20) $$dW_o = -\begin{bmatrix} G^{-1} - H^T Q H \\ Q H \\ -L^{-1} K H \end{bmatrix} F_{x_o}^1 dx_o = W_{X_o} dx_o \quad ,$$

Applying (5.10) and (5.11) yields

(5.21) $$\sigma_{W_o W_o^T} = \begin{bmatrix} G^{-1} - H^T Q H \\ Q H \\ -L^{-1} K H \end{bmatrix} F_{x_o}^1 \sigma F_{x_o}^{1T} \begin{bmatrix} G^{-1} - H^T Q H \\ Q H \\ -L^{-1} K H \end{bmatrix}^T ,$$

which reduces to

(5.22) $$\sigma_{W_o W_o^T} = \begin{bmatrix} G^{-1} - H^T Q H & 0 & -H^T K^T L^{-1} \\ 0 & Q & 0 \\ -L^{-1} K H & 0 & L^{-1} \end{bmatrix} = \begin{bmatrix} \sigma_{\Lambda^1 \Lambda^{1T}} & \sigma_{\Lambda^1 \Delta^T} & \sigma_{\Lambda^1 \Lambda^{2T}} \\ \sigma_{\Delta\Lambda^{1T}} & \sigma_{\Delta\Delta^T} & \sigma_{\Delta\Lambda^{2T}} \\ \sigma_{\Lambda^2 \Lambda^{1T}} & \sigma_{\Lambda^2 \Delta^T} & \sigma_{\Lambda^2 \Lambda^{2T}} \end{bmatrix} .$$

Comparing this result with (4.7) shows that the covariance matrix of the unknowns can be obtained directly from the inverse of the normal equation coefficient matrix. Notice also that the covariance matrices of the vectors Λ^1, Δ and Λ^2, Δ are zero. Since by (3.6) v is a function of Λ alone, it follows that the residuals and parameter corrections, are mutually independent.

To derive the covariance matrix of the residuals we use (3.6), (4.1), (4.3), (4.8) and (4.15) to write

(5.23) $$v = \sigma F_{x_o}^{1T} \Lambda^1 = D^T \Lambda^1 = -D^T((G^{-1} - H^T Q H)F_o^1 - (L^{-1}KH)F_o^2) .$$

Hence by (5.17) and (5.18)

(5.24) $$dv = -D^T(G^{-1} - H^T Q H)F_o^1 dx_o = v_{x_o} dx_o ,$$

which leads to

(5.25) $$\sigma_{vv^T} = v_{x_o} \sigma v_{x_o}^T = D^T(G^{-1} - H^T Q H)D = D^T \sigma_{\Lambda^1 \Lambda^{1T}} D .$$

From this and (5.24) we see that

(5.26) $$v_{x_o} \sigma = \sigma v_{x_o}^T = -v_{x_o} \sigma v_{x_o}^T = -\sigma_{vv^T} .$$

We shall next determine the covariance matrix of the adjusted observations. This matrix is of primary importance, since comparing it with the covariance matrix of the original observations enables one to gage the improvement effected by the adjustment. By (2.2) we may write

(5.27) $$x = x_o + v$$

where x denotes the vector of adjusted Observations. Differentiation gives (5.28)

(5.28) $$dx = (I + v_{x_o})dx_o$$

and using (5.26) the covariance matrix of x turns out to be

(5.29) $$\sigma_{xx^T} = \sigma - \sigma_{vv^T}$$

This shows that the covariance matrix of the adjusted observations is equal to the covariance matrix of the original observations minus that of the residuals.

Once the adjustment is completed it is often required that functions of the adjusted observations, or more generally of the adjusted observations and parameters, be evaluated and that their variances be determined. The latter can be achieved through a reinterpretation of (5.6) - (5.11).

Let

(5.30) $$u_i = u_i(x_1, x_2, \cdots, x_n, \alpha_1, \alpha_2, \cdots, \alpha_p), \qquad i = 1, 2, \cdots, t$$

be arbitrary functions of the adjusted observations and parameters and let u be the vector if the u_i. Then the obvious notation

(5.31) $$dU = U_X dX + U_A dA = \begin{pmatrix} U_X & U_A \end{pmatrix} \begin{pmatrix} dX \\ dA \end{pmatrix}$$

and since by (2.3) $dA = d\Delta$ we have

(5.32) $$\sigma_{UU^T} = \begin{pmatrix} U_X & U_A \end{pmatrix} \sigma_{\begin{pmatrix} X \\ \Delta \end{pmatrix} \begin{pmatrix} X \\ \Delta \end{pmatrix}^T} \begin{pmatrix} U_X & U_A \end{pmatrix}^T$$

