189

Manifestations of the Schur Complement*

RICHARD W. COTTLE
Stanford University
Stanford, California

Recommended by Gene Golub

ABSTRACT

This expository paper describes the ways in which a matrix theoretic construct called the Schur complement arises. Properties of the Schur complement are shown to have use in computing inertias of matrices, covariance matrices of conditional distributions, and other information of interest.

1. INTRODUCTION

In recent years, the designation "Schur complement" has been applied to any matrix of the form $D - CA^{-1}B$. These objects have undoubtedly been encountered from the time matrices were first used. But today under this new name and with new emphasis on their properties, there is greater awareness of the widespread appearance and utility of Schur complements. The purpose of this paper is to highlight some of the many ways that Schur complements arise and to illustrate how their properties assist one in efficiently computing inertias of matrices, covariance matrices of conditional distributions, and other important information.

Why Schur? Complement of what? The full definition, introduced by E. V. Haynsworth [15], answers the second question. Over any field,

if A is a nonsingular leading submatrix[1] of the block matrix

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \qquad (1)$$

then $D - CA^{-1}B$ is the *Schur complement of A in M* and is denoted by (M/A). The name Schur is suggested by the well known determinantal formula (for the case where M is square)

$$\det M = \det A \det(D - CA^{-1}B). \qquad (2)$$

This relation was remarked in 1917 by I. Schur [26, p. 217] within the proof of a matrix-theoretic lemma. Gantmacher [14, p. 46] refers to it as one of the "formulas of Schur."

Matrices of the form $D - CA^{-1}B$ are very common; perhaps their most frequently encountered manifestation is in ordinary or "generalized" Gaussian elimination [14, p. 45]. For instance, consider a system of linear equations $Mz = 0$ where M has been partitioned as in Eq. (1) and z has been partitioned conformally into the direct sum of x and y. The system then becomes

$$Ax + By = 0, \qquad (3a)$$

$$Cx + Dy = 0. \qquad (3b)$$

Recall that we assumed the matrix A (called the *pivot block*) is nonsingular; hence elimination of x from Eq. (3b) is legitimate and yields the system

$$(D - CA^{-1}B)y = 0.$$

Thus, the matrix of coefficients in the reduced form of (3b) is just (M/A).

Closely related to the system (3) and the associated pivot operation is the system

---

[1] Having the nonsingular matrix A in the upper left-hand corner of M is merely for notational convenience. The lower left-hand corner would be equally good in this respect. Indeed, one can consider the Schur complement of any nonsingular submatrix A in the ambient matrix M. However, suitable permutations of the rows and columns of M can be used to shift A to one of these corners. This amounts to pre- and post-multiplication of M by permutation matrices. When M is square and P is a permutation matrix of the same order, PMP^T is called a *principal rearrangement* of M. Such a matrix is obtained by "simultaneous" permutation of the rows and columns of M.

$$u = Ax + By, \tag{3'a}$$

$$v = Cx + Dy. \tag{3'b}$$

If A is nonsingular, one can solve the system (3') for x and v in terms of u and y. Doing this one obtains

$$x = A^{-1}u - A^{-1}By, \tag{3''a}$$

$$v = CA^{-1}u + (D - CA^{-1}B)y. \tag{3''b}$$

When

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

is square, the operation transforming it to

$$\bar{M} = \begin{pmatrix} A^{-1} & -A^{-1}B \\ CA^{-1} & D - CA^{-1}B \end{pmatrix},$$

is called a *principal pivot* [29] or a *gyration* [11]. But whether M is square or rectangular, the Schur complement shows up again.

Next, let us suppose M is square and nonsingular. If M can be partitioned as in Eq. (1) with A and D nonsingular, then

$$M^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}. \tag{4}$$

The inverses of the Schur complements (M/A) and (M/D) exist by virtue of Eq. (2) and the nonsingularity assumptions imposed above. The correctness of Eq. (4) can be verified by multiplication. Its usefulness will be exemplified subsequently.

The Schur complement has also been generalized by Carlson, Haynsworth, and Markham [8] in terms of the Moore-Penrose inverse. Another development of this type has been carried out by W. N. Anderson, Jr. [2] who uses the term *shorted operator* instead of generalized Schur complement. Interest in the latter stems from electrical network theory. See [3], [4], and [12].

2. THE QUOTIENT PROPERTY

This brief section is concerned with a particularly nice property of the Schur complement called the quotient property [9]. It says that if

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} E & F \\ G & H \end{bmatrix},$$

where A and E are nonsingular, then

$$(M/A) = [(M/E)/(A/E)]. \tag{5}$$

As mentioned once above, the nonsingularity of (A/E) follows from that of A and E via Schur's determinantal formula (2). Moreover, it turns out that (A/E) is the leading block of (M/E) so the grand Schur complement on the right-hand side of Eq. (5) is well-defined. One implication of the quotient property is that under the given hypotheses on A and E, the calculation of the Schur complement of A in M can be carried out in two stages.

We have already briefly noted the appearance of the Schur complement in Gaussian elimination. It is not surprising then to find Schur complements in the pivotal algebra of mathematical programming. See Tucker [28], Parsons [24], and Wendler [31] in this regard. Indeed, Theorem 6 in Tucker's paper can be construed as a precursor of the Crabtree-Haynsworth Theorem [9] announcing the quotient formula (5). Recently, Ostrowski [23] has published a "new" proof. The Schur complement terminology and notation aside, there is no telling how long this relationship has been known.[2]

3. A CONNECTION WITH STATISTICS

The multivariate normal distribution in mathematical statistics provides a magnificent example of how the Schur complement and the quotient property arise naturally.

