

ON DERIVING THE INVERSE OF A SUM OF MATRICES

by

H. V. Henderson* and S. R. Searle

Biometrics Unit, Cornell University, Ithaca, New York

ABSTRACT

Available expressions are reviewed and new ones derived for the inverse of the sum of two matrices, one of them being non-singular. Particular attention is given to $(\underline{A} + \underline{UBV})^{-1}$, where \underline{A} is nonsingular and \underline{U} , \underline{B} and \underline{V} may be rectangular; generalized inverses of $\underline{A} + \underline{UBV}$ are also considered. Several statistical applications are discussed.

Paper No. BU-647-M in the Biometrics Unit Series, January 1980.

* Supported by a New Zealand National Research Advisory Council Fellowship.

1. INTRODUCTION

Motivation for many special cases and variants of the result

$$(\underset{\sim}{A} + \underset{\sim}{U}\underset{\sim}{B}\underset{\sim}{V})^{-1} = \underset{\sim}{A}^{-1} - \underset{\sim}{A}^{-1}\underset{\sim}{U}(\underset{\sim}{I} + \underset{\sim}{B}\underset{\sim}{V}\underset{\sim}{A}^{-1}\underset{\sim}{U})^{-1}\underset{\sim}{B}\underset{\sim}{V}\underset{\sim}{A}^{-1} \quad (1)$$

available in the literature of the last sixty years has come from a variety of sources: from inverting partitioned matrices, from applications in statistics and, quite recently, directly from inverting a sum of two matrices.

The first of these, inverting a partitioned matrix, is reviewed in Bodewig [1947, 1959] and Ouellette [1978], and is extended here in Section 3. Although initially a technique for developing equivalent expressions for the inverse of a matrix, it also yields, through equating two such expressions, the special case of (1) when $\underset{\sim}{B} = -\underset{\sim}{D}^{-1}$ is nonsingular, namely

$$(\underset{\sim}{A} - \underset{\sim}{U}\underset{\sim}{D}^{-1}\underset{\sim}{V})^{-1} = \underset{\sim}{A}^{-1} + \underset{\sim}{A}^{-1}\underset{\sim}{U}(\underset{\sim}{D} - \underset{\sim}{V}\underset{\sim}{A}^{-1}\underset{\sim}{U})^{-1}\underset{\sim}{V}\underset{\sim}{A}^{-1} . \quad (2)$$

This is discussed in Section 1.3.

The idea of modifying a matrix $\underset{\sim}{A}$, of known inverse, by simply adding another matrix to it, has also been explored. For example, the inverse of $\underset{\sim}{A} + \underset{\sim}{b}\underset{\sim}{u}\underset{\sim}{v}'$, where $\underset{\sim}{u}$ and $\underset{\sim}{v}'$ are column and row vectors, respectively, is

$$(\underset{\sim}{A} + \underset{\sim}{b}\underset{\sim}{u}\underset{\sim}{v}')^{-1} = \underset{\sim}{A}^{-1} - \frac{\underset{\sim}{b}}{1 + \underset{\sim}{b}\underset{\sim}{v}'\underset{\sim}{A}^{-1}\underset{\sim}{u}} \underset{\sim}{A}^{-1}\underset{\sim}{u}\underset{\sim}{v}'\underset{\sim}{A}^{-1} . \quad (3)$$

Identities (1) and (2) are broad generalizations of this sort of

modification, the importance of which is attested to by Householder [1957, p. 168] who declares (in our notation) that "all methods of inverting matrices reduce to successive applications of (2). The methods differ, however, in the way in which (2) is applied. In the method of modification as such one concentrates on a matrix $\underline{\underline{UD}}^{-1}\underline{\underline{V}}$ to be added to a matrix $\underline{\underline{A}}$ ".

Identities (1), (2) and (3) facilitate inverting many forms of patterned matrices occurring in statistics and other applications. For example, results in Roy and Sarhan [1956] for particular patterns (see also Graybill [1969, Chapter 8]) are readily obtainable from (1). An important case is the dispersion matrix for a multinomial random variable, namely

$$\underline{\underline{V}} = \begin{bmatrix} p_1(1-p_1) & -p_1p_2 & \cdots & -p_1p_k \\ -p_1p_2 & p_2(1-p_2) & \cdots & -p_2p_k \\ \vdots & \vdots & \ddots & \vdots \\ -p_1p_k & -p_2p_k & \cdots & p_k(1-p_k) \end{bmatrix},$$

which may be written as $\underline{\underline{V}} = \text{dg}(\underline{\underline{p}}) - \underline{\underline{p}}\underline{\underline{p}}'$ where $\text{dg}(\underline{\underline{p}})$ denotes a diagonal matrix with elements of $\underline{\underline{p}} = (p_1, \dots, p_k)'$ along the diagonal. Using $[\text{dg}(\underline{\underline{p}})]^{-1}\underline{\underline{p}} = \underline{\underline{1}}_k$ in (3) yields $\underline{\underline{V}}^{-1} = [\text{dg}(\underline{\underline{p}})]^{-1} + s\underline{\underline{J}}_k$, where $\underline{\underline{1}}_k$ is a vector of k ones, $s = 1/(1 - \underline{\underline{1}}_k'\underline{\underline{p}})$, and $\underline{\underline{J}}_k = \underline{\underline{1}}_k\underline{\underline{1}}_k'$ is a $k \times k$ matrix with all elements unity.