The covariance matrix of the vector $(X^T \Delta^T)^T$ is readily found to be—

(5.33) $$\sigma_{\begin{pmatrix} X \\ \Delta \end{pmatrix} \begin{pmatrix} X \\ \Delta \end{pmatrix}^T} = \begin{bmatrix} \sigma - \sigma_{VV^T} & -Q^T HTQ \\ -QHD & Q \end{bmatrix} = \begin{bmatrix} \sigma_{XX^T} & \sigma_{X\Delta^T} \\ \sigma_{\Delta X^T} & \sigma_{\Delta\Delta^T} \end{bmatrix}$$

and this result along with (5.32) may be used to determine the covariance matrix of any vector of functions of the adjusted observations and parameters.

6. DIRECT ADJUSTMENT OF FUNCTIONS OF OBSERVATIONS

In section 2 it was assumed that the x_i^o were the original observations. We now relax this requirement by considering the x_i^o as independent functions of (perhaps) more elemental observations $\hat{x}_i^o$ which have a known multinormal distribution. Accordingly we write

(6.1) $$x_i^o = \phi_i(\hat{x}_1^o, \hat{x}_2^o, \cdots, \hat{x}_k^o), \qquad i = 1, 2, \cdots, n\ ,\ \ n \le k$$

Differentiating this gives

(6.2) $$dx_i^o = \sum_{j=1}^{n} \phi_{ij} d\hat{x}_j^o$$

in which $d\hat{x}_i^o$ may be regarded as the error in the derived observation x_i^o resulting from errors $d\hat{x}_j^o$ in the elemental observations $\hat{x}_i^o$. Writing (6.2) in terms of residuals rather than differentials gives

(6.3) $$v_i = \sum_{j=1}^{n} \phi_{ij} \hat{v}_j \quad ,$$

or in matrix notation

(6.4) $$V = \Phi \hat{V} \quad .$$

As in the original case we assume that the adjusted, derived observations must satisfy the condition equations (2.1). However since the distribution of the derived observations is not now known, we must use the distribution of the elemental observations to obtain the adjustment. From (2.13) and (6.4) it follows that the linearized condition equations in terms of the elemental residuals $\hat{v}_i$ are

(6.5) $$F_{x_o} \Phi \hat{V} + F_{A_o} \Delta + F_o = 0$$

and the most probable elemental residuals are obtained by minimizing, subject to the constraint (6.5),

(6.6) $$S = \hat{V}^T \hat{\sigma}^{-1} \hat{V}$$

in which $\hat{\sigma}$ is the covariance matrix of the elemental observations.

Following the procedure of section 3 with regard to (6.5) and (6.6) we arrive at the following results. The most probable elemental residuals are

(6.7) $$\hat{V} = \hat{\sigma}\ \Phi^T F_{x_o}^T$$

and Λ and Δ are obtained from the normal equations

(6.8) $$\begin{bmatrix} F_{x_o} \Phi \hat{\sigma}\ \Phi^T F_{x_o}^T & F_{A_o} \\ F_{A_o}^T & 0 \end{bmatrix} \cdot \begin{bmatrix} \Lambda \\ \Delta \end{bmatrix} + \begin{bmatrix} F_o \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

From (6.4) and (6.7) the derived residuals are

(6.9) $$V = \Phi \hat{\sigma}\ \Phi^T F_{x_o}^T \Lambda$$

and since the covariance matrix of the derived observations is

(6.10) $$\sigma = \Phi \hat{\sigma}\ \Phi^T$$

equations (6.8) and (6.9) can be written

(6.11) $$\begin{bmatrix} F_{x_o} \sigma F_{x_o}^T & F_{A_o} \\ F_{A_o}^T & 0 \end{bmatrix} \cdot \begin{bmatrix} \Lambda \\ \Delta \end{bmatrix} + \begin{bmatrix} F_o \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

(6.12) $$\nu = \sigma F_{x_o}^T \Lambda$$

But these results are identical respectively with (3.6) and (3.8) which were obtained by minimizing $V^T \sigma^{-1} V$ subject to the same condition equations. Hence it is possible to adjust derived observations directly and without modification by the procedures developed earlier. From this we might infer that the derived observations also have a multinormal density with covariance matrix given by (6.10). A direct proof of this is given in the next section.