Let S denote a symmetric positive definite (here we are working over the real field) matrix of order n and let $\bar{z}$ represent a fixed n-vector. Then

$$f(z) = \left(\frac{\det S^{-1}}{(2\pi)^n} \right)^{1/2} \exp\{- \tfrac{1}{2}(z - \bar{z})^T S^{-1}(z - \bar{z})\}$$

is the density function for the n-variate normal distribution with vector mean $\bar{z}$ and covariance matrix S. (See [1] and [20].)

---

[2] The reader may like to know of an amusing coincidence: Immediately preceding Schur's article [26] in *J. Reine Angew. Math.* 147 (1917) there appears the paper "Über sogenannte perfekte Körper" by A. Ostrowski. Thus these two papers appeared 54 years before Ostrowski [23].

The matrix S and its principal submatrices, being positive definite, are all invertible. A formula for the inverse, akin to the one given in Eq. (4), is therefore applicable to S; moreover, as we shall see shortly, extra benefits accrue from the information that S is symmetric.

Now assume

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{and} \quad \bar{z} = \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix},$$

where x and $\bar{x}$ belong to R^{n_1}, y and $\bar{y}$ belong to R^{n_2}. With the matrix S partitioned conformally as

$$S = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix},$$

it can be shown that the density for the (marginal) distribution of y is given by

$$f(y) = \left(\frac{\det C^{-1}}{(2\pi)^{n_2}}\right)^{1/2} \exp\{-\tfrac{1}{2}(y - \bar{y})^T C^{-1}(y - \bar{y})\}.$$

To come now to the point, the conditional distribution of x given y has density

$$f(x|y) = \frac{f(x, y)}{f(y)},$$

which is multivariate normal of dimension n_1 with mean vector

$$\bar{x} + BC^{-1}(y - \bar{y}),$$

and covariance matrix

$$(S/C) = A - BC^{-1}B^T.$$

Getting the right coefficient in front of the exponential is chiefly a matter of observing that by the counterpart to Eq. (2),

$$\frac{\det S^{-1}}{\det C^{-1}} = \frac{\det C}{\det S} = \frac{1}{\det(S/C)} = \det(S/C)^{-1}.$$

The quotient of the exponentials reduces to looking at the difference of two quadratic forms. This is where the inverse formula and the symmetry of S enter. The following lemma, which makes no use of positive definite-

ness, pinpoints the relationship between the required difference of quadratic forms and the Schur complement (S/C).

LEMMA. *Let the matrix*

$$M = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

be symmetric and nonsingular. If A and C are also nonsingular then

$$
\begin{aligned}
Q(u, v) &\equiv \begin{pmatrix} u \\ v \end{pmatrix}^T \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}^{-1} \begin{pmatrix} u \\ v \end{pmatrix} - v^T C^{-1} v \\
&= (u - BC^{-1}v)^T (A - BC^{-1}B^T)^{-1}(u - BC^{-1}v) \\
&\equiv Q^*(u - BC^{-1}v).
\end{aligned}
\tag{6}
$$

Proof. From the formula (4) for the inverse of M, it follows that

$$
\begin{aligned}
Q(u, v) = u^T(A - BC^{-1}B^T)^{-1}u - 2u^T A^{-1}B(C - B^T A^{-1}B)^{-1}v \\
+ v^T(C - B^T A^{-1}B)^{-1}v - v^T C^{-1}v.
\end{aligned}
$$

The symmetry of M implies that

$$
\begin{aligned}
Q^*(u - BC^{-1}v) &= u^T(A - BC^{-1}B^T)^{-1}u - 2u^T(A - BC^{-1}B^T)^{-1}BC^{-1}v \\
&\quad + v^T C^{-1}B^T(A - BC^{-1}B^T)^{-1}BC^{-1}v, \\
&= u^T(A - BC^{-1}B^T)^{-1}u - 2u^T C^{-1}B(C - B^T A^{-1}B)^{-1}v \\
&\quad + v^T C^{-1}B^T(A - BC^{-1}B^T)^{-1}BC^{-1}v.
\end{aligned}
$$

The first two major terms of $Q(u, v)$ equal those of $Q^*(u - BC^{-1}v)$. The remainder of the proof consists of showing that

$$(C - B^T A^{-1}B)^{-1} - C^{-1} = C^{-1}B^T(A - BC^{-1}B^T)^{-1}BC^{-1}.$$

This identity is well known and is attributed to Woodbury [32]. Again, from the symmetry of M, it is equivalent to showing that

$$(C - B^T A^{-1}B)^{-1} - C^{-1} = C^{-1}B^T A^{-1}B(C - B^T A^{-1}B)^{-1}.$$

However, this follows from

$$(C - B^T A^{-1}B)(C - B^T A^{-1}B)^{-1} = I.$$

The proof is now complete. ■

In this setting, there is an interesting interpretation of the quotient formula. Consider the normally distributed vector

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad \text{with mean} \quad \begin{pmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{pmatrix} \quad \text{and covariance matrix} \quad S = \begin{bmatrix} A & B & C \\ B^T & D & E \\ C^T & E^T & F \end{bmatrix},$$

The marginal distribution of $\begin{pmatrix} y \\ z \end{pmatrix}$ has mean $\begin{pmatrix} \bar{y} \\ \bar{z} \end{pmatrix}$ and covariance matrix

$$G = \begin{bmatrix} D & E \\ E^T & F \end{bmatrix}.$$

The conditional distribution of x given y and z has covariance matrix (S/G). The quotient formula

$$(S/G) = \frac{(S/F)}{(G/F)}$$

says that the conditional distribution of $\left[x \text{ given } \begin{pmatrix} y \\ z \end{pmatrix} \right]$ is that of $\left[\begin{pmatrix} x \\ y \end{pmatrix} \text{given} z \right]$ given $[y$ given $z]$. (See T. W. Anderson [1, pp. 33–34].)

