SRC TR 87-119

Restricted Quadratic Forms, Inertia Theorems and the Schur Complement

by

J. H. Maddocks

Restricted Quadratic Forms, Inertia Theorems
and the Schur Complement

J. H. Maddocks†

Department of Mathematics
and
Institute for Physical Sciences and Technology
University of Maryland
College Park, MD 20742

ABSTRACT: The starting point of this investigation is the properties of restricted quadratic forms, $x^T Ax$, $x \in S \subset \mathbb{R}^m$, where A is an $m \times m$ real symmetric matrix, and S is a subspace. The index theory of Hestenes (1951) and Maddocks (1985) that treats the more general Hilbert space version of this problem is first specialized to the finite-dimensional context, and appropriate extensions, valid only in finite-dimensions, are made. The theory is then applied to obtain various inertia theorems for matrices and positivity tests for quadratic forms. Expressions for the inertias of divers symmetrically partitioned matrices are described. In particular, an inertia theorem for the generalized Schur complement is given. The investigation recovers, links and extends several, formerly disparate, results in the general area of inertia theorems.

†Research supported by the Air Force Office of Scientific Research under Grants #86-0097 and #87-0073.

§1. Introduction

The main subject of this presentation is the properties of a quadratic form defined on $\mathbb{R}^m$ by an $m\times m$ real symmetric matrix A when attention is restricted to a given subspace $S\subset\mathbb{R}^m$. If the subspace S is actually the whole of $\mathbb{R}^m$, the essential properties of the quadratic form are encapsulated in the inertia of the matrix A, denoted $\operatorname{In} A$, namely the triple comprising the number of positive, negative, and zero eigenvalues. It will here be shown that this concept of inertia can be usefully generalized to obtain a triple $\operatorname{In}^*(S;A)$, depending on both the matrix A and the subspace S, that captures the properties of the quadratic form $x^T Ax$ restricted to S.

It is then shown that $\operatorname{In}^*(S;A)$ is intimately connected to $\operatorname{In} A$ and $\operatorname{In}^*(S^A;A)$, where S^A is the subspace that is A-orthogonal to S. The main idea is that the properties of $x^T Ax$ on S are completely determined if the properties on $\mathbb{R}^m$ and on S^A are known. This result is then applied to obtain inertia theorems for matrices of the type $B^T AB$, and for symmetrically partitioned matrices. Particular emphasis is given to the matrix construction known as the generalized Schur complement. Tests concerning positivity of A on subspaces are also described.

Some of the results stated here are necessarily complicated. They comprise equalities between several different indices or dimensions, and it seems unlikely that many of these indices will be readily calculable in concrete applications. However, all of the results obtained here include known results as special cases.

Invariably, these prior results comprise situations in which several of the indices are known. Thus the theory developed here casts light on the necessity of hypotheses and assumptions made in previous analyses. Moreover, a considerable unification is achieved by the construction of connections between previously unrelated works.

In a similar vein, it should be stressed that the proofs given here are not necessarily the simplest for any given result. However, *all* the proofs are based on some version of A-orthogonality. Accordingly, several, apparently disparate, theorems are revealed as different manifestations of one central result.

The presentation is structured as follows. In §2, the index theory of Hestenes (1951) and Maddocks (1985), which was derived in the context of the isoperimetric calculus of variations, is specialized to finite-dimensions, and the appropriate extensions are made in order to obtain inertia theorems. Then, in §3, the theory is reformulated in terms of the Moore-Penrose, or generalized, inverse of A. Connections between $\operatorname{In} A$ and $\operatorname{In} B^T AB$ are described in §4. Of course, the classic result known as Sylvester's Law of Inertia is recovered as a particular case. Attention is turned to partitioned matrices in §§5 and 6. Preliminary results are given in §5, and the generalized Schur complement is discussed in §6. Theorems of Morse (1971) and of Han (1986) are also discussed here. Finally, in §7, the particular question of positivity on a subspace is treated.

It should be emphasized that the theory given here overlaps with, and builds on, the analysis of many authors. Because there are so many connections with prior works, I do not attempt detailed attributions here in the introduction. Full discussion and references are given at the appropriate junctures throughout the body of the text. The interrelations of this work and prior analyses are also summarized in §8.

§2. General Results Involving A-orthogonal Complements

The content of this section is a finite-dimensional version of the theory of restricted quadratic forms derived in Maddocks (1985, §2), which in turn is a development of the theory of Hestenes (1951). Considerable changes in notation and emphasis are made in order to facilitate applications to, and comparison with the theory of symmetric real matrices. Consideration of the finite-dimensional case also allows various extensions of the theory. The scope of the development given here is limited to the material required to understand the statement of the main result, namely Theorem 2.6. In particular the complete proof of Theorem 2.6 is not given. The steps omitted here can be found in Maddocks (op. cit.).

Consider a real symmetric $m \times m$ matrix A. We shall study the properties of the quadratic form $Q(x) = x^T Ax$ for $x \in S$ a subspace of $\mathbb{R}^m$. The main focus of our attention is the connection between the properties of Q restricted to S and of Q restricted to other related subspaces.

It is apparent that the properties of Q on S are intimately connected to the inertia, $\mathrm{In}(M) = (\pi(M), \nu(M), \delta(M))$, i.e. the triple comprising the number of positive, negative, and zero eigenvalues, of the symmetric matrix $M = B^T AB$. Here B is any $m \times m$ matrix whose range is S. Accordingly, the results presented here can later be couched in terms of inertia theorems for certain matrices.

The following notion of A-orthogonality will appear throughout.

Definition 2.1: Two vectors $x_1, x_2 \in \mathbb{R}^m$ are termed A-orthogonal if

$$x_1^T Ax_2 = 0.$$

Remarks.

(a) As A is symmetric, the relation is symmetric.

(b) Any vector is A-orthogonal to any element of the kernel or nullspace of A, which is here denoted N(A).

(c) If either of the vectors is an eigenvector of A not in N(A), then orthogonality and A-orthogonality are equivalent.

(d) Whenever A is neither positive nor negative semidefinite, there exist vectors x such that $Ax \neq 0$, but $x^T Ax = 0$.

(e) The concept of A-orthogonality extends to subspaces in the obvious way. The A-orthogonal complement of a subspace S will be denoted S^A. A useful characterization is provided by

$$S^A = \{y\colon Ay \in S^{\perp}\} = \{y\colon x^T Ay = 0,\ \forall x \in S\} = (AS)^{\perp}.$$

It should be noted that for any subspace S, $N(A) \subset S^A$.

(f) Because of remark (d) above it is possible that $S \cap S^A \cap R(A) \neq \{0\}$. Here $R(A)$ denotes the range of A.

(g) The properties of S^A can be used to derive simple results, such as

$$S^{AA} = S + N(A),$$

and

$$(S^A \cap T^A)^A = S + T + N(A),$$

where T is another subspace.

<u>Lemma 2.1</u>. Any subspace of $\mathbb{R}^m$ has a mutually A-orthogonal basis.

<u>Proof</u>. Consider any basis w_i, $i = 1,\ldots,p$. Then, as A is symmetric, the matrix $W = \{w_{ij}\}$

$$w_{ij} = w_i^T A\ w_j,$$

is symmetric. Therefore there exists an orthogonal $p \times p$ matrix $R = \{r_{ij}\}$ such that $P^T WP$ is diagonal. The set

$$\left\{u_i\colon\ u_i = \sum_j r_{ji} w_j\right\}$$

is a basis because R is nonsingular, and by construction the basis is also mutually A-orthogonal.