If the derived residuals are obtained from (6.11) and (6.12), the problem remains of determining the elemental residuals. This can be accomplished by multiplying (6.9) by $\hat{\sigma}\ \Phi^T(\Phi\hat{\sigma}\ \Phi^T)^{-1}$ which gives, according to (6.7),

(6.13) $$\hat{V} = \hat{\sigma}\ \Phi^T(\Phi\hat{\sigma}\ \Phi^T)^{-1}V$$

The interesting feature of this result is that it is the same thing as the least squares adjustment of the elemental observations subject to condition equations given by (6.4) in which the vector of derived residuals is assumed to be known. Hence once the derived residuals have been obtained, the elemental residuals can be determined from a new least squares adjustment in which the condition equations are the relations between the elemental and derived observations.

The results of this section free us of the sometimes cumbersome restriction of having to adjust the elemental observations directly. Moreover, by judicious formulation of a given problem and choice of derived observations efficient solutions can often be developed. Sometimes, for instance, it is possible to calculate a single covariance matrix for the derived observations, which can be used for a series of different, successive adjustments. In other cases efficient approximation solutions can be developed. For example, if the diagonal terms of σ strongly predominate, it may be possible to ignore the nondiagonal terms completely without altering the final results significantly. An application of this principle in the field of photogrammetric triangulation is given by Brown [15].

7. JOINT DISTRIBUTION OF INDEPENDENT LINEAR COMBINATIONS OF VAREATES FROM A MULTINORMAL DENSITY

For present purposes we shall consider $\hat{V}$ and V as vectors of the actual errors in the elemental and derived observations respectively. By hypothesis the distribution of the elemental errors is

(7.1) $$h(V_1,V_2,\cdots,V_n)=\sqrt{\frac{|\sigma^{-1}|}{2\pi}}\cdot e^{-\frac{1}{2}V^T\sigma^{-1}V}$$

We shall show that the joint, marginal distribution of the derived errors, $V=\Phi\hat{V}$, is a multinormal density with covariance matrix given by (6.10). Equation (6.4) may be rewritten as

(7.2) $$V=\Phi_1\hat{V}_1+\Phi_2\hat{V}_2$$

in which Φ_1 is the square matrix defined by the partition

(7.3) $$\Phi=\left(\Phi_1 \mid \Phi_2\right)=\left[\begin{array}{cccc|ccc} \phi_{1,1} & \phi_{1,2} & \cdot & \cdot & \phi_{1,n} & \phi_{1,n+1} & \cdot & \phi_{1,k} \\ \phi_{2,1} & \phi_{2,2} & \cdot & \cdot & \phi_{2,n} & \phi_{2,n+1} & \cdot & \phi_{2,k} \\ \cdot & \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \phi_{n,2} & \phi_{n,2} & \cdot & \cdot & \phi_{n,n} & \phi_{n,n+1} & \cdot & \phi_{n,k} \end{array}\right]$$

and

(7.4) $$V^T=\left(\hat{V}_1^T \mid \hat{V}_2^T\right)^T=\left(\hat{\nu}_1 \quad \hat{\nu}_2 \quad \cdot \quad \cdot \quad \hat{\nu}_n \mid \hat{\nu}_{n+1} \quad \cdot \quad \hat{\nu}_k\right)^T$$

Since the ν 's are independent, we may assume that the variates have been so ordered that Φ_1 is nonsingular. Solving (7.2) for $\hat{V}_1$ in terms of the n derived errors and remaining $n-k$ elemental errors gives

(7.5) $$\hat{V}_1=\Phi_1^{-1}\left(V-\Phi_2\hat{V}_2\right)$$

and according to (7.4) this allows us to write

(7.6)
$$\hat{V} = \begin{bmatrix} \hat{V}_1 \\ \hat{V}_2 \end{bmatrix} = \begin{bmatrix} \Phi_1^{-1}\left(V - \Phi_2 \hat{V}_2\right) \\ \hat{V}_2 \end{bmatrix} = \begin{bmatrix} \Phi_1^{-1} & -\Phi_1^{-1}\Phi_2 \\ 0 & I \end{bmatrix} \begin{bmatrix} V \\ \hat{V}_2 \end{bmatrix}$$

Letting

(7.7)
$$\frac{\partial}{\partial V} = \left(\frac{\partial}{\partial V_1} \quad \frac{\partial}{\partial V_2} \quad \cdot \quad \cdot \quad \frac{\partial}{\partial V_n} \right)^T$$

the Jacobian of the transformation (7.5) is

(7.8)
$$\left| \frac{\partial}{\partial V} \hat{V}_1^T \right| = \left| \Phi_1^{-1} \right|$$

