NASA CONTRACTOR REPORT

NASA CR-2892

THE INCOMPLETE INVERSE AND ITS APPLICATIONS TO THE LINEAR LEAST SQUARES PROBLEM

Georg E. Morduch

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • AUGUST 1977

1. Report No. NASA CR-2892	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle THE INCOMPLETE INVERSE AND ITS APPLICATIONS TO THE LINEAR LEAST SQUARES PROBLEM		5. Report Date August 1977
		6. Performing Organization Code
7. Author(s) Georg E. Morduch		8. Performing Organization Report No.
9. Performing Organization Name and Address Old Dominion Systems, Inc. Gaithersburg, Maryland		10. Work Unit No.
		11. Contract or Grant No. NAS 5-23587
12. Sponsoring Agency Name and Address National Aeronautics & Space Administration Washington, D. C. 20546		13. Type of Report and Period Covered
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

A modified matrix product is defined and it is shown that this product defines a group. The inverse of the group is called the incomplete inverse. Algorithms for computing the incomplete inverse are provided.

The incomplete inverse has an important application to the least squares problem. It is shown that the incomplete inverse of an augmented normal matrix includes all the quantities (including the effect of "consider" parameters) associated with the least squares solution. In particular, an answer is provided to the problem that occurs when, (I) the data residuals are too large, (II) there is insufficient data to justify augmenting the model by more than one term. A simple computation involving the incomplete inverse will tell which term will yield the best improvement in the data fit. This is of special interest to the processing of satellite data, where the model may always be augmented by any number of geopotential terms. The incomplete inverse may thus be used to determine which geopotential terms most influence an orbit.

17. Key Words (Selected by Author(s)) Least Squares Incomplete inverse Algorithms Matrix	18. Distribution Statement Unclassified – Unlimited Cat. 67		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 84	22. Price* $5.00

ABSTRACT

A modified matrix product is defined and it is shown that
this product may be used to define a group. Each matrix in such a
group possesses an inverse. To distinguish it from the regular ma-
trix inverse, it is called the generalized inverse. The general so-
lution of a matrix equation is expressed in terms of the incomplete
inverse (a special form of the generalized inverse). A minimum norm
solution is derived, and it is shown that the associated matrix (the
Penrose pseudo-inverse) is a generalized inverse.

Algorithms for computing the incomplete inverse of a sym-
metric matrix are derived. These algorithms, which are little more
complicated than those for the regular inverse, utilize matrix sym-
metry so that the matrix may be stored in upper triangular form.
Similar use of matrix symmetry is made in the computation of the
pseudo-inverse (the method thus requires only half as much computer
core storage as the commonly used Andree algorithm).

The most important application of the incomplete inverse
is to the least squares problem. It is shown that the incomplete
inverse of an augmented normal matrix includes all the quantities
(including the effect of 'consider' parameters) associated with the
least squares solution. In particular, an answer is provided to the
problem that occurs when, (i) the data residuals are too large, and
(ii) the mathematical model may theoretically be augmented by a
large number of terms, but (iii) there is insufficient data to jus-
tify augmenting the model by more than one term. A simple computa-

tion involving the incomplete inverse will tell which term will yield the best improvement in the data fit. This is of special interest to the processing of satellite data, where the model may always be augmented by any number of geopotential terms. The incomplete inverse may thus be used to determine which geopotential terms most influence some observed orbit.

PREFACE

.

The incomplete inverse was first introduced in "An Extension of NAP3.1F" (Morduch, 1975) in connection with the application of 'consider' parameters to the least squares solution. In that report it was defined in a rather complicated manner involving matrix partitions and orthogonal transformations. For that reason the derivations and proofs of formulae involving the incomplete inverse tended to be long and laborious, although the formulae themselves were relatively simple. The generalized inverse is defined in this report as the true inverse of a matrix in a group defined by a generalized matrix product, and includes the incomplete inverse as a special case. The reason for this report is not merely to give much simpler proofs for previously derived formulae, but also to describe the application of the incomplete inverse to the general and minimum norm solutions of matrix equations.

TABLE OF CONTENTS

vii

1.0 INTRODUCTION

The reduction of observed data using the method of least squares is often fraught with difficulty. If this is caused by some unpredictable malfunction of a measurement instrument, then, provided that this is recognized, the offending data is excluded from the reduction. However, more frequently, the cause of the difficulty lies in the inadequacy of the mathematical modelling. This is often due not to any lack of knowledge of theory on the part of the investigator, but rather to the impracticability of having to solve for an extremely large set of unknown parameters. The concept of 'consider' parameters has been used by several investigators to tackle the problem when this is so.

The purpose of this report is to introduce a generalized inverse A^* of a matrix A, a particular form of which, the incomplete inverse, has the property that its elements contain all the results of the least squares reduction including the effects of the 'consider' parameters.

The generalized inverse is defined in Section 2 of this report, as is its particular form, the incomplete inverse. The general solution of a matrix equation is expressed in terms of the incomplete inverse. This is used to derive a minimum norm solution. Although the method may inherently be less accurate than, say, the methods of Andree of Gram-Schmidt (Lefferts, 1969), all computations involve only symmetric matrices so that the computer storage requirement is much less than that of either of the two

other methods. It is also shown that the matrix associated with the minimum norm solution belongs to the set of generalized inverses.

An algorithm for the computation of the incomplete inverse of a symmetric matrix is derived in Section 3. It is a generalization of the particular method of Gaussian elimination derived by the author for obtaining the usual inverse. (This method is used in the Navigation Analysis Program, NAP3.1F, used at Goddard Space Flight Center for orbit determination. Although the method involves exactly the same number of arithmetic computations as any other method of Gaussian elimination, which take advantage of the matrix symmetry, it is much simpler, and therefore easier to program, when the matrix to be inverted contains patterned zeroes.)

Section 4 deals with the linear least squares problem. It is shown, amongst other things, that both the weighted least squares estimate, $\hat{x}$, and the weighted sum of the squares of the residuals, $C(\hat{x})$, are contained within the incomplete inverse, M^*, of a matrix M. The effect of consider parameters is taken into account in the derivation of a formula for the expected value of $C(\hat{x})$. ($C(\hat{x})$ is a measure of the goodness of the data fit and is therefore an important quantity to consider in the least squares reduction.) It is also shown that the change in the value of $C(\hat{x})$ when a parameter is switched from a 'consider' to a 'solve' mode is given by a simple expression involving one multiplication and one division. The last mentioned result is of considerable practical significance for the following reason. Suppose that for some data reduction a given mathematical model results in an unacceptably bad data fit. Provided that the investigator knows how to

2

improve his model, he will then selectively (to augment the model by a large number of parameters in one step is not, in general desirable, since the results may be quite meaningless if too many parameters are solved for) augment his model by various parameters and then choose the one which gives the best fit. Such a task may be very arduous. However, by including as 'consider' parameters all such parameters as he might wish to include in his model, he may then select the one, which when switched from the 'consider' to the 'solve' mode yields the best improvement in the data fit. In other words, there is no need to perform a large number of data reductions using different mathematical models. One will do.

The most important formulae are summarized in Section 5.

2.0 DEFINITION OF A
GENERALIZED MATRIX INVERSE AND
ITS APPLICATIONS TO THE SOLUTION
OF MATRIX EQUATIONS

In this section we shall define a generalized matrix product between two square matrices. It will be shown that based on this generalized product a set of n×n matrices form a group. The generalized inverse of a matrix will then be defined as the inverse of a matrix in this group.

2.1   The Generalized Matrix Product of Two Square Matrices.

Definition: The generalized matrix product of any two square matrices (of equal dimensions) A and B is denoted by A·B and is defined by

$$A \cdot B \ = \ -ARB + A(I-R) + (I-R)B, \tag{2.1}$$

where I is the identity matrix and R is a square matrix satisfying

$$R \ = \ R^{T} \tag{2.2}$$

and

$$RR \ = \ R \tag{2.3}$$

It can easily be shown that for any three n×n matrices A, B and C

$$(A \cdot B) \cdot C \ = \ A \cdot (B \cdot C) \tag{2.4}$$

4

so that the generalized product is <u>associative</u>.

It can also be shown that

$$A \cdot (-R) = (-R) \cdot A = A, \qquad (2.5)$$

for any n×n matrix A. Hence (-R) is the <u>identity element</u>.

The <u>inverse</u> of A, which we shall denote by A^*, must satisfy

$$A \cdot A^* = -R \qquad (2.6)$$

As is well known from group theory, it follows from equation (2.6) that

$$A^* \cdot A = -R \qquad (2.7)$$

It follows from the above that based on the generalized product, all n×n matrices that possess an inverse form a group.

We hence conclude that the inverse is unique and also that

$$(A^*)^* = A, \qquad (2.8)$$

and

$$(A \cdot B)^* = B^* \cdot A^* \qquad (2.9)$$

It follows from equations (2.1) and (2.2) that

$$(A \cdot B)^T = -B^T R A^T + (I-R)A^T + B^T (I-R),$$

i.e.,

$$(A \cdot B)^T = B^T \cdot A^T \qquad (2.10)$$

Hence taking the transpose of equation (2.5) we deduce that

$$(A^*)^\mathsf{T} \cdot A^\mathsf{T} \;=\; -R$$

Since by definition

$$(A^\mathsf{T})^* \cdot A^\mathsf{T} \;=\; -R,$$

we conclude that

$$(A^*)^\mathsf{T} \;=\; (A^\mathsf{T})^* \tag{2.11}$$

It follows from the above equation that if A is symmetric then so is A^*.

2.2 The Generalized Inverse.

A^* has been defined as the inverse of A with respect to the generalized product based on R. Henceforth, it will be referred to simply as the generalized inverse of A. However, when any risk of confusion arises, it will be referred to as the generalized inverse of A with respect to R and will be denoted by $A^*_{,R}$.

