ON DERIVING THE INVERSE OF A SUM OF MATRICES

by

BU-647-M June, 1978

H. V. Henderson* and S. R. Searle
Biometrics Unit, Cornell University, Ithaca, New York

Abstract

Alternative expressions are derived for the inverse of the sum of two matrices, when at least one is nonsingular. A variety of expressions are then formulated for $(\tilde{A} + \tilde{U}\tilde{B}\tilde{V})^{-1}$ as $\tilde{A}^{-1}$ plus a matrix depending on $\tilde{A}^{-1}$ and the possibly rectangular $\tilde{U}$, $\tilde{B}$ and $\tilde{V}$. Whereas some of these expressions seem to be new, others are well-known, but for all of them our proofs contain the novelty that they are by derivation, rather than by verification. Applications are indicated for partitioned matrices, for certain special sums of matrices, for maximum likelihood estimation of variance components, and for multivariate analysis.

1. Introduction

Various expressions for special cases of $(\tilde{A} + \tilde{U}\tilde{B}\tilde{V})^{-1}$, when $\tilde{A}$ is nonsingular, are available in the literature. Most of them have been developed independently, during the last 40 years, in areas of application such as inverting a partitioned matrix, inverting a sum of matrices, and in statistics in variance component estimation and in multivariate analysis (e.g., discriminant analysis).

Paper No. BU-647-M in the Biometrics Unit Mimeo Series.

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

One of the earliest, but implicit references is Frazer et al. [1938, p. 113] who give, when $\underset{\sim}{A}$ is nonsingular but $\underset{\sim}{D}$ possibly singular,

$$\begin{bmatrix} A & U \\ V & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}U(D - VA^{-1}U)^{-1}VA^{-1} & -A^{-1}U(D - VA^{-1}U)^{-1} \\ -(D - VA^{-1}U)^{-1}VA^{-1} & (D - VA^{-1}U)^{-1} \end{bmatrix} . \tag{1}$$

Hotelling [1943], when both $\underset{\sim}{A}$ and $\underset{\sim}{D}$ are nonsingular, obtains

$$\begin{bmatrix} A & U \\ V & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - UD^{-1}V)^{-1} & -A^{-1}U(D - VA^{-1}U)^{-1} \\ -D^{-1}V(A - UD^{-1}V)^{-1} & (D - VA^{-1}U)^{-1} \end{bmatrix} . \tag{2}$$

Equating the leading submatrices in (1) and (2) yields

$$(A - UD^{-1}V)^{-1} = A^{-1} + A^{-1}U(D - VA^{-1}U)^{-1}VA^{-1} , \tag{3}$$

a result which Hotelling [1943] perhaps surprisingly, seemed to be unaware of. Duncan [1944], on the other hand, does give (3), this apparently being the first reference to such a result. His main interest is not in (3), but in alternative and equivalent forms for the inverse of a partitioned matrix, including (1) and (2) which, he states, have "been given by Dr. A. C. Aitken of Edinburgh in lectures to his students."

Direct development of results like (3) seems to have started with Sherman and Morrison [1949, 1950] who consider inverting a matrix when elements of a row or column are altered. Bartlett [1951], from his interest in discriminant analysis, generalized this to adding a degenerate matrix $\underset{\sim}{uv}'$ of rank one, so as to obtain $(\underset{\sim}{A} + \underset{\sim}{uv}')^{-1}$ from $\underset{\sim}{A}^{-1}$, in the form

$$(A + uv')^{-1} = A^{-1} - \frac{A^{-1}uv'A^{-1}}{1 + v'A^{-1}u} . \tag{4}$$

Woodbury [1950] generalizes (4) to inverting a modified matrix, giving

$$(\underset{\sim}{A} + \underset{\sim}{UBV})^{-1} = \underset{\sim}{A}^{-1} - \underset{\sim}{A}^{-1}\underset{\sim}{UB}(\underset{\sim}{B} + \underset{\sim}{BVA}^{-1}\underset{\sim}{UB})^{-1}\underset{\sim}{BVA}^{-1} \tag{5}$$

and

$$(\underset{\sim}{A} + \underset{\sim}{UBV})^{-1} = \underset{\sim}{A}^{-1} - \underset{\sim}{A}^{-1}\underset{\sim}{U}(\underset{\sim}{B}^{-1} + \underset{\sim}{VA}^{-1}\underset{\sim}{U})^{-1}\underset{\sim}{VA}^{-1} , \tag{6}$$

when $\underset{\sim}{A}$ and $\underset{\sim}{B}$ are nonsingular, possibly of different order. He makes the interesting comment that "none of the proofs are in the least difficult and are omitted but a considerable amount of exploration was necessary before the useful forms were discovered."