and the joint distribution of the ν 's and the last $n-k$ ν 's is thus

(7.9)
$$\begin{aligned} g(V_1, V_2, \cdots, V_n, \hat{V}_{n+1}, \cdots, \hat{V}_k) &= h(\hat{V}_1, \hat{V}_2, \cdots, \hat{V}_k) \left| \frac{\partial}{\partial V} \hat{V}_1^T \right| \\ &= \left(\frac{1}{2\pi} \right)^{\frac{k}{2}} \left(\left| \Phi_1^{-1} \right|^2 \left| \hat{\sigma}^{-1} \right| \right)^{\frac{1}{2}} e^{-\frac{1}{2}\hat{V}^T \hat{\sigma}^{-1} \hat{V}} \end{aligned}$$

which according to (7.6) may be written

(7.10)
$$g = \left(\frac{1}{2\pi} \right)^{\frac{k}{2}} \left(\left| \Phi_1^{-1} \right|^2 \left| \hat{\sigma}^{-1} \right| \right)^{\frac{1}{2}} e^{-\frac{1}{2}(V^T \hat{V}_2^T)\Omega^{-1}(V^T \hat{V}_2^T)^T}$$

This is a multinormal density in which

(7.11)
$$\Omega^{-1} = \sigma^{-1}_{\begin{bmatrix} V \\ \hat{V}_2 \end{bmatrix}\begin{bmatrix} V \\ \hat{V}_2 \end{bmatrix}^T} = \begin{bmatrix} \Phi_1^{-1} & -\Phi_1^{-1}\Phi_2 \\ 0 & I \end{bmatrix}^T \begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} \\ \hat{\sigma}_{21} & \hat{\sigma}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \Phi_1^{-1} & -\Phi_1^{-1}\Phi_2 \\ 0 & I \end{bmatrix}$$

is the inverse of the covariance matrix of the vector $(V^T \hat{V}_2^T)^T$. The middle matrix in (7.11) is merely $\hat{\sigma}^{-1}$ partitioned to be conformable with the submatrices of its factors.

The marginal distribution of the ν 's in (7.10) can be obtained by integrating out the ν 's. As is well known (Mood [16]) this will lead to a new multinormal distribution the covariance matrix of which can be obtained by striking out the last $n-k$ rows and columns of Ω. To obtain Ω from Ω^{-1} we observe that

(7.12)
$$\begin{bmatrix} \Phi_1^{-1} & -\Phi_1^{-1}\Phi_2 \\ 0 & I \end{bmatrix} = \begin{bmatrix} \Phi_1 & \Phi_2 \\ 0 & I \end{bmatrix}$$

and invert (7.11) applying the reversal rule. The reduced result is

(7.13)
$$\Omega = \begin{bmatrix} \Phi_1\hat{\sigma}_{11}\Phi_1^T + \Phi_2\hat{\sigma}_{21}\Phi_1^T + \Phi_1\hat{\sigma}_{12}\Phi_2^T + \Phi_2\hat{\sigma}_{22}\Phi_2^T & \Phi_1\hat{\sigma}_{12} + \Phi_2\hat{\sigma}_{22} \\ \hat{\sigma}_{21}\Phi_1^T + \hat{\sigma}_{22}\Phi_2^T & \hat{\sigma}_{22} \end{bmatrix}$$

which by the partitioning of Ω and $\hat{\sigma}$ can be written more compactly as

(7.14)
$$\Omega = \begin{bmatrix} \Phi\,\hat{\sigma}\,\Phi^T & \Phi\,\hat{\sigma}_2 \\ \hat{\sigma}_2^T\Phi^T & \hat{\sigma}_{22} \end{bmatrix}$$

where $\hat{\sigma}_2$ is defined by

(7.15)
$$\hat{\sigma}_2 = \begin{bmatrix} \hat{\sigma}_{12} \\ \hat{\sigma}_{22} \end{bmatrix}$$

Hence by (7.14) the covariance matrix of the ν 's in the marginal distribution is

(7.16)
$$\sigma = \Phi\,\hat{\sigma}\,\Phi^T$$

which agrees with (6.10). The marginal distribution of the derived errors is thus

(7.17)
$$h(V_1, V_2, \cdots, V_n) = \left(\frac{1}{2\pi}\right)^{\frac{k}{2}} \left(\left|\sigma^{-1}\right|\right)^{\frac{1}{2}} e^{-\frac{1}{2}V^T\sigma^{-1}V}$$

which is the result we set out to establish. Since the covariance matrix of the set of variates is unique, this result could have been established directly from (6.10) and the fact stated above that a marginal distribution from a multinormal distribution is also a multinormal distribution.