We shall next derive some useful formulae involving A^*. At this point it is convenient to define the square matrix S by

$$S \;=\; I - R \tag{2.12}$$

It follows from the above definition and from equations (2.2) and (2.3) that

$$S \;=\; S^\mathsf{T}, \tag{2.13}$$

$$SR \;=\; RS \;=\; 0, \tag{2.14}$$

and

$$SS \;=\; S \tag{2.15}$$

6

Equation (2.1) may then be rewritten in the form

$$A \cdot B = -ARB + AS + SB \qquad (2.16)$$

From the above and equation (2.6) we deduce that

$$ARA^* = R + AS + SA^*, \qquad (2.17)$$

whence

$$A^* = (AR - S)^{-1} (R + AS) \qquad (2.18)$$

Note that it follows from the uniqueness of A^* that if A^* exists then $(AR-S)$ must be non-singular. From equations (2.16) and (2.7) we similarly find that

$$A^* RA = R + A^* S + SA, \qquad (2.19)$$

whence

$$A^* = (R + SA)(RA - S)^{-1} \qquad (2.20)$$

It follows from equations (2.8) and (2.18) that

$$A = (A^* R - S)^{-1} (R + A^* S), \qquad (2.21)$$

and similarly from equations (2.8) and (2.20) that

$$A = (R + SA^*)(RA^* - S)^{-1} \qquad (2.22)$$

Postmultiplying equation (2.17) by R we obtain

$$ARA^* R = R + SA^* R, \qquad (2.23)$$

whence

$$(AR - S) A^* R = R \qquad (2.24)$$

Since $(AR-S)S = -S$ and $R+S = I$, we therefore conclude

7

that

$$(AR-S)(A^*R-S) = I \qquad (2.25)$$

Hence also,

$$(A^*R-S)(AR-S) = I \qquad (2.26)$$

Premultiplying equation (2.23) by R we find that

$$(RAR)(RA^*R) = R, \qquad (2.27)$$

whence

$$(RAR+S)(RA^*R+S) = I \qquad (2.28)$$

Since

$$AR-S = A-(A+I)S,$$

we deduce from equation (2.25) that

$$A(A^*R-S) = I + (A+I)(SA^*R-S) \qquad (2.29)$$

2.3   The Rank of A when $SA^*S = 0$.

In general, if M, B and C are three n×n matrices satisfying

$$M = B + C, \qquad (2.30)$$

and

$$BC = 0, \qquad (2.31)$$

then clearly

$$rank(M) \le rank(B) + rank(C) \qquad (2.32)$$

and

$$\text{rank}(B) \leq n\text{-rank}(C) \qquad (2.33)$$

If furthermore, M is non-singular so that rank(M) = n, then it follows from (2.32) that

$$\text{rank}(B) \geq n\text{-rank}(C) \qquad (2.34)$$

Comparing inequalities (2.33) and (2.34) we deduce that

$$\text{rank}(B) = n\text{-rank}(C) \qquad (2.35)$$

Since

$$I = R+S \text{ and } RS = 0,$$

we therefore conclude that

$$\text{rank}(S) = n\text{-rank}(R) \qquad (2.36)$$

Since for any matrix M,

$$(I-RMS)(I+RMS) = I, \qquad (2.37)$$

it follows that I-RMS is non-singular.

Since

$$I-RMS = S-RMS + R$$

we deduce that for any matrix M

$$\text{rank}(S-RMS) = n\text{-rank}(R) \qquad (2.38)$$

It follows from equation (2.28) that

9

RAR + S   is non-singular.

Since

$$RAR + S \ = \ (S-RAS) + RA,$$

we hence deduce that

$$rank(RA) \ = \ n-rank(S-RAS)$$

Since equation (2.38) is valid for any matrix M it follows from the above that

$$rank(RA) \ = \ rank(R) \qquad\qquad (2.39)$$

Since quite obviously

$$rank(RA) \ \leq \ rank(A),$$

we conclude that

$$rank(R) \ \leq \ rank(A) \qquad\qquad (2.40)$$

Similarly

$$rank(R) \ \leq \ rank(A^{*}) \qquad\qquad (2.41)$$

We shall now make use of the assumed relation

$$SA^{*}S \ = \ 0 \qquad\qquad (2.42)$$

Postmultiplying equation (2.17) by S, we deduce with the aid of the above equation that

$$A(S-RA^{*}S) \ = \ 0 \qquad\qquad (2.43)$$

Hence

$$\text{rank}(A) \leq n - \text{rank}(S - RA^*S) \qquad (2.44)$$

Letting $M = A^*$ in equation (2.38) we obtain

$$\text{rank}(S - RA^*S) = n - \text{rank}(R) \qquad (2.45)$$

Hence

$$\text{rank}(A) \leq \text{rank}(R) \qquad (2.46)$$

Comparing inequalities (2.40) and (2.46) we conclude that

$$\text{rank}(A) = \text{rank}(R) \qquad (2.47)$$

Note that inequalities (2.40), (2.41) and (2.45) follow from the existence of A^*, whereas inequality (2.46) is a consequence of equation (2.42).

2.4   An Orthogonal Transformation of the Generalized Inverse.

Let v be an orthogonal matrix satisfying

$$vv^T = v^T v = I \qquad (2.48)$$

Let A_o, R_o and S_o be defined by

$$A_o = vAv^T, \qquad (2.49)$$

$$R_o = vRv^T, \qquad (2.50)$$

and

$$S_o = vSv^T. \qquad (2.51)$$

11

It follows from the above that

$$R_o = R_o'$$ (2.52)

$$R_o R_o = R_o$$ (2.53)

and
$$S_o = I - R_o$$ (2.54)

Writing equation (2.17) in the form

(2.55)

We deduce after premultiplying the equation by v and postmutiplying it by v^T, and inserting $v^T v (=I)$ between all matrices that

$$(A_o R_o - S_o) v (A_o^* R) v^T = R_o + A_o S_o ,$$

whence

$$v(A_o^* R) v^T = (A_o R_o - S_o)^{-1} (R_o + A_o S_o)$$ (2.56)

Since the right hand side of the above equation equals A_o^*, R_o by definition, it follows that

$$A_o^*, R_o = v(A_o^* R) v^T$$ (2.57)

2.5   The Incomplete Inverse.

The incomplete inverse is defined as the generalized inverse with respect to R, where R is a diagonal matrix.

It thus follows from equation (2.3) that the elements of R are either zero or one. The elements of S are similarly either

12

zero or one.

The reason for calling the inverse incomplete is as follows. Inverting equation (2.21) we obtain

$$A^{-1} = (R + A^*S)^{-1} (A^*R - S) \qquad (2.58)$$

The inverse of A may thus always be obtained directly from the generalized inverse. However, the total number of computations in obtaining A^{-1} is independent of whether or not incomplete inverses are obtained in intermediate steps.

Let us now assume that we are given matrices A and R, where R is a diagonal matrix, whose elements are either zero or one.

We can then find an orthogonal matrix v satisfying equation (2.48) such that R_o, as given by equation (2.50), may be written in partitioned form as

$$R_o = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} \qquad (2.59)$$

Correspondingly let A_o and S_o (as defined by equations (2.49) and (2.51)) be written as

$$A_o = \begin{bmatrix} G & H \\ K & L \end{bmatrix} \qquad (2.60)$$

and

$$S_o = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \qquad (2.61)$$

13

It follows from the above that

$$(A_o R_o - S_o) = \begin{bmatrix} -I & H \\ 0 & L \end{bmatrix} \tag{2.62}$$

and

$$(R_o + A_o S_o) = \begin{bmatrix} G & 0 \\ K & I \end{bmatrix} \tag{2.63}$$

Hence,

$$(A_o R_o - S_o)^{-1} = \begin{bmatrix} -I & HL^{-1} \\ 0 & L^{-1} \end{bmatrix} \tag{2.64}$$

Since

$$A_o^*, R_o = (A_o R_o - S_o)^{-1}(R_o + A_o S_o),$$

we obtain, when substituting the expressions from equations (2.63) and (2.64) on the right hand side of the above equation

$$A_o^*, R_o = \begin{bmatrix} (HL^{-1}K-G) & HL^{-1} \\ L^{-1}K & L^{-1} \end{bmatrix} \tag{2.65}$$

Since by equations (2.48) and (2.57)

$$A^*, R = v^T (A_o^*, R_o) v, \tag{2.66}$$

it follows from equation (2.65) that

$$A_o^* R = v^T \begin{bmatrix} (H^{-1}LK-G) & HL^{-1} \\ L^{-1}K & L^{-1} \end{bmatrix} v \tag{2.67}$$

where

$$\begin{bmatrix} G & H \\ K & L \end{bmatrix} = vAv^\mathsf{T} \qquad (2.68)$$

It can be seen from the above that if A is appropriately defined then A^* will yield matrices of the form L^{-1}, $L^{-1}x$ and $L^{-1}B$ even if in a jumbled (to the extent that v differs from the identity matrix) form. The application of the incomplete inverse to the solution of matrix equations involving consider parameters will be dealt with in a later section.

2.6 The General Solution of the Matrix Equation Ax = y, Where A is a Symmetric Matrix.

Let us consider the matrix equation

$$Ax = y, \qquad (2.69)$$

where A is a symmetric n×n matrix and x and y are n-vectors. Let us assume that rank(A) = r. Let R be defined as a diagonal matrix, whose elements are either zero or one. We can clearly choose R such that the non-zero columns of AR are linearly independent and the remaining columns of A are linear functions of the columns of AR. In other words, there exists an n×n matrix P such that

$$AS = ARP, \qquad (2.70)$$

and

$$SP = 0, \qquad (2.71)$$

15

where

$$S = I-R.$$

Since the non-zero columns of AR are linearly independent it follows that if there exists a vector u such that

$$ARu = 0, \quad \text{then } Ru = 0. \tag{2.72}$$

We shall now show that AR-S is non-singular. Let us assume that there exists a vector u such that

$$(AR-S)u = 0 \tag{2.73}$$

Premultiplying the above equation by R and S, respectively we deduce that

$$RARu = 0 \tag{2.74}$$

and

$$(SAR-S)u = 0 \tag{2.75}$$

Postmultiplying the transpose of equation (2.70) by Ru we obtain

$$SARu = P^{T}RARu,$$

whence by equation (2.74),

$$SARu = 0 \tag{2.76}$$

Adding equations (2.74) and (2.76) we obtain

$$ARu = 0.$$

It hence follows from equation (2.72) that

$$Ru = 0 \qquad (2.77)$$

Adding equations (2.76) and (2.77) and subtracting equation (2.75) from the result we find that

$$u = 0 \qquad (2.78)$$

We have just shown that $(AR-S)u = 0$, implies that $u = 0$. Consequently $(AR-S)$ is non-singular. It hence follows from equation (2.18) that A^* exists.

Subtracting equation (2.71) from (2.70) we obtain

$$(AR-S)P = AS, \qquad (2.79)$$

whence

$$P = (AR-S)^{-1} AS \qquad (2.80)$$

The right hand side of the above equation equals A^*S, as may be readily verified by postmultiplying equation (2.18) by S. Hence

$$P = A^*S \qquad (2.81)$$

It then follows from equation (2.71) that

$$SA^*S = 0 \qquad (2.82)$$

Hence equation (2.43) applies, i.e.,

$$A(S-RA^*S) = 0 \qquad (2.43)$$

Taking the transpose of the above equation we obtain

$$(S-SA^*R)A = 0 \qquad (2.83)$$

Therefore, in order that equation (2.69) have any solution at all, it is necessary that

$$(S-SA^*R)y = 0 \qquad (2.84)$$

From the above and equation (2.29) it then follows that

$$A(A^*R-S)y = y \qquad (2.85)$$

Comparing equations (2.69) and (2.85) we conclude that

$$x_o = (A^*R-S)y \qquad (2.86)$$

is a solution of equation (2.69). In view of equation (2.43) the general solution of equation (2.86) may be written as

$$x = x_o + (S-RA^*S)a, \qquad (2.87)$$

where the vector a satisfies

$$Ra = 0, \qquad (2.88)$$

but is otherwise arbitrary.

Adding equations (2.84) and (2.86) we obtain the following alternative form for x_o.

$$x_o = RA^*Ry \qquad (2.86)'$$

That x as given by equation (2.87) is the most general solution of equation (2.69) can be seen from the following argument. Since by assumption rank$(A) = r$, the general solution of equation

(2.69) must contain an arbitrary linear combination of (n-r) linearly independent vectors. Since by equation (2.47) rank(R) = rank(A) it follows that rank(R) = r, whence by equation (2.45) rank($S-RA^*S$) = n-r. x therefore is of the required form and is hence the general solution.

