Properties of the matrix A-XY

Steerneman, Ton; van Perlo -ten Kleij, Frederieke

Published in:
Linear Algebra and Its Applications

DOI:
10.1016/j.laa.2004.10.028

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Linear Algebra and its Applications 410 (2005) 70–86
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

Properties of the matrix $A - XY^*$

Ton Steerneman, Frederieke van Perlo-ten Kleij *

Department of Econometrics, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands

Received 8 November 2002; accepted 25 October 2004
Available online 18 December 2004
Submitted by H.J. Werner

Abstract

The main topic of this paper is the matrix $V = A - XY^*$, where A is a nonsingular complex $k \times k$ matrix and X and Y are $k \times p$ complex matrices of full column rank. Because properties of the matrix V can be derived from those of the matrix $Q = I - XY^*$, we will consider in particular the case where $A = I$. For the case that $Y^*X = I$, so that Q is singular, we will derive the Moore–Penrose inverse of Q. The Moore–Penrose inverse of V in case $Y^*A^{-1}X = I$ then easily follows. Finally, we will focus on the eigenvalues and eigenvectors of the real matrix $D - xy'$ with D diagonal.

Keywords: Moore–Penrose inverse; Idempotent matrix; Projection matrix; Eigenvalues; Eigenvectors

1. Introduction

In statistics, econometrics and linear algebra we often encounter a matrix of the type $V = A - XY^*$, where A is some nonsingular $k \times k$ complex matrix and X, Y are $k \times p$ complex matrices of rank $p < k$. Of special interest is the case $p = 1$. Well-known is the centering operator with the matrix $H = I - k^{-1}\iota\iota'$, where ι is a $k \times 1$ vector of ones. This operator maps a vector $(x_1, \ldots, x_k)'$ to $(x_1 - \bar{x}, \ldots, x_k - \bar{x})'$, where $\bar{x}$ denotes the mean of the x_i. The matrix H is idempotent being the

* Corresponding author.
E-mail addresses: a.g.m.steerneman@eco.rug.nl (T. Steerneman), f.van.perlo-ten.kleij@eco.rug.nl (F. van Perlo-ten Kleij).

0024-3795/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2004.10.028

orthogonal projector on the subspace orthogonal to the vector ι. Another well-known example is $R = \Pi - \pi\pi'$, where $\pi = (\pi_1, \ldots, \pi_k)'$ with $\pi_i > 0$, $\iota'\pi = 1$, and $\Pi = \text{diag}(\pi)$. Note that nR is the covariance matrix of the multinomial distribution with parameters n and π. A third example is $V = \text{diag}(d) + a\iota'$, where $d, a \in \mathbb{R}^k$ have positive elements. This matrix was originally studied by Vermeulen [18], because of a physical investigation on the electronic properties of particle-counting diamonds.

The matrix V, with A symmetric and $p = 1$, was studied in its general form by Trenkler [17]. He generalized results previously obtained by Vermeulen [18], Klamkin [10], Tanabe and Sagae [16], Neudecker [12], and Watson [20]. We will continue this line of research by, for example, dropping the assumption of symmetry of A and allowing $p > 1$. We also consider the more special case $V = D - ab'$, where $D = \text{diag}(d)$ and $a, b, d \in \mathbb{R}^k$.

In our notation, Vermeulen [18] showed that the eigenvalues of $D - a\iota'$, where $D = \text{diag}(d)$, $d_1, \ldots, d_k > 0$ and $a_1, \ldots, a_k < 0$, are real. Klamkin [10] gave a more elementary derivation. Moreover, he gave simple bounds for the eigenvalues. He derived the characteristic polynomial

$$|\lambda I - (D - a\iota')| = \left\{1 + \sum_{j=1}^{k} \frac{a_j}{\lambda - d_j}\right\} \prod_{i=1}^{k} (\lambda - d_j). \tag{1.1}$$

From this result, he observed that if $0 < d_1 < d_2 < \cdots < d_k$, then the eigenvalues $\lambda_1, \ldots, \lambda_k$ are obtained by solving

$$1 + \sum_{j=1}^{k} \frac{a_j}{\lambda - d_j} = 0, \tag{1.2}$$

and if some of the d_i coincide, then there will be eigenvalues equal to the d_i that coincide. This was generalized to $D - ab'$ by Trenkler [17]. In Section 5, we will return to these results of Trenkler [17] and give alternative proofs that exploit the original ideas of Vermeulen [18] and Klamkin [10]. Moreover, we will also derive the eigenvectors using some ideas of Watson [20].

De Boer and Harkema [2] were interested in the maximum likelihood estimation of sum-constrained linear models: $Y \sim \mathcal{N}_k(\mu, \Omega)$, $\iota'Y = c$, so that $\Omega\iota = 0$, where a certain structure will be imposed on μ. Such models are of interest in modelling demand systems, brand choice, and so on. In case of relatively small samples, the model has to be parsimonious, especially with regard to the parameterization of Ω. De Boer and Harkema [2] suggested the specification

$$\Omega = D - \frac{1}{\iota'd} dd', \tag{1.3}$$

where $D = \text{diag}(d)$ and $d \in \mathbb{R}^k$, $d_i \neq 0$, $i = 1, \ldots, k$. Because of the constraint, they deleted one component of Y and the $(k-1) \times (k-1)$ covariance matrix obtained became nonsingular. Wansbeek [19] assumed $d_1 < \cdots < d_k$ and obtained the following results. One eigenvalue of Ω is equal to zero, the other eigenvalues λ satisfy

$$\sum_{i=1}^{k} \frac{d_i}{\lambda - d_i} = 0. \tag{1.4}$$

This follows from the characteristic equation he derived in the following way:

$$\begin{aligned} 0 = |\lambda I - \Omega| &= |\lambda I - D| \left(1 + \frac{1}{\iota' d} d'(\lambda I - D)^{-1} d\right) \\ &= \frac{1}{\iota' d} |\lambda I - D| \Big(\iota'(\lambda I - D)(\lambda I - D)^{-1} d + d'(\lambda I - D)^{-1} d\Big) \\ &= |\lambda I - D| \Big\{ \frac{\lambda}{\iota' d} \iota'(\lambda I - D)^{-1} d \Big\}. \end{aligned}$$