A further application of (3) is

$$\tilde{V}^{-1} = \begin{bmatrix} a & b & \cdots & b \\ b & a & \cdots & b \\ \vdots & \vdots & \ddots & \vdots \\ b & b & \cdots & a \end{bmatrix}^{-1} = \left[(a-b)\tilde{I}_k + b\tilde{J}_k \right]^{-1} = \frac{1}{a-b} \left[\tilde{I}_k - \frac{b}{a+(k-1)b} \tilde{J}_k \right]$$

where \tilde{V} has the pattern of an intraclass correlation matrix, in terminology dating back to Fisher [1925].

Another occurrence of these identities arises in the estimation of variance components, using the general linear model familiarly represented as $\underline{y} = \underline{X}\underline{\beta} + \underline{Z}\underline{u} + \underline{e}$, where $\underline{\beta}$ is a vector of fixed effects, \underline{u} and \underline{e} are independent vectors of random effects and residual error terms with zero means and dispersion matrices \underline{D} and \underline{R} , respectively. Then \underline{y} has mean $\underline{X}\underline{\beta}$ and dispersion matrix $(\underline{R} + \underline{Z}\underline{D}\underline{Z}')$. The generalized least squares estimator of $\underline{\beta}$, say $\hat{\underline{\beta}}$, is the solution to

$$\underline{X}'(\underline{R} + \underline{Z}\underline{D}\underline{Z}')^{-1}\underline{X}\hat{\underline{\beta}} = \underline{X}'(\underline{R} + \underline{Z}\underline{D}\underline{Z}')^{-1}\underline{y}.$$

A difficulty with these equations is that $\underline{R} + \underline{Z}\underline{D}\underline{Z}'$ is, in many applications (e.g., in genetics), often large and non-diagonal, to the extent that even with today's computing facilities, inverting it is quite impractical. An alternative set of equations is

$$\begin{aligned} \underline{X}'\underline{R}^{-1}\underline{X}\hat{\underline{\beta}} + \underline{X}'\underline{R}^{-1}\underline{Z}\hat{\underline{u}} &= \underline{X}'\underline{R}^{-1}\underline{y} \\ \underline{Z}'\underline{R}^{-1}\underline{X}\hat{\underline{\beta}} + (\underline{Z}'\underline{R}^{-1}\underline{Z} + \underline{D}^{-1})\hat{\underline{u}} &= \underline{Z}'\underline{R}^{-1}\underline{y} \end{aligned}$$

as discussed in Henderson et al. [1959]. These equations are easy to write in many practical situations, where \underline{R} and \underline{D} are diagonal,

and have the important property that $\tilde{\beta} = \hat{\beta}$. This is established by eliminating \tilde{u} to obtain

$$\tilde{X}'\tilde{W}\tilde{\beta} = \tilde{X}'\tilde{W}y ,$$

where

$$\tilde{W} = \tilde{R}^{-1} - \tilde{R}^{-1}\tilde{Z}(\tilde{Z}'\tilde{R}^{-1}\tilde{Z} + \tilde{D}^{-1})^{-1}\tilde{Z}'\tilde{R}^{-1} .$$

The symmetric counterpart of the inversion formula (2) gives

$$\tilde{W} = (\tilde{R} + \tilde{Z}\tilde{D}\tilde{Z}')^{-1} \text{ which immediately yields } \tilde{\beta} = \hat{\beta} .$$

Having indicated some applications of identities like (1) and its variants, we now review the development of such identities, beginning with associated results for determinants, which are often the origin of matrix theory.

1.1. The Determinant of a Partitioned Matrix

Aitken [1937] indicates that the determinant of a bordered matrix was considered by Cauchy and by Darboux [1874]. However, according to Ouellette [1978], it was probably Frobenius [1908 and 1968, p. 405] who first gave the result in the now familiar form

$$\begin{vmatrix} \tilde{A} & \tilde{u} \\ \tilde{v}' & \tilde{d} \end{vmatrix} = \tilde{d}|\tilde{A}| - \tilde{v}'(\text{adj}\tilde{A})\tilde{u} = |\tilde{A}|(\tilde{d} - \tilde{v}'\tilde{A}^{-1}\tilde{u}) , \quad (4)$$

where $\text{adj}\tilde{A}$ is the adjoint matrix of \tilde{A} and \tilde{u} and \tilde{v}' are vectors as in (3). The second equality in (4) holds only for nonsingular \tilde{A} and was extended by Schur [1917, p. 217], who studied under Frobenius, to

$$\begin{vmatrix} \underline{\underline{A}} & \underline{\underline{U}} \\ \underline{\underline{V}} & \underline{\underline{D}} \end{vmatrix} = |\underline{\underline{A}}| |\underline{\underline{D}} - \underline{\underline{V}} \underline{\underline{A}}^{-1} \underline{\underline{U}}| = |\underline{\underline{D}}| |\underline{\underline{A}} - \underline{\underline{U}} \underline{\underline{D}}^{-1} \underline{\underline{V}}|, \quad (5)$$

The second equality in (5) is not explicitly in Schur [1917], even though Gantmacher [1959, p. 46] refers to (5) as "formulas of Schur". A special case of (5) is

$$|\underline{\underline{I}} + \underline{\underline{V}} \underline{\underline{U}}| = |\underline{\underline{I}} + \underline{\underline{U}} \underline{\underline{V}}|, \quad (6)$$

a result which is often attributed to Sylvester, as in Press [1972, p. 20].

