NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Technical Memorandum 33-807

The Covariance Matrix for the Solution Vector of an Equality-Constrained Least-Squares Problem

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

December 15, 1976

PREFACE

The work described in this report was performed by the Information Systems Division of the Jet Propulsion Laboratory.

CONTENTS

THE COVARIANCE MATRIX FOR THE SOLUTION VECTOR OF AN EQUALITY-CONSTRAINED LEAST SQUARES PROBLEM

1. INTRODUCTION

Consider the linear least squares problem

$$Ex \cong f$$

subject to the linear equality constraints

$$Cx = d$$

We refer to this as Problem LSE denoting Lease Squares with Equality constraints.

Methods for solving Problem LSE are described in Chapters 20 - 22 of Ref. (1).

In this note we describe methods for computing the covariance matrix V for the solution vector x. Different methods of computing V will be discussed which are convenient for use with each of the different solution algorithms given in Ref. (1). Any reference to a Chapter, Section, or Page without further qualification is to be understood to refer to Ref. (1).

We assume throughout that the covariance matrix of f is the identity matrix. If the covariance matrix of f is known to be something other than the identity matrix then a preliminary left multiplication of E and f by an appropriate matrix will produce the desired standard situation. (See Chapter 25, Section 2.)

We assume that E, C, and d are known exactly, or at least that their errors are very small relative to those of f.

Let C be an $m_1 \times n$ matrix and let E be $m_2 \times n$. We assume that

$$m_1 < n$$

$$m_1 + m_2 \geq n$$

$$\text{Rank}\ (C) = m_1$$

$$\text{Rank}\left(\begin{bmatrix} C \\ E \end{bmatrix}\right) = n$$

With these assumptions Problem LSE has a unique solution vector and all of the solution methods to be discussed apply without the need to consider unusual special cases.

As a small numerical example to illustrate the computational methods to be presented we use the same problem that was used in Chapters 20 - 22. (See p. 140).

$$C = \begin{bmatrix} 0.4087 & 0.1593 \end{bmatrix} \qquad d = 0.1376$$

$$E = \begin{bmatrix} 0.4302 & 0.3516 \\ \\ 0.6246 & 0.3384 \end{bmatrix} \qquad f = \begin{bmatrix} 0.6593 \\ \\ 0.9666 \end{bmatrix}$$

The computations described in Chapters 20 - 22 were done using a relative precision of 10^{-8} whereas intermediate and final results were rounded to about four decimal places for publication. In this note we begin with the published intermediate results when applicable and compute using a pocket calculator.

2. SOLUTION METHOD USING A BASIS OF THE NULL SPACE

This solution method is described in Chapter 20, pages 134-141. It may be summarized as follows.

Apply Householder orthogonal transformations to C from the <u>right</u> to reduce C to <u>lower</u> triangular form. Apply these same transformations to E from the right. Denoting the product of these orthogonal transformations by the $n \times n$ orthogonal matrix K these operations may be represented by the equation:

$$\begin{bmatrix} C \\ E \end{bmatrix} K = \begin{array}{c} \left[\begin{array}{cc} \tilde{C}_1 & O \\ \tilde{E}_1 & \tilde{E}_2 \end{array}\right] \begin{array}{l} \}m_1 \\ \}m_2 \end{array} \\ \begin{array}{cc} \underbrace{\quad}_{m_1} & \underbrace{\quad}_{n-m_1} \end{array} \end{array} \tag{1}$$

Solve the following lower triangular system for y_1:

$$\tilde{C}_1 \, y_1 = d$$

Compute:

$$\tilde{f} = f - \tilde{E}_1 \, y_1$$

Solve the least squares problem:

$$\tilde{E}_2 \, y_2 \cong \tilde{f} \tag{2}$$

Compute:

$$x = K \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

To compute the covariance matrix, V, for x, first compute the covariance matrix S for y_2:

$$S = (\tilde{E}_2^T \, \tilde{E}_2)^{-1} \tag{3}$$

Then the covariance matrix for

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

is

$$U = \begin{bmatrix} 0 & 0 \\ 0 & S \end{bmatrix} \tag{4}$$

and the covariance matrix for x is

$$V = KUK^T \tag{5}$$

Consider the numerical example given on pp 140-141.