Wansbeek [19] observed from (1.4) that if d_i and d_{i+1} are of the same sign, then there lies an eigenvalue between them. We will use the same method in Section 5 to obtain the characteristic polynomial of $D - ab'$. Wansbeek [19] also derived the Moore–Penrose inverse of Ω, namely $\Omega^+ = HD^{-1}H$. Since the matrix Ω is symmetric and should be positive semi-definite, he concluded that $0 < d_2 < \cdots < d_k$ is a necessary condition. In case $d_1 < 0 < d_2 < \cdots < d_k$ he used the Moore–Penrose inverse to establish that it is necessary that $\sum_{i=1}^{k} d_i < 0$. This can, however, more easily be shown by observing that it is necessary that the (1, 1) element of Ω should be nonnegative. This amounts to

$$d_1 - \frac{1}{\iota' d} d_1^2 > 0$$

and hence $d_1 < 0$ implies that $\iota' d < 0$.

A matrix that is very similar to Ω is the matrix R we already discussed, since the covariance matrix of the multinomial distribution is based upon $R = \Pi - \pi\pi'$, where $\Pi = \operatorname{diag}(\pi)$, $\pi_1, \ldots, \pi_k > 0$ and $\iota'\pi \leqslant 1$. If there are $k + 1$ possible categories, then one may wish to count only the number of outcomes in the first k categories, because the number of outcomes in category $k + 1$ uniquely follows from the total number of outcomes in the remaining categories. In this case $\iota'\pi < 1$. The matrix R has been studied under the condition $\iota'\pi \leqslant 1$ by Tanabe and Sagae [16]. They obtained, among other things, the square-root free Cholesky decomposition, the Moore–Penrose inverse in case $\iota'\pi = 1$, namely $R^+ = H\Pi^{-1}H$, and the inverse in case $\iota'\pi < 1$, that is, $R^{-1} = \Pi^{-1} + (1 - \iota'\pi)^{-1}\iota\iota'$. Neudecker [12] offered more elegant proofs and presented some new results. Watson [20] assumed $\iota'\pi = 1$ and showed how the eigenvalues and eigenvectors can be obtained. He showed that an eigenvalue not equal to any of the π_i satisfies

$$\sum_{i=1}^{k} \frac{\pi_i^2}{\pi_i - \lambda} = 1. \tag{1.5}$$

This equation is very similar to (1.2) and (1.4). One eigenvalue is equal to zero and the other eigenvalues $\lambda_1, \ldots \lambda_{k-1}$ satisfy

$$\pi_1 \leqslant \lambda_1 \leqslant \pi_2 \leqslant \lambda_2 \leqslant \pi_3 \leqslant \cdots \leqslant \lambda_{k-1} \leqslant \pi_k$$

with strict inequalities if the π_i's are all distinct. Similar observations have been made by Klamkin [10] and Wansbeek [19]. Watson furthermore derived how to obtain the eigenvectors. The product of the nonzero eigenvalues of R was obtained by Tanabe and Sagae [16] and Neudecker [12].

Dol [3] and Dol et al. [4] studied the Horvitz-Thompson estimator. Consider a finite population $Y_1, \ldots, Y_N$. A fixed effective sample design of size n can be interpreted as a probability distribution on the set of all subsets of n elements from the labels $\{1, \ldots, N\}$. Let S denote the random set of n labels that occur in the sample. The indicators $E_1, \ldots, E_N$ are defined by $E_i = 1$ if $i \in S$, and $E_i = 0$ if $i \notin S$. The first order inclusion probability is $\pi_i = \mathrm{P}(S \ni i)$ for $i = 1, \ldots, N$. It is assumed that $\pi_i = \mathrm{E}E_i$ is positive. The Horvitz-Thompson estimator $\bar{Y}_{HT}$ for the population mean $\bar{Y}$ is

$$\bar{Y}_{HT} = \frac{1}{N}\sum_{i\in S}\frac{Y_i}{\pi_i} = \frac{1}{N}\sum_{i=1}^{N} E_i\frac{Y_i}{\pi_i}.$$

This is a famous unbiased estimator. In order to give the variance, the second order inclusion probabilities are needed: $\pi_{ij} = \mathrm{P}(S \ni i, j) = \mathrm{E}E_iE_j$, for $i, j = 1, \ldots, N$. Note that $\pi_{ii} = \pi_i$. We define $\pi = (\pi_1, \ldots, \pi_N)'$, $\Pi = \mathrm{diag}(\pi)$ and $\Pi_2 = (\pi_{ij})$. It is easy to see that $\pi'\iota = n$ and $\Pi_2\iota = n\pi$. The well-known expression for the variance of the Horvitz-Thompson estimator is

$$\mathrm{Var}\,\bar{Y}_{HT} = N^{-2}Y'\Pi^{-1}(\Pi_2 - \pi\pi')\Pi^{-1}Y,$$

where $Y = (Y_1, \ldots, Y_N)'$. The matrix $\Pi_2 - \pi\pi'$ looks similar to R, but it is more complicated. In order to derive bounds for this variance, Dol [3] and Dol et al. [4] obtained the following Moore–Penrose inverse:

$$(\Pi_2 - \pi\pi')^+ = H\Pi_2^{-1}H.$$

Inspired by Trenkler [17], we will derive the (Moore–Penrose) inverse of

$$V = A - XY^*, \tag{1.6}$$

where A is a nonsingular $k \times k$ complex matrix and X and Y are $k \times p$ complex matrices of full column rank $p < k$. Thus, we will generalize Trenkler's results in two ways. First, the matrix A is only restricted to be nonsingular, symmetry is not necessary. Secondly, the vectors a and b in the matrix $A - ab'$ examined by Trenkler can be replaced by matrices of full column rank. Because $|V| = |A|\,|I - Y^*A^{-1}X|$ (see e.g. [7]), an interesting case is $Y^*A^{-1}X = I$, so that V is singular. We call this the singular case and it will be discussed in Section 4. Note that the assumption that $Y^*A^{-1}X = I$ implies that both X and Y are of full-column rank. If $|I - Y^*A^{-1}X| \neq 0$, then V is invertible, and we will refer to this as the nonsingular case. It will be discussed in Section 3. We will not consider the mixture case where $Y^*A^{-1}X \neq I$ and $|I - Y^*A^{-1}X| = 0$.