4. USAGE IN A DETERMINANTAL TEST

In some circumstances, it is desirable to know whether the leading principal minors of a square matrix $M = [m_{ij}]$ are all nonzero (or, of a particular sign, say positive). This can be determined by pivoting and use of the quotient formula.

Let M be of order n and denote by $M[1, \ldots, k]$ its leading principal submatrix of order k where $k = 1, \ldots, n$:

$$M[1, \ldots, k] = \begin{bmatrix} m_{11} & \cdots & m_{1k} \\ \vdots & & \vdots \\ m_{k1} & \cdots & m_{kk} \end{bmatrix}.$$

Now suppose m_{11} is nonzero (positive). Using m_{11} as the pivot leads to the Schur complement

$$M^{(1)} = (M/m_{11}),$$

in which the leading entry is

$$m_{11}^{(1)} = m_{22} - \frac{m_{12}m_{21}}{m_{11}} = \frac{M[1,2]}{M[1]}.$$

Moreover,

$$m_{11}^{(1)} = \det m_{11}^{(1)} = \frac{\det M[1,2]}{\det M[1]}.$$

In general, if the procedure is not interrupted by the discovery of a leading entry (i.e., leading principal minor) of zero (nonpositive) value, then after $k < n$ steps

$$m_{11}^{(k)} = \frac{M[1,\ldots,k+1]}{M[1,\ldots,k]}$$

and

$$m_{11}^{(k)\cdot} = \det m_{11}^{(k)} = \frac{\det M[1,\ldots,k+1]}{\det M[1,\ldots,k]}.$$

It should be emphasized here that at each stage the entries of the new Schur complement are easily computed from the matrix currently at hand. This is done by pivoting just as in the case of Gaussian elimination. With $M = M^{(0)}$, the individual entries of $M^{(k)}$ are given by the formula

$$m_{ij}^{(k)} = m_{i+1,j+1}^{(k-1)} - \frac{m_{i+1,1}^{(k-1)}m_{1,j+1}^{(k-1)}}{m_{11}^{(k-1)}}.$$

The procedure above has an obvious application to the well known determinantal test for positive definiteness. (See Stiefel [27, p. 68].) If M is symmetric, it is *positive definite* if and only if its leading principal minors are all positive. Notice, it is not really necessary to know the *values* of these determinants: only their *signs* matter.

The method can also be used to ascertain whether a square (not necessarily symmetric) matrix has the so-called *Minkowski property.* This is the case when its off-diagonal entries are nonpositive and all its principal minors are positive. Checking the signs of the off-diagonal entries is no problem, but for a large matrix, verifying the positivity of every principal minor is a disagreeable task. Fortunately, one can do less. Only the *leading* principal minors need be checked for positivity. This is proved by Fiedler and Pták [13, p. 387] in an omnibus theorem on this subject.

5. THE INERTIA FORMULA

The second property of the Schur complement to be discussed and applied here concerns the concept of inertia. The formulation is exclusively in terms of real symmetric matrices. With a little extra effort, the case of complex Hermitian matrices can also be covered.

The *inertia* of a real symmetric $n \times n$ matrix M is a triple

$$\text{In } M = [\pi(M), \nu(M), \delta(M)],$$

where $\pi(M)$ = number of positive eigenvalues of M, $\nu(M)$ = number of negative eigenvalues of M, $\delta(M)$ = number of zero eigenvalues of M.

The three components of In M are sometimes called the *positivity*, *negativity* and *nullity* of M, respectively. They are related to the *rank* $\rho(M)$ and *signature* $\sigma(M)$ of M through the equations

$$\rho(M) = \pi(M) + \nu(M),$$

$$\sigma(M) = \pi(M) - \nu(M).$$

The inertia of M can be inferred from knowledge of its rank, signature, and order.

The inertia of a nonsingular matrix and of its inverse are the same since their eigenvalues are all nonzero and reciprocals of each other. The inertias of similar matrices are the same because their eigenvalues are identical. The inertias of congruent matrices are the same; this is just another way of stating Sylvester's famous Law of Inertia.

Now, if

$$M = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

and A is nonsingular, the facts on inverses and congruent matrices cited in the preceding paragraph can easily be used to show that

$$\text{In } M = \text{In } A + \text{In}(M/A). \tag{7}$$

This form of the *inertia formula* is proved by Haynsworth in [15, p. 75]. It is used there for obtaining inertia results for partitioned matrices.

In a later work, Haynsworth and Ostrowski [16, p. 302] prove a special inertia formula for matrices of the form

$$M = \begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix}, \quad \begin{array}{l} A \text{ arbitrary but symmetric,} \\ B \text{ nonsingular of order } k. \end{array} \tag{8}$$

They attribute to Carlson and Schneider [7] the observation that

$$\text{In} \begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix} = (k, k, 0), \tag{9}$$

and then give an alternate proof. In the next paragraph, we sketch another alternate proof that seems to retain a little bit more of the Schur complement flavor.