Note that we have shown that

(i)   The general solution of equation (2.69) is given by equation (2.87), where x_0 is any solution of equation (2.69),

(ii)   Given that $SA^*S = 0$ it follows that $(A^*R-S)y$ is a solution of equation (2.69)

and   (iii)   There exists an R and a corresonding S, both of which are diagonal matrices, and which satisfy $SA^*S = 0$.

It will later be shown (see Section 3, equations (3.10)', (3.12)', and (3.13)') that if we define the symmetric matrices M and R_M by

$$M = \begin{bmatrix} C & y^T \\ y & A \end{bmatrix} \qquad (2.89)$$

and

$$R_M = \begin{bmatrix} 0 & 0 \\ 0 & R \end{bmatrix} \qquad (2.90)$$

then

$$M^* = \begin{bmatrix} y^T R x_0 - C & x_0^T \\ x_0 & A^* \end{bmatrix}, \qquad (2.91)$$

19

where x_o is given by equation (2.86) or equivalently be equation (2.86)'. Since it follows from equation (2.86)' that

$$Rx_o = x_o \qquad (2.92)$$

equation (2.91) may also be written in the form

$$M^* = \begin{bmatrix} y^\tau x_o - C & x_o^\tau \\ x_o & A^* \end{bmatrix} \qquad (2.93)$$

The significance of the term $(y^\tau x_o - C)$ will be explained later in Section 4 dealing with least squares parameter estimates.

Finally we note that it follows from equations (2.84) and (2.87) that

$$y^\tau x = y^\tau x_o \qquad (2.94)$$

In other words, the product $y^\tau x$ is the same for all solutions x of equation (2.69)

2.6.1   The Minimum Norm Solution

If A in equation (2.69) is singular then S is not equal to zero and equation (2.87) will yield an infinite number of solutions for equation (2.69). Of all possible solutions, the minimum norm solution is the one that minimizes the norm $N_Q(x)$ given by

$$N_Q(x) = x^\tau Q^{-1} x, \qquad (2.95)$$

where Q is a diagonal positive-definite matrix, which defines the

20

norm. Let us denote the value of x that minimizes $N_Q(x)$ by x_Q. Since x is a function of a it follows that the parital derivatives of $N_Q(x)$ with respect to a must vanish. We hence find with the aid of equation (2.87) that

$$(S-RA^*S)^T Q^{-1} [x_0 + (S-RA^*S)a] = 0,$$

i.e.,

$$(S-SA^*R)Q^{-1} [x_0 + (S-RA^*S)a] = 0 \qquad (2.96)$$

Since S, R and Q^{-1} are diagonal matrices, they commute and the above equation may be simplified to

$$(SA^*RQ^{-1}RA^*S + Q^{-1}S)a = -(S-SA^*R)Q^{-1}x_0 \qquad (2.97)$$

Since it follows from equation (2.88) that

$$Sa = a \qquad (2.98)$$

equation (2.97) may be rewritten as

$$(SA^*RQ^{-1}RA^*S + Q^{-1})a = -(S-SA^*R)Q^{-1}x_0 \qquad (2.99)$$

In order to simplify the algebra it is convenient to define

$$H = RA^*S \qquad (2.100)$$

It follows from the definition and the properties of R and S that

$$HH = SH = HR = 0, \qquad (2.101)$$

and

$$RH = HS = H \qquad (2.102)$$

21

Equation (2.99) may now be re-expressed in the form

$$(H^T Q^{-1} H + Q^{-1})a = -(S-H^T)Q^{-1}x_o. \qquad (2.103)$$

$(H^T Q^{-1} H + Q^{-1})$ being the sum of a positive-definite and a semi-positive-definite matrix must be non-singular. Hence

$$a = -(H^T Q^{-1} H + Q^{-1})^{-1} (S-H^T)Q^{-1}x_o \qquad (2.104)$$

From the above and equations (2.87) and (2.100) it follows that

$$x_Q = Nx_o, \qquad (2.105)$$

where

$$N = I -(S-H)(H^T Q^{-1} H + Q^{-1})^{-1} (S-H^T)Q^{-1} \qquad (2.106)$$

An alternative form for N is obtained as follows. By definition of the inverse

$$(H^T Q^{-1} H + Q^{-1})^{-1} (H^T Q^{-1} H + Q^{-1}) = I \qquad (2.107)$$

Postmultiplying the above equation by QH^T we obtain

$$(H^T Q^{-1} H + Q^{-1})^{-1} H^T Q^{-1} (HQH^T + Q) = QH^T, \qquad (2.108)$$

whence

$$B = QH^T (HQH^T + Q)^{-1} \qquad (2.109)$$

where

$$B = (H^T Q^{-1} H + Q^{-1})^{-1} H^T Q^{-1} \qquad (2.110)$$

From equations (2.107) and (2.110) we deduce that

$$(H^T Q^{-1} H + Q^{-1})^{-1} = Q -BHQ \qquad (2.111)$$

22

It follows from equation (2.109) that

$$HB = I - Q(HQH^T + Q)^{-1} \qquad (2.112)$$

Writing equation (2.106) in the form

$$N = I - S(H^T Q^{-1} H + Q^{-1})^{-1} SQ^{-1} + (S-H)(H^T Q^{-1} H + Q^{-1})H^T Q^{-1}$$
$$+ H(H^T Q^{-1} H + Q^{-1})^{-1} SQ^{-1},$$

we deduce with the aid of equations (2.110) and (2.111) that

$$N = I - S(Q-BHQ)SQ^{-1} + (S-H)B + QB^T SQ^{-1} \qquad (2.113)$$

Since S and Q^{-1} commute it follows from the above and equations (2.109) and (2.112) that

$$N = I - S + BH + B - HB + QB^T Q^{-1}$$
$$= -S + QH^T(HQH^T + Q)^{-1}(I+H) + Q(HQH^T + Q)^{-1}$$
$$+ Q(HQH^T + Q)^{-1} H$$

Hence

$$N = -S + (Q + QH^T)(HQH^T + Q)^{-1}(I + H) \qquad (2.114)$$

Note that since it follows from equation (2.86)' that $Sx_o = 0$, whence also $Hx_o = 0$, equations (2.105), (2.106), and (2.114) may be replaced by

$$x_Q = x_o + (S-H)(H^T Q^{-1} H + Q^{-1})^{-1} H^T Q^{-1} x_o \qquad (2.115)$$

and the alternative form

$$x_Q = (Q + QH^T)(HQH^T + Q)^{-1} x_o \qquad (2.116)$$

The choice as to which equation to use may be made by noting that $(H^T Q^{-1} H + Q^{-1})$ is effectively an $(n-r) \times (n-r)$ matrix and $(HQH^T + Q)$ is effectively an $r \times r$ matrix. (Both matrices are $n \times n$ matrices and r is the rank of A).

2.6.2 Relationship of the Minimum Norm Solution to the Generalized Inverse.

It follows from equation (2.110) that $BS = 0$. We therefore deduce from equation (2.112) that

$$(HQH^T + Q)^{-1} S = Q^{-1} S \qquad (2.117)$$

Since $(I+H)(S-H) = S$ it follows from the above and equation (2.114) that

$$N(S-H) = -S + (Q+QH^T)Q^{-1} S,$$

i.e.,

$$N(S-H) = 0 \qquad (2.118)$$

From the above and equation (2.106) we deduce that

$$NN = N \qquad (2.119)$$

Since be equations (2.43) and (2.100)

$$A(S-H) = 0, \qquad (2.120)$$

we similarly deduce from equation (2.106) that

$$AN = A \qquad (2.121)$$

Since by equations (2.84) and (2.100)

$$(S-H^T)y = 0, \qquad (2.122)$$

we also deduce that

$$N^T y = y \qquad (2.123)$$

Let us now define Q_o as the diagonal matrix, whose elements equal the square-root of the corresponding elements of Q, so that

$$Q_o Q_o = Q \qquad (2.124)$$

Also define

$$R_o = Q_o^{-1} N Q_o \qquad (2.125)$$

It can be seen from equation (2.106) that R_o is symmetric, i.e.,

$$R_o^T = R_o \qquad (2.126)$$

It can easily be seen from the definition of R_o and from equation (2.119) that

$$R_o R_o = R_o \qquad (2.127)$$

Defining S_o by

$$S_o = I - R_o, \qquad (2.128)$$

it follows that

$$S_o S_o = S_o \qquad (2.129)$$

and

$$S_o R_o = R_o S_o = 0 \qquad (2.130)$$

Let L be defined by

$$L = R_o Q_o^{-1} RA^*RQ_o^{-1} R_o, \qquad (2.131)$$

and let us consider the expression

$$E = (Q_o A Q_o R_o - S_o)(LR_o - S_o)$$

It follows from equations (2.125) through (2.132) that

$$E = Q_o A Q_o L + S_o \qquad (2.132)$$

$$= Q_o ANRA^*RQ_o^{-1} R_o + S_o$$

From the above and equations (2.121) and (2.23) we deduce that

$$E = Q_o (R + SA^*R)Q_o^{-1} R_o + S_o$$

$$= R_o + S_o + Q_o (H^\mathsf{T} - S)Q_o^{-1} R_o$$

$$= I + \left[R_o Q_o^{-1} (H-S)Q_o \right]^\mathsf{T}$$

$$= I + \left[Q_o^{-1} N(H-S)Q_o \right]^\mathsf{T}$$

$$= I, \text{ by equation (2.118)}$$

It follows from the above and equation (2.132) that

$$(Q_o A Q_o R_o - S_o)^{-1} = LR_o - S_o \qquad (2.133)$$

From equations (2.121) and (2.125) we deduce that

26

$$(Q_o A Q_o) R_o = Q_o A Q_o, \qquad (2.134)$$

whence

$$(Q_o A Q_o) S_o = 0 \qquad (2.135)$$

From the above and equation (2.18) we conclude that the generalized inverse of $(Q_o A Q_o)$ with respect to R_o is given by

$$(Q_o A Q_o)^*_{;R_o} = (L R_o - S_o) R_o,$$

whence by equations (2.127), (2.130), and (2.131)

$$(Q_o A Q_o)^*_{;R_o} = L \qquad (2.136)$$

We shall next show that $x_Q = (Q_o L Q_o) y$.

It follows from equations (2.123) and (2.125) that

$$Q_o^{-1} R_o Q_o y = y \qquad (2.137)$$

From the above and equation (2.131) we deduce that

$$LQ_o y = R_o Q_o^{-1} R A^* R y$$
$$= R_o Q_o^{-1} x_o, \qquad (2.138)$$
$$\text{by equation (2.86)'}$$

From the above and equation (2.125) we find that

$$LQ_o y = Q_o^{-1} N x_o$$
$$= Q_o^{-1} x_Q$$
$$\text{by equation (2.105)}$$

Hence

$$x_Q = (Q_o L Q_o) y, \qquad (2.139)$$

27

where L is given by equation (2.136) as the generalized inverse of $Q_o AQ_o$ with respect to R_o. In other words we have shown that there exists a generalized inverse, which is directly related to the minimum norm solution x_Q.

It is interesting to note that L is the Penrose pseudo-inverse of $(Q_o AQ_o)$. (See e.g., (Lefferts, 1969)).