It is worthwhile to first consider a special case of (1.6), namely $A = I$, because properties of the matrix V can be derived from those of the matrix $Q = I - XY^*$. If $Y^*X = I$, then the matrix Q is idempotent, since $Q^2 = Q$, hence

$$\operatorname{rank}(Q) = \operatorname{tr}(Q) = k - \operatorname{tr}(XY^*) = k - \operatorname{tr}(Y^*X) = k - p.$$

In particular, if $V = A - XY^*$ with $Y^*A^{-1}X = I$, it is not difficult to see that the rank of V also equals $k - p$, since $V = A - XY^* = A(I - A^{-1}XY^*)$, and $\operatorname{rank}(V) = \operatorname{rank}(I - A^{-1}XY^*) = k - p$ as we showed above.

In the literature, considerable attention has been paid to the (generalized) inverse of a sum of matrices. Henderson and Searle [7] reviewed and derived expressions for the (generalized) inverse of matrices of the form $A + UBV$, where A is nonsingular and U, B and V may be rectangular. Riedel [15] considered the matrix $A + (V_1 + W_1)G(V_2 + W_2)^*$ with A square and singular, G nonsingular and V_i and W_i possibly rectangular, $i = 1, 2$. Under the conditions $\mathscr{R}(V_1) \subseteq \mathscr{R}(A)$, $\mathscr{R}(W_1) \perp \mathscr{R}(A)$, $\mathscr{R}(V_2) \subseteq \mathscr{R}(A^*)$, $\mathscr{R}(W_2) \perp \mathscr{R}(A^*)$, W_1 is of full rank, W_2 is of full rank and $\mathscr{R}(W_1) = \mathscr{R}(W_2)$, they were able to find an explicit expression for $(A + (V_1 + W_1)G(V_2 + W_2)^*)^+$ in terms of A^+, G, V_1, V_2, W_1, and W_2. Fill and Fishkind [5] exploit this result to find a neat relationship between the Moore–Penrose inverse of $A + B$ and the Moore–Penrose inverse of the individual terms A and B, provided that $\operatorname{rank}(A + B) = \operatorname{rank}(A) + \operatorname{rank}(B)$. Moreover, they also showed that this rank-additivity hypothesis cannot be avoided in any proof that uses the result by Riedel, since rank additivity is shown to be implied by the hypotheses of Riedel's theorem. Because $\operatorname{rank}(A - XY') = k - p$ while $\operatorname{rank}(A) + \operatorname{rank}(XY') \geqslant k$, the result of Fill and Fishkind does not apply to our situation, and consequently, neither does the result of Riedel. Kala and Klaczyński [9] established the representation of various generalized inverses of the sum of two (i) rectangular matrices of the form $A + BDC^*$, and (ii) Hermitian positive semi-definite matrices of the form $A + BDB^*$, where A and D are Hermitian positive semi-definite matrices, thereby extending the results of Riedel. However, by checking the conditions Kala and Klaczyński impose on the matrix $A + BDC^*$, it follows that their results are also not applicable to obtain the Moore–Penrose inverse of our matrix $A - XY^*$. For results on generalized inverses of a sum of two weakly bicomplementary matrices, we refer to Werner [21] and Jain et al. [8].

In Section 2 we present some preliminaries. Subsequently, we will shortly address the case where V is nonsingular in Section 3. Section 4 deals with the singular case, where we will first consider the specific situation where $A = I$. The general case then easily follows. In Sections 2–4, the matrices may be complex. Finally, Section 5 focuses on the eigenvalues and eigenvectors of the real matrix $D - xy'$ with D diagonal.

2. Some preliminaries

Let A be a $k \times p$ complex matrix and consider a $p \times k$ complex matrix X which satisfies one ore more of the following properties:

(1) $AXA = A$,
(2) $XAX = X$,

(3) XA is Hermitian,
(4) AX is Hermitian.

If X satisfies (1), then X is called a generalized inverse of A, denoted by $X = A^-$. If X satisfies both (1) and (2), then X is called a reflexive generalized inverse of A, which is denoted by $X = A_r^-$. If X satisfies the properties (1), (2), and (3), then we call X a left pseudoinverse of A, denoted by A_L^-, whereas we call X a right pseudoinverse of A, denoted by A_R^-, if it satisfies the properties (1), (2), and (4). Finally, if X satisfies all four properties, then X is called the Moore–Penrose inverse of A which we will denote by A^+. The Moore–Penrose inverse of a matrix is uniquely defined by (1)–(4). For textbooks on generalized inverses we refer to, for example, [1,14].

Lemma 1. *The matrix $A_L^- A A_R^-$ is the Moore–Penrose inverse of A.*

This lemma is easily proved by checking the four conditions the Moore–Penrose inverse has to satisfy [1, Chapter 1].

As already mentioned in Section 1, the matrix $Q = I - XY^*$ is idempotent if $Y^*X = I$. A typical example is $Y^* = X^+ = (X^*X)^{-1}X^*$. In this case $Q = I - XX^+ = I - X(X^*X)^{-1}X^*$ is the very familiar Hermitian, idempotent matrix to be denoted by Q_X: the orthogonal projector on the orthogonal complement of the column space of X.

Since we also use the properties of idempotent matrices, we recall the most important facts. A $k \times k$ complex matrix Q is idempotent if $Q^2 = Q$. In statistics and econometrics, Q will often also be Hermitian, but this is not necessary as we have remarked. If Q is idempotent, then $I - Q$ is also idempotent and $Q(I - Q) = (I - Q)Q = 0$. For idempotent matrices, the rank is equal to the trace.