It is not obvious that (5) requires the nonsingularity of $\underset{\sim}{B}$ and in fact Woodbury overlooked this. Our development will emphasize the need for it, which is illustrated by observing that for $\underset{\sim}{I}$ an identity matrix and $\underset{\sim}{J}$ a square matrix with every element unity, with both matrices of order n,

$$(\underset{\sim}{I} + \underset{\sim}{J})^{-1} \neq \underset{\sim}{I} - \underset{\sim}{J}(\underset{\sim}{J} + \underset{\sim}{J}^2)^{-1}\underset{\sim}{J} .$$

Equation (5) with $\underset{\sim}{A}$, $\underset{\sim}{U}$ and $\underset{\sim}{V}$ identity matrices and $\underset{\sim}{B} = \underset{\sim}{J}$, would have this as an equality, which is not so because $\underset{\sim}{J} + \underset{\sim}{J}^2 = (n + 1)\underset{\sim}{J}$ which is singular. Indeed, as is well-known, $(\underset{\sim}{I} + \underset{\sim}{J})^{-1} = \underset{\sim}{I} - \underset{\sim}{J}/(n + 1)$. Woodbury [1950] also overlooked the equivalence of (6) to Duncan's [1944] result (3), being apparently unaware of this work in his discussion of applications of (6) to inverting partitioned matrices. On the other hand, he appears to have been the first to explicitly state that inverting partitioned matrices reduces to inverting sums of matrices. He also notes possible computational advantages of (6) when $\underset{\sim}{A}^{-1}$ and $\underset{\sim}{B}^{-1}$ are known and $(\underset{\sim}{B}^{-1} + \underset{\sim}{VA}^{-1}\underset{\sim}{U})^{-1}$ has smaller order than $(\underset{\sim}{A} + \underset{\sim}{UBV})^{-1}$, with (4) illustrating an extreme case of this.

Recently, Press [1972, p. 23] has called (5) the Binomial Inverse Theorem, stating without comment that $\underset{\sim}{A}$ and $\underset{\sim}{B}$ are both nonsingular.

The symmetric case of $(A + UBV)^{-1}$ with $A = A'$, $B = B'$ and $V = U'$ occurs as a dispersion matrix in many mixed models of the analysis of variance, with notation $V = R + ZDZ'$. An early treatment in this context is Henderson et al. [1959], where A and B are both taken as nonsingular, giving the symmetric counterpart of (6),

$$(A + UBU')^{-1} = A^{-1} - A^{-1}U(B^{-1} + U'A^{-1}U)^{-1}U'A^{-1} . \tag{7}$$

Lindley and Smith [1972] develop this, as they say, as an "unexpected byproduct" of a Bayesian process. In a discussion of their paper, Kempthorne [1972] draws attention to the earlier result and proof, which was by verification, of Henderson et al. [1959].

Harville [1976, 1977] gives an expression similar to (7) suited to B being singular, namely

$$(A + UBU')^{-1} = A^{-1} - A^{-1}UB(I + U'A^{-1}UB)^{-1}U'A^{-1} ,$$

and also considers a generalized inverse of $A + UBV$. His results are of particular use in the maximum likelihood estimation of variance components.

2. Useful Identities

Results like (3) - (7) are usually proved by verification, through checking that an appropriate product reduces to an identity matrix. In contrast, we derive results of this nature using identities obtained from the "add and subtract" principle that effects factorizations that lead to desired, and in some cases new, identities, new in particular for $(A + UBV)^{-1}$ when A is nonsingular and U, B and V are possibly rectangular.

The starting point is $I + P$, for some P such that $I + P$ is nonsingular. Then

$$(I + P)^{-1} = (I + P - P)(I + P)^{-1} = I - P(I + P)^{-1} . \tag{8}$$

Similarly,

$$(\underset{\sim}{I} + \underset{\sim}{P})^{-1} = (\underset{\sim}{I} + \underset{\sim}{P})^{-1}(\underset{\sim}{I} + \underset{\sim}{P} - \underset{\sim}{P}) = \underset{\sim}{I} - (\underset{\sim}{I} + \underset{\sim}{P})^{-1}\underset{\sim}{P} \,. \tag{9}$$

Note that equating (8) and (9) gives $(\underset{\sim}{I} + \underset{\sim}{P})^{-1}\underset{\sim}{P} = \underset{\sim}{P}(\underset{\sim}{I} + \underset{\sim}{P})^{-1}$. A generalization of this identity is based on

$$\underset{\sim}{P}(\underset{\sim}{I} + \underset{\sim}{QP}) = (\underset{\sim}{I} + \underset{\sim}{PQ})\underset{\sim}{P} \,. \tag{10}$$