1.2. Inverting a Partitioned Matrix

Schur [1917] established the first equality in (5) by taking determinants of the identity

$$\begin{bmatrix} \underline{\underline{A}}^{-1} & \underline{\underline{O}} \\ \underline{\underline{-VA}}^{-1} & \underline{\underline{I}} \end{bmatrix} \begin{bmatrix} \underline{\underline{A}} & \underline{\underline{U}} \\ \underline{\underline{V}} & \underline{\underline{D}} \end{bmatrix} = \begin{bmatrix} \underline{\underline{I}} & \underline{\underline{A}}^{-1} \underline{\underline{U}} \\ \underline{\underline{O}} & \underline{\underline{D}} - \underline{\underline{VA}}^{-1} \underline{\underline{U}} \end{bmatrix}. \quad (7)$$

But he seems to have missed (perhaps surprisingly) the opportunity available from (7) of deriving the inverse of a partitioned matrix by inverting the right-hand side of (7) to obtain, when $\underline{\underline{A}}$ is non-singular but $\underline{\underline{D}}$ possibly singular,

$$\begin{aligned} \begin{bmatrix} \underline{\underline{A}} & \underline{\underline{U}} \\ \underline{\underline{V}} & \underline{\underline{D}} \end{bmatrix}^{-1} &= \begin{bmatrix} \underline{\underline{I}} & \underline{\underline{-A}}^{-1} \underline{\underline{U}} (\underline{\underline{D}} - \underline{\underline{VA}}^{-1} \underline{\underline{U}})^{-1} \\ \underline{\underline{O}} & (\underline{\underline{D}} - \underline{\underline{VA}}^{-1} \underline{\underline{U}})^{-1} \end{bmatrix} \begin{bmatrix} \underline{\underline{A}}^{-1} & \underline{\underline{O}} \\ \underline{\underline{-VA}}^{-1} & \underline{\underline{I}} \end{bmatrix} \\ &= \begin{bmatrix} \underline{\underline{A}}^{-1} + \underline{\underline{A}}^{-1} \underline{\underline{U}} (\underline{\underline{D}} - \underline{\underline{VA}}^{-1} \underline{\underline{U}})^{-1} \underline{\underline{VA}}^{-1} & \underline{\underline{-A}}^{-1} \underline{\underline{U}} (\underline{\underline{D}} - \underline{\underline{VA}}^{-1} \underline{\underline{U}})^{-1} \\ \underline{\underline{-(D}} - \underline{\underline{VA}}^{-1} \underline{\underline{U}})^{-1} \underline{\underline{VA}}^{-1} & (\underline{\underline{D}} - \underline{\underline{VA}}^{-1} \underline{\underline{U}})^{-1} \end{bmatrix}. \quad (8) \end{aligned}$$

Apparently Schur was very close to displaying (8), so close that it is often attributed to him, e.g., Bodewig [1947, 1959] and Marsaglia and Styan [1974]. Others were also close: Bodewig [1959, pp. 217-218] suggests that Boltz [1923] was, and Aitken [1939, p. 67] certainly was in that he has (7) but not (8).

The first explicit presentation of (8) appears to be Banachiewicz [1937a,b], using "crocovian-notation" in which columns are multiplied by columns. The result is discussed by Stankiewicz [1938a,b] and its proximity to Schur's work is noted by Bodewig [1947], who (on p. 49) is concerned about "the question of the priority" of (7) and (8), thus prompting Ouellette [1978, p. 3] to call (8) the "Schur-Banachiewicz inverse formula". Independently, Frazer et al. [1938, p. 113] and Waugh [1945] also established (8), and Jossa [1940] derived equivalent scalar expressions.

A first alternative to (8) is from Hotelling [1943a,b] which, in contrast to the need in (8) for $\underset{\sim}{A}$ to be nonsingular, demands that both $\underset{\sim}{A}$ and $\underset{\sim}{D}$ be nonsingular:

$$\begin{bmatrix} \underset{\sim}{A} & \underset{\sim}{U} \\ \underset{\sim}{V} & \underset{\sim}{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\underset{\sim}{A} - \underset{\sim}{U}\underset{\sim}{D}^{-1}\underset{\sim}{V})^{-1} & -\underset{\sim}{A}^{-1}\underset{\sim}{U}(\underset{\sim}{D} - \underset{\sim}{V}\underset{\sim}{A}^{-1}\underset{\sim}{U})^{-1} \\ -\underset{\sim}{D}^{-1}\underset{\sim}{V}(\underset{\sim}{A} - \underset{\sim}{U}\underset{\sim}{D}^{-1}\underset{\sim}{V})^{-1} & (\underset{\sim}{D} - \underset{\sim}{V}\underset{\sim}{A}^{-1}\underset{\sim}{U})^{-1} \end{bmatrix}. \quad (9)$$

Next is Duncan [1944] who, in addition to (8) and (9), also gives

$$\begin{bmatrix} \underset{\sim}{A} & \underset{\sim}{U} \\ \underset{\sim}{V} & \underset{\sim}{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\underset{\sim}{A} - \underset{\sim}{U}\underset{\sim}{D}^{-1}\underset{\sim}{V})^{-1} & -(\underset{\sim}{A} - \underset{\sim}{U}\underset{\sim}{D}^{-1}\underset{\sim}{V})^{-1}\underset{\sim}{U}\underset{\sim}{D}^{-1} \\ -(\underset{\sim}{D} - \underset{\sim}{V}\underset{\sim}{A}^{-1}\underset{\sim}{U})^{-1}\underset{\sim}{V}\underset{\sim}{A}^{-1} & (\underset{\sim}{D} - \underset{\sim}{V}\underset{\sim}{A}^{-1}\underset{\sim}{U})^{-1} \end{bmatrix}. \quad (10)$$