3. The nonsingular case

Let A be a nonsingular complex $k \times k$ matrix and X and Y complex $k \times p$ matrices. It is well-known that the matrix

$$S = \begin{pmatrix} A & X \\ Y^* & I \end{pmatrix}$$

can be written as

$$S = \begin{pmatrix} I & 0 \\ Y^*A^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & I - Y^*A^{-1}X \end{pmatrix} \begin{pmatrix} I & A^{-1}X \\ 0 & I \end{pmatrix}. \tag{3.1}$$

This representation is very instructive, since it immediately follows that S is nonsingular if and only if $I - Y^*A^{-1}X$ is nonsingular. Moreover, Eq. (3.1) also shows that $|S| = |A|\,|I - Y^*A^{-1}X|$. If S is nonsingular, we know from the standard results on inverses of partitioned matrices that S^{11}, the upper left-hand block of S^{-1}, can be written in two ways:

$$S^{11} = A^{-1} + A^{-1}X(I - Y^*A^{-1}X)^{-1}Y^*A^{-1} \tag{3.2}$$
$$= (A - XY^*)^{-1}, \tag{3.3}$$

and we have a well-known expression for $(A - XY^*)^{-1}$, also known as the Sherman–Morrison–Woodbury formula. Henderson and Searle [7] gave an excellent review of the Sherman–Morrison–Woodbury formula and related formulas. If S is singular, it is tempting to replace the inverses by Moore–Penrose inverses. According to Corollary 4.4 from [13], we have the following result.

Theorem 1. *If A is a $k \times k$ complex matrix, X and Y are $k \times p$ complex matrices with $k \geqslant p$, and if*

$$\operatorname{rank}\begin{pmatrix} A \\ Y^* \end{pmatrix} = \operatorname{rank}(A, X) = \operatorname{rank} A = \operatorname{rank}(A - XY^*)$$

and

$$\operatorname{rank}(I - Y^*A^+X) = p,$$

then

$$(A - XY^*)^+ = A^+ + A^+X(I - Y^*A^+X)^{-1}Y^*A^+. \tag{3.4}$$

On account of Theorem 4.6 from [13], which originates from [11, p. 439], we know that we need $I - Y^*A^+X$ to be nonsingular in order to have results similar to (3.2) and (3.4).

However, in the sequel, we will focus on the case that $Y^*A^+X = I$. In particular, we already assumed that A is nonsingular. Note that theorem 1 also does not apply to the mixture case where $Y^*A^{-1}X \neq I$ and $|I - Y^*A^{-1}X| = 0$, as mentioned in Section 1.

4. The singular case

In this section we will be interested in obtaining the Moore–Penrose inverse of $V = A - XY^*$, where the $k \times k$ complex matrix A is nonsingular and the $k \times p$ complex matrices X and Y satisfy the condition $Y^*A^{-1}X = I$. As we observed, this implies that X and Y are of full column rank $p < k$. We will first derive the Moore–Penrose inverse in the special case that $A = I$ and then derive the more general result in a constructive way.

Let $Q = I - XY^*$ where X and Y are $k \times p$ matrices with $Y^*X = I$. Some useful properties are:

$$QX = 0, \qquad Y^*Q = 0, \tag{4.1}$$
$$QQ_X = Q, \qquad Q_YQ = Q, \tag{4.2}$$
$$Q_XQ = Q_X, \qquad QQ_Y = Q_Y, \tag{4.3}$$

$$QY = -Q_Y XY^*Y, \quad X^*Q = -X^*XY^*Q_X, \tag{4.4}$$

$$QQ = Q. \tag{4.5}$$

Obviously, we have from (4.2) and (4.5) that $QQ_XQ_YQ = QQ = Q$, so that Q_XQ_Y is a generalized inverse of Q. Next, we observe that $Q_YQQ_X = Q$. Hence $Q_XQ_YQQ_XQ_Y = Q_XQQ_Y = Q_XQ_Y$. Moreover, $Q_XQ_YQ = Q_XQ = Q_X$ and $QQ_XQ_Y = Q_Y$ are Hermitian matrices. These observations prove the following theorem.

Theorem 2. *Let $Q = I - XY^*$, where X and Y are $k \times p$ complex matrices with $Y^*X = I$. Then*

$$Q^+ = Q_XQ_Y.$$

We can now use Theorem 2 and Lemma 1 to prove our main result as given in the following theorem.

Theorem 3. *Let $V = A - XY^*$, where A is a nonsingular complex $k \times k$ matrix, and X and Y are $k \times p$ complex matrices with $Y^*A^{-1}X = I$. Define $K = A^{-1}X$ and $L = (A^{-1})^*Y$, then*

$$V^+ = Q_KA^{-1}Q_L$$

is the Moore–Penrose inverse of V.

Proof. We will derive the Moore–Penrose inverse by using a left and right pseudoinverse of V, cf. Lemma 1. Note that

$$\begin{aligned} A - XY^* &= A(I - A^{-1}XY^*) \\ &= (I - XY^*A^{-1})A. \end{aligned}$$

This suggests to consider $(I - A^{-1}XY^*)^+A^{-1}$ and $A^{-1}(I - XY^*A^{-1})^+$ to be denoted by V_L^- and V_R^-, respectively. Obviously, V_L^- is indeed a left pseudoinverse of V and V_R^- is a right pseudoinverse of V. Lemma 1 states that the Moore–Penrose of V can now be computed as $V^+ = V_L^-VV_R^-$. From Theorem 2 we know that $(I - A^{-1}XY^*)^+ = Q_KQ_Y$ and $(I - XY^*A^{-1})^+ = Q_XQ_L$. On account of (4.2), it follows that

$$\begin{aligned} V^+ &= Q_KQ_YA^{-1}(A - XY^*)A^{-1}Q_XQ_L \\ &= Q_KQ_Y(I - KY^*)A^{-1}Q_XQ_L \\ &= Q_K(I - KY^*)A^{-1}Q_XQ_L \\ &= Q_KA^{-1}(I - XL^*)Q_XQ_L \\ &= Q_KA^{-1}(I - XL^*)Q_L \\ &= Q_KA^{-1}Q_L. \qquad \square \end{aligned}$$

Taking $X = a$ and $Y = -b$, we have the result derived in [17]. If we compare the expression of Trenkler for the Moore–Penrose inverse of $A + ab'$ with V^+, then we see that our result is a straightforward generalization. We therefore could have guessed this solution and verify the four conditions the Moore–Penrose inverse has to satisfy, just as we did in the proof of Theorem 2. Anyway, the basic properties (4.1)–(4.5) of idempotent matrices like Q are needed. We think, however, that the proof as given above is nicer, because it is more constructive.