In this example

$$\tilde{E}_2 = \begin{bmatrix} 0.1714 \\ 0.0885 \end{bmatrix}$$

and

$$K = \begin{bmatrix} -0.9317 & -0.3632 \\ -0.3632 & 0.9317 \end{bmatrix}$$

Thus using Eq. (3) - (5) we obtain

$$S = (0.037210)^{-1} = 26.874$$

and

$$V = \begin{bmatrix} 3.545 & -9.094 \\ -9.094 & 23.33 \end{bmatrix} \tag{6}$$

Note that although Eq. (3) is a valid mathematical definition of S it does not represent the most stable way to compute S. If Problem (2) is solved using Householder transformations, then one would have an upper triangular matrix R such that

$$Q\tilde{E}_2 = \begin{bmatrix} R \\ 0 \end{bmatrix} \tag{7}$$

where Q is $m_2 \times m_2$ orthogonal.

Then, as is described in Chapter 12, one could compute S as

$$S = R^{-1} (R^{-1})^T \tag{8}$$

3. SOLUTION METHOD USING DIRECT ELIMINATION

This solution method is described in Chapter 21, pp 144-147. It may be summarized as follows.

Assume column interchanges have been done in the augmented matrix

$$\begin{bmatrix} C \\ E \end{bmatrix},$$

if necessary, to assure that the first m_1 columns of C are linearly independent.

Use Gaussian elimination to zero all elements below the diagonal in the first m_1 columns of

$$\begin{bmatrix} C \\ E \end{bmatrix},$$

$$G \begin{bmatrix} C & d \\ E & f \end{bmatrix} = \begin{array}{c} \left[\begin{array}{ccc} \tilde{C}_1 & \tilde{C}_2 & \tilde{d} \\ 0 & \tilde{E}_2 & \tilde{f} \end{array} \right] \begin{array}{l} \}\, m_1 \\ \}\, m_2 \end{array} \\ \begin{array}{ccc} \underbrace{\quad}_{m_1} & \underbrace{\quad}_{n-m_1} & \underbrace{\quad}_{1} \end{array} \end{array}$$

Solve the least squares problem:

$$\tilde{E}_2 x_2 \cong \tilde{f} \tag{9}$$

Solve for x_1 in

$$\tilde{C}_1 x_1 = \tilde{d} - \tilde{C}_2 x_2 \tag{10}$$

Then the solution vector is

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

To compute the covariance matrix of x introduce the $m_1 \times (n - m_1)$ matrix H, obtained by solving

$$\tilde{C}_1 H = \tilde{C}_2$$

Then from Eq. (10) we may write

$$x_1 = \tilde{C}_1^{-1} \tilde{d} - H x_2 \tag{11}$$

Let $\mathcal{E}$ denote the expected value operator. Introduce the mean values

$$\bar{x}_1 = \mathcal{E}(x_1)$$

and

$$\bar{x}_2 = \mathcal{E}(x_2)$$

These mean values satisfy Eq. (11), i.e.,

$$\bar{x}_1 = \tilde{C}_1^{-1} \tilde{d} - H \bar{x}_2 \tag{12}$$

Subtract Eq. (12) from Eq. (11) obtaining

$$(x_1 - \bar{x}_1) = -H(x_2 - \bar{x}_2) \tag{13}$$

from which we may write

$$x - \bar{x} = \begin{bmatrix} x_1 - \bar{x}_1 \\ x_2 - \bar{x}_2 \end{bmatrix} = \begin{bmatrix} -H \\ I \end{bmatrix} \cdot (x_2 - \bar{x}_2) \tag{14}$$

Let W denote the $(n - m_1) \times (n - m_1)$ covariance matrix of x_2, which from Eq. (9) may be defined as

$$W = (\tilde{E}_2^T \tilde{E}_2)^{-1} \tag{15}$$

Then using Eq. (14) the covariance matrix V of x can be written as

$$V = \begin{bmatrix} -H \\ \\ I \end{bmatrix} W \begin{bmatrix} -H^T & I \end{bmatrix} = \begin{bmatrix} HWH^T & -HW \\ \\ -WH^T & W \end{bmatrix}$$

Consider the same numerical example as before, solved by this method. (See p. 147). We have

$$\tilde{E}_2 = \begin{bmatrix} 0.1839 \\ \\ 0.0949 \end{bmatrix}$$

and

$$H = \tilde{C}_1^{-1} \tilde{C}_2 = (0.4087)^{-1}(0.1593) = 0.38977$$

We compute

$$W = (\tilde{E}_2^T \tilde{E}_2)^{-1} = (0.042825)^{-1} = 23.351$$

and

$$V = \begin{bmatrix} 3.548 & -9.101 \\ \\ -9.101 & 23.351 \end{bmatrix}$$

Note that Eq. (15) is a valid mathematical definition of W but not a recommended computational formula. See the remark at the end of Sec. 2 for suggestions for a more stable way of computing W.