Since it was shown that E in equation (2.132) equals the identity matrix it follows that

$$(Q_o AQ_o)L = R_o \qquad (2.140)$$

Since R_o is symmetric we deduce that

$$(Q_o AQ_o)L = L(Q_o AQ_o) \qquad (2.141)$$

From equation (2.131) we find that

$$LR_o = L,$$

whence we deduce from equation (2.140) that

$$L(Q_o AQ_o)L = L \qquad (2.142)$$

Similarly we deduce from the transpose of equation (2.134) and from equation (2.140) that

$$(Q_o AQ_o)L(Q_o AQ_o) = Q_o AQ_o \qquad (2.143)$$

Since equations (2.141) through (2.143) are the equations satisfied by the Penrose pseudo-inverse of the symmetric matrix

$(Q_o A Q_o)$ it follows that L is the Penrose pseudo-inverse of $(Q_o A Q_o)$.

2.7 A Relationship Between Two Incomplete Inverses.

Let U and V be defined by

$$U = A_i^* R \quad , \tag{2.144}$$

and

$$V = A_i^* R' \quad , \tag{2.145}$$

where R and R' are diagonal matrices, whose elements are either zero or one. Let us assume that R' has more non-zero elements than R, so that we may write

$$R' = R + K, \tag{2.146}$$

where the elements of K are either zero or one, and

$$RK = 0 \tag{2.147}$$

It follows from equations (2.21) and (2.18), respectively, that

$$A = (UR - S)^{-1} (R + US), \tag{2.148}$$

and

$$V = (AR' - S')^{-1} (R + AS') \tag{2.149}$$

From equation (2.148) we deduce that

$$AR'-S' = (UR-S)^{-1} \left[(R + US)R' -(UR-S)S' \right], \qquad (2.150)$$

and

$$AS' + R' = (UR-S)^{-1} \left[(R + US)S' + (UR-S)R' \right] \qquad (2.151)$$

It follows from the above and equation (2.149) that

$$V = \left[U(SR'-RS') + (RR' + SS') \right]^{-1} \left[U(RR' + SS')-(SR'-RS') \right] \qquad (2.152)$$

Since

$$S' = I-R'$$

we deduce from equation (2.146) that

$$S' = S-K \qquad (2.153)$$

It follows from equation (2.147) that

$$SK = K \qquad (2.154)$$

We hence deduce that

$$RR' + SS' = R + S-K$$

$$= I-K,$$

and

$$SR'-RS' = K,$$

whence

$$V = \left[UK + (I-K) \right]^{-1} \left[U(I-K)-K \right]. \qquad (2.155)$$

30

The above equation gives the desired relationship between the two incomplete inverses defined by equations (2.144) and (2.145).

Note that if $K = S$, so that $R' = I$ and hence $V = A^{-1}$, then

$$V = (US + R)^{-1} (UR-S) \qquad (2.155)'$$

The above equation could also have been obtained directly by inverting both sides of equation (2.148).

3.0 AN ALGORITHM FOR COMPUTING

THE INCOMPLETE INVERSE

Let M, M^*, R, and S be partitioned according to

$$M = \begin{bmatrix} C & Y^T \\ Y & A \end{bmatrix}, \qquad M^* = \begin{bmatrix} D & X^T \\ X & B \end{bmatrix}, \tag{3.1}$$

$$R = \begin{bmatrix} R_c & 0 \\ 0 & R_a \end{bmatrix}, \qquad S = \begin{bmatrix} S_c & 0 \\ 0 & S_a \end{bmatrix}, \tag{3.2}$$

where M and M^* are symmetric matrices and R and S are diagonal matrices, whose elements equal either one or zero, and $R + S = I$.

Since by definition the incomplete inverse M^* must satisfy equation (2.17) it follows that

$$\begin{bmatrix} C & Y^T \\ Y & A \end{bmatrix} \begin{bmatrix} R_c & 0 \\ 0 & R_a \end{bmatrix} \begin{bmatrix} D & X^T \\ X & B \end{bmatrix}$$

$$= \begin{bmatrix} R_c & 0 \\ 0 & R_a \end{bmatrix} + \begin{bmatrix} C & Y^T \\ Y & A \end{bmatrix} \begin{bmatrix} S_c & 0 \\ 0 & S_a \end{bmatrix} + \begin{bmatrix} S_c & 0 \\ 0 & S_a \end{bmatrix} \begin{bmatrix} D & X^T \\ X & \cdot B \end{bmatrix} \tag{3.3}$$

Equating the partitions we hence obtain

$$CR_c D + Y^T R_a X = R_c + CS_c + S_c D, \tag{3.4}$$

$$YR_c D + AR_a X = YS_c + S_a X, \tag{3.5}$$

32

$$YR_c X^T + AR_a B = R_a + AS_a + S_a B \qquad (3.6)$$

Note that since matrices M, M^*, R, and S are symmetric, the equation for the fourth partition of equation (3.3) is redundant.

Since by equation (2.25)

$$(AR_a - S_a)^{-1} = A^* R_a - S_a , \qquad (3.7)$$

we deduce from equation (3.5) that

$$X = (A^* R_a - S_a) Y (S_c - R_c D) \qquad (3.8)$$

In order to simplify the algebra let us write

$$F = -(A^* R_a - S_a) Y, \qquad (3.9)$$

so that

$$X = F(R_c D - S_c) \qquad (3.10)$$

Substituting the expression for X from equation (3.10) on the left hand side of equation (3.4) we obtain

$$(C + Y^T R_a F)(R_c D - S_c) = R_c + S_c D \qquad (3.11)$$

Comparing equations (3.11) and (2.20) we see that

$$C + Y^T R_a F = D^*$$

We hence deduce with the aid of equation (2.8) that

$$D = (C + Y^T R_a F)^* \qquad (3.12)$$

33

It follows from equation (3.6) that

$$(AR_a - S_a)B = R_a + AS_a - YR_c X^\Gamma,$$

whence by equations (2.18) and (3.7)

$$B = A^* - (A^* R_a - S_a) Y R_c X^T$$

From the above and equation (3.9) we finally obtain

$$B = A^* + FR_c X^T \qquad (3.13)$$

Note that if C, D, R_c and S_c are scalars, X and Y vectors and

$$R_c = 0 \quad \text{and} \quad S_c \cdot = 1$$

then equations (3.9), (3.10), (3.12), and (3.13) reduce to

$$X = (A^* R_a - S_a) Y \quad , \qquad (3.10)'$$

$$D = Y^T R_a X - C \quad , \qquad (3.12)'$$

and

$$B = A^* \qquad (3.13)'$$

3.1   The Computation of the Incomplete Inverse

Let us define the symmetric matrix $\widetilde{M}$ by

$$\widetilde{M} = \begin{bmatrix} M' & F^T \\ F & A^* \end{bmatrix} \quad , \qquad (3.14)$$

where

$$M' = C + Y^T R_a F \qquad (3.15)$$

34

It follows from the above and equations (3.10), (3.12) and (3.13) that the partitions of

$$M^* = \begin{bmatrix} D & X^\mathsf{T} \\ X & B \end{bmatrix},$$

are given by,

$$D = (M')^* \tag{3.16}$$

$$X^\mathsf{T} = (DR_c - S_c)F^\mathsf{T} \tag{3.17}$$

and

$$B = A^* + FR_c X^\mathsf{T} \tag{3.13}$$

It can thus be seen that M^* may be computed directly from $\widetilde{M}$. $(M')^*$ may be computed by partitioning M' similarly to the way M was partitioned. This process may be incorporated in a procedure as follows.

If M is a symmetric $n \times n$ matrix, define M_n by

$$M_n = M \tag{3.18}$$

Then for $r = n$, $n-1$, $n-2,\ldots 2$ partition each symmetric $r \times r$ matrix M_r according to

$$M_r = \begin{bmatrix} C_r & Y_r^\mathsf{T} \\ Y_r & A_r \end{bmatrix} \tag{3.19}$$

Also define

$$R^n = R, \quad S^n = S \tag{3.20}$$

and

$$R^r = \begin{bmatrix} R_c^r & 0 \\ 0 & R_a^r \end{bmatrix}, \quad S^r = \begin{bmatrix} S_c^r & 0 \\ 0 & S_a^r \end{bmatrix} \tag{3.21}$$

35

Corresponding to equations (3.9) and (3.15) define

$$F_r^T = -Y_r^T (R_a^r A_r^* - S_a^r) \qquad (3.22)$$

$$M_{r-1} = C_r + Y_r^T R_a^r F_r, \qquad (3.23)$$

Also define

$$R^{r-1} = R_c^r \quad \text{and} \quad S^{r-1} = S_c^r \qquad (3.24)$$

and A_1 by

$$A_1 = M_1 , \qquad (3.25)$$

whence

$$M_1^* = A_1^* \qquad (3.26)$$

Equations (3.18) through (3.26) are referred to as the _elimination_ equations. The inversion of M is completed through the use of _back-substitution_ equations (3.27) through (3.29).

For

$$r = 2, 3, \ldots, n$$

$$M_r^* = \begin{bmatrix} M_{r-1}^* & X_r^T \\ X_r & B_r \end{bmatrix} , \qquad (3.27)$$

where corresponding to equations (3.17) and (3.13)

$$X_r^T = (M_{r-1}^* R_c^r - S_c^r) F_r^T , \qquad (3.28)$$

and

$$B_r = A_r^* + F_r R_c^r X_r^T \qquad (3.29)$$

36

Finally, since $M_n = M$ it follows that

$$M^* = M_n^*$$ (3.30)

<u>NOTE 1</u>

If M^* is calculated on a digital computer, then since all matrices M_r are symmetric, only the upper triangular portion of each matrix need be stored in computer memory. Also, storage space may be shared by the following matrices:

$$(Y_r^T, F_r^T, X_r^T), \quad (A_r, A_r^*, B_r) \text{ and } (C_r, M_{r-1}, M_{r-1}^*).$$

In view of equation (3.28) n temporary storage locations must however be allocated. The total number of storage locations required to compute M^* from M is thus $n(n+3)/2$.

<u>NOTE 2</u>

Since R_a^r is a 1×1 matrix and the elements of R are either zero or one, equations (3.22) and (3.23) may be simplified for the two cases $R_a^r = 1$ and $R_a^r = 0$. Similarly A_r^* as given by equation (2.18) may be simplified. Corresponding to the two cases we obtain:

If $R_a^r = 1$ then

$$A^* = 1/A_r$$ (3.31W)

$$F_r^T = -Y_r^T A_r^*$$ (3.22W)

and

$$M_{r-1} = C_r + Y_r^T F_r \quad ,$$ (3.23W)

and if $R_a^r = 0$ then

$$A_r^* = -A_r,$$ (3.31V)

37

$$F_r^T \;=\; Y_r^T \;, \hspace{4cm} (3.22V)$$

and
$$M_{r-1} \;=\; C_r \hspace{4cm} (3.23V)$$

NOTE 3

Although A_r and B_r in equations (3.19) and (3.27) are assumed to be 1×1 matrices, equations (3.18) through (3.30) remain valid if e.g., they are $k \times k$ matrices. This fact may be utilized, when the matrix M is so large, that only part of it may be stored in main computer memory.

NOTE 4

If, e.g.,

$$M \;=\; \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad R \;=\; \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

then $M^* = M^{-1} = M$. However, the above inversion scheme will obviously break down. It will next be shown that the inversion scheme can always be applied when the matrix (RMR+S) is positive-definite.

3.2 A Sufficient Condition for the Applicability of the Inversion Scheme.

It will now be shown that if the matrix (RMR+S) is positive-definite, then the inversion scheme is always applicable. First we shall show that M^* exists if and only if (RMR+S) is non-singular.