5. Eigenvalues and eigenvectors of $D - xy'$

Throughout this section we will only deal with real matrices and vectors. We will study the eigenvalues and eigenvectors of the matrix $D - xy'$, where $D = \mathrm{diag}(d)$ is a nonsingular diagonal real matrix and x and y are $k \times 1$ real vectors. Similar problems have been studied by Vermeulen [18], Klamkin [10], Wansbeek [19], Watson [20] and Trenkler [17]. Trenkler noted that we need not restrict ourselves to diagonal D. If we look at the eigenvalues of $A - xy'$, where A is a nonsingular symmetric matrix, then there exists an orthogonal matrix C, such that $A = C\Lambda C'$, where Λ is a nonsingular diagonal matrix with the eigenvalues of A along its diagonal. Since $A - xy'$ and $\Lambda - C'xy'C$ have the same eigenvalues, we might as well study the matrix $\Lambda - C'xy'C$. The condition of symmetry can be replaced by the requirement that A is similar to a diagonal matrix Λ, that is, $A = U\Lambda U^{-1}$ for some $k \times k$ matrix U. In this case, the eigenvalues of $A - xy'$ coincide with those of $\Lambda - U^{-1}xy'U$.

5.1. Eigenvalues

Consider the matrix $D - xy'$, where $D = \mathrm{diag}(d)$ is a nonsingular diagonal matrix and d, x and y are $k \times 1$ vectors. We are interested in the eigenvalues of this matrix. Inspired by Vermeulen [18], we now present the following theorem.

Theorem 4. *If $D = \mathrm{diag}(d)$ is a nonsingular diagonal $k \times k$ matrix and x and y are $k \times 1$ vectors with $x_i y_i \neq 0$, $i = 1, \ldots, k$, then*

$$|D - xy'| = (-1)^{s(x,y)}|D_{xy} - vv'|,$$

where

$$\begin{aligned}
S_x &= \mathrm{diag}(\mathrm{sgn}\, x_1, \ldots, \mathrm{sgn}\, x_k),\\
S_y &= \mathrm{diag}(\mathrm{sgn}\, y_1, \ldots, \mathrm{sgn}\, y_k),\\
D_{xy} &= DS_xS_y,\\
v &= \left(|x_1y_1|^{\frac{1}{2}}, \ldots, |x_ky_k|^{\frac{1}{2}}\right)',\\
s(x, y) &= \#\left\{i = 1, \ldots, k \mid x_iy_i < 0\right\}.
\end{aligned}$$

Proof. Let $D_x = \operatorname{diag}(|x_1|^{\frac{1}{2}}, \ldots, |x_k|^{\frac{1}{2}})$ and $D_y = \operatorname{diag}(|y_1|^{\frac{1}{2}}, \ldots, |y_k|^{\frac{1}{2}})$. Then

$$\begin{aligned}|D - xy'| &= |D_x|\,|D_x^{-1} D D_y^{-1} - D_x^{-1} xy' D_y^{-1}|\,|D_y| \\ &= |D_y D_x^{-1} D D_y^{-1} D_x - D_y D_x^{-1} xy' D_y^{-1} D_x| \\ &= |D - S_x vv' S_y| \\ &= |S_x|\,|S_x D S_y - vv'|\,|S_y| \\ &= (-1)^{s(x,y)} |D_{xy} - vv'|. \qquad \square\end{aligned}$$

If, for some index j, we have $x_j y_j = 0$, then we can expand $|D - xy'|$ along its j^{th} row or column:

$$|D - xy'| = d_j |D_{jj} - (xy')_{jj}|,$$

where D_{jj} and $(xy')_{jj}$ denote the matrices obtained by deleting the jth row and the jth column of D, respectively xy'. We can continue this process until none of the $x_i y_i$ is equal to zero and then apply Theorem 4 to the remaining part of the matrix $D - xy'$.

From Theorem 4, we see that $|D - xy'|$ is the same as $|D_{xy} - vv'|$, except possibly for a difference in sign. Likewise,

$$|\lambda I - (D - xy')| = (-1)^{s(x,y)} |(\lambda I - D) S_x S_y + vv'| \tag{5.1}$$

with S_x, S_y and v as defined in Theorem 4. Eq. (5.1) implies that if $S_x S_y = I$, that is if $x_i y_i > 0$ for $i = 1, \ldots, k$, then the roots of the characteristic equation of $D - xy'$ and those of the symmetric matrix $D - vv'$ are the same. If, on the other hand $S_x S_y = -I$, that is if $x_i y_i < 0$ for $i = 1, \ldots, k$, then the roots of the characteristic equation $D - xy'$ and those of the symmetric matrix $D + vv'$ are the same. Because the roots of a symmetric matrix are always real, we have shown that if all $x_i y_i$ have the same sign, then the eigenvalues of $D - xy'$ are real.

Theorem 5. *If $D = \operatorname{diag}(d)$ is a nonsingular diagonal $k \times k$ matrix and x, y are $k \times 1$ vectors such that $x_i y_i < 0$ for $i = 1, \ldots, k$ or $x_i y_i > 0$ for $i = 1, \ldots, k$, then $Q = D - xy'$ has real eigenvalues.*

Vermeulen [18] showed that the roots of the determinantal equation

$$|\lambda I + \operatorname{diag}(e) + a\iota'|, \tag{5.2}$$

with all a_i and e_i strictly positive are real. By constructing a difference equation for the determinantal equation, he also showed that these roots are negative. These results immediately follow from Theorem 4, because

$$|\lambda I + \operatorname{diag}(e) + a\iota'| = |\lambda I + \operatorname{diag}(e) + vv'|,$$

with $v = (\sqrt{a_1}, \ldots, \sqrt{a_k})'$, so that the eigenvalues of $\operatorname{diag}(e) + a\iota'$ are the same as the eigenvalues of the symmetric matrix $\operatorname{diag}(e) + vv'$. These eigenvalues are positive, because $\operatorname{diag}(e) + vv'$ is positive definite. This implies that the roots of the determinantal equation (5.2) are real and negative.

Trenkler [17] remarked in his paper that the matrix $T = A + ab'$, with A being symmetric and nonsingular, always has real eigenvalues. However, the condition of Theorem 5 that all $a_i b_i$ should have the same sign is vital more or less, as can be seen from the following example.