When $\underset{\sim}{P}$ and $\underset{\sim}{Q}$, singular or nonsingular, are such that $\underset{\sim}{I} + \underset{\sim}{QP}$ is nonsingular, then so is $\underset{\sim}{I} + \underset{\sim}{PQ}$ (because these matrices have equal determinants), in which case (10) gives

$$(\underset{\sim}{I} + \underset{\sim}{PQ})^{-1}\underset{\sim}{P} = \underset{\sim}{P}(\underset{\sim}{I} + \underset{\sim}{QP})^{-1} \,. \tag{11}$$

Note the nature of this equality: the right-hand side is derived by the apparent operation of shifting the operator $(\underset{\sim}{I} + \cdot)^{-1}$ one position to the right in the sequence $\underset{\sim}{PQP}$, thus preserving the sequence of matrices without $\underset{\sim}{I}$. Repeated use of (11) is made in the sequel.

3. Inverting $\underset{\sim}{A} + \underset{\sim}{X}$

Although $\underset{\sim}{A} + \underset{\sim}{X}$ can be nonsingular when both $\underset{\sim}{A}$ and $\underset{\sim}{X}$ are singular, identities for $(\underset{\sim}{A} + \underset{\sim}{X})^{-1}$ involving functions of $\underset{\sim}{A}$ and $\underset{\sim}{X}$ are of interest only when one of them is nonsingular. We choose to consider $\underset{\sim}{A}$ as being nonsingular. Then when $\underset{\sim}{A} + \underset{\sim}{X}$ is nonsingular so also are $\underset{\sim}{I} + \underset{\sim}{A}^{-1}\underset{\sim}{X}$ and $\underset{\sim}{I} + \underset{\sim}{XA}^{-1}$, and

$$(\underset{\sim}{A} + \underset{\sim}{X})^{-1} = (\underset{\sim}{I} + \underset{\sim}{A}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{A}^{-1} \tag{12}$$

$$= \underset{\sim}{A}^{-1}(\underset{\sim}{I} + \underset{\sim}{XA}^{-1})^{-1} \,. \tag{13}$$

Then, $\underset{\sim}{P} = \underset{\sim}{A}^{-1}\underset{\sim}{X}$ used in (9) and substituted into (12), followed by repeated applications of (11), gives

$$(\underset{\sim}{A} + \underset{\sim}{X})^{-1} = \underset{\sim}{A}^{-1} - (\underset{\sim}{I} + \underset{\sim}{A}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{A}^{-1}\underset{\sim}{X}\underset{\sim}{A}^{-1} , \tag{14}$$

$$= \underset{\sim}{A}^{-1} - \underset{\sim}{A}^{-1}(\underset{\sim}{I} + \underset{\sim}{X}\underset{\sim}{A}^{-1})^{-1}\underset{\sim}{X}\underset{\sim}{A}^{-1} , \tag{15}$$

$$= \underset{\sim}{A}^{-1} - \underset{\sim}{A}^{-1}\underset{\sim}{X}(\underset{\sim}{I} + \underset{\sim}{A}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{A}^{-1} , \tag{16}$$

$$= \underset{\sim}{A}^{-1} - \underset{\sim}{A}^{-1}\underset{\sim}{X}\underset{\sim}{A}^{-1}(\underset{\sim}{I} + \underset{\sim}{X}\underset{\sim}{A}^{-1})^{-1} . \tag{17}$$

For the special case when $\underset{\sim}{X}$ is nonsingular we also have

$$(\underset{\sim}{I} + \underset{\sim}{X}\underset{\sim}{A}^{-1})^{-1} = (\underset{\sim}{A}^{-1} + \underset{\sim}{X}^{-1})^{-1}\underset{\sim}{X}^{-1} = \underset{\sim}{X}(\underset{\sim}{X} + \underset{\sim}{X}\underset{\sim}{A}^{-1}\underset{\sim}{X})^{-1} . \tag{18}$$

Substituting this in (15) gives

$$(\underset{\sim}{A} + \underset{\sim}{X})^{-1} = \underset{\sim}{A}^{-1} - \underset{\sim}{A}^{-1}(\underset{\sim}{A}^{-1} + \underset{\sim}{X}^{-1})^{-1}\underset{\sim}{A}^{-1} , \tag{19}$$

$$= \underset{\sim}{A}^{-1} - \underset{\sim}{A}^{-1}\underset{\sim}{X}(\underset{\sim}{X} + \underset{\sim}{X}\underset{\sim}{A}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{X}\underset{\sim}{A}^{-1} . \tag{20}$$

Using $(\underset{\sim}{I} + \underset{\sim}{A}^{-1}\underset{\sim}{X})^{-1}$, in a manner similar to (18), in (16) leads to the same results. We think it might be more appropriate to call (19) the Binomial Inverse Theorem than the case when $\underset{\sim}{X} = \underset{\sim}{U}\underset{\sim}{B}\underset{\sim}{V}$, referred to by Press [1972].