It may readily be verified that

$$(R-S + SMR)(R-S + SMR) = I,$$

so that (R-S + SMR) is non-singular. Since

$$(MR-S) = (RMR + S)(R-S + SMR),$$

it follows that (MR-S) is non-singular if (RMR+S) is. We hence, deduce with the aid of equation (2.18) that M^* exists if (RMR+S) is non-singular. Since it follows from equation (2.28) that (RMR+S) is non-singular if M^* exists, we conclude that M^* exists, if and only if (RMR+S) is non-singular.

To show that the inversion scheme **always can** be applied when (RMR+S) is positive-definite, it is, in view of the recursive nature of the scheme, sufficient to show that, when M is partitioned according to equation (3.1), both $(R_a AR_a + S_a)$ and $(R_c M'R_c + S_c)$ $\Big[$where M' is given by equation (3.15)$\Big]$ are positive-definite.

It follows from equations (3.1) and (3.2) that

$$RMR+S = \begin{bmatrix} R_c CR_c + S_c & R_c Y^T R_a \\ R_a YR_c & R_a AR_a + S_a \end{bmatrix} \tag{3.32}$$

Since RMR+S is positive-definite, it follows that

$$R_a AR_a + S_a$$

and

$$G = R_c CR_c + S_c - R_c Y^T R_a (R_a AR_a + S_a)^{-1} R_a YR_c \tag{3.33}$$

are positive-definite. $\Big\{$That G must be positive-definite can be seen from the following. Let the vector v be defined by

$v^T = [u^T, -u^T R_c Y^T R (R_a A R_a + S_a)^{-1}]$, whence $v^T (RMR + S)v = u^T Gu$.
The result thus follows} .

An alternative expression for G is obtained with the aid
of equation (2.28) as

$$G = R_c C R_c + S_c - R_c Y^T R_a (A^* R_a + S_a) R_a Y R_c ,$$

i.e.,

$$G = R_c [C - Y^T R_a A^* R_a Y] R_c + S_c$$

$$= R_c [C - Y^T R_a (A^* R_a - S_a) Y] R_c + S_c$$

$$= R_c [C + Y^T R_a F] R_c + S_c ,$$

by equation (3.9)

It follows from the above and equation (3.15) that
$G = R_c M' R_c + S_c$. We have hence shown that both $R_a A R_a + S_a$ and
$R_c M' R_c + S_c$ are positive-definite and that therefore, the inversion
procedure always can be applied when RMR+S is positive-definite.

3.3   Computation of the Determinant of the Normalized Inverse.

In this sub-section we shall compute the determinant of

$$\bar{M} = \Lambda^{-1} (RMR + S)^{-1} \Lambda^{-1} , \tag{3.34}$$

where Λ is the diagonal matrix, whose non-zero elements equal the
square-root of the corresponding elements of $(RMR+S)^{-1}$. It follows
from equation (3.34) that

$$\det(\bar{M}) = \det(RMR + S)^{-1} / (\det \Lambda)^2 ,$$

and hence that

$$\det(\overline{M}) \;=\; 1/\det(RMR + S)\det(\wedge^2) \tag{3.35}$$

It follows from the definition of $\wedge$, that $\wedge^2$ is the diagonal matrix, whose non-zero elements equal the corresponding diagonal elements of $(RMR+S)^{-1}$. Since by equation (2.28), $(RMR+S)^{-1} = (RM^*R+S)$, it follows that

$$\det(\overline{M}) \;=\; \det(RM^*R + S)/\det(\wedge^2), \tag{3.36}$$

where $\det(\wedge^2)$ equals the product of the diagonal elements of $(RM^*R + S)$.

It follows from equations (3.1) and (3.2) that

$$RM^*R + S \;=\; \begin{bmatrix} R_c\,DR_c + S_c & R_c\,X^T R_a \\ R_a\,X\,R_c & R_a\,BR_a + S_a \end{bmatrix}, \tag{3.37}$$

whence by equation (3.13)

$$RM^*R + S \;=\; \begin{bmatrix} R_c\,DR_c + S_c & \cdot \quad 0 \\ R_a\,X\,R_c & R_a\,A^*R_a + S_a \end{bmatrix}$$

$$+ \;\; \begin{bmatrix} 0 & R_c\,X^T R_a \\ 0 & R_a\,FR_c\,X^T R_a \end{bmatrix} \tag{3.38}$$

Since $X = F(R_c\,D - S_c)$ it follows that

$$R_c\,X^T R_a \;=\; R_c\,DR_c\,F^T R_a,$$

$$R_a FR_c X^I R_a = R_a FR_c DR_c F^T R_a,$$

and

$$R_a XR_c = R_a FR_c DR_c,$$

whence

$$\begin{bmatrix} R_c X^T R_a \\ R_a FR_c X^T R_a \end{bmatrix} = \begin{bmatrix} R_c DR_c + S_c \\ R_a XR_c \end{bmatrix} R_c F^T R_a \qquad (3.39)$$

Since the value of a determinant is unaffected by the addition to any one column of a linear combination of the remaining columns, we deduce from equations (3.38) and (3.39) that

$$\det(RM^*R + S) = \det \begin{bmatrix} R_c DR_c + S_c & 0 \\ R_a XR_c & R_a A^*R_a + S_a \end{bmatrix}.$$

Hence

$$\det(RM^*R + S) = \det(R_c DR_c + S_c)\det(R_a A^*R_a + S_a).$$

From the above and equation (3.16) we deduce that

$$\det(RM^*R + S) = \det(R_c (M')^*R_c + S_c)\det(R_a A^*R_a + S_a) \qquad (3.40)$$

Applying equation (3.40) to matrix M_r as defined by equation (3.19) we obtain with the aid of equations (3.21) through (3.24)

$$\det(R^r M_r^* R^r + S^r) = \det(R^{r-1} M_{r-1}^* R^{r-1} + S^{r-1})$$
$$\times \det(R_a^r A_r^* R_a^r + S_a^r) \qquad (3.41)$$

42

Since $M_n = M$ we deduce from the above and equation (3.26) that

$$\det(RM^*R + S) = \prod_{r=1}^{n} \det(R'_a A^*_r R'_a + S'_a), \qquad (3.42)$$

where

$$R^1_a = R^1 \quad \text{and} \quad S^1_a = S^1$$

Since R'_a, S'_a and A^*_r are 1×1 matrices, the above may be simplified to

$$\det(RM^*R + S) = \prod_{r=1}^{n} (R'_a A^*_r R'_a + S'_a) \qquad (3.43)$$

It follows from equations (3.26) and (3.27) that $\det(\wedge^2)$, which equals the product of the diagonal elements of $(RM^*R + S)$, is given by

$$\det(\wedge^2) = (R^1_a A^*_1 R^1_a + S^1_a) \prod_{r=2}^{n} (R'_a B_r R'_a + S'_a) \qquad (3.44)$$

It hence follows from equations (3.36), (3.43), and (3.44) and, since R'_a equals either one or zero, that

$$\det(\overline{M}) = \prod_{r=2}^{n} Q_r, \qquad (3.45)$$

where

$$Q_r = \begin{cases} A^*_r / B_r & \text{if } R^1_a = 1 \\ 1 & \text{if } R^1_a = 0 \end{cases} \qquad (3.46)$$

Note that, since for large matrices $\det(\overline{M})$ may be extremely small, $\det(\overline{M})$ is best computed logarithmically.

43

3.4   The Determination of R when M is a Semi-positive-definite Matrix.

In this subsection it will be shown how a diagonal matrix R is determined such that the non-zero columns of MR are linearly independent, but the remaining columns of M are linear functions of the columns of MR.

It was previously shown (see Section 2.6) that if R is chosen with the above property then M^* exists. It then follows from equations (3.18) through (3.30) that A_r^* must exist. We therefore conclude from equations (3.31W) and (3.31V) that if $A_r = 0$ (or in practice less than some small constant multiplied by the corresponding element of M) then $R_a^r = 0$, but otherwise $R_a^r = 1$. (Note that, as can be seen from equations (3.22) and (3.23), R_a^r is not used in any computations till A_r^* has been determined). Since the r-th diagonal element of R equals R_a^r (by definition of R_a^r), it follows that the determination of all matrices R_a^r in effect determines R.

3.5   The Computation of the Incomplete Inverse in the Presence of Patterned Zeroes.

Let us consider the symmetric matrix

$$M = \begin{bmatrix} C & Y_1^\top & Y_2^\top \\ Y_1 & A_1 & 0 \\ Y_2 & 0 & A_2 \end{bmatrix} \qquad (3.47)$$

and its incomplete inverse

44

$$M^* = \begin{bmatrix} D & X_1^\intercal & X_2^\intercal \\ X_1 & B_1 & L_{12}^\intercal \\ X_2 & L_{12} & B_2 \end{bmatrix} \qquad (3.48)$$

Examination of elimination equations (3.18) through (3.26) show that A_2 and Y_2 do not modify A_1 and Y_1. It can also be seen from the back-substitution equations (3.27) through (3.30) that B_1 and X_1 do not contribute anything to the computation of B_2 and X_2. It follows that the partitions of M^* (except for L_{12}) may equivalently be computed by the following scheme:

(i) Consider $M_1 = \begin{bmatrix} C & Y_2^\intercal \\ Y_2 & A_2 \end{bmatrix}$ $\qquad (3.49)$

and apply the elimination equations through A, obtaining

$$M_2 = \begin{bmatrix} \overline{C} & \overline{Y}_2^\intercal \\ \overline{Y}_2 & \overline{A}_2 \end{bmatrix} \qquad (3.50)$$

(ii) Save $\overline{Y}_2$, $\overline{A}_2$ and

(iii) form $M_3 = \begin{bmatrix} \overline{C} & Y_1^\intercal \\ Y_1 & A_1 \end{bmatrix}$ $\qquad (3.51)$

(iv) Obtain incomplete inverse of M_3

$$M_3^* = \begin{bmatrix} D & X_1^\intercal \\ X_1 & B_1 \end{bmatrix} \qquad (3.52)$$

(v) Save X_1, B_1 and form

$$M_4 = \begin{bmatrix} D & \overline{Y}_2^T \\ \overline{Y}_2 & \overline{A}_2 \end{bmatrix} \tag{3.53}$$

(vi)   Apply elimination equations through $\overline{A}_2$ to obtain

$$M_5 = \begin{bmatrix} D & X_2^T \\ X_2 & B_2 \end{bmatrix} \tag{3.54}$$

The scheme as outlined above may also be applied when M is of the form

$$M = \begin{bmatrix} C & Y_1^T & Y_2^T & Y_3^T & \cdots & Y^T \\ Y_1 & A_1 & 0 & 0 & \cdots & 0 \\ Y_2 & 0 & A_2 & 0 & \cdots & 0 \\ Y_3 & 0 & 0 & A_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ Y & 0 & 0 & 0 & \cdots & A \end{bmatrix} \tag{3.55}$$

Furthermore, the scheme may, in a fairly obvious way, be extended for the case of each A_i (for $i = 1, 2, \ldots, n$) itself being of the form (3.55).