8. EVALUATION OF THE QUADRATIC FORM OF THE RESIDUALS

While S can always be computed directly from the definition (2.12) , this is not always convenient, especially if σ is nondiagonal and difficult to invert or if the vector of residuals is not otherwise required. It is possible to derive alternate expressions for S which involve neither V nor σ and which are essentially byproducts of the solution of the normal equations. Thus, starting with the definition (2.12) and employing (5.23) and (4.9) yields

(8.1) $$S = V^T \sigma^{-1} V = (\Lambda^{1T} F^1_{x_o} \sigma)\sigma^{-1}(\sigma F^{1T}_{x_o} \Lambda^1) = \Lambda^{1T} G \Lambda^1$$

A simpler form results from this if we successively employ the relations

(8.2) $$G\Lambda^1 = -(F^1_o + F^1_{A_o}\Delta), \quad \Lambda^{1T} F^1_{A_o} = -\Lambda^{2T} F^2_{A_o}, \quad F^2_{A_o}\Delta = -F^2_o$$

which follow directly from the set of normal equations (4.6). The reduced result is

(8.3) $$S = -\Lambda^{1T} F^1_o - \Lambda^{2T} F^2_o = -\Lambda^T F_o$$

which corresponds to the expression given by Deming [14] (equation 17, p. 57) for the general uncorrelated case. This result provides a convenient starting point for the derivation of still other expressions for S which may be useful in special cases. For the general problem, however, it is doubtful whether a simpler or more convenient formula than (8.3) exists.

When the original condition equations have been linearized by Taylor's series, a sequence of iterations of the solution may be necessary to remove the influence of neglected higher order terms. In this case the final value of the vector V is the sum of the initial V and those obtained from successive iterations. The same is true of the vectors $\Lambda,\ \Delta$ and F_o . Equation (8.3) is not strictly valid when Λ and F_o result from iterations. It should, however, provide an accurate approximation in most cases. An exact expression for S when N iterations have been performed may be derived from the second equality of (8.1). Let $(\Lambda^{1T})_i$ and $(F^1_{x_o})_i$ be the values of Λ^{1T} and $F^1_{x_o}$ corresponding to the i^{th} iteration with the values for the initial solution corresponding to $i = 0$. Then

(8.4) $$S = B^T \sigma B$$

where B is the vector

(8.5)
$$B^T = \sum_{i-0}^{N} (\Lambda^{1T})_i (F^1_{x_o})_i$$

The principal merit of this result is that it does not require the inversion of σ. If this is not a serious problem, it may be preferable to employ the first equality of (8.1), using the final vector of resid

9. TESTS OF SIGNIFICANCE FOR THE ADJUSTMENT

Deming (1935) [13] has shown that the quadratic form of the residuals obtained from the least squares adjustment of uncorrelated, normally distributed Observations has a χ^2 distribution with $r = n - n_o$ degrees of freedom (see first paragraph of section 2 of this report). By means of a transformation of the residuals it can be shown that this result holds for the correlated case as well. This fact may be used to provide a test of significance for the adjustment. Essentially, the test determines whether the estimate of unit variance obtained from the adjustment is compatible with the pre-established value. We set

(9.1) $$\chi^2 = S$$

and determine the probability $P(\chi^2 \geq \chi_i^2 \ ; \ \text{r degrees of freedom})$ from a table of the χ^2 distribution. If this probability is unreasonably small (or large, though this would rarely occur in practice), a poor adjustment is indicated and an effort should be made to determine and, if possible, correct the cause. Among the principal reasons for an unsatisfactory χ^2 result are:

(a) Computational errors: Though mention of this possibility may seem trivial, it is felt that the correctness of the computations should be established before seeking other explanations. This is especially true if the adjustment is routine and has been consistently successful before, or if $P(\chi^2 \geq \chi_i^2)$ turns out to be so extreme that the other possibilities to be mentioned seem unlikely.

(b) Uncorrected systematic errors in the observations: An investigation into all phases of the measuring operation-is necessary to evaluate this possibility. Special instrumental calibrations provide the means for correcting such 'errors'. It should be pointed out that a satisfactory χ^2 result does not, in itself, preclude the existence of systematic errors, especially if they are constant or nearly so.

(c) Inadequate or incorrect condition equations: The condition equations represent a mathematical model of a physical situation and as such are satisfactory only if they actually approximate the true situation to a degree compatible with the accuracy of the measurements. In many cases, as measuring accuracy increases, more complex models become necessary in order to account for previously insignificant factors. If a model is inadequate, it may well show up

in the χ^2 test. However, since inadequate models can result from systematic errors, the remarks of (b) hold here also.

An adequate set of condition equations may lead to an inadequate set of linearized condition equations due to the influence of neglected higher order terms. The residuals or parameter corrections may be so large that ordinary iterations actually cause the solution to diverge or to converge to an incorrect result. When such difficulties result from poor approximation parameters alone, the method of 'damped least squares' developed by Levenberg (1944) [17] may be useful (this is discussed somewhat in the next section), while if the residuals also are too large, a gradient method of minimization described by Curry (1944) [18] may lead to a satisfactory solution.