He states (p. 661) that these expressions have "been given by Dr. A. C. Aitken of Edinburgh in lectures to his students", and (on p. 666) "were communicated [to him] by Dr. A. C. Aitken". Aitken [1946, pp. 138-139], in a revision of his book, also has (10) as an "additional example" and gives yet another form when $\underline{\underline{A}}$, $\underline{\underline{B}}$, $\underline{\underline{U}}$ and $\underline{\underline{V}}$ are nonsingular:

$$\begin{bmatrix} \underline{\underline{A}} & \underline{\underline{U}} \\ \underline{\underline{V}} & \underline{\underline{D}} \end{bmatrix}^{-1} = \begin{bmatrix} (\underline{\underline{A}} - \underline{\underline{U}}\underline{\underline{D}}^{-1}\underline{\underline{V}})^{-1} & (\underline{\underline{V}} - \underline{\underline{D}}\underline{\underline{U}}^{-1}\underline{\underline{A}})^{-1} \\ (\underline{\underline{U}} - \underline{\underline{A}}\underline{\underline{V}}^{-1}\underline{\underline{D}})^{-1} & (\underline{\underline{D}} - \underline{\underline{V}}\underline{\underline{A}}^{-1}\underline{\underline{U}})^{-1} \end{bmatrix}. \quad (11)$$

Note that $\underline{\underline{S}}^{-1} = (\underline{\underline{A}} - \underline{\underline{U}}\underline{\underline{D}}^{-1}\underline{\underline{V}})^{-1}$ is the leading sub-matrix of (9), (10) and (11); and that similar inverses are involved elsewhere in (8) - (11). Thus it is, that inverting a partitioned matrix leads to inverting the sum of two matrices, $\underline{\underline{A}}$ and $(-\underline{\underline{U}}\underline{\underline{D}}^{-1}\underline{\underline{V}})$. Haynsworth [1968] calls $\underline{\underline{S}}$ a "Schur complement". Cottle [1974] and Ouellette [1978] survey the history and applications of Schur complements.

We now discuss the development of alternate expressions for the inverse of a Schur complement, and their connection with inverting a sum of matrices.

1.3. Inverting $\underline{\underline{A}} - \underline{\underline{U}}\underline{\underline{D}}^{-1}\underline{\underline{V}}$

An immediate consequence of (8) and (9) obtained by equating their leading terms is

$$(\underline{\underline{A}} - \underline{\underline{U}}\underline{\underline{D}}^{-1}\underline{\underline{V}})^{-1} = \underline{\underline{A}}^{-1} + \underline{\underline{A}}^{-1}\underline{\underline{U}}(\underline{\underline{D}} - \underline{\underline{V}}\underline{\underline{A}}^{-1}\underline{\underline{U}})^{-1}\underline{\underline{V}}\underline{\underline{A}}^{-1}. \quad (12)$$

It is surprising that Hotelling [1943a], in deriving (9), apparently

seemed unaware of (12), whereas Duncan [1944], with his interest being alternatives to (8), does give (12), in what seems to be its first appearance. Equation (12) also occurs in Guttman [1946] who is interested in factor analysis and who acknowledges Duncan [1944]; and it also appears in Bodewig [1959], but not in Bodewig [1947].

A further consequence of the equality of (8) and (9), noted by Duncan [1944, (4.9)] and Guttman [1946] is

$$(\underline{D} - \underline{V}\underline{A}^{-1}\underline{U})^{-1}\underline{V}\underline{A}^{-1} = \underline{D}^{-1}\underline{V}(\underline{A} - \underline{U}\underline{D}^{-1}\underline{V})^{-1}. \quad (13)$$

This result is more easily derived directly from the identity

$$\underline{V}\underline{A}^{-1}(\underline{A} - \underline{U}\underline{D}^{-1}\underline{V}) = (\underline{D} - \underline{V}\underline{A}^{-1}\underline{U})\underline{D}^{-1}\underline{V}.$$

1.4. Early Development of $(\underline{A} + \underline{U}\underline{B}\underline{V})^{-1}$

Prior to 1950 the development of (12) was indirect, coming from equating submatrices in equivalent expressions for the inverse of a partitioned matrix. Direct development seems to have started with Sherman and Morrison [1949, 1950], who consider inverting a matrix when elements of one of its rows or columns are altered. Bartlett [1951], from his interest in discriminant analysis, generalized this to adding a degenerate matrix $\underline{u}\underline{v}'$ to \underline{A} so as to obtain $(\underline{A} + \underline{u}\underline{v}')^{-1}$ from \underline{A}^{-1} , in the form

$$(\underline{A} + \underline{u}\underline{v}')^{-1} = \underline{A}^{-1} - \frac{\underline{A}^{-1}\underline{u}\underline{v}'\underline{A}^{-1}}{1 + \underline{v}'\underline{A}^{-1}\underline{u}}. \quad (14)$$

Woodbury [1950], apparently being unaware of earlier work like that of Duncan [1944], generalizes (14) in the form

$$(\underline{A} + \underline{UBV})^{-1} = \underline{A}^{-1} - \underline{A}^{-1} \underline{UB} (\underline{B} + \underline{BVA}^{-1} \underline{UB})^{-1} \underline{BVA}^{-1} \quad (15)$$

$$= \underline{A}^{-1} - \underline{A}^{-1} \underline{U} (\underline{B}^{-1} + \underline{VA}^{-1} \underline{U})^{-1} \underline{VA}^{-1}, \quad (16)$$

for \underline{A} and \underline{B} nonsingular, possibly of different order. It is perhaps not obvious that (15) requires the nonsingularity of \underline{B} , a fact which Woodbury overlooked. Its necessity is demonstrated by observing that for the right-hand side of (15) to exist the following determinant must be non-zero: $|\underline{B} + \underline{BVA}^{-1} \underline{UB}| = |\underline{B}| |\underline{I} + \underline{VA}^{-1} \underline{UB}|$, which can also be expressed as $|\underline{A}|^{-1} |\underline{A} + \underline{UBV}|$, on using (6). Hence $|\underline{B}| \neq 0$. Therefore (15) and (16) are equivalent to (12) with $\underline{B} = -\underline{D}^{-1}$, nonsingular. Relaxing this requirement leads to (1) and its many variants that are discussed in the sequel.