Note that the scheme just outlined is from a programming point of view far simpler (particularly when the matrix is stored in upper triangular form) than the conventional way of computing the true inverse (which is a special case of the incomplete inverse) through the scheme:

46

(i) Compute A_2^{-1} , $A_2^{-1} Y_2$ and $\bar{C} = C - Y_2 A_2^{-1} Y_2^T$,   (3.56)

(ii) Compute A_1^{-1} , $A_1^{-1} Y_1$ and $D = (\bar{C} - Y_1^T A_1^{-1} Y_1)^{-1}$ (3.57)

(iii) Compute $X_1 = -A_1^{-1} Y_1 D$ and $B_1 = A_1^{-1} - X_1 (A_1^{-1} Y_1)^T$ (3.58)

(iv) Compute $X_2 = -A_2^{-1} Y_2 D$ and $B_2 = A_2^{-1} - X_2 (A_2^{-1} Y_2)^T$ (3.59)

When each A_i itself is of the (3.55), the attractiveness of using the scheme based on equations (3.49) through (3.54) becomes even more apparent.

4.0 THE APPLICATION OF THE GENERALIZED INVERSE TO THE SOLUTION OF THE LINEAR LEAST SQUARES PROBLEM

Let us consider the linear matrix equation

$$u = A_o x + \varepsilon, \tag{4.1}$$

where

the n-vector u represents a set of n observations,

the m-vector x represents a set of m unknown parameters,

the n-vector ε represents the measurement noise of each observation, and

the n m matrix A_o represents the mathematical modelling of the observations.

The noise vector ε is unknown. We do, however, assume that its expected value (denoted by the operator E) vanishes, i.e.,

$$E(\varepsilon) = 0 \tag{4.2}$$

It is also assumed that the covariance of the noise is known. Denoting this covariance by W^{-1}, we thus have

$$E(\varepsilon \varepsilon^T) = W^{-1}. \tag{4.3}$$

48

4.1   The Solution

It can be shown that if the measurement noise is random Gaussian then the most likely solution of equation (4.1) is obtained by minimizing the expression

$$C(x) = (A_o x - u)^T W(A_o x - u) \qquad (4.4)$$

If $\hat{x}$ is the value of x, which minimizes $C(x)$, then it follows that the partial derivations of $C(x)$ with respect to x vanish when $x = \hat{x}$ and

$$A_o^T W(A_o \hat{x} - u) = 0 \qquad (4.5)$$

Defining A and y, respectively, by

$$A = A_o^T W A_o, \qquad (4.6)$$

and

$$y = A_o^T W u, \qquad (4.7)$$

it follows that

$$A\hat{x} = y, \qquad (4.8)$$

and hence that

$$\hat{x} = A^{-1} y \qquad (4.9)$$

$\hat{x}$ is known as the weighted least squares estimate. If the measurement noise is random Gaussian then it is also the maximum likelihood estimate.

4.2 The Error in the Solution.

It follows from equation (4.1), (4.6), and (4.7) that

$$y = Ax + A_o^T W \varepsilon, \qquad (4.10)$$

where x represents the true parameter vector. From the above and equation (4.9) we obtain

$$\hat{x} - x = A^{-1}(A_o^T W \varepsilon) \qquad (4.11)$$

We hence deduce from equation (4.2) that

$$E(\hat{x}) = x \qquad (4.12)$$

Since $E(\varepsilon \varepsilon^T) = W^{-1}$ it follows from equation (4.6) that

$$E\left[(A_o^T W \varepsilon)(A_o^T W \varepsilon)^T\right] = A.$$

We hence deduce from equation (4.12) that

$$E\left[(\hat{x} - x)(\hat{x} - x)^T\right] = A^{-1},$$

or

$$cov(\hat{x}) = A^{-1} \qquad (4.12)$$

4.3 The Value of $C(\hat{x})$.

It follows from equations (4.4) and (4.5) that

$$C(\hat{x}) = u^T W u - u^T W A_o \hat{x} \qquad (4.14)$$

50

Defining C_o by

$$C_o = u^T W u, \tag{4.15}$$

it follows from equations (4.7), (4.9), and (4.14) that

$$C(\hat{x}) = C_o - y^T A^{-1} y \tag{4.16}$$

4.4 The Computation of $\hat{x}$, $cov(\hat{x})$, and $C(\hat{x})$ Using the Incomplete Inverse.

Let the symmetric matrix M be defined by

$$M = \begin{bmatrix} C_o & y^T \\ y & A \end{bmatrix}, \tag{4.17}$$

and R (corresponding to the same partitioning) by

$$R = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} \tag{4.18}$$

It then follows from equations (2.59), (2.60), and (2.65) that

$$M^* = \begin{bmatrix} (y^T A^{-1} y - C_o) & y^T A^{-1} \\ A^{-1} y & A^{-1} \end{bmatrix} \tag{4.19}$$

From the above and equations (4.9), (4.13), and (4.16) we

51

deduce that

$$M^* = \begin{bmatrix} -C(\hat{x}) & \hat{x}^I \\ \hat{x} & \mathrm{cov}(\hat{x}) \end{bmatrix} \qquad (4.2)$$

Note that $-C(\hat{x})$, which is the upper left diagonal element of M^* is obtained at the conclusion of the application of the elimination algorithms.

4.5   The Solution When Some of the Parameter Values Already Are Known.

Suppose A_0 and x in equation (4.1) are partitioned according to

$$A_0 = \begin{bmatrix} A_1 , & A_2 \end{bmatrix} \qquad (4.21)$$

and

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad (4.22)$$

so that equation (4.1) may be written in the form

$$u = A_1 x_1 + A_2 x_2 + \varepsilon \qquad (4.23)$$

Suppose further that we already have an estimate $\hat{x}_2$ for x_2 and that we do not wish to improve on that estimate. To reflect this, let us write equation (4.23) as

$$u' = A_1 x_1 + \varepsilon' , \qquad (4.24)$$

52

where

$$u' = u - A_2 \hat{x}_2 , \qquad (4.25)$$

and

$$\varepsilon' = \varepsilon - A_2 (\hat{x}_2 - x_2), \qquad (4.26)$$

The components of x_2 are generally referred to as consider parameters.

Corresponding to equations (4.6) through (4.9) the weighted least squares solution of equation (4.24) is given by

$$\hat{x}_1 = (A_1^T W A_1)^{-1} (A_1^T W u') \qquad (4.27)$$

Corresponding to equation (4.11) we obtain for the error in the estimate

$$\hat{x}_1 - x_1 = (A_1^T W A_1)^{-1} (A_1^T W \varepsilon') \qquad (4.28)$$

From the above and equation (4.26) we obtain

$$\hat{x}_1 - x_1 = (A_1^T W A_1)^{-1} \left[A_1^T W \varepsilon - A_1^T W A_2 (\hat{x}_2 - x_2) \right] \qquad (4.29)$$

We do not know the correct value of x_2. If we did, then obviously we would have used that value for our estimate $\hat{x}_2$. We do, however, assume that we know $\hat{x}_2$ to some known degree of accuracy. Specifically, we assume that

$$E(\hat{x}_2) = x_2 , \qquad (4.30)$$

53

and

$$E\left[(\hat{x}_2 - x_2)(\hat{x}_2 - x_2) \right] = V_2 V_2 , \tag{4.31}$$

where V_2 is a known diagonal matrix.

It follows from equations (4.29), (4.2), and (4.30) that

$$E(\hat{x}_1) = x_1 . \tag{4.32}$$

The sensitivity matrix Z is defined by

$$Z = -(A_1^T W A_1)^{-1} (A_1^T W A_2) \tag{4.33}$$

Equation (4.29) may then be expressed as

$$\hat{x}_1 - x = (A_1^T W A_1)^{-1} (A_1^T W \epsilon) + Z(\hat{x}_2 - x_2) \tag{4.34}$$

Since clearly $E(\epsilon x_2^T) = 0$ we deduce from equations (4.34), (4.3), and (4.31) that $\text{cov}(\hat{x}_1)$, which is defined by

$$\text{cov}(\hat{x}_1) = E\left[(\hat{x}_1 - x_1)(\hat{x}_1 - x_1)^T \right] , \tag{4.36}$$

is given by

$$\text{cov}(\hat{x}_1) = (A_1^T W A_1)^{-1} + Z V_2 V_2 Z^T \tag{4.37}$$

The Alias matrix L is defined by

$$L = Z V_2$$

Hence equation (4.36) may also be written as

54

$$\text{cov}(\hat{x}_1) = (A_1^\mathsf{T} W A_1)^{-1} + LL^\mathsf{T} \tag{4.37}$$

Let us now consider the more general form of equation (4.22) that generally occurs in practice.

As before let us assume that we already have an estimate for some of the components of x, and also that we know the standard deviation of the error in each estimate.

Let S be defined as the diagonal matrix, whose diagonal elements equal either one or zero, such that an element equals one if the corresponding parameter is to be estimated but otherwise equals zero.

If futher R is defined by

$$R = I - S, \tag{4.38}$$

then we obtain corresponding to equation (4.23)

$$u = A_0 Rx + A_0 Sx + \varepsilon, \tag{4.39}$$

and corresponding to equations (4.24) through (4.26)

$$\bar{u} = A_0 R(x - x_A) + \bar{\varepsilon} \tag{4.40}$$

where

$$\bar{u} = u - A_0 x_A \tag{4.41}$$

and

$$\bar{\varepsilon} = \varepsilon + A_0 S(x - x_A), \tag{4.42}$$

55

where x_A is the initial or a priori estimate of x.

Minimizing

$$C(x) = \left[\bar{u} - A_o R(x-x_A) \right]^\mathsf{T} W \left[\bar{u} - A_o R(x-x_A) \right] \quad , \qquad (4.43)$$

with respect to x, we obtain, when as before denoting the corresponding value of x by $\hat{x}$.

$$RA_o^\mathsf{T} W \left[\bar{u} - A_o R(\hat{x}-x_A) \right] = 0 \qquad (4.44)$$

It follows from the above and equation (4.6) that

$$R \left[\bar{y} - AR(\hat{x}-x_A) \right] = 0 \quad , \qquad (4.45)$$

where

$$\bar{y} = A_o^\mathsf{T} W\bar{u} \qquad (4.46)$$

Since premultiplying equation (2.19) by R and then postmultiplying the resulting equation by R, yields,

$$RA^*RA = R + RA^*S. \qquad (4.47)$$

and

$$RA^*RAR = R \qquad (4.48)$$

we deduce after premultiplying equation (4.45) by RA^* that

$$R(\hat{x}-x_A) = RA^*R\bar{y} \qquad (4.49)$$

Since we already know $S\hat{x} = (Sx_A)$, equation (4.49) yields

56

the required least squares solution of equation (4.39).

Let us next investigate the error in the estimate $R\hat{x}$. It follows from equations (4.46), (4.40), and (4.42) that

$$\bar{y} = A_o^\mathsf{T} W \left[A_o (x-x_A) + \epsilon \right] ,$$

whence by equation (4.6),

$$\bar{y} = A(x-x_A) + A_o^\mathsf{T} W\epsilon \tag{4.50}$$

From the above and equations (4.47) through (4.49) we deduce that

$$R(\hat{x}-x_A) = (R + RA^*S)(x-x_A) + RA^*R\, A_o^\mathsf{T} W\epsilon ,$$

and hence that

$$R(\hat{x}-x) = RA^*R\, A_o^\mathsf{T} W\epsilon - RA^*S(x_A-x) \tag{4.51}$$

As before let us assume that

$$E(Sx_A) = Sx , \tag{4.52}$$

and

$$E\left[S(x_A-x)(x_A-x)^\mathsf{T} S \right] = SVVS , \tag{4.53}$$

where SV is a known diagonal matrix.