4. Inverting $\underset{\sim}{A} + \underset{\sim}{U}\underset{\sim}{B}\underset{\sim}{V}$

We now derive six alternative forms of $(\underset{\sim}{A} + \underset{\sim}{U}\underset{\sim}{B}\underset{\sim}{V})^{-1}$, in a sequence that displays an interesting pattern. $\underset{\sim}{A}$ is taken as nonsingular and $\underset{\sim}{U}$, $\underset{\sim}{B}$ and $\underset{\sim}{V}$ as rectangular (or square) of order $n \times p$, $p \times q$ and $q \times n$, respectively. In (14) put $\underset{\sim}{X} = \underset{\sim}{U}\underset{\sim}{B}\underset{\sim}{V}$ and then repeatedly apply (11) to obtain

$$(\mathbf{A} + \mathbf{UBV})^{-1} = \mathbf{A}^{-1} - (\mathbf{I} + \mathbf{A}^{-1}\mathbf{UBV})^{-1}\mathbf{A}^{-1}\mathbf{UBVA}^{-1} , \tag{21}$$

$$= \mathbf{A}^{-1} - \mathbf{A}^{-1}(\mathbf{I} + \mathbf{UBVA}^{-1})^{-1}\mathbf{UBVA}^{-1} , \tag{22}$$

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

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

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

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

These six expressions differ from those available in the literature and discussed earlier, in that they require neither symmetry nor squareness of $\mathbf{U}$, $\mathbf{B}$ or $\mathbf{V}$, let alone nonsingularity of $\mathbf{B}$. Furthermore, expressions (23) and (24) have not been seen elsewhere; their existence is assured by arguments similar to those used in establishing (11). For example, in (23), $(\mathbf{I} + \mathbf{BVA}^{-1}\mathbf{U})^{-1}$ exists because its determinant is non-zero:

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

An important feature of (23) and (24) is that their use can be computationally advantageous. All of the expressions (21) - (26), other than (23) and (24), involve inverses of order n, the order of $\mathbf{A}$. But, apart from $\mathbf{A}^{-1}$, (23) and (24) involve inverses only of order p and q, respectively, which 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 $\mathbf{A}^{-1}$, $\mathbf{U}$, $\mathbf{B}$, $\mathbf{V}$ and $\mathbf{A}^{-1}$ in that sequence, together with an inverse matrix which is the inverse of $\mathbf{I}$ plus a permuted form of $\mathbf{A}^{-1}\mathbf{UBV}$. The exact form is determined by the position of the inverse matrix in the product and is such that the sequence of matrices, without $\underset{\sim}{I}$, is $\underset{\sim}{A}^{-1}\underset{\sim}{UBVA}^{-1}\underset{\sim}{UBVA}^{-1}$, a very easy memory crutch.

5. Special Cases

Of the many special cases that could be considered, at least six are worthy of mention.

(1) Nonsingularity of $\underset{\sim}{B}$ implies that both (23) and (24) can be rewritten as (5) or (6), which are Woodbury's [1950] results.

(2) The symmetric case with $\underset{\sim}{A}$ and $\underset{\sim}{B}$ symmetric and $\underset{\sim}{V} = \underset{\sim}{U}'$ does, of course, give $(\underset{\sim}{A} + \underset{\sim}{UBV})^{-1}$ symmetric. Despite this, none of the right-hand sides of the symmetric versions of (21) - (26) appear to be symmetric. This has been noted by Harville [1977], although he gives only one of our six results, namely (24), under symmetry. If, in addition, $\underset{\sim}{B}$ is nonsingular, (23) and (24) become (5) or (6), as already noted, and (6) then reduces further, to (7) of Henderson et al. [1959], in which the symmetry is plainly evident.