<u>Remark</u>. Because A is not necessarily definite, the obvious generalization of the Gram-Schmidt procedure does not work. In particular it is not in general possible to construct an A-orthonormal basis.

Lemma 2.2. Let B be an $m \times n$ matrix of rank p. Let $\{w_i\}$ $i = 1,\ldots,p$, $w_i \in \mathbb{R}^m$, be any basis for $S = R(B)$, the range of B. Define the $p \times p$ matrix W by $W = \{w_{ij}\} = \{w_i^T A w_j\}$. Then

$$\mathrm{In}\, B^T AB = \mathrm{In}\, W + (0,0,n-p). \tag{2.1}$$

In particular, if $\{w_i\}$ is a mutually A-orthogonal basis, then the inertia of the $n \times n$ symmetric matrix $B^T AB$ is determined by the inertia of the $p \times p$ diagonal matrix $\mathrm{Diag}\,\{w_i^T A w_i\}$.

Proof. The Poincaré variational characterization of the eigenvalues of a self-adjoint operator (e.g. Weinberger, 1974) implies that $\nu(M)$, the number of negative eigenvalues of a symmetric matrix M, coincides with the largest dimension k such that $x^T Mx$ is negative for all x in a subspace of dimension k. This result can be applied both to the $n \times n$ matrix $B^T AB$ and to the $p \times p$ matrix $W = \{v_j^T A v_i\}$. But any k dimensional subspace of $\mathbb{R}^n$ on which $B^T AB$ is negative, provides a k dimensional subspace of $\mathbb{R}^p$ on which W is negative and vice versa, via the following construction. Suppose $\{x_i\}$ $i = 1,\ldots,k$, $x_i \in \mathbb{R}^n$ is a basis of a negative subspace of $B^T AB$. Then $Bx_i \neq 0$, and as $\{v_j\}$ is a basis of $R(B)$, $\exists$ an $m \times p$ matrix $K = \{k_{ij}\}$ of rank k such that $Bx_i = \sum k_{ij} v_j$. By construction, the rows of K span a negative subspace of dimension k in $\mathbb{R}^p$. Thus $\nu(B^T AB) = \nu(W)$.

In a similar fashion it can be shown that $\pi(B^T AB) = \pi(W)$. The last equality contained in (2.1) is then easily obtained once it is remarked that for any symmetric $q \times q$ matrix M

$$\pi(M) + \nu(M) + \delta(M) = q. \tag{2.2}$$

Remark. When the properties of the quadratic form $x^T Ax$, $x \in R(B)$ are under consideration, it is natural to consider the eigenvalue problem

$$B^T ABy = \lambda B^T By, \quad By \neq 0.$$

It is eigenvalues in this sense that determine whether $x^T Ax$ is positive definite on $R(B)$. Thus, zero eigenvalues of $B^T AB$ that have an eigenvector $y \in N(B)$ are not of concern. The modified eigenvalue problem leads to a modified inertia

$$\mathrm{In}^*(B^T AB) = \mathrm{In}(B^T AB) - (0,0, \dim N(B)).$$

Accordingly, equation (2.1) appearing in Lemma 2.2 can be rewritten in the form

$$\mathrm{In}^*(B^T AB) = \mathrm{In}\, W. \tag{2.3}$$

Although we shall not pause to prove it, the modified inertia also measures the number of positive, negative and zero eigenvalues λ of the problem

$$x \in R(B), \; Ax - \lambda x \in N(B^T).$$

As an immediate Corollary to Lemma 2.2, we have

Corollary 2.3. If B_1 is $m \times n$, B_2 is $m \times \ell$ and $R(B_1) = R(B_2)$ then

$$\mathrm{In}\, B_1^T AB_1 + (0,0, \ell - n) = \mathrm{In}\, B_2^T AB_2$$

Remark. If the matrices B_1 and B_2 are square and of full rank, i.e. $m = n = \ell = p$, Corollary (2.3) reduces to the celebrated result known as Sylvester's Law of Inertia.

Because of Lemma 2.2 it is practical to associate an inertia with the subspace S.

Definition 2.2. The inertia of a subspace S, denoted $\mathrm{In}^*(S;A)$ or $\mathrm{In}^*(S)$, is defined by

$$\mathrm{In}^*(S;A) = \mathrm{In}\, W,$$

where W is the matrix defined in Lemma 2.2. By the remark following Lemma 2.2 and equation (2.3)

$$\mathrm{In}^*(S;A) = \mathrm{In}^*(B^T AB)$$

where B is any $m \times n$ matrix whose range is S. Obviously the quadratic form $Q(x) = x^T Ax$ satisfies $x^T Ax \geq 0$, $\forall x \in S$ if and only if $\nu^*(S;A) = 0$, with analogous results holding for positivity, nonpositivity etc.

Definition 2.3. For a subspace $S \subset \mathbb{R}^m$ the *relative nullity of* A *on* S, denoted $d^\circ(S;A)$ or just $d^\circ(S)$, is defined by

$$d^\circ(S;A) = \dim \{S \cap S^A \cap R(A)\}. \tag{2.4}$$

Remark. In Maddocks (1985) the relative nullity is given a different definition and (2.4) has the status of a theorem. In the work of Hestenes as is reported in Gregory (1980, §2), the relative nullity is given yet another equivalent definition.

Lemma 2.4. For any $m \times m$ matrix A, and subspace $S \subset \mathbb{R}^m$ the relative nullity $d^\circ(S;A)$ and $\delta^*(S;A)$, (defined in Definition 2.2) are related by

$$\delta^*(S;A) = \dim\{S \cap S^A\} = d^\circ(S;A) + \dim\{N(A) \cap S\} \tag{2.5}$$

Remark. The quantity $\dim\{S \cap S^A\}$ is sometimes called the nullity of A on S.

Proof. The second equality in (2.5) follows immediately from Definition 2.3 and the facts that $R(A) = \{N(A)\}^{\perp}$ and $N(A) \subset S^A$.