It then follows that

$$E(R\hat{x}) = Rx \quad , \tag{4.54}$$

and

$$\text{cov}(R\hat{x}) = (RA^*R)A(RA^*R) + (RA^*S)SVVS(SA^*R)$$

The last equation may with the aid of equation (4.48) be written as

$$\text{cov}(R\hat{x}) = RA^*R + LL^\top \tag{4.55}$$

where the Alias matrix L is given by

$$L = -RA^*SV \tag{4.56}$$

Finally we deduce from equations (4.43), (4.44), and (4.46) that

$$C(\hat{x}) = \bar{u}^\top W\bar{u} - \bar{y}^\top R(\hat{x}-x_A) \tag{4.57}$$

4.5.1 . The Computation of the Solution, The Alias and Covariance Matrices Using the Incomplete Inverse.

Let the symmetric matrix M be defined by

$$M = \begin{bmatrix} C_0 & \bar{y}^\top \\ \bar{y} & A \end{bmatrix} \tag{4.58}$$

where·

$$C_0 = \bar{u}^\top W\bar{u}. \tag{4.59}$$

Let $\bar{R}$ and $\bar{S}$, when partitioned similarly to M in equation (4.58) be defined by

$$\bar{R} = \begin{bmatrix} 0 & 0 \\ 0 & R \end{bmatrix}, \quad \text{and} \quad \bar{S} = \begin{bmatrix} 1 & 0 \\ 0 & S \end{bmatrix} \tag{4.60}$$

Comparing the above with equations (3.1), (3.2), (3.9), (3.10), (3.12), and (3.13) we deduce that

$$M^* = \begin{bmatrix} D & \bar{x}^{\scriptscriptstyle T} \\ \bar{x} & B \end{bmatrix}, \tag{4.61}$$

where

$$\bar{x} = (A^* R - S)\bar{y} \tag{4.62}$$

$$D = -C_o + \bar{y}^{\scriptscriptstyle T} R\bar{x} \tag{4.63}$$

and

$$B = A^* \tag{4.64}$$

Premultiplying equation (4.62) by R we deduce, with the aid of equation (449) that

$$R(\hat{x} - x_A) = R\bar{x} \ , \tag{4.65}$$

and hence from equations (4.63) and (4.57) that

$$D = -C(\hat{x}) \tag{4.66}$$

It follows from the above that

$$M^* = \begin{bmatrix} -C(\hat{x}) & \bar{x}^\mathsf{T} \\ \bar{x} & A^* \end{bmatrix} \tag{4.67}$$

The Alias matrix L and the covariance matrix $\mathrm{cov}(R\hat{x})$ may be computed from equations (4.56) and (4.55), respectively. Note that since R and S are diagonal matrices, whose elements equal either zero or one, the elements of RA^*R and RA^*S are also elements of A^*. Also note that it can easily be shown that the full covariance of $\hat{x}$ is given by

$$\mathrm{cov}(\hat{x}) = RA^*R + (SV+L)(SV+L)^\mathsf{T} \tag{4.68}$$

4.5.2   The Change in the Solution When the Number of Parameters Being Estimated is Increased.

In this subsection we shall investigate how the least squares solution of matrix equation (4.39) is changed, when one of the consider parameters is included in the set of parameters being estimated.

Specifically, we shall consider how the solution is changed when S, as defined in the previous subsection, is changed to S^1 such that $S-S^1$ is a diagonal matrix with only one non-zero element, which equals one.

Corresponding to equation (4.38) let us define

$$R^1 = I - S^1 \tag{4.69}$$

and corresponding to equation(4.60)

60

$$\overline{R}^1 = \begin{bmatrix} 0 & 0 \\ 0 & R^1 \end{bmatrix} \text{ and } \overline{S}^1 = \begin{bmatrix} 1 & 0 \\ 0 & S^1 \end{bmatrix} \qquad (4.70)$$

It then follows from the definition of S^1 that

$$\overline{R}^1 = \overline{R}+K \quad , \qquad (4.71)$$

where K is a diagonal matrix with only one non-zero element, which equals one. Let us assume that this is the (k+1)-th diagonal element, so that it is the k-th parameter that is now included in the set of parameters being estimated.

Since $\overline{u}$ and, hence C_o and $\overline{y}$ are independent of S, it follows that M as given by equation (4.58) is independent of S. However, M^* as given by equation (4.67) is the incomplete inverse of M with respect to $\overline{R}$, so that M^* will change as $\overline{R}$ changes from $\overline{R}^1$. Retaining the notation $M^* = M_i^* \overline{R}$, we deduce, by comparison with equations (2.144) through (2.146) and equation (2.155) that

$$M_i^* \overline{R}^1 = \left[M^*K + (I-K) \right]^{-1} \left[M^*(I-K) -K \right] \qquad (4.72)$$

Before futher considering equation (4.72), let us write M^* in partitioned form as

$$
M^* = \begin{bmatrix} M_{11} & m_1 & M_{21}^\mathsf{T} \\ m_1^\mathsf{T} & \mu & m_2^\mathsf{T} \\ M_{21} & m_2 & M_{22} \end{bmatrix}
\tag{4.73}
$$

where the vector $(m_1^\mathsf{T}, \mu, m_2^\mathsf{T})$ occupies the $(k+1)$th row of M^*. It then follows that

$$
M^* K + (I-K) = \begin{bmatrix} I & m_1 & 0 \\ 0 & \mu & 0 \\ 0 & m_2 & I \end{bmatrix}
\tag{4.74}
$$

and

$$
M^*(I-K) - K = \begin{bmatrix} M_{11} & 0 & M_{21}^\mathsf{T} \\ m_1^\mathsf{T} & -1 & m_2^\mathsf{T} \\ M_{21} & 0 & M_{22} \end{bmatrix}
\tag{4.75}
$$

Inverting equation (4.74) we obtain

$$
\left[M^* K + (I-K) \right]^{-1} = \begin{bmatrix} I & -\bar{m}_1 & 0 \\ 0 & -\bar{\mu} & 0 \\ 0 & -\bar{m}_2 & I \end{bmatrix}
\tag{4.76}
$$

where

$$
\bar{m}_1 = m_1/\mu
\tag{4.77}
$$

$$
\bar{m}_2 = m_2/\mu
\tag{4.78}
$$

and

$$
\bar{\mu} = -1/\mu
\tag{4.79}
$$

62

It follows from equations (4.72), (4.75), and (4.76) that

$$
M^*_{,\,\overline{R}^1} = \begin{bmatrix} (M_{11} - \overline{m}_1 m_1^T) & \overline{m}_1 & (M_{21}^T - \overline{m}_1 m_2^T) \\ \overline{m}_1^T & \overline{\mu} & \overline{m}_2^T \\ (M_{21} - \overline{m}_2 m_1^T) & \overline{m}_2 & (M_{22} - \overline{m}_2 m_2^T) \end{bmatrix} \qquad (4.80)
$$

Equation (4.80) may be interpreted as follows. Let the estimate corresponding to R^1 be denoted by $\hat{x}'$ and let $A^*_{,\,R^1}$ be denoted by A^{*1}. It then follows from equations (4.64), (4.67), (4.73), and (4.77) through (4.80) that

$$
(A^{*1})_{kk} = -1/(A^*)_{kk} \qquad (4.81)
$$

$$
(\hat{x}' - x_A)_k = (\overline{x})_k / (A^*)_{kk} \qquad (4.82)
$$

and

$$
C(\hat{x}') = C(\hat{x}) + (\hat{x}' - x_A)_k (\overline{x})_k \qquad (4.83)
$$

Of the three equations (4.81) through (4.83) the last one is of the greatest practical interest, because it tells us which parameter, if included in the set of parameters to be estimated, yields the least value for $C(\hat{x}')$. In other words, it tells us which additional parameter to estimate in order to achieve the maximum improvement in the data fit.

4.6   The Solution When the Matrix is Computationally Singular.

Two cases will be considered here:

(i)   Initial (or a priori) estimates are available for all parameters, and

(ii) Initial (or a priori) estimates may be available for some parameters, but certainly not for all.

In the next subsection, it will be shown how, in case (i), the computational singularity may be avoided.

4.6.1   Avoidance of the Computational Singularity When Independent A Priori Estimates are Available for All Parameters.

Let us consider matrix equations (4.40) through (4.42). Since the a priori estimates x_A in a mathematical sense are equivalent to direct measurements of x, they are included in the vector $\bar{u}$ representing the observations. To reflect this, equations (4.39) through (4.42) may be written in partitioned form as

$$\begin{bmatrix} u' \\ Rx_A \end{bmatrix} = \begin{bmatrix} A_o' \\ R \end{bmatrix} Rx + \begin{bmatrix} A_o' \\ R \end{bmatrix} Sx + \begin{bmatrix} \varepsilon' \\ R(x_A - x) \end{bmatrix}, \qquad (4.84)$$

$$\begin{bmatrix} \bar{u}' \\ 0 \end{bmatrix} = \begin{bmatrix} A_o' \\ R \end{bmatrix} R(x - x_A) + \begin{bmatrix} \bar{\varepsilon}' \\ R(x_A - x) \end{bmatrix} \qquad (4.85)$$

where

$$\bar{u}' = u' - A_o' x_A \qquad (4.86)$$

and

$$\bar{\varepsilon}' = \varepsilon' + A_o' S(x - x_A) \qquad (4.87)$$

Comparing equations (4.85) and (4.40), we thus see that

$$\bar{u} = \begin{bmatrix} \bar{u}' \\ 0 \end{bmatrix} \quad \text{and} \quad A_o = \begin{bmatrix} A_o' \\ R \end{bmatrix} \tag{4.88}$$

It also follows that the weighting matrix W is of the form,

$$W = \begin{bmatrix} W' & 0 \\ 0 & V^{-1} V^{-1} \end{bmatrix}, \tag{4.89}$$

where consistent with equations (4.52) and (4.53)

$$E(x_A) = x \tag{4.90}$$

and

$$E\left[(x_A - x)(x_A - x)^T\right] = VV \tag{4.91}$$

From the above and equations (4.6) and (4.7) we deduce that

$$A = A' + V^{-1} V^{-1}, \tag{4.92}$$

where

$$A' = (A_o')^T W' A_o'. \tag{4.93}$$

We similarly deduce from equation (4.46) that

$$\bar{y} = (A_o')^T W' \bar{u}' \tag{4.94}$$

Since by assumption the estimates of the components of x are independent it follows that V^{-1} is a diagonal matrix.

65

It was previously established (Subsection 3.4) that the matrix being inverted is computationally singular when $A_r = 0$ in equation (3.19). Since the matrix must be at least semi-positive definite without any a priori estimates being available, A_r should therefore be tested for being less than or equal to zero (or, in practice, some small predetermined constant. See (Morduch, 1975)) before the addition of the corresponding element of $V^{-1}V^{-1}$, and if found to satisfy the test then both A_r and y_r should be set to zero before the addition of the corresponding element of $V^{-1}V^{-1}$. y_r should be set to zero, since if $A_r = 0$ then the matrix cannot be semi-positive-definite unless $y_r = 0$.

Note that the procedure indicated above is mathematically equivalent to switching a parameter from being a 'solve for' to being a 'consider' parameter in order to be able to obtain a solution.