   Alternate Proof of Eq. (9). We use induction on k. For $k = 1$, the formula is an easy consequence of the definition of an eigenvalue. Assume it is true for B of order less than k. When B has order k, we can isolate a 2×2 principal submatrix of the form

$$C = \begin{bmatrix} a_{rr} & b_{rs} \\ b_{rs} & 0 \end{bmatrix}, \qquad b_{rs} \neq 0.$$

As remarked in Sec. 1, it is not necessary for C to be a leading principal submatrix in order to have a Schur complement. In the present circumstances, it appears advantageous to perform a principal rearrangement of M making it permissible to assume $r = k$, $s = 1$. Hence (M/C) is found by first pivoting on $b_{rs} = c_{12} = m_{k,k+1} \neq 0$ and then on the element that corresponds to c_{21} in (M/c_{12}). Because of the block of zeros in M, these operations produce a new matrix, the Schur complement (M/C), having the same block form

$$\tilde{M} = \begin{bmatrix} \tilde{A} & \tilde{B} \\ \tilde{B}^T & 0 \end{bmatrix}, \qquad \begin{array}{l} \tilde{A} \text{ symmetric,} \\ \tilde{B} \text{ nonsingular of order } k - 1 \end{array}$$

as M. The formula (9) now follows from the inertia formula, the case for $k = 1$, and the inductive hypothesis. ■

7. AN ALGORITHM FOR COMPUTING INERTIA

   The inertia formula engenders a nice algorithm for computing the inertia of a real symmetric matrix, say M. Notice M has exactly one of the following properties:

   (P1)   diag $M \neq 0$;

   (P2)   diag $M = 0$, $M \neq 0$;

   (P3) $M = 0$.

In case (P1) holds, we may use principal rearrangement of M to guarantee that $m_{11} \neq 0$. Then

$$\text{In } m_{11} = \begin{cases} (1, 0, 0), & \text{if } m_{11} > 0, \\ (0, 1, 0), & \text{if } m_{11} < 0. \end{cases} \tag{10}$$

The Schur complement of m_{11} in M is defined and is symmetric. If (P2) holds, we may use principal rearrangement to insure that $m_{12} \neq 0$, whence

$$\text{In } \begin{bmatrix} 0 & m_{12} \\ m_{21} & 0 \end{bmatrix} = (1, 1, 0). \tag{11}$$

The Schur complement of

$$\begin{bmatrix} 0 & m_{12} \\ m_{21} & 0 \end{bmatrix}$$

in M is defined and symmetric. In the third case, (P3), if M is of order n,

$$\text{In } M = (0, 0, n). \tag{12}$$

The same analysis can be applied to the Schur complements constructed in cases (P1) and (P2). This is the basis of the algorithm; the details follow.

Formally, let $\emptyset$ stand for the empty matrix and take $\text{In } \emptyset = (0, 0, 0)$. Let $M^{(k)}$ denote kth Schur complement computed, and let M_k denote the kth pivot (block). We adopt the following conventions:

$$(M/\emptyset) = M, \qquad (M/M) = \emptyset; \tag{13}$$

$$M^{(0)} = M, \qquad\qquad M_0 = \emptyset. \tag{14}$$

At the outset, we have trivially,

$$\text{In } M = \text{In } M_0 + \text{In } M^{(0)}.$$

The steps listed below pertain only to $M^{(k)}$ which has order $n - k$. Start with $k = 0$.

Step 0. So far, we have

$$\text{In } M = \sum_{i=0}^{k} \text{In } M_i + \text{In } M^{(k)}.$$

If $M^{(k)}$ has property (Pi), $i = 1, 2, 3$, go to Step i.

Step 1.   We may assume that $m_{11}^{(k)} \neq 0$. Let

$$M_{k+1} = m_{11}^{(k)},$$

$$M^{(k+1)} = (M^{(k)}/M_{k+1}).$$

Use Eq. (10) to obtain In M_{k+1} and then,

$$\text{In } M = \sum_{i=0}^{k+1} \text{In } M_i + \text{In } M^{(k+1)}.$$

If $M^{(k+1)} = \emptyset$, stop. Otherwise, return to Step 0 with k replaced by $k + 1$.

Step 2.   We may assume that $m_{12}^{(k)} = m_{21}^{(k)} \neq 0$. Let

$$M_{k+1} = \emptyset,$$

$$M_{k+2} = \begin{pmatrix} 0 & m_{12}^{(k)} \\ m_{12}^{(k)} & 0 \end{pmatrix},$$

$$M^{(k+1)} = (M^{(k)}/M_{k+1}) = M^{(k)},$$

$$M^{(k+2)} = (M^{(k+1)}/M_{k+2}).$$

Use Eq. (11) to obtain In M_{k+2}, and then

$$\text{In } M = \sum_{i=0}^{k+2} \text{In } M_i + \text{In } M^{(k+2)}.$$

If $M^{(k+2)} = \emptyset$, stop. Otherwise, return to Step 0 with k replaced by $k + 2$.

Step 3.   In this case, $M^{(k)} = 0$. It is of order $n - k$, so

$$\text{In } M^{(k)} = (0, 0, n - k),$$

and hence

$$\text{In } M = \sum_{i=0}^{k} \text{In } M_i + (0, 0, n - k).$$

Stop.   The inertia of M has been found.

This algorithm can be regarded as the logical extension of the one suggested by Stiefel (loc. cit.).   Stiefel's concern there is to determine whether a given real symmetric matrix is positive definite, that is, whether

it has inertia $(n, 0, 0)$. He does this by a pivoting scheme closely related to the one above but not described explicitly in terms of the Schur complement or inertia.

Such a method is reminiscent of the technique of *completing the square* often attributed to Lagrange.[3] To take the simplest example, consider the binary quadratic form

$$Q(x, y) = ax^2 + 2bxy + cy^2,$$

in which the coefficient a is nonzero. Then

$$Q(x, y) = a\left(x^2 + 2\frac{b}{a}xy\right) + cy^2,$$

$$= a\left(x + \frac{b}{a}y\right)^2 + \left(c - \frac{b^2}{a}\right)y^2,$$

which reveals that the coefficient of y^2 is just the Schur complement of a in the coefficient matrix $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$ of the form Q. This observation extends to n-ary quadratic forms. Unfortunately, the complication presented by the absence of squared terms in the portion of the form that remains to be expressed as the sum of squares of linear forms makes the connection with the Schur complement concept just a bit less compelling. However, it should be emphasized that in the technique of reducing a quadratic form to a sum of squares of linear forms, one demands more information than simply the inertia of the coefficient matrix. In some applications, such as quadratic programming, only the inertia matters; not everyone wants to reduce to diagonal form!