To prove the first equality consider an A-orthogonal basis $\{s_i\}$, $i = 1,\ldots,p$ of S, the existence of which is guaranteed by Lemma 2.1. Then, by definition,

$$\delta^*(S;A) = \delta(W),$$

where $\delta(W)$ is the number of basis vectors s_i satisfying

$$s_i^T A s_i = 0.$$

Contrariwise, any $x \in S \cap S^A$ can be written in terms of the basis $\{s_i\}$ of S as

$$x = \sum_i \alpha_i s_i,$$

but for x to be in S^A

$$\alpha_i s_i^T A s_i = 0, \ \forall i.$$

Thus, those s_i satisfying $s_i^T A s_i = 0$ comprise a basis of $S \cap S^A$, and consequently,

$$\delta(W) = \dim(S \cap S^A).$$

Remark. Because $S^{AA} = S + N(A)$, it can be seen that

$$\delta^*(S^A;A) = \dim(S \cap S^A) + d^\circ(S;A).$$

Lemma 2.5. Let S and Y be subspaces of $\mathbb{R}^m$ with $S \cap R(A) \subset Y^A \cap R(A)$ and $Y \cap Y^A \cap R(A) = \{0\}$. Then

$$d^\circ(S;A) = d^\circ(S^A \cap Y^A;A).$$

Proof. By definition

$$d^\circ(S;A) = \dim\{S \cap S^A \cap R(A)\}$$

and

$$d^\circ(S^A \cap Y^A;A) = \dim\{S^A \cap Y^A \cap (S^A \cap Y^A)^A \cap R(A)\}.$$

It was earlier remarked that

$$(S^A \cap Y^A)^A = S + Y + N(A),$$

and by hypothesis $S \cap R(A) \subset Y^A \cap R(A)$, and $Y \cap Y^A \cap R(A) = \{0\}$. Consequently,

$$\begin{aligned} d^\circ(S^A \cap Y^A;A) &= \dim\{(S + Y) \cap S^A \cap Y^A \cap R(A)\} = \dim\{S \cap S^A \cap Y^A \cap R(A)\} \\ &= \dim\{S \cap S^A \cap R(A)\} = d^\circ(S;A). \end{aligned}$$

Remark. The hypotheses on the subspace Y appear unduly restrictive, but they encompass the important case $Y = \{0\}$, in which we obtain the result

$$d^\circ(S;A) = d^\circ(S^A;A).$$

Theorem 2.6. Let A be an $m \times m$ real symmetric matrix, and let S and Y be subspaces of $\mathbb{R}^m$, with $S \subset Y^A$, and $Y \cap Y^A \cap R(A) = \{0\}$, ie. $d^\circ(Y;A) = 0$. Then

$$\begin{aligned} In^*(Y^A;A) = In^*(S;A) &+ In^*(S^A \cap Y^A;A) \\ &+ (d^\circ(S), d^\circ(S), - d^\circ(S) - \dim\{S \cap S^A\}). \end{aligned} \tag{2.6}$$

The notations S^A, In^* and d° were introduced in Definitions 2.1, 2.2 and 2.3

Proof. Identity (2.6) asserts three equalities. The equality

$$\nu^*(Y^A;A) = \nu^*(S;A) + \nu^*(S^A \cap Y^A;A) + d^\circ(S) \tag{2.7}$$

is a restatement of Theorem 2 of Maddocks (1985) applied to the operator $L = PAP$, where P is orthogonal projection onto the subspace Y^A. The equivalence is apparent once it is remarked that the number of negative eigenvalues of PAP, denoted $\sigma^-(PAP)$, is just $\nu^*(Y^A;A)$, that the index $d^-(S)$ is equivalent to $\nu^*(S)$, that

$$S^{PAP} = (S^A \cap Y^A) \cup (Y^A)^{\perp}, \tag{2.8}$$

and that

$$S \cap S^A = S \cap S^A \cap Y^A = S \cap S^{PAP}. \tag{2.9}$$

Similarly, consideration of the operator $L = -PAP$ provides the equation

$$\pi^*(Y^A;A) = \pi^*(S;A) + \pi^*(S^A \cap Y^A;A) + d^\circ(S). \tag{2.10}$$

The final equation can be obtained as follows.

Three applications of Lemma 2.4, yield

$$\delta^*(Y^A;A) = d^\circ(Y^A;S) + \dim\{N(A) \cap Y^A\} \tag{2.11}$$

$$\delta^*(S;A) = \dim\{S \cap S^A\} \tag{2.12}$$

and

$$\delta^*(S^A \cap Y^A;A) = d^\circ(S^A \cap Y^A;A) + \dim\{N(A) \cap S^A \cap Y^A\}. \tag{2.13}$$

Now, the A-orthogonal complement of any subspace contains $N(A)$, and by hypothesis

$$\dim(Y \cap Y^A \cap R(A)) = d^\circ(Y^A;S) = 0.$$

Accordingly (2.11) and (2.13) reduce to

$$\delta^*(Y^A;A) = \dim N(A),$$

and

$$\delta^*(Y^A;A) = \delta^*(S^A \cap Y^A;A) - d^\circ(S^A \cap Y^A;A).$$

Lemma (2.5) and equation (2.12) then imply the equation

$$\delta^*(Y^A;A) = \delta^*(S;A) + \delta^*(S^A \cap Y^A;A) - d^\circ(S;A) - \dim\{S \cap S^A\}$$

as required.

Remark. Theorem 2.6 will be applied in §6, and the full result stated above will be required. Nevertheless, for most applications the case $Y = \{0\}$ suffices. The special case is sufficiently ubiquituous in the sequel that it is formalized as:

Corollary 2.7. Let A be an $m\times m$ real symmetric matrix, and let S be a subspace of $\mathbb{R}^m$. Then

$$\operatorname{In} A = \operatorname{In}^*(S;A) + \operatorname{In}^*(S^A;A) + (d^\circ(S), d^\circ(S), -d^\circ(S) - \dim\{S\cap S^A\}).$$

Proof. The subspace $Y = \{0\}$ satisfies the hypotheses of Theorem 2.6, for then $Y^A = \mathbb{R}^m$. Moreover, the inertia of the subspace $\mathbb{R}^m$ coincides with the inertia of the matrix A.

Remark. Han & Fujiwara (1985, Theorem 2.3), obtained Corollary 2.7 in the further special case $d^\circ(S) = 0$. Their result includes the additional hypothesis

$$S\cap S^A \subset N(A)$$

so that

$$d^\circ(S) = \dim\{S\cap S^A \cap R(A)\} = 0.$$

Han & Fujiwara's development was independent of the works of Hestenes (1951) and Maddocks (op. cit.), and they actually adopt a different definition for the quantity that is here denoted $\operatorname{In}^*(S^A;A)$.

Example 2.8.

$m = 2$, $S = (s,0)$, $s \in \mathbb{R}$, where $(s,0)$ denotes a *column* vector, and $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

Then

$S^A = \operatorname{span}(s,0)$, $R(A) = \mathbb{R}^2$, $d^\circ(S;A) = 1 = \dim(S\cap S^A)$,

$(1,0)^T A(1,0) = (1,0)^T(0,1) = 0$, and

$\operatorname{In}^*(S;A) = \operatorname{In}^*(S^A;A) = (0,0,1)$.

Consequently, Corollary 2.7 predicts that

$$\operatorname{In} A = (0,0,1) + (0,0,1) + (1,1,-2) = (1,1,0),$$

and of course the eigenvalues of A are ± 1.

Corollary 2.8 can be applied to clarify the geometrical role played by $d^{\circ}(S;A)$. Because the sum of the components of an inertia equal the dimension of the space, we find from Corollary 2.8 that

$$m = \dim S + \dim S^{A} + d^{\circ}(S) - \dim\{S \cap S^{A}\}.$$

Consequently, $\mathbb{R}^{m}$ can be decomposed as the sum of S and S^{A} precisely if $d^{\circ}(S) = 0$. The decomposition is a direct sum if the stronger condition $\dim\{S \cap S^{A}\} = 0$ is satisfied.

§3. Results Utilizing the Generalized Inverse of A

The results of §2 are direct in the sense that they do not involve any inversion of the matrix A. It will here be shown that when the Moore-Penrose, or generalized, inverse of A is introduced, Corollary 2.7 can be reformulated to emphasize the symmetry, or duality, between the subspace S and the subspace $S^{\perp}$, the usual orthogonal complement of S.