Example 1. Consider the matrix $D - xy'$ with $D = \operatorname{diag}(1, 2)$, $x = (-1, -1)'$ and $y = (1, -3)'$. It is easily derived that in this case $|\lambda I - (D - xy')| = \lambda^2 - \lambda + 1$, so that both eigenvalues are complex.

The eigenvalues of $D - xy'$, with D nonsingular, can be determined from the characteristic equation

$$\begin{aligned} 0 &= |\lambda I - (D - xy')| \\ &= |(\lambda I - D)(I + (\lambda I - D)^{-1} xy')| \\ &= |\lambda I - D|(1 + y'(\lambda I - D)^{-1} x) \\ &= \left[\prod_{i=1}^{k} (\lambda - d_i)\right]\left[1 + \sum_{i=1}^{k} \frac{x_i y_i}{\lambda - d_i}\right] \\ &= \prod_{i=1}^{k} (\lambda - d_i) + \sum_{i=1}^{k} x_i y_i \prod_{j \neq i} (\lambda - d_j), \end{aligned} \tag{5.3}$$

a result which was also given by Graybill [6] in Theorem 8.5.2. The following theorem immediately follows from (5.3). It covers Theorems 2, 3, and 4 of [17].

Theorem 6. *Consider $D - xy'$, where $D = \operatorname{diag}(d)$ is a nonsingular diagonal matrix and x and y are $k \times 1$ vectors.*

(i) *If all d_i are different and all $x_i y_i \neq 0$, then none of the d_i is an eigenvalue of $D - xy'$. In this case, λ is an eigenvalue if and only if*

$$1 + \sum_{i=1}^{k} \frac{x_i y_i}{\lambda - d_i} = 0. \tag{5.4}$$

(ii) *If $x_j y_j = 0$, then d_j is an eigenvalue of $D - xy'$.*

(iii) *If some of the d_i's coincide, then d_i is an eigenvalue of $D - xy'$.*

Note that we can find all eigenvalues of $D - xy'$ by combining the different cases considered in this theorem. Suppose d_j has multiplicity $k_j \geqslant 1$. Without loss of generality we assume that D is partitioned as follows:

$$D = \operatorname{diag}(d_1 I, d_2 I, \ldots, d_r I), \tag{5.5}$$

where d_jI is of order $k_j \times k_j$ and $\sum_{j=1}^r k_j = k$. We partition x and y accordingly:

$$\begin{aligned} x &= (x'_{k_1}, x'_{k_2}, \dots, x'_{k_r})', \\ y &= (y'_{k_1}, y'_{k_2}, \dots, y_{k'_r})'. \end{aligned} \tag{5.6}$$

Eq. (5.3) gives

$$\begin{aligned} |\lambda I - (D - xy')| &= \prod_{j=1}^r (\lambda - d_j)^{k_j} + \sum_{j=1}^r (x'_{k_j} y_{k_j})(\lambda - d_j)^{k_j - 1} \prod_{h \neq j} (\lambda - d_h)^{k_h} \\ &= \left[\prod_{j=1}^r (\lambda - d_j)^{k_j - 1} \right] \\ &\quad \times \left[\prod_{j=1}^r (\lambda - d_j) + \sum_{j=1}^r (x'_{k_j} y_{k_j}) \prod_{h \neq j} (\lambda - d_h) \right]. \end{aligned} \tag{5.7}$$

From (5.7) we observe that in this case $\lambda = d_j$ has at least multiplicity $k_j - 1$. Moreover, the remaining eigenvalues can be found by putting the second factor on the right-hand side of (5.7) equal to zero. The equation to be solved is then exactly of the type (5.3), so that the remaining eigenvalues can be determined from (5.4). Note that $\lambda = d_j$ can have multiplicity k_j if and only if $x'_{k_j} y_{k_j} = 0$.

If all d_i coincide, that is, if $D = dI$, then Eq. (5.3) gives

$$|\lambda I - (D - xy')| = (\lambda - d)^{k-1} \left[(\lambda - d) + \sum_{i=1}^k x_i y_i \right] = 0,$$

so that $\lambda = d$ is an eigenvalue with multiplicity $k - 1$ at least, and the remaining eigenvalue equals $d - y'x$, which is equal to d if $y'x = 0$ and in this case $\lambda = d$ has multiplicity k.

In the situation that all d_i are different and all $x_i y_i \neq 0$, case (i) of Theorem 6, the eigenvalues of $D - xy'$ can be determined by solving Eq. (5.4). Note that, in this case, $\lambda = 0$ is a solution of (5.4) if and only if $y'D^{-1}x = 1$, that is, $D - xy'$ is singular. Moreover, if $y'D^{-1}x = 1$, Eq. (5.3) simplifies to

$$\begin{aligned} 0 &= |\lambda I - (D - xy')| \\ &= |\lambda I - D| \left(1 + y'(\lambda I - D)^{-1} x \right) \\ &= |\lambda I - D| \, (y'D^{-1}(\lambda I - D)(\lambda I - D)^{-1} x + y'(\lambda I - D)^{-1} x) \\ &= |\lambda I - D| \, (y'(\lambda D^{-1} - I)(\lambda I - D)^{-1} x + y'(\lambda I - D)^{-1} x) \\ &= |\lambda I - D| \, \lambda y' D^{-1} (\lambda I - D)^{-1} x. \end{aligned}$$

Thus, apart from Eq. (5.4), the eigenvalues different from zero also satisfy

$$\sum_{i=1}^{k} \frac{x_i y_i}{d_i(\lambda - d_i)} = 0.$$

Let us assume that $d_1 < d_2 < \cdots < d_k$. Just as Klamkin [10], Wansbeek [19] and Trenkler [17], we want to pay attention to the location of the eigenvalues for this special case. We restrict ourselves to the situation in which all $x_i y_i$ have the same sign, because then the eigenvalues are real. So, we are dealing with case (i) of Theorem 6. In line with Klamkin [10], consider the graph of

$$f(\lambda) = 1 + \frac{x_1 y_1}{\lambda - d_1} + \frac{x_2 y_2}{\lambda - d_2} + \cdots + \frac{x_k y_k}{\lambda - d_k}. \tag{5.8}$$

This graph is continuous except at the points $\lambda = d_1, d_2, \ldots, d_k$, which correspond to vertical asymptotes. It follows by continuity that there are k real roots such that an eigenvalues lies between every two successive d_i. If $x_i y_i > 0$ for $i = 1, \ldots, n$, this implies that

$$\lambda_1 < d_1 < \lambda_2 < \cdots < d_{k-1} < \lambda_k < d_k. \tag{5.9}$$

If $x_i y_i < 0$ for $i = 1, \ldots, n$, then, by using a similar argument, we observe that

$$d_1 < \lambda_1 < d_2 < \cdots < \lambda_{k-1} < d_k < \lambda_k. \tag{5.10}$$

A typical graph for $k = 3$ is shown in Fig. 1.