4.6.2   The Minimum Norm Solution

If it is found impossible to compute a definite solution to the least squares problem, then it is often desirable to obtain a minimum norm solution. The computational procedure for a minimum norm solution is given in Subsection 2.6.1. In this subsection we shall discuss both the form of the norm and some justification for choosing a minimum norm solution.

The norm of the estimate $\hat{x}$, in accordance with equation (2.95) is given by

$$N_Q(\hat{x}) \;=\; \hat{x}^T Q^{-1} \hat{x} \qquad\qquad (4.95)$$

In as far as the components of x represent physical quantities, the terms of the sum appearing on the right hand side of equation (4.95) must have physically the same dimensions. It should mention that this condition is not, in general, met by the frequently used Euclidean norm (Q=I). A suitable norm is given by Q being the diagonal matrix, whose diagonal elements satisfy

$$Q_{ii} = A_{ii} \qquad (4.96)$$

That this leads to a dimensionally correct norm is obvious since it has the same dimensions as $C(x)$, the quantity being minimized for the least squares solution. It can be shown that the physical interpretation of the norm given by equation (4.96) is that not only is the weighted sum of the squares of the residuals minimized, but also the weighted sum of the squares of all individual contributions to the observation vector $\bar{u}$. The proposed norm is quite general. Other norms that better fit particualr physical situations may, however, also be considered.

Let us now discuss the justification for choosing a minimum norm solution. First, let us say that if the only requirement is that the weighted sum of the squares of the residuals should be a minimum and no other information is given, then any solution of the least squares problem is as likely to be correct as any other. Such, however, is never the case when we are dealing with physical quantitities, which are always bounded. Also, by minimizing the corrections to the unknown parameter values, we ensure that, to the extent that they are in error, we also minimize any adverse effects

those parameters might have, when used to predict another set of observations.

4.7   The Expected Value of the Sum of the Squares of the Weighted Residuals.

In this subsection we shall consider the expected value of $C(\hat{x})$ as given by equation (4.57), viz.

$$C(\hat{x}) = \bar{u}^T W\bar{u} - \bar{y}^T R(\hat{x}-x_A) \tag{4.57}$$

It follows from the above and equations (4.49) and (4.46) that

$$C(\hat{x}) = \bar{u}^T W\left[\bar{u} - A_o RA^* RA_o^T W\bar{u}\right] \tag{4.97}$$

Since by equations (4.40) and (4.42)

$$\bar{u} = A_o(x-x_A) + \varepsilon \ , \tag{4.98}$$

we deduce with the aid of equations (4.6) and (4.47) that

$$C(\hat{x}) = \bar{u}^T W\left[A_o(x-x_A) + \varepsilon - A_o(RA^*S+R)(x-x_A)\right. $$
$$\left. - A_o RA^* RA_o^T W\varepsilon\right]$$

i.e.,

$$C(\hat{x}) = \bar{u}^T W\left[A_o(S-RA^*S)(x-x_A) + \varepsilon - A_o RA^* RA_o^T W\varepsilon\right] \tag{4.99}$$

68

Since, if for any two matrices P and Q both PQ and QP are defined, then trace(PQ) = trace (QP), it follows that

$$C(\hat{x}) = \text{trace} \left\{ (x-x_A)\bar{u}^T WA_o (S-RA^*S) \right\} + \text{trace} \left\{ \varepsilon \bar{u}^T W \right\}$$
$$- \text{trace} \left\{ A_o^I W \varepsilon \bar{u}^T WA_o RA^*R \right\} \qquad (4.100)$$

From the above and equations (4.98), (4.2), (4.3), (4.6), and (4.53) we obtain;

$$E\left(C(\hat{x}) \right) = \text{trace} \left\{ VVA(S-RA^*S) \right\} + \text{trace} \left\{ W^{-1} W \right\}$$
$$- \text{trace} \left\{ A \; RA^*R \right\} \qquad (4.101)$$

Post-multiplying equation (2.17) by S we deduce that

$$AS - ARA^*S = -SA^*S \qquad (4.102)$$

Since the diagonal elements of SA^*R vanish we find with the aid of equation (2.23) that

$$\text{trace} \left\{ ARA^*R \right\} = \text{trace} \left\{ R \right\} \qquad (4.103)$$

Hence,

$$E\left(C(\hat{x}) \right) = \text{trace} \left\{ W^{-1} W \right\} - \text{trace} \left\{ R \right\} - \text{trace} \left\{ VV(SA^*S) \right\} \qquad (4.104)$$

If the number of observations equals n and the number of parameters being estimated is m, then

$$\text{trace}\left\{W^{-1}W\right\} = n \quad \text{and} \quad \text{trace}\left\{R\right\} = m. \quad (4.105)$$

We hence obtain the result

$$E\left(C(\hat{x})\right) = n-m - \text{trace}\left\{VV(S\overset{*}{A}S)\right\} \quad (4.106)$$

5.0 SUMMARY

The generalized matrix product of two square matrices A and B, denoted by A·B, is defined by

$$A \cdot B = -ARB + AS + SB \quad , \qquad (2.16)$$

where

$$R = R^{\mathsf{T}} \qquad (2.2)$$

$$RR = R \qquad (2.3)$$

and

$$S = I - R \qquad (2.12)$$

It is shown that for any three square matrices A, B, and C

$$(A \cdot B) \cdot C = A \cdot (B \cdot C) \qquad (2.4)$$

and for any square matrix A

$$A \cdot (-R) = (-R) \cdot A = A \qquad (2.5)$$

The generalized inverse of A with respect to R is denoted by $A^{*}_{;R}$ or, when no risk of confusion arises, simply by A^{*}. It satisfies

$$A \cdot A^{*} = A^{*} \cdot A = -R \qquad (2.6)$$

Also

$$(A^*)^* = A \qquad (2.8)$$

and

$$(A^*)^T = (A^T)^* \qquad (2.11)$$

When R defining the generalized inverse is diagonal, the inverse is referred to as the <u>incomplete inverse</u> with respect to R.

Given the matrix equation

$$Ax = y \qquad (2.69)$$

where A is a symmetric matrix and x and y are vectors, it is shown that there exists an R such that $A^*(=A^*_\gamma R)$ is defined and

$$SA^*S = 0 \qquad (2.82)$$

The general solution of equation (2.69) is given by

$$x = x_o + (S-RA^*S)a \quad , \qquad (2.87)$$

where

$$x_o = RA^*Ry \quad , \qquad (2.86)'$$

and the vector a satisfies Ra = 0, but is otherwise arbitrary. (Note that if A is non-singular then R = I, S = 0 and $A^* = A^{-1}$).

The minimum norm solution x_Q is obtained by minimizing the

72

norm

$$N_Q(x) = x' Q^{-1} x \quad , \tag{2.95}$$

with respect to a. It is given by

$$x_Q = x_o + (S-H)(H^T Q^{-1} H + Q^{-1})^{-1} H^T Q^{-1} x_o \quad , \tag{2.115}$$

or alternatively by

$$x_Q = (Q + QH^T)(HQH^T + Q)^{-1} x_o \quad , \tag{2.116}$$

where

$$H = RA^* S$$

A procedure for obtaining the incomplete inverse of a symmetric matrix M is given in Section 3. It is shown that the procedure may always be applied when $(RMR+S)$ is positive-definite.

Consider the linear matrix equation

$$\bar{u} = A_o R(x-x_A) + \bar{\varepsilon} \quad , \tag{4.40}$$

where

$$\bar{u} = u - A_o x_A \quad , \tag{4.41}$$

$$\bar{\varepsilon} \doteq \varepsilon + A_o S(x-x_A) \quad , \tag{4.42}$$

73

u is a vector of observed measurements, ε is the measurement noise, x is a parameter vector, x_A is the a priori estimate of x, R and S are diagonal matrices satisfying equations (2.3) and (2.12), respectively. The components of x are either 'solve' or 'consider' parameters. R and S are further defined such that

$$R_{ii} = 1 \quad \text{and} \quad S_{ii} = 0 \qquad \text{if } x_i \text{ is a 'solve' parameter,}$$

and

$$R_{ii} = 0 \quad \text{and} \quad S_{ii} = 1 \qquad \text{if } x_i \text{ is a 'consider' parameter.}$$

The weighted sum of the square of the residuals is given by

$$C(x) = \left[\bar{u} - A_o R(x-x_A)\right]^T W\left[\bar{u} - A_o R(x-x_A)\right] \qquad (4.43)$$

where

$$W^{-1} = E(\varepsilon\varepsilon^T) \qquad (4.3)$$

C(x) is a minimum when $x = \hat{x}$. $\hat{x}$ is the weighted least squares solution of equation (4.40).

The symmetric matrix M is defined by

$$M = \begin{bmatrix} C_o & \bar{y}^T \\ \bar{y} & A \end{bmatrix}, \qquad (4.58)$$

where

$$C_o = \bar{u}^T W\bar{u} \quad , \qquad (4.59)$$

$$\overline{y} = A_o^T W \overline{u} \tag{4.46}$$

and

$$A = A_o^T W A_o \tag{4.6}$$

Corresponding to M, $\overline{R}$ and $\overline{S}$ are defined by

$$\overline{R} = \begin{bmatrix} 0 & 0 \\ 0 & R \end{bmatrix} \text{ and } \overline{S} = \begin{bmatrix} 1 & 0 \\ 0 & S \end{bmatrix} \tag{4.60}$$

It is shown that

$$M^* = \begin{bmatrix} -C(\hat{x}) & \overline{x}^T \\ \overline{x} & A^* \end{bmatrix} , \tag{4.61}, \tag{4.64}, \tag{4.66}$$

$$R(\hat{x} - x_A) = R\overline{x} , \tag{4.65}$$

and

$$cov(R\hat{x}) = RA^*R + LL^T , \tag{4.55}$$

where

$$L = -RA^* SV , \tag{4.56}$$

and

$$SVVS = E\left[S(x_A - x)(x_A - x)^T S \right] \tag{4.53}$$

75

It is further shown that

$$E\left[C(\hat{x})\right] \;=\; n\text{-}m \;-\; \text{trace}\left\{\; VV(SA^{*}S) \;\right\} \qquad (4.106)$$

where n is the number of observations and m is the number of 'solve' parameters.   Note that a priori estimates for 'solve' parameters should be counted as observations   .

If, e.g., x_k is switched from being a 'consider' to a 'solve' parameter, and the new estimate for x is denoted by $\hat{x}'$, then

$$C(\hat{x}') \;=\; C(\hat{x}) \;+\; (\bar{x})_k^2 \;/\; (A^{*})_{kk} \qquad \begin{matrix}(4.83)\\(4.82)\end{matrix}$$

ACKNOWLEDGEMENT

REFERENCES

(Lefferts, 1969)   An Experimental Comparison of Several Approach-
                      es to the Linear Least Squares Problem,
                      E.J. Lefferts and D.P. Muhonen, Goddard Space
                      Flight Center, X-551-69-442, September 1969.

(Morduch, 1975)   An Extension of NAP3.1F to Handle 'Consider'
                      Parameters and Increasing the Maximum Number of
                      Parameters to 600, G.E. Morduch, Old Dominion
                      Systems Inc., Gaithersburg, Maryland, NASA Con-
                      tract NAS5-20532, Task 11, December 1975.