Definition 3.1. For an $m \times n$ matrix C, the *generalized*, or *Moore-Penrose*, inverse is the unique $n \times m$ matrix C^{+} satisfying the four conditions:

(i) $CC^{+}C = C$

(ii) $C^{+}CC^{+} = C^{+}$

(iii) $(CC^{+})^{T} = CC^{+}$

(iv) $(C^+C)^T = C^+C$.

Geometrically, C^+ can be regarded as the inverse of the invertible operator that is obtained when C is restricted to domain $R(C^T)$, and range $R(C)$. It should be remarked that

$$N(C^+) = N(C^T) \quad \text{and} \quad R(C^+) = R(C^T).$$

As a simple consequence of the definition we have that

$$(C^T)^+ = (C^+)^T.$$

When A is symmetric, A^+ is also symmetric, and conditions (iii) and (iv) both reduce to

$$A^+A = AA^+.$$

Moreover, direct consideration of the standard eigenvalue problem demonstrates that

$$\text{In}\, A = \text{In}\, A^+ \tag{3.1}$$

<u>Theorem 3.1</u>. Let A be an $m\times m$ real symmetric matrix, and let S be a subspace of $\mathbb{R}^m$. Then

$$\text{In}\, A = \text{In}^*(S;A) + \text{In}^*(S^\perp;A^+) + (d,d,-2d), \tag{3.2}$$

where

$$d = \dim\{S\cap A^+S^\perp\} = \dim\{AS\cap S^\perp\}. \tag{3.3}$$

The triple In^* was introduced in Definition 2.2.

<u>Proof</u>. Consider Corollary 2.7 and notice that

$$\text{In}^*(\mathbb{R}^m;A) = \text{In}\, A$$

and that

$$S^A = N(A) \oplus A^+S^\perp, \tag{3.4}$$

the sum being direct. Then (2.6) can be written as

$$\text{In}\,A = \text{In}^*(S;A) + \text{In}^*(N(A) + A^+S^\perp;A) + (d,d,-d - \dim\{S\cap(N(A) + A^+(S^\perp))\}) \tag{3.5}$$

where d is given by (3.3). Because the subspaces $N(A)$ and $A^+S^\perp$ are mutually orthogonal and mutually A-orthogonal,

$$\text{In}^*(N(A) + A^+S^\perp;A) = \text{In}^*(A^+S^\perp;A) + (0,0,\dim N(A)) \tag{3.6}$$

Moreover, by Definition 2.2 of In^*, and property (ii) of the definition for A^+,

$$\text{In}^*(A^+S^\perp;A) = \text{In}^*(S^\perp;A^+) - (0,0,\dim\{N(A)\cap S^\perp\}). \tag{3.7}$$

Now

$$\dim\{S\cap(N(A) + A^+S^\perp)\} = \dim\{S\cap N(A)\} + \dim\{S\cap A^+S^\perp\},$$

and

$$\dim\{N(A)\cap S^\perp\} + \dim\{N(A)\cap S\} = \dim N(A).$$

Consequently, substitution of (3.6) and (3.7) in (3.5) yields (3.2) as required, where d is defined by the first equality in (3.3). The second equality follows because $S\cap A^+S^\perp \subset R(A)$, and when restricted to the range of A, AA^+ acts as the identity. That is

$$\dim\{S\cap A^+S^\perp\} = \dim\{AS\cap AA^+S^\perp\} = \dim\{AS\cap S^\perp\}.$$

Remarks.

(a) Han & Fujiwara (1985, Corollary 2.5), obtained the special case of Theorem 3.1 that arises when the additional two conditions of A being invertible and d vanishing, are imposed as hypotheses. Han (1986b, Theorem 4.3) obtained the special case of Theorem 3.1 that arises when the additional condition of d

vanishing is imposed as a hypothesis.

(b) The roles of S and $S^{\perp}$, and A and A^{+} can be permuted in the proof of Theorem (3.1) to yield three other analogous equations. However, in light of (3.1) and (3.3) no new information is obtained.

§4. Connections between $\mathrm{In}A$ and $\mathrm{In}\, B^{T}AB$

As before A denotes a real $m\times m$ symmetric matrix, and B denotes a real $m\times n$ matrix. Recall that $\mathrm{In}(B^{T}AB)$ is the triple comprising the number of positive, negative and zero eigenvalues of the standard eigenvalue problem

$$B^{T}ABx = \lambda Ix, \; x \in \mathbb{R}^{n},$$

whereas $\mathrm{In}^{*}(B^{T}AB)$ is the triple comprising the number of positive, negative and zero eigenvalues of the eigenvalue problem

$$B^{T}ABy = \lambda B^{T}By, \; y \in \mathbb{R}^{n}, \; By \neq 0.$$

According to the theory of §2, with B $m\times n$,

$$\mathrm{In}(B^{T}AB) = \mathrm{In}^{*}(B^{T}AB) + (0,0,\, n-p), \tag{4.1}$$

where p is the rank of B, i.e. $p = \dim R(B)$.

Furthermore, if $S = R(B)$, the inertia $\mathrm{In}^{*}(S;A)$ associated with the subspace S satisfies

$$\mathrm{In}^{*}(S;A) = \mathrm{In}^{*}(B^{T}AB).$$

Consequently, Theorem (2.5) and (3.1) can be applied to obtain inertia theorems for certain matrices.

In effect it only remains to obtain the most explicit expressions for quantities such as S^A and $S \cap S^A$.

Suppose that $S = R(B)$; then

$$S^A = \{y:\ y^T Ax = 0,\ \forall x \in R(B)\} = \{y:\ y^T ABz = 0,\ \forall z \in \mathbb{R}^n\}.$$

That is $R(B)^A = (R(AB))^{\perp} = N(B^T A)$. It is also convenient to note that

$$N(B^T A) = A^{+}N(B^T) \oplus N(A), \tag{4.2}$$

the sum being direct. Here A^+ is the generalized inverse defined in §3. Equation (4.2) is merely a restatement of (3.4), because $N(B^T) = R(B)^{\perp}$. Consequent upon (4.2)

$$\dim\{R(B) \cap R(B)^A\} = \dim\{R(B) \cap (A^{+}N(B^T) \oplus N(A))\}$$

and

$$d^{\circ}(R(B)) = \dim\{R(B) \cap A^{+}N(B^T)\}.$$

Application of Corollary 2.7 then provides

<u>Corollary 4.1</u>. Let A be an $m \times m$ real symmetric matrix, let B be an $m \times n$ real matrix, and let C be any $m \times q$ matrix whose range is $N(B^T A)$. Then

$$\operatorname{In} A = \operatorname{In}^{*} B^T AB + \operatorname{In}^{*} C^T AC + (d, d, -d-e) \tag{4.3}$$

where

$$d = \dim\{R(B) \cap A^{+}N(B^T)\} \tag{4.4}$$

and

$$e = \dim\{R(B) \cap N(B^T A)\}. \tag{4.5}$$

<u>Remarks</u>.

(a) A result in terms of $\operatorname{In} B^T AB$ and $\operatorname{In} C^T AC$ is easily obtained by exploitation of (4.1).