Now we assume $d_1 \leqslant d_2 \leqslant \cdots \leqslant d_k$. In case of situation (ii) or (iii) of Theorem 6, where some of the d_j coincide or $x_j y_j = 0$, we know that d_j is an eigenvalue of $D - xy'$. The other eigenvalues are located as before, so that, if all $x_i y_i$ are nonnegative, then

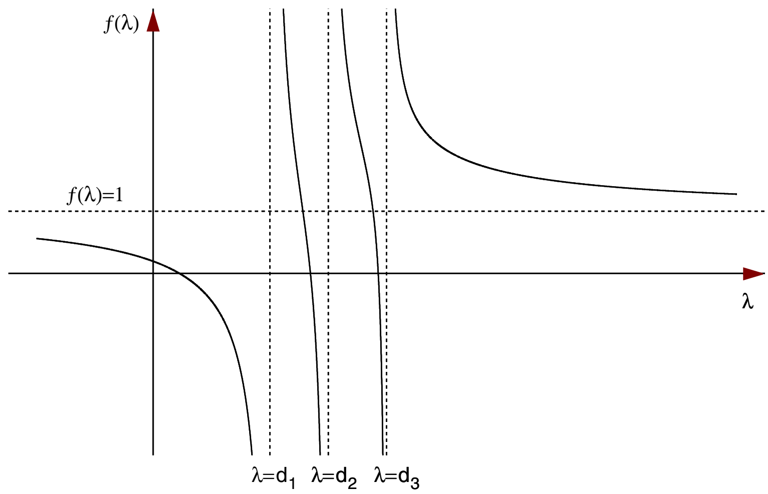

Fig. 1. Example of the roots of the characteristic equation $|\lambda I - (D - xy')| = 0$ for $k = 3$ and $x_i y_i > 0, i = 1, 2, 3$.

$$\lambda_1 \leqslant d_1 \leqslant \lambda_2 \leqslant \cdots \leqslant d_{k-1} \leqslant \lambda_k \leqslant d_k. \tag{5.11}$$

If all $x_i y_i$ are nonpositive, we obtain the inequality

$$d_1 \leqslant \lambda_1 \leqslant d_2 \leqslant \cdots \leqslant \lambda_{k-1} \leqslant d_k \leqslant \lambda_k. \tag{5.12}$$

To show that we cannot locate the eigenvalues among the d_i as easily as in Eq. (5.12) if we do not impose any restrictions on the sign of the $x_i y_i$, consider the following example.

Example 2. We continue with Example 1 and consider the graph of (5.8), that is, the graph of

$$f(\lambda) = 1 - \frac{1}{\lambda - 1} + \frac{3}{\lambda - 2}.$$

We know that there are no real roots, so that the graph never intersects the x-axis. The corresponding graph is shown in Fig. 2.

5.2. *Eigenvectors*

Watson [20] considered the eigenvectors of the matrix $\Pi - \pi\pi'$, with $\Pi = \mathrm{diag}(\pi)$ and $\sum \pi_i = 1$. Somewhat more general, we will consider in this section the eigenvectors of $D - xy'$.

We begin with the case where all d_i are different and all $x_i y_i \neq 0$, case (i) of Theorem 6. The elements of the eigenvector $v_j = (v_{1j}, \ldots, v_{kj})'$ of $D - xy'$ corresponding to λ_j must satisfy

$$d_i v_{ij} - (y'v_j)x_i = \lambda_j v_{ij}, \quad i = 1, \ldots, k,$$

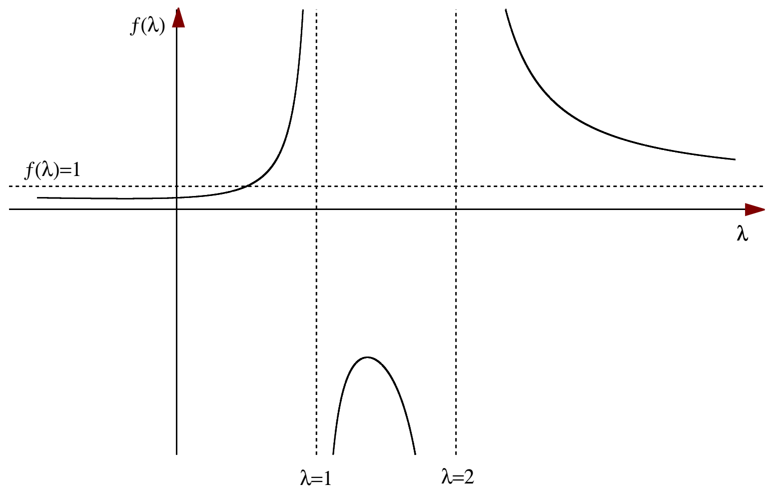

Fig. 2. Graph corresponding to Example 2.

so that

$$v_{ij} = (y'v_j)\frac{x_i}{d_i - \lambda_j}. \tag{5.13}$$

Note that $\lambda_j \neq d_i$. In order to have a true eigenvector, we must have $y'v_j \neq 0$. If we choose

$$v_{ij} = \frac{x_i}{d_i - \lambda_j} \neq 0, \tag{5.14}$$

then

$$y'v_j = \sum_{i=1}^{k} \frac{x_i y_i}{d_i - \lambda_j} = 1,$$

because λ_j satisfies (5.4). This shows that we can indeed find the elements of the eigenvector v_j by means of Eq. (5.14). Note that we only need the vector y to determine the eigenvectors of $D - xy'$ via the eigenvalues of $D - xy'$.