4. SOLUTION BY WEIGHTING

This solution method is described in Chapter 22. It may be summarized as follows:

Suppose the data are scaled so that the elements of largest magnitude in the matrices C and E are approximately the same size. Introduce a scale factor, ϵ, such that ϵ^2 is smaller than the working precision. For instance set $\epsilon < 10^{-4}$ for Univac single precision arithmetic and $\epsilon < 10^{-9}$ for Univac double precision.

Solve the least squares problem

$$\begin{bmatrix} C \\ \\ \epsilon E \end{bmatrix} x \cong \begin{bmatrix} d \\ \\ \epsilon f \end{bmatrix} \tag{16}$$

using Householder or Givens orthogonal transformations.

Solving the problem by either of these methods involves triangularization by left multiplication by an orthogonal matrix Q:

$$Q \begin{bmatrix} C & d \\ \\ \epsilon E & \epsilon f \end{bmatrix} = \begin{array}{l} \left.\begin{bmatrix} \tilde{C} & \tilde{d} \\ \\ \epsilon\tilde{E}_1 & \epsilon\tilde{f}_1 \\ \\ 0 & \epsilon\tilde{f}_2 \end{bmatrix}\begin{matrix} \}m_1 \\ \\ \}n - m_1 \\ \\ \}m_1 + m_2 - n \end{matrix}\right. \\ \quad\ \underbrace{\quad}_{n} \quad\ \underbrace{\quad}_{1} \end{array}$$

Then x is obtained by solving the upper triangular system

$$\begin{bmatrix} \tilde{C} \\ \\ \epsilon\tilde{E}_1 \end{bmatrix} x = \begin{bmatrix} \tilde{d} \\ \\ \epsilon\tilde{f}_1 \end{bmatrix}$$

The condition number of this problem is very large (about ϵ^{-1}) however this does not affect the accuracy of the solution because of the special structure of the matrix and right-side vector.

The covariance matrix, V, of x is

$$V = \epsilon^2 \begin{bmatrix} \tilde{C} \\ \\ \epsilon\tilde{E}_1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \tilde{C} \\ \\ \epsilon\tilde{E}_1 \end{bmatrix}^{-1^T} \tag{17}$$

Even though the triangular matrix $\begin{bmatrix} \tilde{C} \\ \epsilon\tilde{E}_1 \end{bmatrix}$ has a large condition number its inverse can be computed without numerical difficulty.

Consider the example used before. The weighted problem to be solved (see p. 156) is

$$\begin{bmatrix} 0.4087 & 0.1593 \\ \\ 0.4302\epsilon & 0.3516\epsilon \\ \\ 0.6246\epsilon & 0.3384\epsilon \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \\ x_2 \end{bmatrix} \cong \begin{bmatrix} 0.1376 \\ \\ 0.6593\epsilon \\ \\ 0.9666\epsilon \end{bmatrix} \tag{18}$$

Since we will be using 4 or 5 place decimal arithmetic we could choose any value of $\epsilon < 10^{-3}$. The point is that for any two numbers, a and b, of comparable magnitude ϵ should be small enough relative to the computational precision so that the computed value of $a^2 + (\epsilon b)^2$ will just be a^2. For our numerical example we will not assign a specific value to ϵ but will use the computational rule that the computed value of an expression of the form $a^2 + (\epsilon b)^2$ is a^2 when a and b are of the same order of magnitude.