8. REMARKS

The algorithm described in the preceding section is appropriate only for exact arithmetic. Where finite precision arithmetic is involved, as in electronic digital computers, the method is not numerically stable. In [6], this problem is overcome by Bunch and Parlett who exhibit a numerically stable principal or (as they call it) *diagonal* pivoting procedure

---

[3] Debreu [10, p. 299] remarks that Lagrange discussed only binary and ternary quadratic forms, whereas the first treatment of the general case was given by F. Brioschi [5].

for achieving a factorization

$$M = LDL^T,$$

where L is unit lower triangular, D is the direct sum of 1×1 or 2×2 matrices, and $l_{i+1,i} = 0$ if $d_{i+1,i} \neq 0$.

The Bunch-Parlett method uses an unpublished proposal made in 1965 by W. Kahan (see [6, pp. 646–647]) that pertains to Lagrange's reduction method. As noted at the end of the preceding section, the coefficient matrix of a reduced quadratic form in Lagrange's method can be obtained from the Schur complement of a nonzero diagonal entry in the matrix of the current nonreduced quadratic form. If there are no such entries and the quadratic form is not identically zero, one uses a transformation invented by Lagrange to create them and then proceed as before. Bunch and Parlett acknowledge Kahan's contribution by saying [6, p. 646] "Kahan made the important observation that Lagrange's transformation could be interpreted as generalizing the notion of a pivot from a 1×1 to a 2×2 principal submatrix." They also describe and comment on Kahan's pivotal strategies.

To assure the numerical stability of a method like the one in Sec. 7, one must occasionally resort to 2×2 pivots even when nonzero diagonal entries are available. Nevertheless, it is easy to keep track of the inertia. {Indeed, if the 2×2 pivot block is

$$P = \begin{bmatrix} a & b \\ b & c \end{bmatrix},$$

its inertia is In $P = (1, 1, 0)$ if $ac - b^2 < 0$. On the other hand, if $ac - b^2 > 0$, then $ac > 0$. Hence, In $P = (2, 0, 0)$ if $a > 0$ and In $P = (0, 2, 0)$ if $a < 0$.} Details on how and when to use 2×2 block pivots are carefully explained in [6].

In addition to being used in computing the number of positive and negative eigenvalues of a matrix (i.e., its positivity and negativity) Schur complements are found in theorems for *estimating the eigenvalues* themselves. Such a result is due to Rutishauser [25, p. 51] who shows that if

$$M = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

and A is a positive definite principal submatrix of M of maximal order, then the smallest eigenvalue of (M/A) is negative and is a lower bound for the smallest eigenvalue of M.

9. THE RESTRICTION OF A QUADRATIC FORM

In this section we develop a second interpretation of the Schur complement, as the coefficient matrix of a quadratic form restricted to the null space of a matrix.

To begin, let Q be a quadratic form in n variables, say,

$$Q(z) = z^T M z = z^T \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} z,$$

where A is nonsingular. We may then write

$$Q(z) = P(x, y) = x^T A x + 2x^T B y + y^T C y. \tag{15}$$

Suppose Q is constrained by the system of equations

$$Ax + By = 0, \tag{16}$$

the solutions of which are the null space L of $[A, B]$. [One would be led to such a consideration by minimizing $P(x, y)$ with respect to x.]

To represent $Q_L = Q|L$, we may use Eq. (16) to eliminate x:

$$x = -A^{-1}By. \tag{17}$$

From Eqs. (15) and (17) we have

$$Q_L(y) = y^T B^T A^{-1} A A^{-1} B y - 2y^T B^T A^{-1} B y + y^T C y,$$

$$= y^T (C - B^T A^{-1} B) y, \tag{18}$$

and the symmetric coefficient matrix of Q_L is just (M/A). Using the substitution (17) amounts to choosing the columns of

$$\begin{bmatrix} -A^{-1}B \\ I \end{bmatrix}$$

as a base for L.

In the preceding instance, L was specified as the null space of a set of rows of M. It is also possible to associate a Schur complement with the restriction of a quadratic form to a null space given "externally." For example, let Q be written as before and suppose N is a matrix of order $r \times n$ and rank r. Let Q_L be the restriction of Q to

$$L = \{z : Nz = 0\}.$$

Since N is of rank r, its columns may be permuted so that the first r of them are linearly independent. To compensate for this, it is necessary to take the corresponding principal rearrangement of M. Assume this is already done and that D is nonsingular. Let M be partitioned conformally, and then consider the symmetric bordered matrix

$$F = \begin{bmatrix} 0 & D & E \\ D^T & A & B \\ E^T & B^T & C \end{bmatrix}.$$

Since D is nonsingular, so is

$$G = \begin{bmatrix} 0 & D \\ D^T & A \end{bmatrix}.$$

Thus, G is a nonsingular leading principal submatrix of F and its Schur complement, (F/G), is defined.