(b) Notice that $d = e$ precisely if $R(B) \cap N(A) = \{\underset{\sim}{0}\}$. In particular if A is invertible, $d = e$.

(c) One possible choice for C is the $m \times m$ matrix $(I_m - (B^T A)^+ (B^T A))$.

(d) Dancis (1986, Theorem 3.1) obtained related inequalities that can be rederived from Corollary 4.1.

When B has rank m, Corollary (4.1) can be simplified. For then, $N(B^T) = \{\underset{\sim}{0}\}$ and $R(B) = \mathbb{R}^m$. Accordingly, $d = 0$, $e = \delta(A)$, and $\mathrm{In}^* C^T AC = (0,0, \delta(A))$.

Thus we have

<u>Corollary 4.2</u>. Let A be an $m \times m$ real symmetric matrix, and B be an $m \times n$, $n \geq m$, real matrix of rank m. Then

$$\mathrm{In}\, A = \mathrm{In}\, B^T AB + (0,0, m-n).$$

<u>Remarks</u>. (a) The case $m = n$ is Sylvester's Law of Inertia, and the case $n > m$ is merely a trivial extension that could be obtained directly. The more complicated result embodied in Corollary 4.1 indicates that there can be no direct and simple extension of Sylvester's Law of Inertia to the case where B is not of rank m.

(b) In contrast to the case covered by Corollary 4.2, the case $m > n$ with B of rank m provides no signficant simplification of (4.3).

In a similar way Theorem 3.1 can be translated into purely matrix form.

Corollary 4.3. Let A be an $m\times m$ symmetric matrix, let B be any $m\times n$ matrix, and let C be any $m\times q$ matrix whose range is $N(B^T)$. Then

$$\operatorname{In} A = \operatorname{In}^* B^T AB + \operatorname{In}^* C^T A^+ C + (d,d,-2d),$$

where

$$d = \dim\{R(B)\cap A^+ N(B^T)\} = \dim\{AR(B)\cap N(B^T)\}.$$

Remarks.

(a) One possible choice for C is the $m\times m$ matrix $(I - (B^T)^+ B^T) = I - BB^+,$.

(b) As before, a result in terms of $\operatorname{In} B^T AB$ and $\operatorname{In} C^T A^+ C$ is easily obtained via (4.1), namely

$$\operatorname{In} A = \operatorname{In} B^T AB + \operatorname{In} C^T A^+ C + (d,d,-2d) + (0,0,m-n-q).$$

§5. Applications to Partitioned Matrices

We shall here apply the theory of §§3 and 4 to obtain results relating the inertias of certain partitioned matrices to the inertias of submatrices.

Lemma 5.1. Let A be an $m\times m$ symmetric matrix, and let L denote the $2m\times 2m$ symmetric matrix

$$L = \begin{bmatrix} A & I \\ I & 0 \end{bmatrix}$$

Then

$$\operatorname{In} L = (m,m,0). \tag{5.1}$$

Moreover,

$$L^{-1} = \begin{bmatrix} A & I \\ I & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & I \\ I & -A \end{bmatrix}. \tag{5.2}$$

Proof. Denote the eigenvalues of A by μ_i, $i = 1,\ldots,m$, and denote the corresponding eigenvectors by u_i. Then L has m positive eigenvalues λ_i^+ given by the formula

$$\lambda_i^+ = \frac{\mu_i + \sqrt{\mu_i^2 + 4}}{2},$$

with corresponding eigenvectors $(\lambda_i^+ y, y)$.[1] Similarly, L has m negative eigenvalues

$$\lambda_i^- = \frac{\mu_i - \sqrt{\mu_i^2 + 4}}{2}$$

with eigenvectors $(\lambda_i^- y, y)$.

Remark. Equation (5.1) is widely known, see for example Chabrillac & Crouzeix (1984), Cottle (1974) and references therein. The explicit proof by exhibition of the eigenvalues and eigenfunctions that is given here will be exploited later.

Lemma 5.2. Let A be an $m\times m$ symmetric matrix and let B be an $m\times n$ matrix of rank p. Then

$$\mathrm{In}\begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix} = (m-d,\ m-d,\ n-p+2d) - \mathrm{In}^*(N(B^T);\ -A), \tag{5.3}$$

where

$$\begin{aligned} d &= \dim\{N(A)\cap N(B^T)\} + \dim\{AN(B^T)\cap R(B)\} \\ &= \dim\{N(B^T)\cap N(B^T)^A\} = \delta^*(N(B^T);A) \end{aligned} \tag{5.4}$$

---

[1]Here and throughout we adopt the convention that when x and y are column vectors, (x,y) is also a column vector.

Proof. Apply Theorem 3.1 with the matrix L of Lemma 5.1 playing the role of A and the subspace S being $(\mathbb{R}^m, R(B))$. Then

$$\operatorname{In} L = \operatorname{In}^* \begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix} + \operatorname{In}^*((\underset{\sim}{0}_m, N(B^T)); L^{-1}) + (d, d, -2d) \qquad (5.5)$$

where

$$d = \dim\{(\mathbb{R}^m, R(B)) \cap L^{-1}(\underset{\sim}{0}_m, N(B^T))\}.$$

But Lemma 5.1 provides L^{-1} explicitly, and consequently, for $y \in N(B^T)$, $L^{-1}(\underset{\sim}{0}_m, y) \in (\mathbb{R}^m, R(B))$ whenever $y \in N(A)$ or $Ay \in R(B)$. Consequently, d is given by the first equality in (5.4). The second equality follows from (3.4), and the third equality is then given by (2.5).

Furthermore,

$$\operatorname{In}^*((\underset{\sim}{0}_m, N(B^T)); L^{-1}) = \operatorname{In}^*(N(B^T); -A),$$

from (4.1)

$$\operatorname{In}^* \begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix} = \operatorname{In}^* \begin{bmatrix} I_m & 0 \\ 0 & B^T \end{bmatrix} \begin{bmatrix} A & I_m \\ I_m & 0 \end{bmatrix} \begin{bmatrix} I_m & 0 \\ 0 & B \end{bmatrix} = \operatorname{In} \begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix} - (0, 0, n-p),$$

and from Lemma 5.1

$$\operatorname{In} L = (m, m, 0).$$

Substitution into (5.5) then yields (5.3).

Remark. Chabrillac & Crouzeix (1984) obtained a result related to Lemma 5.2. Detailed discussion of the connection is deferred to §7. Han & Fujiwara (1985, Theorem 3.4) and Han (1986b, Theorem 4.6) obtained the special case with $d = 0$ by imposing additional hypotheses. Then

$$\operatorname{In}^*(N(B^T); -A) = (\nu^*(N(B^T); A), \pi^*(N(B^T); A), 0),$$

and because $\nu^*(N(B^T);A) + \pi^*(N(B^T);A) = m-p$, equation (5.3) can be rewritten

$$\mathrm{In}\begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix} = (p,p,n-p) + \mathrm{In}^*(N(B^T);A) = (p+\pi^*, p+\nu^*, n-p),$$

<u>Example 5.3</u>. Take $A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$.