Aside from inadequate condition equations, which may be regarded as poor approximations, incorrect equations, which do not approximate the physical situation at all, may be included among the set of condition equations. Incorrect condition equations result from outright mistakes or from faulty analysis and can be expected to affect the χ^2 result adversely.

In complex measuring situations the most difficult problem may not lie in the actual adjustment, for this can be done straightforwardly by the methods of this paper, nor in obtaining correct condition equations, but rather in the determination of the degrees of freedom of a set of observations and hence of the number of condition equations. We may speak of an incomplete or overcomplete set of condition equations according to whether less or more than the correct number are chosen. Assuming the individual condition equations to be correct, an overcomplete set will lead to a singular set of normal equations. An incomplete set, on the other hand, may yield a solution and may or may not result in a poor χ^2 test, depending upon the importance of the omitted condition equations.

(d) Inaccurate covariance matrix of the observations: In order to make the χ^2 test it is necessary to assume that σ is accurately known. Actually, in practice, an estimate of σ, generally derived from replicated observations, is used. In order to have confidence in such an estimate the degrees of freedom upon which the estimates of the elements of σ are based should be reasonably large, say greater than 20. In many problems, the elements of the covariance matrix of the observations may be known precisely to a constant multiple, σ_{oo}, the variance of unit weight: that is $\sigma = \sigma_{oo}\sigma_o$; where σ_o is known precisely. In this case only σ_{oo} need be

estimated, and an alternate test of significance described below may be used to compare the least squares estimate of unit variance with a given prior estimate.

From the least squares adjustment under consideration we may obtain an estimate $\left(\sigma_{oo}\right)_i$ of σ_{oo} (note that if σ_{oo} is unknown; $S = V^T \sigma_o^{-1} V = \sigma_{oo} V^T \sigma^{-1} V$)

(9.2)
$$(\sigma_{oo})_i = \frac{S}{r} = \frac{\chi_i^2 \sigma_{oo}}{r}$$

Hence

(9.3)
$$\frac{\chi_i^2}{r_i} = \frac{\left(\sigma_{oo}\right)_i}{\sigma_{oo}}$$

in which, for the sake of uniformity, we have used r_i rather than r to denote the degrees of freedom. Now let $\left(\sigma_{oo}\right)_o$ be an estimate of σ_{oo} obtained from a least squares adjustment independent of that from which $\left(\sigma_{oo}\right)_i$ was obtained. (Note that the usual, straightforward method of computing the variance of a set of repeated observations from the sum of squares of deviations from the mean is essentially an estimate based upon a least squares adjustment.) Letting the degrees of freedom associated with such an independent estimate be r_o, we may write

(9.4)
$$\frac{\chi_o^2}{r_o} = \frac{\left(\sigma_{oo}\right)_o}{\sigma_{oo}}$$

We then form the ratio for the F distribution

(9.5)
$$F_o = \frac{\chi_i^2 / r_i}{\chi_o^2 / r_o} = \frac{\left(\sigma_{oo}\right)_i}{\left(\sigma_{oo}\right)_o}$$

The value $P(F > F_o \ ; \ r_i, r_o \text{ degrees of freedom})$ provides a test of the compatibility of the two estimates of unit variance and should be used in place of the χ^2 test when the unit variance is not accurately known beforehand. When r_o is large the χ^2 and *F* tests lead to similar results.

When it is not possible to obtain an independent estimate of σ_{oo}, the above tests cannot be applied. In such cases the adjustment is employed to obtain the estimate given by (9.2). Indeed, this may even be the primary purpose for the adjustment. It must be kept in mind, however, that for such an estimate to be valid it must be known that such factors as have been mentioned above do not influence the result significantly.

Mistakes can often be localized and the nature of systematic errors and model deformities revealed, through a study of the individual residuals and parameters.

Although the χ^2 and F tests described are applicable to the great majority of problems involving physical measurements, more general methods are required when the conditions underlying their application are not fulfilled. In this regard we shall merely mention that such methods are provided by tests based upon the Wishart and related distributions. A derivation of the Wishart distribution together with a study of its properties and applications is given by Wilks (1943) [19].

10. GENERAL REMARKS

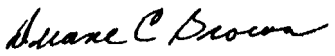
Consideration of the general least squares adjustment and related problems in terms of matrix algebra provides a broad, uncluttered concept of the procedures and operations necessary in the reduction. Since all problems in least squares are merely special cases of the general problem, there is no need for the classification of problems into distinct categories. It is probably this compartmenting of least squares which so long delayed the solution of the general curve fitting problem and has otherwise retarded the application and interpretation of the method.