One merit for (16), and consequently for (12), noted by Guttman [1946] and Woodbury [1950] and important at the time their papers were published, is that when \underline{A}^{-1} and \underline{B}^{-1} are known, (16) is computationally advantageous to the extent that \underline{B} has smaller order than \underline{A} . This is still the case today, if the order of \underline{A} is very large and that of \underline{B} is quite modest, because $(\underline{B}^{-1} + \underline{VA}^{-1} \underline{U})^{-1}$ in the right-hand side of (16) will then be more readily computable than $(\underline{A} + \underline{UBV})^{-1}$ of the left-hand side.

Aspects of the history of (16) are discussed by Householder [1953, pp. 78-85; 1957, pp. 166-8; and 1964, pp. 123-4 and 141-2], to whom (15) and (16) are sometimes attributed (e.g., Aoki [1967, pp. 80-81]), despite his [1953, p. 84] references to Bodewig [1947] and Woodbury [1950]. More recently, Press [1972, p. 23] has called

(15) the Binomial Inverse Theorem and Ouellette [1978] has considerable discussion of both (15) and (16).

1.5. The Symmetric Case

The symmetric case of $(\underline{A} + \underline{UBV})^{-1}$ with $\underline{A} = \underline{A}'$, $\underline{B} = \underline{B}'$ and $\underline{V} = \underline{U}'$ occurs as a dispersion matrix in many mixed models of the analysis of variance as discussed in Section 1, in which context Henderson et al. [1959] give the symmetric counterpart of (12),

$$(\underline{A} + \underline{UBU}')^{-1} = \underline{A}^{-1} - \underline{A}^{-1}\underline{U}(\underline{B}^{-1} + \underline{U}'\underline{A}^{-1}\underline{U})^{-1}\underline{U}'\underline{A}^{-1}, \quad (17)$$

where \underline{A} and \underline{B} are both nonsingular. Lindley and Smith [1972] develop (17) as an "unexpected by-product" of a Bayesian process using a probabilistic argument. In the discussion of their paper, Kempthorne [1972] draws attention to the earlier result of Henderson et al. [1959]. Harville [1977], in considering maximum likelihood estimation of variance components, gives an expression similar to (17) but suited to \underline{B} being singular, namely

$$(\underline{A} + \underline{UBU}')^{-1} = \underline{A}^{-1} - \underline{A}^{-1}\underline{UB}(\underline{I} + \underline{U}'\underline{A}^{-1}\underline{UB})^{-1}\underline{U}'\underline{A}^{-1}. \quad (18)$$

A special case of (17) has \underline{A} diagonal and $\underline{D} = \underline{I}$, for which Guttman [1940] and Guttman and Cohen [1943] consider (12); and Ledermann [1938, 1939] and Guttman [1940, p. 91] discuss (13).

2. SIMPLE IDENTITIES

We have shown that expressions like (12) and (16) originated from considering the inverse of a partitioned matrix. Once available,

they are usually verified by showing that appropriate products simplify to $\underline{\underline{I}}$. In contrast, we develop such expressions directly, by applications of the identity $\underline{\underline{I}} = \underline{\underline{I}} + \underline{\underline{P}} - \underline{\underline{P}}$. This development yields not only established results but also new ones, and demonstrates that all the results can apply (for nonsingular $\underline{\underline{A}}$) for $\underline{\underline{U}}$, $\underline{\underline{B}}$ and $\underline{\underline{V}}$ being rectangular, in contrast to (12) and (15) - (17) where $\underline{\underline{B}}$ is necessarily square and nonsingular.

We begin by noting that for any matrix $\underline{\underline{P}}$ with $\underline{\underline{I}} + \underline{\underline{P}}$ nonsingular the identity $\underline{\underline{I}} = \underline{\underline{I}} + \underline{\underline{P}} - \underline{\underline{P}}$ immediately gives

$$(\underline{\underline{I}} + \underline{\underline{P}})^{-1} = \underline{\underline{I}} - \underline{\underline{P}}(\underline{\underline{I}} + \underline{\underline{P}})^{-1} = \underline{\underline{I}} - (\underline{\underline{I}} + \underline{\underline{P}})^{-1}\underline{\underline{P}}. \quad (19)$$

Obviously $(\underline{\underline{I}} + \underline{\underline{P}})^{-1}\underline{\underline{P}} = \underline{\underline{P}}(\underline{\underline{I}} + \underline{\underline{P}})^{-1}$, which is a special case of (13), as is

$$(\underline{\underline{I}} + \underline{\underline{PQ}})^{-1}\underline{\underline{P}} = \underline{\underline{P}}(\underline{\underline{I}} + \underline{\underline{QP}})^{-1}, \quad (20)$$

also.