THEOREM. *In terms of the base*

$$\begin{bmatrix} -D^{-1}E \\ I \end{bmatrix}$$

for L, the quadratic form Q_L has the coefficient matrix

$$(F/G) = \left(\begin{bmatrix} 0 & D & E \\ D^T & A & B \\ E^T & B^T & C \end{bmatrix} \middle/ \begin{bmatrix} 0 & D \\ D^T & A \end{bmatrix} \right).$$

Proof. Q can be expressed as in Eq. (15). The equation

$$Nz = Dx + Ey = 0$$

yields $x = -D^{-1}Ey$ which can be substituted into the expression for Q to give

$$Q_L(y) = y^T E^T (D^{-1})^T A D^{-1} E y - 2y^T E^T (D^{-1})^T B y + y^T C y,$$

$$= y^T [C - 2E^T (D^{-1})^T B + E^T (D^{-1})^T A D^{-1} E] y.$$

On the other hand,

$$G^{-1} = \begin{bmatrix} (D^T)^{-1} A D^{-1} & (D^T)^{-1} \\ D^{-1} & 0 \end{bmatrix};$$

so, by definition,

$$(F/G) = C - [E^T, B^T] \begin{bmatrix} - (D^T)^{-1}AD^{-1} & (D^T)^{-1} \\ D^{-1} & 0 \end{bmatrix} \begin{bmatrix} E \\ B \end{bmatrix},$$

$$= C - E^T(D^T)^{-1}B - B^TD^{-1}E + E^T(D^T)^{-1}AD^{-1}E.$$

This means the quadratic form associated with (F/G) is just Q_L. ∎

There is another variation on this theme. Suppose F (as defined above) is the coefficient matrix of a quadratic form Q; i.e.,

$$Q(x, y, z) = y^TAy + z^TCz + 2x^TDy + 2x^TEz + 2y^TBz.$$

Suppose further that D is nonsingular and Q is to be represented on the null space L of $[0, D, E]$. This can be done by using the equation

$$0x + Dy + Ez = 0$$

to eliminate y from the expression for Q. Since

$$\begin{bmatrix} 0 & - D^{-1}E \\ I & 0 \\ 0 & I \end{bmatrix}$$

is a base for L, the resulting quadratic form would be expected to involve x and z; only $y = - D^{-1}Ez$ is eliminated. Curiously enough, the substitution process *appears* automatically to eliminate x as well as y. That is to say, x does not show up in the expression for Q_L. Indeed, the suggested elimination produces

$$Q_L(x, z) = z^T[C - 2E^T(D^{-1})^TB + E^T(D^{-1})^TAD^{-1}E]z.$$

From this and the theorem above, it follows that

$$Q_L(x, z) = (x^T, z^T) \begin{bmatrix} 0 & 0 \\ 0 & (F/G) \end{bmatrix} \begin{pmatrix} x \\ z \end{pmatrix}.$$

The subspace L generated by a given set of linearly independent n-vectors can always be viewed as the null space of a matrix appropriately constructed from the components of the vectors themselves. The restriction of a quadratic form Q to the subspace L can then be written in terms of a Schur complement. This may seem an awkward thing to do if only

the values of Q_L really matter; for if V denotes the $n \times r$ matrix whose columns are the given base for L, then elements of L are of the form Vu and hence

$$Q_L(u) = u^T V^T M V u.$$

The advantage of the approach to Q_L via the Schur complement lies in what can be said about the inertia of its coefficient matrix. This will be brought out in the next section.

It might be mentioned at this point that determining the inertia of a quadratic form restricted to a subspace is an important problem which bears on the second-order conditions for a constrained local minimum. See Debreu [10] for a discussion of this point and a determinantal approach to testing Q_L for positive semidefiniteness.

10. APPLICATION TO SOME THEOREMS OF M. MORSE

In [21] and [22], M. Morse considers a quadratic form restricted to a subspace. His results, which could be classified as inertia theorems, have technical applications in the "study of critical extremals and extended Sturm theorems" (see [22, pp. 560 and 569]). In this section, we employ the interpretation of the Schur complement given in Sec. 9 in conjunction with the inertia formula (7) to rederive Morse's findings. We caution the reader that some of Morse's notation has been drastically modified, but we believe no content has been lost thereby.

The formulation is as follows. Suppose we are given a real quadratic form

$$Q(z) = z^T M z$$

in n variables. For some integer r, $0 < r < n$, let $s = n - r$ and assume z is expressed as the direct sum of an r-vector x and an s-vector y. Thus

$$(z_1, \ldots, z_n) = (x_1, \ldots, x_r; y_1, \ldots, y_s).$$

This induces a corresponding decomposition of M into a block matrix

$$M = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}.$$

Accordingly, we write

$$Q(z) \equiv P(x, y) \equiv x^T A x + 2 x^T B y + y^T C y.$$

Morse terms the quadratic form $P(x, 0) = x^T A x$ *subordinate* to $Q(z)$. The central point of his study is to obtain a formula, when $[A, B]$ is of rank r, for the difference

$$\text{index } P(x, y) - \text{index } P(x, 0).$$

In this context, the *index* of a quadratic form is the negativity of the corresponding coefficient matrix. Morse gives his formula in terms of the index and nullity of a class of "congruent quadratic forms" called *complementary* to $P(x, 0)$. The latter stem from the restriction of Q to the null space L of $[A, B]$. By the rank assumption, L is of dimension s. This rank property can be assured if either A or M is nonsingular, but these sufficient conditions are not necessary.

A quadratic form complementary to $P(x, 0)$ is just obtained by representing Q in terms of a base for L. The coefficient matrices resulting from such choices are all congruent and consequently have the same inertia. From these considerations, it is evident that it suffices to extract a base for L from $[A, B]$ itself.