In some problems the distinction between observations and parameters is not clear cut. A parameter may have a physical interpretation and be capable of direct measurement. In such cases an approximation value of the parameter is sometimes obtained from a direct measurement. But if such a quantity is actually measured and has a probability distribution, it should, strictly speaking, be treated as an observation. In practice, however, a measured quantity is often regarded as a parameter, rather than an observation, when it has a very large variance compared with that which would result from calculating the quantity from the other observations. It might thus be conjectured that in a least squares adjustment a parameter may be regarded as an observation with infinitely large variance. It turns out that this consideration leads to the same results as the original development, providing it be postulated that an observation of infinite variance contributes nothing to the degrees of freedom of a set of observations.(Since variates must have finite variances in order to have the multinormal distribution, this discussion should be considered only in a heuristic sense). This concept allows the formulation of approximation procedures in which observations with relatively large variances are treated as parameters. Conversely, treating parameters as observations with relatively large variances can lead to useful results. Levenberg (1944) [17], for example, used this concept implicitly in deriving the method of 'damped least squares,' which is useful when the usual least squares solution fails to converge due to poor approximation parameters. In essence, Levenberg showed that by treating the parameters as observations with appropriate variances (a method for calculating optimum variances was given) the solution can be made to converge to the correct result. Normally, with each iteration the variances of the parameter residuals increase until ultimately they no longer influence the solution. Although Levenberg derived damped least squares only for the special case in which a single,

independent observation appears in each condition equation (σ diagonal; F_{x_o} square, diagonal), the method can readily be extended to hold for the general case. Incidentally, this provides an illustration of the fact that many results which have been proven for a specific area of least squares actually hold (perhaps with slight modification) for the general case as well.

The least squares adjustment is capable of the following geometrical interpretation. Consider an n-dimensional coordinate system with orthogonal axes $v_1, v_2, \cdots, v_n$. Then the quadratic form $S = \boldsymbol{v}^T \boldsymbol{\sigma}^{-1} \boldsymbol{v}$, being positive definite, will represent an n-dimensional ellipsoid (a detailed study of the n-dimensional ellipsoid is given by Wilks (1943) [19]). The ellipsoid is centered at the origin. If σ is diagonal, the axes of the ellipsoid will coincide with the coordinate axes, while for σ nondiagonal the ellipsoid will be in a tilted orientation. It is clear that by a rotational coordinate transformation, a tilted ellipsoid can be reoriented into a standard position. Such a transformation is specified by $v = R \cdot \hat{v}$ where R is an $n x n$ matrix whose rows (or columns) are composed of the normalized characteristic vectors of σ. Thus a problem involving correlated observations can be reduced to one involving derived observations which are uncorrelated. The dimensions of the hyperellipsoid are, of course, unaffected by a rotation. In fact, the lengths of the axes are directly proportional to the square roots of the characteristic roots of σ. The constant of proportionality which determines their absolute dimensions is simply $S^{1/2}$. It thus follows that the volume of the ellipsoid is directly proportional to $S^{n/2}$ (the complete expression for the volume is given by Burington and May (l953)[20]). Therefore, minimizing S is equivalent to minimizing the volume of the ellipsoid, it being understood, naturally, that the condition equation constraints must be satisfied by some point on the ellipsoid. To simplify matters we may assume that any parameters have been eliminated from the linearized condition equations, leaving r relations between the residuals alone. Each condition equation then represents a hyperplane, and the residuals must lie on the intersection of the r hyperplanes. Now consider the family of hyperellipsoids defined by varying S The orientation and relative dimensions of such ellipsoids will be constant, and all will be centered at the origin. We may think of the family as being formed by the balloonlike expansion of an initial infinitesimal ellipsoid. Let the ellipsoid expand until it becomes tangent to the intersection of the condition equation hyperplanes. For this point all the condition equations are satisfied and the volume of the ellipsoid (and consequently S) is obviously minimum. Hence the coordinates of the point of tangency give the most probable residuals. Thus, from a geometrical point of view the only difference between the adjustment of correlated and

uncorrelated observations lies in the orientation of the hyperellipsoid relative to the coordinate axes. To go somewhat further with the interpretation we may suppose that the ellipsoid has been rotated into a standard position. Then the ellipsoid can be transformed to a hypersphere by a simple stretch transformation. The most probable-set of residuals is then given by the coordinates of the point on the intersection of condition equation hyperplanes which is closest to the origin. The distance of this point from the origin represents the sum of the squares of the residuals. This interpretation is especially convenient for problems in conventional least squares, since the ellipsoids are in standard position to begin with and the necessary stretch is readily accomplished by scaling the observations.