Then $m = 3$, $n = 1$, and $p = 1$. Moreover, $N(B^T) = (\alpha,0,\beta)$, $R(B) = (0,\gamma,0)$ $\alpha,\beta,\gamma \in \mathbb{R}$ and $AN(B^T) = (0,\alpha+\beta,0)$. Furthermore, $d = 2$ because $N(A) = (\alpha,0,-\alpha)$, $\alpha \in \mathbb{R}$. It is also apparent that $\mathrm{In}^*(N(B^T);A) = (0,0,2)$, because $(\alpha,0,\beta)^T A(\alpha,0,\beta) = 0$. Thus Lemma 5.2 can be applied to find that

$$\mathrm{In}\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} = (3-2,\ 3-2,\ 1-1+4) - (0,0,2) = (1,1,2),$$

which result can be verified directly.

Lemma 5.2 is somewhat unsatisfactory in the sense that the quantity $\mathrm{In}^*(N(B^T);-A)$ is in general undetermined. Accordingly, special cases are of interest.

<u>Corollary 5.4</u>. Add the condition $p = m$ to the hypotheses of Theorem 5.2. Then

$$\mathrm{In}\begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix} = (m,m,n-m). \tag{5.6}$$

Proof. When $p = m$, $N(B^T) = \{0\}$.

The case $n = m$ is well known. The case $n > m$ can be obtained directly, without recourse to Lemma (5.2). See, for example, Cottle (1974) and references therein. A less trivial special case is

Corollary 5.5. Add the condition $N(B^T) \subset N(A)$ to the hypotheses of Theorem 5.2. Then

$$\mathrm{In}\begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix} = (p, p, n + m - 2p).$$

Proof. When $N(B^T) \subset N(A)$,

$$\mathrm{In}^*(N(B^T); -A) = (0, 0, \dim N(B^T)).$$

and $d = \dim N(B^T) = m - p$.

Consider now a $(m+n)\times(m+n)$ symmetric matrix M partitioned as

$$M = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \tag{5.7}$$

with A $m\times m$. The objective here is to apply Corollary 2.7 to obtain inequalities between InM and InA. These inequalities recover results of Dancis (1987). The more intricate analysis described in §6 below will provide *equalities* involving InA and InM.

Corollary 2.7 implies that for any symmetric matrix M and any subspace S

$$\pi(M) = \pi^*(S; M) + \pi^*(S^M; M) + \dim\{S \cap S^M \cap R(M)\}.$$

Trivially, $\pi^*(S^M;M)$ is nonnegative, so

$$\pi(M) - \pi^*(S;M) \geq \dim\{S \cap S^M \cap R(M)\}.$$

Moreover, from the properties of $In^*(S^M;M)$

$$\pi^*(S^M;M) + \nu^*(S^M;M) + \delta^*(S^M;M) = \dim S^M$$

and

$$\delta^*(S^M;M) = \delta(M) + \dim\{S \cap S^M \cap R(M)\}.$$

Consequently,

$$\pi^*(S^M;M) \leq \dim S^M - \delta^*(S^M;M)$$

and

$$\dim S^M - \delta(M) \geq \pi(M) - \pi^*(S;M) \geq \dim\{S \cap S^M \cap R(M)\}. \tag{5.8}$$

When the subspace S is characterized as the orthogonal complement of another subspace T, ie. $S = T^\perp$, (5.8) can be rewritten as

$$\dim\{MT\} \geq \pi(M) - \pi^*(T^\perp;M) \geq \dim\{MT^\perp \cap T\}. \tag{5.9}$$

When $T = (\underset{\sim}{0}_m, \mathbb{R}^n) \subset \mathbb{R}^{m+n}$, and M is of the form (5.7), we have that $\pi^*(T^\perp;M) = \pi(A)$, and (5.9) becomes

$$\begin{aligned} \dim\{R(B) \cup R(C)\} &= n - \dim\{N(B) \cap N(C)\} \\ &\geq \pi(M) - \pi(A) \geq \dim\{N(A) \cap R(B)\} \end{aligned} \tag{5.10}$$

The quantity $\nu(M) - \nu(A)$ satisfies the same inequalities.

§6. Partitioned Matrices and the Generalized Schur Complement

In this section M denotes an $\ell\times\ell$ real symmetric matrix that is partitioned in the form

$$M = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}, \tag{6.1}$$

where A and C are $m\times m$ and $n\times n$ real symmetric matrices, and B is an $m\times n$ matrix of rank p. Following Carlson et al (1974), we have

Definition 6.1. The generalized Schur complement of A in M, denoted M/A, is

$$M/A = C - B^T A^+ B, \tag{6.2}$$

where, as before, A^+ denotes the Moore-Penrose inverse of A. Whenever A is invertible $A^+ = A^{-1}$, and (6.2) reduces to the standard definition of the Schur complement of a symmetrically partitioned matrix.

In this section the preceding development will be applied to obtain certain equalities involving the inertias of M, A and M/A. These equalities subsume inequalities obtained by Carlson et al (op. cit.), as well as the classic result that applies when A is nonsingular (vide infra). Theorems due to Morse (1971) and to Han (1986) will also be discussed. It should be remarked that the papers of Ouellette (1981) and Styan (1985) provides a comprehensive survey of results involving the generalized Schur complement.

We first apply Corollary 2.7, with $S = \{R(A), 0_n\}$. Here, as before, (x,y) denotes a column vector. Thus

$$S^M = \{(x,y):\ Ax + By \in N(A)\}$$

or

$$S^M = (-A^+B\,\mathbb{R}^n, \mathbb{R}^n) \oplus (N(A), \underset{\sim}{0}_n),$$

and $S \cap S^M = \{\underset{\sim}{0}\}$, so $d^\circ(S) = 0$. Consequently,

$$\operatorname{In} M = \operatorname{In}^*(R(A);M) + \operatorname{In}^*(S^M;M). \tag{6.3}$$

It is apparent from Definition 2.2, with the underlying choice of basis taken from eigenvectors of A, that

$$\operatorname{In}^*(R(A);M) = (\pi(A),\ \nu(A),\ 0). \tag{6.4}$$

Theorem (2.6) is next applied to the subspace

$$Q = (-A^+BR(M/A),\ R(M/A))$$

regarded as a subspace of S^M. Consequently, the subspace $S^M \cap Q^M$ must be determined. The calculation is as follows. An element of $S^M \cap Q^M$ is of the form

$$(x - A^+Bz, z),\ x \in N(A),\quad z \in \mathbb{R}^n,$$

where

$$(x - A^+Bz, z)^T M(-A^+By,\ y) = 0,\quad \forall y \in R(M/A). \tag{6.5}$$

Block multiplication and the properties of A^+ demonstrate that (6.5) is equivalent to

$$y^T(B^T x + (M/A)z) = 0,\quad \forall y \in R(M/A).$$

Thus

$$B^T x + (M/A)z \in N(M/A),$$

or

$$z + (M/A)^+B^T x \in N(M/A).$$

Consequently,

$$S^M \cap Q^M = \{(x - A^+Bz, z):\ x \in N(A) \ \text{ and } \ z + (M/A)^+B^T x \in N(M/A)\} \tag{6.6}$$

Moreover, $S \cap S^M \cap Q^M = \{\underset{\sim}{0}\}$, so $d^{\circ}(Q) = 0$. Theorem 2.5 therefore allows (6.3) to be written in the form

$$\mathrm{In}\, M = \mathrm{In}^*(R(A);M) + \mathrm{In}^*(Q;M) + \mathrm{In}^*(S^M \cap Q^M;M) \tag{6.7}$$

It has already been noted in (6.4) that $\mathrm{In}^*(R(A);M)$ has a particularly simple form. It is next shown that

$$\mathrm{In}^*(Q;M) = (\pi(M/A),\ \nu(M/A), 0). \tag{6.8}$$

To appreciate this fact calculate that the restricted quadratic form

$$w^T Mw,\ w \in Q,$$

is identical with

$$(-A^+By,\ y)^T\ M(-A^+By,\ y), \qquad y \in R(M/A)$$

which, because $A^+AA^+ = A^+$, is equivalent to

$$y^T (M/A)y,\ y \in R(M/A).$$

As before, Definition (2.2) then implies (6.8).