In the case where A is nonsingular, the required complementary form Q_L is given by the matrix $(M/A) = C - B^T A^{-1} B$; the inertia formula applies and yields

$$\text{index } P(x, y) - \text{index } P(x, 0) = \text{index } Q_L(y), \tag{19}$$

$$\text{nullity } P(x, y) = \text{nullity } Q_L(y). \tag{20}$$

These two equations are just (3.17) and (3.18) of [22]. They are indicative of the general results to follow.

The case where A is not necessarily of full rank is, of course, more complicated. The assumption that $[A, B]$ has rank r means it contains a nonsingular submatrix D or order r. Since the particular choice of D (equivalently, base for L) has no bearing on the inertia of the coefficient matrix of interest, it may as well be chosen conveniently.

To this end, let A have rank t. Note: $0 \leqslant t \leqslant r$. If $t = 0$, then $A = 0$; we do not exclude this case from the present discussion. It follows that A contains a principal submatrix of order t and rank t. Moreover, no larger principal submatrix of A is nonsingular. By a principal rearrangement of the first r rows and columns of M, we may assume this $t \times t$ submatrix of A is A_{11} where

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

A consequence of this assumption is

$$(A/A_{11}) = 0 \qquad \text{(or } \emptyset \text{ if } A = A_{11}), \tag{21}$$

for otherwise A contains a nonsingular principal submatrix of order $t + 1$ or $t + 2$.

This decomposition of A induces one of M, say

$$M = \begin{bmatrix} A_{11} & A_{12} & B_1 \\ A_{21} & A_{22} & B_2 \\ B_1{}^T & B_2{}^T & C \end{bmatrix}.$$

Next, it follows that the $r - t$ rows of B_2 are linearly independent. Hence the last s rows and columns of M can be permuted so that

$$M = \begin{bmatrix} A_{11} & A_{12} & B_{11} & B_{12} \\ A_{21} & A_{22} & B_{21} & B_{22} \\ B_{11}^T & B_{21}^T & C_{11} & C_{12} \\ B_{12}^T & B_{22}^T & C_{21} & C_{22} \end{bmatrix}, \tag{22}$$

where B_{21} is nonsingular. Now select

$$D = \begin{bmatrix} A_{11} & B_{11} \\ A_{21} & B_{21} \end{bmatrix}$$

as the nonsingular $r \times r$ submatrix of $[A, B]$. In agreement with the decomposition (22) of M, write

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \qquad y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix};$$

thus L is specified by the equations

$$A_{11}x_1 + A_{12}x_2 + B_{11}y_1 + B_{12}y_2 = 0, \tag{23}$$

$$A_{21}x_1 + A_{22}x_2 + B_{21}y_1 + B_{22}y_2 = 0. \tag{24}$$

We compute the restriction of Q to L in two stages. By eliminating x_1 from Q [i.e., restricting Q to the solution set L_1 of Eq. (23)] we obtain the quadratic form Q_{L_1} with coefficient matrix

$$(M/A_{11}) = \begin{bmatrix} \bar{A}_{22} & \bar{B}_{21} & \bar{B}_{22} \\ \bar{B}_{21}^T & \bar{C}_{11} & \bar{C}_{12} \\ \bar{B}_{22}^T & \bar{C}_{21} & \bar{C}_{22} \end{bmatrix}.$$

Actually, $\bar{A}_{22} = (A/A_{11}) = 0$ (or perhaps $\bar{A}_{22} = \emptyset$), and $\bar{B}_{21} = (D/A_{11})$ is nonsingular. Moreover, $\bar{A}_{22}$, $\bar{B}_{21}$, and $\bar{B}_{22}$ are the coefficients of x_2, y_1, and y_2, respectively, after the elimination of x_1 from Eq. (24). Finally, Q_L is just the restriction of Q_{L_1} to L_2 the solution set of Eq. (24) in revised form, i.e.,

$$\bar{B}_{21}y_1 + \bar{B}_{22}y_2 = 0.$$

But this quadratic form corresponds to the matrix

$$\left(\begin{bmatrix} 0 & \bar{B}_{21} & \bar{B}_{22} \\ \bar{B}_{21}^T & \bar{C}_{11} & \bar{C}_{12} \\ \bar{B}_{22}^T & \bar{C}_{21} & \bar{C}_{22} \end{bmatrix} \Big/ \begin{bmatrix} 0 & \bar{B}_{21} \\ \bar{B}_{21}^T & \bar{C}_{11} \end{bmatrix} \right).$$

The net effect of all this is that (relative to the choice of D) the restriction of Q to L has the coefficient matrix (M/E) where

$$E = \begin{bmatrix} A_{11} & A_{12} & B_{11} \\ A_{21} & A_{22} & B_{21} \\ B_{11}^T & B_{21}^T & C_{11} \end{bmatrix}.$$

The order of (M/E) is just the dimension of y_2. This would suggest there is a discrepancy in identifying Q_L with (M/E) because only x_1 and y_1 are eliminated. Strictly speaking, the proper coefficient matrix of Q_L is

$$\begin{bmatrix} 0 & 0 \\ 0 & (M/E) \end{bmatrix}.$$

Its nullity is evidently $r - t$. From the inertia formula (7), we see that

$$\text{In } M = \text{In } E + \text{In}(M/E), \tag{25}$$

$$\text{In } E = \text{In } A_{11} + \text{In}(E/A_{11}). \tag{26}$$

Since $\bar{B}_{21}$ is nonsingular, Eq. (9) applies to (E/A_{11}) and gives us

$$\text{In}(E/A_{11}) = (r - t, r - t, 0). \tag{27}$$

Substituting Eq. (27) into Eq. (26) and the result into Eq. (25) we get

$$\text{In } M = \text{In } A_{11} + (r - t, r - t, 0) + \text{In}(M/E). \tag{28}$$

This equation and the association between forms and matrices imply

index $P(x, y)$ — index $P(x, 0)$ = nullity $Q_L(x_2, y_2)$ + index $Q_L(x_2, y_2)$.

$$(29)$$

Now, in general,

$$\text{nullity } Q_L(x_2, y_2) = r - t + \text{nullity } (M/E),$$

$$= r - t + \text{nullity } M.$$

If we assume M is nonsingular, then nullity $Q_L(x_2, y_2) = r - t$, and upon substitution, Eq. (29) becomes the conclusion reached by Morse [22, Theorem 1.1] under just this assumption.