Duane C. Brown

REFERENCES

[1] P. S. Dwyer, "A matrix presentation of least squares and correlation theory with matrix justification of improved methods of solution," Annals of Math. Stat., vol. 15 (1944), pp. 82-89.

[2] A. C. Aitken, "On least squares and linear combinations of observations," Proc. Roy. Soc. Edinburgh, Vol. 55 (1935), pp. 42-48.

[3] E. R. Cohen, "The basis of the criterion of least squares," Rev. Mod. Physics, Vol. 25 (1953), pp. 709-713.

[4] E. Gotthardt, "Ableitung der Grundformeln der Ausgleichungsrechnung mit Hilfe der Matrizenrechnung," Deutche Geodatische Kommission, Reihe A: Hohere Geodasie Veroffentlichung Nr 4, Bamberg (1952).

[5] M. Merriman, "A TEXTBOOK ON THE METHOD OF LEAST SQUARES," J. Wiley and Sons (1884).

[6] L. D. Weld, "THEORY OF ERRORS AND LEAST SQUARES," Macmillan (1916).

[7] 0. M. Leland, "PRACTICAL LEAST SQUARES," McGraw-Hill (1921).

[8] N. Arley and K. R. Duch, "INTRODUCTION TO THE THEORY OF PROBABILITY AND STATISTICS," Wiley (1950), ch. 12.

[9] F. Helmert, "Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate," Leipzig (1872), pp. 215-222 (second edition (1907), pp. 285-293).

[10] Jordan-Eggert, "Handbuch der Vermessungskunde," Vol. 1, Stuttgart (1939), PP. 184-185.

[11] W. E. Deming, "On the application of least squares," Phil. Mag., Vol. 11 (1931), pp. 346-158.

[12] W. E. Doming, "On the application of least squares -- II," Phil. Mag., Vol. 17 (1934), pp. 804-829.

[13] W. E. Deming, "On the application of least squares -- III. A new property of least squares," Phil. Mag., Vol. 19 (1935), pp. 389-402.

[14] W. E. Deming, "STATISTICAL ANALYSIS OF DATA,":Wiley (1943).

[15] D. C. Brown, "Adjustment and triangulation of fixed camera observations," Ballistic Research Laboratories Report No. 960 (1955).

[16] A. M. Mood, "INTRODUCTION TO THE THEORY OF STATISTICS," McGraw-Hill (1950), p. 181.

[17] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," Quart. Appl. Math., Vol. 2 (1944), PP. 164-168.

[18] H. B. Curry, "The method of steepest descent for non-linear minimization problems," Quart. Appl. Math., Vol. 2 (1944), pp. 258-261.

[19] S. S. Wilks, "MATHEMATICAL STATISTICS," Princeton University Press (1944), ch. XI.

[20] B. S. Burington and D. C. May, "HANDBOOK OF PROBABILITY AND STATISTICS WITH TABLES," Handbook Publishers, Inc. (1953), pp. 245-246.

SUPPLEMENTARY REMARKS

After the present report had been readied for publication, the author encountered the two additional matrix treatments of least squares which follow:

[21] C. R. Rao, "ADVANCED STATISTICAL METHODS IN BICMETRIC RESEARCH," Wiley (1952), ch. 2,3.

[22] O. Kempthorne, "THE DESIGN AND ANALYSIS OF EXPERIMENTS," Wiley (1952), PP. 54-66.

Both of these references consider the special case for which F_{x_o} is equal to the unit matrix and σ is merely a multiple of the unit matrix. Rao extends his results to include the case for which some of the condition equations involve parameters only. However, his $F^1_{x_o}$ submatrix for this more general case is again the unit matrix,

G. H. Weiss of BRL has pointed out that the principal result of Section 7 is to be found in the reference

[23] A H. Cramer, "MATHEMATICAL METHODS OF STATISTICS," Princeton Univ. Press (1946), pp. 312-313.

Cramer obtains the result quite simply by showing that the characteristic function of the transformation is that of a multinormal distribution with covariance matrix of the form (7.16).

It has been suggested that a reference on inversion by the method of sub-matrices would be appropriate in connection with the derivation of equation (4.7). This is provided by

[24] Frazer, Duncan, Collar, "ELEMENTARY MATRICES," Cambridge Univ. Press (1950), pp. 112-113.

Also in regard to (4.7) it seems worthwhile to mention that the derivation is considerably simplified if the relations (4.18) - (4.20) are employed. This is also true for the reductions leading to equations (5.22), (5.25) and (5.33).