In light of (6.4) and (6.8) it remains to analyze the term $\mathrm{In}^*(S^M \cap Q^M;M)$ appearing in (6.7). To this end consider the quadratic form

$$z^T (M/A)z, \quad z \in S^M \cap Q^M,$$

which is equivalent to

$$(x - A^+By,\ y)^T M(x-A^+By,\ y), \qquad y = -(M/A)^+B^T x + w,$$
$$x \in N(A),\ w \in N(M/A).$$

or

$$2x^T Bw - x^T B(M/A)^+B^T x, \qquad x \in N(A),\ w \in N(M/A).$$

From this last expression it is apparent that

$$\mathrm{In}^*(S^M \cap Q^M;M) = \mathrm{In}^*(N(T);L) \tag{6.9}$$

where the $\ell \times \ell$ symmetric matrices L and T are defined by

$$L = \begin{bmatrix} -B(M/A)^{+}B^{T} & B \\ B^{T} & 0 \end{bmatrix}, \tag{6.10}$$

and

$$T = \begin{bmatrix} A & 0 \\ 0 & (M/A) \end{bmatrix}. \tag{6.11}$$

The determination of an inertia of the form (6.9) is precisely the question of restricted quadratic forms that will be treated in §7, and either of the two methods described there could in principle be applied. Moreover, the calculations arising are not as formidable as might be imagined, for Corollary 5.5 applies to state that

$$\operatorname{In} L = (p, p, n+m-2p),$$

where p is the rank of B. It can also be easily verified that

$$L^{+} = \begin{bmatrix} 0 & B^{+^{T}} \\ B^{+} & -B(M/A)^{+}B^{T} \end{bmatrix}.$$

Nevertheless, that course is not pursued here. Instead, a direct assault is launched.

<u>Theorem 6.1</u>. Let M and M/A be defined as in (6.1) and (6.2). Then

$$\begin{aligned} \operatorname{In} M = \operatorname{In} A + \operatorname{In}(M/A) &+ \operatorname{In}^{*}(B^{T} N(A) \cap R(M/A); -(M/A)^{+}) \\ &+ (t, t, -t - \dim\{B^{T} N(A)\}) \end{aligned} \tag{6.12}$$

where

$$t = \dim\{B^{T} N(A) \cap N(M/A)\} \tag{6.13}$$

<u>Proof</u>. The matrix L defined in (6.10) can be factored as

$$L = \begin{bmatrix} B & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} -(M/A)^{+} & I \\ I & 0 \end{bmatrix} \begin{bmatrix} B^{T} & 0 \\ 0 & I \end{bmatrix} \equiv \tilde{B}^{T} \tilde{L} \tilde{B}$$

Accordingly,

$$\mathrm{In}^*(N(T);-L) = \mathrm{In}^*(\tilde{B}N(T);\tilde{L}) + (0,0,\ \dim\{N(\tilde{B})\cap N(T)\}) \qquad (6.14)$$

where the matrix T was defined in (6.11). But

$$\begin{aligned}\dim\{N(\tilde{B})\cap N(T)\} = \dim\{N(B^T)\cap N(A)\} &= \delta(A) - \dim\{R(B)\cap N(A)\} \\ &= \delta(A) - \dim\{B^T N(A)\}. \qquad (6.15)\end{aligned}$$

Furthermore, $\mathrm{In}^*(\tilde{B}N(T);\tilde{L})$ can be estimated directly. The objective is to construct a $\tilde{L}$-orthogonal basis for $\tilde{B}N(T)$ and apply Lemma 2.2. The subspace $\tilde{B}N(T)$ can be decomposed as

$$\begin{aligned}\tilde{B}N(T) &= \{B^T N(A)\cap R(M/A), \underset{\sim}{0}\} \oplus \{B^T N(A)\cap N(M/A),\ B^T N(A)\cap N(M/A)\} \\ &\qquad \oplus \{\underset{\sim}{0},\ N(M/A)\cap(B^T N(A))^{\perp}\} \\ &\equiv S_1 \oplus S_2 \oplus S_3.\end{aligned}$$

It is apparent that the subspaces S_i are orthogonal. It can also be verified that they are $\tilde{L}$-orthogonal. Consequently,

$$\mathrm{In}^*(\tilde{B}N(T);\tilde{L}) = \mathrm{In}^*(S_1;\tilde{L}) + \mathrm{In}^*(S_2;\tilde{L}) + \mathrm{In}^*(S_3;\tilde{L}). \qquad (6.16)$$

Now

$$(\underset{\sim}{0},\underset{\sim}{x})^T\ \tilde{L}(0,\underset{\sim}{x}) = 0, \quad \forall \underset{\sim}{x} \in \mathbb{R}^n,$$

so

$$\begin{aligned}\mathrm{In}^*(S_3;\tilde{L}) &= (0,0,\ \dim\{N(M/A)\cap(B^T N(A))^{\perp}\}) \\ &= \delta(M/A) - \dim\{B^T N(A)\cap N(M/A)\}. \qquad (6.17)\end{aligned}$$

Moreover, as was described in the proof of Lemma 5.1, the vectors $(y_i, \pm y_i)$ are eigenvectors of $\tilde{L}$ (with corresponding eigenvalues ± 1) whenever $y_i \in N(M/A)$. Thus the subspace S_2 is mapped into itself by $\tilde{L}$, and

$$\mathrm{In}^*(S_2;\tilde{L}) = (t,t,0) \qquad (6.18)$$

where

$$t = \dim\{B^T N(A) \cap N(M/A)\}. \qquad (6.19)$$

The best that can be said about the remaining term is that

$$\mathrm{In}^*(S_1;\tilde{L}) = \mathrm{In}^*(B^T N(A) \cap R(M/A); -(M/A)^+). \qquad (6.20)$$

Equations (6.4), (6.7), (6.8), (6.9) and (6.14)-(6.20) provide (6.12) and (6.13) as required.

Remarks.

(a) The classic result in this area concerns the case $N(A) = \{0\}$, in which

$$\mathrm{In}\,M = \mathrm{In}\,A + \mathrm{In}(M/A), \qquad (6.21)$$

which formula is due to Haynsworth (1968). Theorem 6.1 demonstrates that the conditons

$$B^T N(A) \subset R(M/A),$$

and

$$\begin{aligned} \mathrm{In}^*(B^T N(A); -(M/A)^+) &= (0,0,\ \dim\{B^T N(A)\}) \\ &= (0,0,\ \dim\{R(B) \cap N(A)\}) \end{aligned}$$

are necessary and sufficient for (6.21) to hold.