3. INVERTING $\underline{\underline{A}} + \underline{\underline{UBV}}$

Six alternative forms of $(\underline{\underline{A}} + \underline{\underline{UBV}})^{-1}$ are now derived using (19) and (20) in a sequence that displays an interesting pattern. $\underline{\underline{A}}$ is taken as nonsingular and $\underline{\underline{U}}$, $\underline{\underline{B}}$ and $\underline{\underline{V}}$ as rectangular (or square) of order $n \times p$, $p \times q$ and $q \times n$, respectively. First, factor out $\underline{\underline{A}}$, to yield $(\underline{\underline{A}} + \underline{\underline{UBV}})^{-1} = (\underline{\underline{I}} + \underline{\underline{A}}^{-1}\underline{\underline{UBV}})^{-1}\underline{\underline{A}}^{-1}$, and to this apply the second equality in (19) to give

$$(\underline{\underline{A}} + \underline{\underline{UBV}})^{-1} = \underline{\underline{A}}^{-1} - (\underline{\underline{I}} + \underline{\underline{A}}^{-1}\underline{\underline{UBV}})^{-1}\underline{\underline{A}}^{-1}\underline{\underline{UBVA}}^{-1}. \quad (21)$$

Then repeatedly apply (20) to obtain

$$(\underline{\underline{A}} + \underline{\underline{UBV}})^{-1} = \underline{\underline{A}}^{-1} - \underline{\underline{A}}^{-1}(\underline{\underline{I}} + \underline{\underline{UBVA}}^{-1})^{-1}\underline{\underline{UBVA}}^{-1}, \quad (22)$$

$$= \underline{\underline{A}}^{-1} - \underline{\underline{A}}^{-1}\underline{\underline{U}}(\underline{\underline{I}} + \underline{\underline{BVA}}^{-1}\underline{\underline{U}})^{-1}\underline{\underline{BVA}}^{-1}, \quad (23)$$

$$= \underline{\underline{A}}^{-1} - \underline{\underline{A}}^{-1}\underline{\underline{UB}}(\underline{\underline{I}} + \underline{\underline{VA}}^{-1}\underline{\underline{UB}})^{-1}\underline{\underline{VA}}^{-1}, \quad (24)$$

$$= \underline{\underline{A}}^{-1} - \underline{\underline{A}}^{-1}\underline{\underline{UBV}}(\underline{\underline{I}} + \underline{\underline{A}}^{-1}\underline{\underline{UBV}})^{-1}\underline{\underline{A}}^{-1}, \quad (25)$$

$$= \underline{\underline{A}}^{-1} - \underline{\underline{A}}^{-1}\underline{\underline{UBVA}}^{-1}(\underline{\underline{I}} + \underline{\underline{UBVA}}^{-1})^{-1}. \quad (26)$$

Expressions (21) - (26) differ from those available in the literature and discussed earlier, in that they require neither symmetry nor squareness of $\underline{\underline{U}}$, $\underline{\underline{B}}$ or $\underline{\underline{V}}$, let alone nonsingularity of $\underline{\underline{B}}$. Furthermore, the existence of the inverses (apart from $\underline{\underline{A}}^{-1}$) is assured. This is so because each is the inverse of $\underline{\underline{I}}$ plus a cyclic permutation of $\underline{\underline{A}}^{-1}\underline{\underline{UBV}}$ and exists because its determinant is non-zero:

$$|\underline{\underline{I}} + \underline{\underline{A}}^{-1}\underline{\underline{UBV}}| = |\underline{\underline{A}}^{-1}| |\underline{\underline{A}} + \underline{\underline{UBV}}| \neq 0.$$

An important feature of (23) and (24), as of (16), is their possible computational advantage: all of (21) - (26) involve inverses of order n , the order of $\underline{\underline{A}}$, but apart from $\underline{\underline{A}}^{-1}$, expressions (23) and (24) involve inverses only of order p and q , respectively. This is attractive whenever p and/or q are less than n , particularly if considerably less.

A noticeable feature of the second term in each of (21) - (26) is that it is the product of matrices $\underline{\underline{A}}^{-1}$, $\underline{\underline{U}}$, $\underline{\underline{B}}$, $\underline{\underline{V}}$ and $\underline{\underline{A}}^{-1}$ in that sequence, together with an inverse matrix which is the inverse of $\underline{\underline{I}}$

plus a permuted form of $\underline{\underline{A}}^{-1}\underline{\underline{UBV}}$. The exact form is determined by the position of the inverse matrix in the product and is such that the sequence of matrices, without $\underline{\underline{I}}$, is $\underline{\underline{A}}^{-1}\underline{\underline{UBVA}}^{-1}\underline{\underline{UBVA}}^{-1}$, a very easy memory crutch.

Many simple cases can, of course, be derived from (21) - (26). For example, putting $\underline{\underline{B}} = \underline{\underline{I}}$ gives $(\underline{\underline{A}} + \underline{\underline{UV}})^{-1}$; and using $\underline{\underline{U}} = \underline{\underline{X}}$ and $\underline{\underline{B}} = \underline{\underline{V}} = \underline{\underline{I}}$ gives various forms of $(\underline{\underline{A}} + \underline{\underline{X}})^{-1}$.

The symmetric case with $\underline{\underline{A}}$ and $\underline{\underline{B}}$ symmetric and $\underline{\underline{V}} = \underline{\underline{U}}'$ does, of course, give $(\underline{\underline{A}} + \underline{\underline{UBV}})^{-1}$ symmetric. Despite this, none of the right-hand sides of the symmetric versions of (21) - (26) appears to be symmetric. This has been noted by Harville [1977] for (18), the symmetric case of (24). If, in addition, $\underline{\underline{B}}$ is nonsingular (23) and (24) become (12), (15) or (16) as already noted, and then reduce further to (17), the Henderson et al. [1959] result in which the symmetry is plainly evident.