(3) Many simple cases can, of course, be derived from (21) - (26). For example, putting $\underset{\sim}{B} = \underset{\sim}{I}$ gives expressions for $(\underset{\sim}{A} + \underset{\sim}{UV})^{-1}$; using $\underset{\sim}{U} = \underset{\sim}{X}$ and $\underset{\sim}{B} = \underset{\sim}{V} = \underset{\sim}{I}$ gives the expressions for $(\underset{\sim}{A} + \underset{\sim}{X})^{-1}$ shown in (14) - (17); and putting $\underset{\sim}{A} = \underset{\sim}{I}$, $\underset{\sim}{U} = \underset{\sim}{P}$ and $\underset{\sim}{B} = \underset{\sim}{V} = \underset{\sim}{I}$ gives results (8) and (9) for $(\underset{\sim}{I} + \underset{\sim}{P})^{-1}$.

(4) Another well-known result is

$$(\underset{\sim}{A} + b\underset{\sim}{uv}')^{-1} = \underset{\sim}{A}^{-1} - \frac{b}{1 + b\underset{\sim}{v}'\underset{\sim}{A}^{-1}\underset{\sim}{u}}\underset{\sim}{A}^{-1}\underset{\sim}{uv}'\underset{\sim}{A}^{-1} \tag{27}$$

which comes from (23), with $\underset{\sim}{B} = b$ a scalar and consequently $\underset{\sim}{U} = \underset{\sim}{u}$ and $\underset{\sim}{V} = \underset{\sim}{v}$ as vectors. With b = 1, (27) reduces to Bartlett's [1951] result (4).

(5) A further case of (27) is with $\underset{\sim}{A} = \underset{\sim}{I}$. Then

$$(\underset{\sim}{I} + b\underset{\sim}{u}\underset{\sim}{v}')^{-1} = \underset{\sim}{I} - \frac{b}{1 + b\underset{\sim}{v}'\underset{\sim}{u}}\underset{\sim}{u}\underset{\sim}{v}'$$

and with b = 1

$$(\underset{\sim}{I} + \underset{\sim}{u}\underset{\sim}{v}')^{-1} = \underset{\sim}{I} - \frac{\underset{\sim}{u}\underset{\sim}{v}'}{1 + \underset{\sim}{v}'\underset{\sim}{u}},$$

as is well-known.

(6) A final special case of (27) is with $\underset{\sim}{u} = \underset{\sim}{v} = \underset{\sim}{1}$, a vector of ones, so that $\underset{\sim}{u}\underset{\sim}{v}' = \underset{\sim}{1}\underset{\sim}{1}' = \underset{\sim}{J}$ and

$$(\underset{\sim}{A} + b\underset{\sim}{J})^{-1} = \underset{\sim}{A}^{-1} - \frac{b}{1 + b\underset{\sim}{1}'\underset{\sim}{A}^{-1}\underset{\sim}{1}}\underset{\sim}{A}^{-1}\underset{\sim}{J}\underset{\sim}{A}^{-1}.$$

For $\underset{\sim}{A} = a\underset{\sim}{I}_n$, we then have the familiar

$$(a\underset{\sim}{I}_n + b\underset{\sim}{J}_n)^{-1} = \frac{1}{a}\underset{\sim}{I}_n - \frac{b}{a(a + bn)}\underset{\sim}{J}_n .$$

References

Bartlett, M. S. [1951]. An inverse matrix adjustment arising in discriminant analysis. Annals of Mathematical Statistics 22:107-111.

Duncan, W. J. [1944]. Some devices for the solution of large sets of simultaneous linear equations. Philosophical Magazine 7th series 35:660-670.

Frazer, R. A., Duncan, W. G., and Collar, A. R. [1938]. Elementary Matrices and Some Applications to Dynamics and Differential Equations. Cambridge University Press. 416 pp.

Harville, D. A. [1976]. Extension of the Gauss-Markov Theorem to include the estimation of random effects. 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.

Henderson, C. R., Kempthorne, O., Searle, S. R., and von Krisgk, C. N. [1959]. Estimation of environmental and genetic trends from records subject to culling. Biometrics 15:192-218.

Hotelling, Harold [1943]. Some new methods in matrix calculation. Annals of Mathematical Statistics 43:1-34.

Kempthorne, Oscar [1972]. Discussion of the paper by Professor Lindley and Dr. Smith. Journal of the Royal Statistical Society, B, 34:33-37.

Lindley, D. V. and Smith A. F. M. [1972]. Bayes estimates for the linear model. Journal of the Royal Statistical Society, B, 34:1-18.

Press, S. James [1972]. Applied Multivariate Analysis. Holt, Rinehart and Winston, Inc.

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). 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. Annals of Mathematical Statistics 21:124-127.

Woodbury, Max [1950]. Inverting modified matrices. Technical Report No. 42, Princeton University, Princeton, N. J. 6 pp.