The data arrays of Eq. (18) can be triangularized by Householder transformations to obtain the equivalent problem

$$\begin{bmatrix} -0.4087 & -0.15930 \\ \\ 0 & -0.20698\epsilon \\ \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \\ x_2 \end{bmatrix} \cong \begin{bmatrix} -0.13760 \\ \\ -0.80403\epsilon \\ \\ 0.43606\epsilon \end{bmatrix} \tag{19}$$

Solving the nonsingular system represented by the first two rows of Eq. (19) given the solution vector

$$x = \begin{bmatrix} -1.1774 \\ \\ 3.8846 \end{bmatrix}$$

Let R denote the leading 2 × 2 triangular matrix in Eq. (19). We compute

$$R^{-1} = \begin{bmatrix} -2.4468 & 1.8831\epsilon^{-1} \\ \\ 0 & -4.8314\epsilon^{-1} \end{bmatrix}$$

Then using Eq. (17) we compute the covariance matrix V of x as

$$V = \epsilon^2 R^{-1}(R^{-1})^T = \begin{bmatrix} 3.5461 & -9.0980 \\ \\ -9.0980 & 23.342 \end{bmatrix}$$

This computational procedure looks peculiar in some ways but it is valid. For example the upper left element of R^{-1}, namely -2.4468, is entirely lost in the roundoff error when the product $R^{-1}(R^{-1})^T$ is computed and this results in the computed V being singular whereas R^{-1} was clearly nonsingular.

This is exactly the right thing to happen, however, since the covariance matrix V for problem LSE should be singular and should not be influenced by the upper left element of R^{-1}.

Close analysis of this weighted method (See Exercise 22.40, p. 157) shows that with sufficiently small ϵ this is just a sneaky way of performing the direct elimination algorithm treated in Sec. 3 of this note (Chap. 21 of the book).

5. ONE MORE APPROACH

Still another way of looking at Problem LSE is presented on pp. 141-143. As is noted there we expect that this approach may not have practical value but may be of theoretical interest.

Let K be the n×n orthogonal matrix defined in Sec. 2 of this note (Chapter 20 of the book). Let K be partitioned as

$$K = \underbrace{\left[K_1\right.}_{m_1} \quad \underbrace{\left.K_2\right]}_{n-m_1}\}n$$

Define

$$\hat{E} = (EK_2)(EK_2)^+ E$$

where the superscript "+" denotes pseudoinverse. Define

$$\hat{A} = \begin{bmatrix} C \\ \hat{E} \end{bmatrix}$$

Then, as is proved in Chapter 20, the least squares solution of

$$\hat{A}x \cong \begin{bmatrix} d \\ f \end{bmatrix} \tag{20}$$

is the same as the solution of problem LSE:

$$\begin{cases} Cx = d \\ Ex \cong f \end{cases}$$

To compute the covariance matrix of x, regarding x as the solution of Eq. (20), we first write

$$x = \hat{A}^+ \begin{bmatrix} d \\ f \end{bmatrix}$$

Assuming the covariance matrix of $\begin{bmatrix} d \\ f \end{bmatrix}$ is

$$\begin{array}{cc} \left.\begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}\begin{array}{l} \}m_1 \\ \}m_2 \end{array}\right. \\ \underbrace{\quad}_{m_1}\ \underbrace{\quad}_{m_2} \end{array}$$

it follows that the covariance matrix, V, of x is

$$V = \hat{A}^+ \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} \hat{A}^{+T} \tag{21}$$

From Eq. (20.30) on p. 141 we know that $\hat{A}^+$ can be written as

$$\hat{A}^+ = \Big[\underbrace{C^+ - K_2(EK_2)^+ EC^+}_{m_1},\ \underbrace{K_2(EK_2)^+}_{m_2}\Big]\Big\}n \tag{22}$$

Substituting Eq. (22) into Eq. (21) gives

$$V = K_2(EK_2)^+ (EK_2)^{+T} K_2^T \tag{23}$$

From Eq. (1) we have

$$EK_2 = \tilde{E}_2$$

and thus Eq. (23) can be written as

$$V = K_2 \tilde{E}_2^+ \tilde{E}_2^{+T} K_2^T$$

$$= K_2(\tilde{E}_2^T E_2)^{-1} K_2^T$$

This last expression is identical to the right-side of Eq. (5). Thus we obtain the same representation of V as in Sec. 2.

REFERENCES

1. C. L. Lawson, and R. J. Hanson, "Solving Least Squares Problems", Prentice-Hall, 1974.