REFERENCES

1 T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, New York (1958).

2 W. N. Anderson, Jr., Shorted operators, *SIAM J. Appl. Math.* 20(1971), 520–525.

3 W. N. Anderson, Jr. and R. J. Duffin, Series and parallel addition of matrices, *J. Math. Anal. Appl.* 26(1969), 576–594.

4 W. N. Anderson, Jr., R. J. Duffin, and G. E. Trapp, Parallel subtraction of matrices, *Proc. Nat. Acad. Sci. U.S.A.* 69(1972), 2530–2531.

5 F. Brioschi, Sur les séries qui donnent le nombre de racines réelles des équations algébriques à une ou plusiers inconnues, *Nouvelles Annales de Mathématiques* 15(1856), 264–286.

6 J. R. Bunch and B. N. Parlett, Direct method for solving symmetric indefinite systems of linear equations, *SIAM J. Numer. Anal.* 8(1971), 639–655.

7 D. Carlson and H. Schneider, Inertia theorems for matrices: the positive semidefinite case, *J. Math. Anal. Appl.* 6(1963), 430–446.

8 D. Carlson, E. Haynsworth, and T. Markham, A generalization of the Schur complement by means of the Moore-Penrose inverse, submitted for publication in *SIAM J. Appl. Math.* (1972).

9 D. E. Crabtree and E. V. Haynsworth, An identity for the Schur complement of a matrix, *Proc. Amer. Math. Soc.* 22(1969), 364–366.

10 G. Debreu, Definite and semidefinite quadratic forms, *Econometrica* 20(1952), 295–300.

11 R. J. Duffin, D. Hazony, and N. Morrison, Network synthesis through hybrid matrices, *J. SIAM Appl. Math.* 14(1966), 390–413.

12 R. J. Duffin and G. E. Trapp, Hybrid addition of matrices — network theory concept, *Applicable Analysis* 2(1972), 241–254.

13 M. Fiedler and V. Pták, On matrices with non-positive off-diagonal elements and positive principal minors, *Czech. Math. J.* **12**(1962), 382–400.

14 F. R. Gantmacher, *The Theory of Matrices*, Vol. I, Chelsea, New York (1960).

15 E. V. Haynsworth, Determination of the inertia of a partitioned Hermitian matrix, *Lin. Alg. and Appl.* **1**(1968), 73–81.

16 E. V. Haynsworth and A. M. Ostrowski, On the inertia of some classes of partitioned matrices, *Lin. Alg. and Appl.* **1**(1968), 299–316.

17 E. V. Haynsworth, Reduction of a matrix using properties of the Schur complement, *Lin. Alg. and Appl.* **3**(1970), 23–29.

18 E. V. Haynsworth, Applications of an inequality for the Schur complement, *Proc. Amer. Math. Soc.* **24**(1970), 512–516.

19 J. L. Lagrange, Recherches sur la méthode de maximis et minimis, *Oeuvres* **1**(1759), 3–20.*

20 R. von Mises, *Mathematical Theory of Probability and Statistics* (H. Geiringer, Ed.), Academic, New York (1964).

21 M. Morse, Subordinate quadratic forms and their complementary forms, *Proc. Nat. Acad. Sci. U.S.A.* **68**(1971), 579.

22 M. Morse, Subordinate quadratic forms and their complementary forms, *Rev. Roum. Math. Pures et Appl.* **16**(1971), 559–569.

23 A. Ostrowski, A new proof of Haynsworth's quotient formula for Schur complements, *Lin. Alg. and Appl.* **4**(1971), 389–392.

24 T. D. Parsons, Applications of principal pivoting, in *Proceedings of the Princeton Symposium on Mathematical Programming* (H. W. Kuhn, Ed.), Princeton Univ. Press, Princeton, N.J. (1970), pp. 567–581.

25 H. Rutishauser, Über eine kubisch konvergente Variante der *LR* Transformation, *Z. Angew. Math. Mech.* **40**(1960), 49–54.

26 I. Schur, Potenzreihen im Innern des Einheitskreises, *J. Reine Angew. Math.* **147**(1917), 205–232.

27 E. L. Stiefel, *An Introduction to Numerical Analysis*, Academic, New York (1963).

28 A. W. Tucker, A combinatorial equivalence of matrices, in *Proceedings of Symposia in Applied Mathematics*, Vol. 10 (R. E. Bellman and M. Hall, Eds.), American Math. Soc., Providence, R.I. (1960), pp. 129–140.

29 A. W. Tucker, Principal pivotal transforms of square matrices, *SIAM Rev.* **5**(1963), 305.

30 L. J. Watford, The Schur complement of a generalized *M*-matrix, *Lin. Alg. and Appl.* **5**(1972), 247–255.

31 K. Wendler, *Hauptaustauschschritte (Principal Pivoting)*, Lecture Notes in Operations Research and Mathematical Systems, Vol. 45, Springer-Verlag, Berlin (1971).

32 M. A. Woodbury, *Inverting Modified Matrices*, Memorandum Report No. 42, Statistical Research Group, Princeton University, Princeton, N.J. (1950).