Again we assume that all d_i are different. If $x_j y_j = 0$, the second case of Theorem 6, then d_j is an eigenvalue of $D - xy'$. First, consider the situation that $y_j = 0$ and x_j is arbitrary. It is straightforward to show that in this case e_j, the jth unit vector, is an eigenvector corresponding to d_j. If, on the other hand $x_j = 0$ and $y_j \neq 0$, then the elements of the eigenvector $v = (v_1, \ldots, v_k)'$ corresponding to d_j must satisfy

$$d_i v_i - (y'v)x_i = d_j v_i, \quad i = 1, \ldots, k. \tag{5.15}$$

For $i \neq j$ this leads to

$$v_i = \frac{(y'v)x_i}{d_i - d_j},$$

so that

$$y'v = y_j v_j + (y'v)\sum_{i \neq j} \frac{x_i y_i}{d_i - d_j},$$

and v_j follows:

$$v_j = \frac{y'v}{y_j}\left(1 - \sum_{i \neq j} \frac{x_i y_i}{d_i - d_j}\right).$$

Because $y'v$ is a constant factor, this shows that we can choose v such that

$$v_i = \frac{x_i}{d_i - d_j} \quad \text{for } i \neq j,$$

$$v_j = \frac{1}{y_j}\left(1 - \sum_{i \neq j} \frac{x_i y_i}{d_i - d_j}\right).$$

Note that $y'v = 1$ with this choice of v.

In case (iii) of Theorem 6, where some of the d_i coincide, partition D, x and y as in (5.5) and (5.6). We know from Theorem 6 that if $k_j > 1$, then d_j is an eigenvalue of $D - xy'$. We partition an eigenvector v corresponding to d_j in a similar fashion as in (5.6). Assume that $x_{k_j} \neq 0$ and $y_{k_j} \neq 0$. It is easy to show that in this case, $v = (v_{k_1}, v_{k_2}, \ldots, v_{k_r})'$ consists of zeros, except for the subvector v_{k_j}. This subvector must satisfy $y'_{k_j} v_{k_j} = 0$. This implies that in this case, d_j has $k_j - 1$ eigenvectors of the form $(0, \ldots, 0, v_{k_j}, 0 \ldots, 0)'$, where the v_{k_j} are orthogonal to $y_{k_j} \neq 0$ and are also mutually orthogonal. If $x_{k_j} = 0$ or $y_{k_j} = 0$, then d_j has multiplicity k_j and we can take for v_{k_j} the $k_j \times 1$ unit vectors $e_1, \ldots, e_{k_j}$.

Acknowledgments

The authors would like to thank Michel van de Velden for providing the reference Trenkler [17] and Amy Wong for her reading of some earlier draft versions and helpful discussions. We are also grateful to an anonymous referee for several remarks to improve the paper.

References

[1] T.L. Bouillion, P.L. Odell, Generalized Inverse Matrices, John Wiley, New York, 1971.
[2] P.M.C. De Boer, R. Harkema, Maximum likelihood estimation of sum-constrained linear models when samples are small, Kwantitatieve Methoden 16 (1984) 97–108.
[3] W. Dol, Small area estimation; a synthesis between sampling theory and econometrics. PhD thesis, Faculty of Economics, University of Groningen, Groningen, 1991.
[4] W. Dol, T. Steerneman, T. Wansbeek, Matrix algebra and sampling theory: the case of the Horvitz-Thompson estimator, Linear Algebra Appl. 237/238 (1996) 225–238.
[5] J.A. Fill, D.E. Fishkind, The Moore–Penrose generalized inverse for sums of matrices, SIAM J. Matrix Anal. Appl. 21 (2) (2000) 629–635.
[6] F.A. Graybill, Matrices with Applications in Statistics, Wadsworth International Group, Belmont, CA, 1983.
[7] H.V. Henderson, S.R. Searle, On deriving the inverse of a sum of matrices, SIAM Rev. 23 (1) (1981) 53–60.
[8] S.K. Jain, S.K. Mitra, H.J. Werner, Extensions of $\mathscr{G}$-based matrix partial orders, SIAM J. Matrix Anal. Appl. 17 (1996) 834–850.
[9] R. Kala, K. Klaczyńiski, Generalized inverses of a sum of matrices, Sankhyā, Series A 56 (1994) 458–464.
[10] M.S. Klamkin, On the roots of a certain determinantal equation, Math. Gazette 54 (387) (1970) 57–58.
[11] G. Marsaglia, G.P.H. Styan, Rank conditions for generalized inverses of partitioned matrices. Sankhyā, Series A 36 (1974) 347–442.
[12] H. Neudecker, Mathematical properties of the variance of the multinomial distribution, J. Math. Anal. Appl. 189 (1995) 757–762.
[13] D.V. Ouellette, Schur complements and statistics, Linear Algebra Appl. 36 (1981) 187–295.
[14] C.R. Rao, S.J. Mitra, Generalized Inverse of Matrices and its Applications, John Wiley, New York, 1971.

[15] K.S. Riedel, A Sherman–Morrison–Woodbury identity for rank augmenting matrices with application to centering, SIAM J. Matrix Anal. Appl. 13 (2) (1992) 659–662.
[16] K. Tanabe, M. Sagae, An exact Cholesky decomposition and the generalized inverse of the variance–covariance matrix of the multinomial distribution, with applications, J. Roy. Statist. Soc. B 54 (1992) 211–219.
[17] G. Trenkler, On a generalisation of the covariance matrix of the multinomial distribution, in: R.D.H. Heijmans, D.S.G. Pollock, A. Satorra, (Eds.), Innovations in Multivariate Statistical Analyses; A Festschrift for Heinz Neudecker, 2000.
[18] L.A. Vermeulen, The solution of a certain polynomial equation, Math. Gazette 51 (1967) 308–309.
[19] T. Wansbeek, Singuliere covariantiematrices en SUR-modellen: Enige opmerkingen over de specificatie van De Boer en Harkema, Kwantitatieve Methoden 18 (1985) 99–102.
[20] G.S. Watson, Spectral decomposition of the covariance matrix of a multinomial, J. Roy. Statist. Soc. B 58 (1996) 289–291.
[21] H.J. Werner, Generalized inversion and weak bi-complementarity, Linear and Multilinear Algebra 19 (1986) 357–372.