4. SOME GENERALIZED INVERSES

Harville [1977] indicates that (18) can be derived as a generalized inverse form of $(\underline{\underline{A}} + \underline{\underline{UBV}})$ given in Harville [1976] provided the column space of $\underline{\underline{UBV}}$ is a subset of that of $\underline{\underline{A}}$, or equivalently, $\underline{\underline{AA}}^{-}\underline{\underline{UBV}} = \underline{\underline{UBV}}$, where $\underline{\underline{A}}^{-}$ is a generalized inverse of $\underline{\underline{A}}$, i.e., $\underline{\underline{AA}}^{-}\underline{\underline{A}} = \underline{\underline{A}}$. He also requires the symmetry of $\underline{\underline{A}}$ and $\underline{\underline{UBV}}$, but this may be replaced by the requirement that the row space of $\underline{\underline{UBV}}$ be a subset of the row space of $\underline{\underline{A}}$, or equivalently, $\underline{\underline{UBVA}}^{-}\underline{\underline{A}} = \underline{\underline{UBV}}$. In the presence of these row and column space requirements (which are satisfied in the special

case of \tilde{A} being nonsingular), we have derived generalized inverses of $\tilde{A} + \tilde{U}\tilde{B}\tilde{V}$ in forms akin to those of (21) - (26) for the regular inverse. They are as follows.

$$G_{\tilde{1}} = \tilde{A}^{-} - \tilde{A}^{-}(\tilde{A}^{-} + \tilde{A}^{-}\tilde{U}\tilde{B}\tilde{V}\tilde{A}^{-})^{-}\tilde{A}^{-}\tilde{U}\tilde{B}\tilde{V}\tilde{A}^{-}$$

$$G_{\tilde{2}} = \tilde{A}^{-} - \tilde{A}^{-}\tilde{U}(\tilde{U} + \tilde{U}\tilde{B}\tilde{V}\tilde{A}^{-}\tilde{U})^{-}\tilde{U}\tilde{B}\tilde{V}\tilde{A}^{-}$$

$$G_{\tilde{3}} = \tilde{A}^{-} - \tilde{A}^{-}\tilde{U}\tilde{B}(\tilde{B} + \tilde{B}\tilde{V}\tilde{A}^{-}\tilde{U}\tilde{B})^{-}\tilde{B}\tilde{V}\tilde{A}^{-}$$

$$G_{\tilde{4}} = \tilde{A}^{-} - \tilde{A}^{-}\tilde{U}\tilde{B}\tilde{V}(\tilde{V} + \tilde{V}\tilde{A}^{-}\tilde{U}\tilde{B}\tilde{V})^{-}\tilde{V}\tilde{A}^{-}$$

$$G_{\tilde{5}} = \tilde{A}^{-} - \tilde{A}^{-}\tilde{U}\tilde{B}\tilde{V}\tilde{A}^{-}(\tilde{A}^{-} + \tilde{A}^{-}\tilde{U}\tilde{B}\tilde{V}\tilde{A}^{-})^{-}\tilde{A}^{-}.$$

$G_{\tilde{3}}$ is the form given by Harville [1976], and it is also the generalized inverse analogue of Woodbury's [1950] result, (15).

A resemblance between $G_{\tilde{1}} - G_{\tilde{5}}$ and (21) - (26) can be noted. In the second terms of each $G_{\tilde{i}}$ the factors multiplying the parenthesized term are similar to those in (21) - (26) in the following way: in $G_{\tilde{1}}$ the pre- and post-multipliers of the ()⁻ term are similar to those in (22) and (21), respectively. In $G_{\tilde{2}}$ the similarity is to (23) and (22), respectively; and so on, until for $G_{\tilde{5}}$ the similarity is to (26) and (25). Finally, for nonsingular \tilde{A} , the generalized inverse $G_{\tilde{1}}$ yields both (21) and (22) and $G_{\tilde{5}}$ yields (25) and (26).

ACKNOWLEDGMENTS

Thanks go to R. D. Anderson, F. Pukelsheim, G. P. H. Styan, R. L. Quaas and M. A. Woodbury for helpful comments on early drafts of this paper.

REFERENCES

- Aitken, A. C. [1939, 1946]. Determinants and Matrices. Oliver and Boyd, Edinburgh. [Fourth edition, revised 1946.]
- Aoki, M. [1967]. Optimization of Stochastic Systems, Topics in Discrete-Time Systems. Academic Press, New York.
- Banachiewicz, T. [1937a]. Sur l'inverse d'un cracovien et une solution générale d'un système d'équations linéaires. Comptes Rendus Mensuels des Séances de la Classe des Sciences Mathématiques et Naturelles de l'Académie Polonaise des Sciences et des Lettres, Avril 1937, No. 4, 3-4.
- Banachiewicz, T. [1937b]. Zur Berechnung der Determinanten, wie auch der Inversen, und zur darauf basierten Auflösung der Systeme linearer Gleichungen. Acta Astronomica, Série C, 3, 41-67.
- Bartlett, M. S. [1951]. An inverse matrix adjustment arising in discriminant analysis. The Annals of Mathematical Statistics, 22, 107-111.
- Bodewig, E. [1947]. Comparison of some direct methods for computing determinants and inverse matrices. Koninklijke Nederlandse Akademie van Wetenschappen. Proceedings of the Section of Sciences, 50, 49-57.
- Bodewig, E. [1959]. Matrix Calculus. North-Holland, Amsterdam. Second Revised and Enlarged Edition.

- Boltz, H. [1923]. Entwicklungs-Verfahren zum ausgleichen geodätischer Netze nach der Methode der kleinsten Quadrate. Veröffentlichung des Preussischen Geodätischen Institutes. Neue Folge, Nr. 90, Druck und Verlag von P. Stankiewicz, Buchdruckerei G.M.B.H., Berlin.
- Cottle, R. W. [1974]. Manifestations of the Schur complement. Linear Algebra and Its Applications, 8, 189-211.
- Darboux, M. G. [1874]. Mémoire sur la théorie algébrique des formes quadratiques. Journal de Mathématiques Pures et Appliquées, Deuxième Série 1, 19, 347-396.
- Duncan, W. J. [1944]. Some devices for the solution of large sets of simultaneous linear equations. (With an appendix on the reciprocation of partitioned matrices.) The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, Seventh Series, 35, 660-670.
- Fisher, R. A. [1925]. Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh.
- Frazer, R. A., Duncan, W. J., and Collar, A. R. [1938]. Elementary Matrices and Some Applications to Dynamics and Differential Equations. Cambridge University Press. 416 pp.
- Frobenius, F. G. [1908, 1968]. Über Matrizen aus positiven Elementen. Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin, 471-476. [Reprinted in Ferdinand Georg Frobenius Gesammelte Abhandlungen, Springer-Verlag, Berlin, III (1968), 404-409.]

- Gantmacher, F. R. [1959]. The Theory of Matrices. Volume One.
Chelsea Publishing Company, New York.
- Graybill, F. A. [1969]. Introduction to Matrices with Applications in Statistics. Wadsworth, Belmont, California.
- Guttman, Louis [1940]. Multiple rectilinear prediction and the resolution into components. Psychometrika, 5, 75-99.
- Guttman, Louis [1946]. Enlargement methods for computing the inverse matrix. The Annals of Mathematical Statistics, 17, 336-343.
- Guttman, Louis and Cohen, Jozef [1943]. Multiple rectilinear prediction and the resolution into components: II. Psychometrika, 8, 169-183.
- Harville, D. A. [1976]. Extension of the Gauss-Markov Theorem to include the estimation of random effects. The Annals of Statistics, 4, 384-395.
- Harville, D. A. [1977]. Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association, 72, 320-338.
- Haynsworth, E. V. [1968]. Determination of the inertia of a partitioned Hermitian matrix. Linear Algebra and Its Applications, 1, 73-81.
- Henderson, C. R., Kempthorne, O., Searle, S. R., and von Kries, C. N. [1959]. Estimation of environmental and genetic trends from records subject to culling. Biometrics, 15, 192-218.
- Hotelling, Harold [1943a]. Some new methods in matrix calculation. The Annals of Mathematical Statistics, 14, 1-34.

- Hotelling, Harold [1943b]. Further points on matrix calculation and simultaneous equations. The Annals of Mathematical Statistics, 14, 440-441.
- Householder, A. S. [1953]. Principles of Numerical Analysis. McGraw-Hill Book Company, New York.
- Householder, A. S. [1957]. A survey of closed methods for inverting matrices. Journal of the Society for Industrial and Applied Mathematics, 5, 155-169.
- Householder, A. S. [1964]. The Theory of Matrices in Numerical Analysis. Blaisdell, New York.
- Jossa, F. [1940]. Risoluzione progressiva di un sistema di equazioni lineari. Analogia con un problema meccanico. Rendiconto dell'Accademia delle Scienze Fisiche e Matematiche, Società Reale di Napoli, Serie 4, 10, 346-352.
- Kemphorne, Oscar [1972]. Discussion of the paper by Professor Lindley and Dr. Smith. Journal of the Royal Statistical Society, Series B, 34, 33-37.
- Ledermann, Walter [1938]. A shortened method of estimation of mental factors by regression. Nature, 141, 650.
- Ledermann, Walter [1939]. On a shortened method of estimation of mental factors by regression. Psychometrika, 4, 109-116.
- Lindley, D. V. and Smith, A. F. M. [1972]. Bayes estimates for the linear model. Journal of the Royal Statistical Society, Series B, 34, 1-18.
- Marsaglia, George and Styan, George P. H. [1974]. Rank conditions for generalized inverses of partitioned matrices. Sankhyā, Series A, 36, 437-442.

- Ouellette, D. V. [1978]. Schur complements and statistics. M.Sc. Thesis, McGill University, Montreal.
- Press, S. James [1972]. Applied Multivariate Analysis. Holt, Rinehart and Winston, Inc.
- Roy, S. N. and Sarhan, A. E. [1956]. On inverting a class of patterned matrices. Biometrika, 43, 227-231.
- Schur, J. (Issai) [1917]. Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind. Journal für die reine und angewandte Mathematik, 147, 205-232.
- Sherman, J. and Morrison, W. J. [1949]. Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix. (Abstract). The Annals of Mathematical Statistics, 20, 621.
- Sherman, J. and Morrison, W. J. [1950]. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. The Annals of Mathematical Statistics, 21, 124-127.
- Stankiewicz, L. [1938a]. Sur les opérations arithmétiques dans le calcul des déterminants d'après les méthodes à cracoviens de M. T. Banachiewicz. Bulletin International de l'Académie Polonaise des Sciences et des Lettres, Série A, 1938, 10-23.
- Stankiewicz, L. [1938b]. Sur les opérations arithmétiques dans la résolution numérique d'un système d'équations linéaires suivant les méthodes à cracoviens de M. T. Banachiewicz. Bulletin International de l'Académie Polonaise des Sciences et des Lettres, Série A, 1938, 24-32.

Waugh, F. V. [1945]. A note concerning Hotelling's method of inverting a partitioned matrix. The Annals of Mathematical Statistics, 16, 216-217.

Woodbury, Max A. [1950]. Inverting modified matrices. Memorandum Report 42, Statistical Research Group, Princeton, New Jersey.