(b) Carlson et al (1974) obtain the inequalities

$$\pi(M) \geq \pi(A) + \pi(M/A)$$

$$\nu(M) \geq \nu(A) + \nu(M/A)$$

that arise from (6.12) when the last two terms are discarded. The inequality

$$\delta(M) \leq \delta(A) + \delta(M/A)$$

is then immediate. They also obtain necessary and sufficient conditions for equality in these inequalities, namely:

$N(M/A) \subset N((I-AA^+)B)$,[2] & $(I-AA^+)B(M/A)^+B^T(I-AA^+) = 0.$

These conditions are equivalent to the necessary and sufficient conditions for equality given in Remark (a) above. Actually, the interests of Carlson et al are not restricted to *symmetric* partitioned matrices, and the method of proof they use is correspondingly more general.

(c) Lemma 5.2 is actually a special case of Theorem 6.1. This fact is not immediately obvious, but is consequent upon the following calculations. When $C = 0$, $(M/A) = -B^T A^+ B$, and

$$\dim\{B^T N(A) \cap R(-B^T A^+ B)\} \leq \{N(A) \cap R(A^+ B)\} = 0.$$

Consequently, $B^T N(A) \subset N(M/A)$, so that

$$\mathrm{In}^*(B^T N(A) \cap R(M/A);\ -(M/A)^+) = (0,0,0),$$

and

$$\begin{aligned} t = \dim\{B^T N(A) \cap N(M/A)\} &= \dim B^T N(A) \\ &= \dim\{N(A) \cap R(B)\}. \end{aligned} \qquad (6.22)$$

Then (6.12) becomes

$$\mathrm{In}\, M = \mathrm{In}\, A + \mathrm{In}(B^T(-A^+)B) + (t,t,-2t). \qquad (6.23)$$

According to equation (4.1), $\mathrm{In}(B^T(-A^+)B)$ can be rewritten as $\mathrm{In}^*(B^T(-A^+)B) + (0,0,n-p)$, and Corollary 4.3 can be applied to eliminate $\mathrm{In}^*(B^T(-A^+)B)$ in favour of a term $\mathrm{In}^*(N(B^T);-A)$. The relation so obtained is

[2] In fact, the paper of Carlson et al contains a typographical error in which the inclusion sign has been replaced by an equality.

$$\operatorname{In} M = \operatorname{In} A + \operatorname{In}(-A^{+}) - \operatorname{In}^{*}(N(B^{T});-A)$$
$$+ (t-d, t-d, -2(t-d) + n-p),$$

where

$$d = \dim\{R(B) \cap AN(B^{T})\}.$$

But,

$$\operatorname{In} A + \operatorname{In}(-A^{+}) = (\pi(A) + \nu(A),\ \pi(A) + \nu(A),\ 2\delta(A))$$
$$= (m-\delta(A),\ m-\delta(A),\ 2\delta(A))$$

and

$$t - \delta(A) = \dim\{N(A) \cap R(B)\} - \dim\{N(A)\} = -\dim\{N(A) \cap N(B^{T})\}.$$

Consequently, equation (5.3) is recovered.

Example 6.2. Consider the matrix

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \\ 0 & 1 & -1 & 1 \end{bmatrix}$$

Here

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$$

$$A^{+} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad (M/A) = C = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix},$$

and

$$B^{T} N(A) = N(M/A) = \operatorname{span}\ (1,1).$$

Then Theorem 6.1 reduces to

$$\operatorname{In} M = (1,0,1) + (1,0,1) + (0,0,0) + (1,1,-2) = (3,1,0),$$

and it can be verified that M has eigenvalues $\pm\sqrt{2}, 1$, and 2.

We now consider a problem involving partitioned matrices that was originally posed by Morse (1971), and which has also been analyzed in terms of the standard, as opposed to generalized, Schur complement by Cottle (1974). The problem is to determine

$$\mathrm{In}^*(N(T);M)$$

where the $(m\times n)\times(m\times n)$ matrix M is partitioned as in (6.1) and

$$T = [A \quad B].$$

That is, the constraint matrix T coincides with the top segment of the partitioned matrix M. One motivation for the study of such problems is the minimization of $z^T Mz = (x,y)^T M(x,y)$ with respect to the variable x only.

Corollary 2.7 is applied with $S = (\mathbb{R}^m, 0)$. Then

$$S^M = \{(x,y): Ax + By = 0\} = N(T).$$

Moreover,

$$S\cap S^M = (N(A),0)$$

so that $d^{\circ}(S) - \dim\{S\cap S^M \cap R(M)\} = \dim\{N(A)\cap R(B)\}$. Consequently, it may be concluded that

$$\mathrm{In}\, M = \mathrm{In}\, A + \mathrm{In}^*(N(T);M) + (d,d,-d-\delta(A)) \tag{6.24}$$

where

$$d = \dim\{N(A)\cap R(B)\} = \dim\{B^T N(A)\}.$$

Morse (op. cit.) obtained results in two special cases, namely M nonsingular, or A nonsingular. When A is nonsingular, (6.24) reduces to

$$\mathrm{In}\, M = \mathrm{In}\, A + \mathrm{In}^*(N(T);M), \qquad \text{with} \quad \delta(A) = 0.$$

As was pointed out by Cottle (op. cit.), this case is particularly straightforward because

$$N(T) = A^{-1}R(B),$$

and the restriction of M to $N(T)$ coincides with the quadratic form associated with the matrix M/A. That is

$$\mathrm{In}^*(N(T);M) = \mathrm{In}(M/A).$$

The second case considered by Morse was M nonsingular. This assumption implies that

$$N(A) \subset R(B),$$

so

$$d = \delta(A).$$

Rearrangement of (6.24) then provides the equations

$$\nu(M) = \nu(A) + \nu^*(N(T);M) + \delta(A),$$
$$\delta(A) = \delta^*(N(T);M),$$

and

$$\pi(M) = \pi(A) + \pi^*(N(T);M) + \delta(A),$$

which results are equivalent to the relations obtained by Morse and rederived by Cottle.

In point of fact, the main thrust of Morse's work was to obtain an expression for the quantity $\mathrm{In}\,M - \mathrm{In}\,A$, and the introduction of the restriction of M to $N(T)$, was one natural way to find such expressions. However, equation (6.12) of Theorem 6.1 provides a different formula for $\mathrm{In}\,M - \mathrm{In}\,A$. Of course, (6.24) and (6.12) are closely related. They coincide exactly in the case that A is invertible. However, in general, the two equations provide distinct information.

Other results are obtained if the quantity $\mathrm{In}\,M - \mathrm{In}\,A$ is eliminated from (6.12) and (6.24) to obtain an expression for $\mathrm{In}^*(N(T);M)$. We shall only consider the case $C = 0$, in which

$\operatorname{In} M - \operatorname{In} A$ can be eliminated between (6.23) and (6.24), to obtain

$$\operatorname{In}^*(N(T);M) = \operatorname{In}(B^T(-A^+)B) + (0,0,\ \delta(A)-t) \tag{6.25}$$

where

$$t = \dim\{N(A)\cap R(B)\}.$$

But

$$\begin{aligned}\operatorname{In}(B^T(-A^+)B) &= \operatorname{In}^*(B^T(-A^+)B) + (0,0,\ n-p)\\ &= \operatorname{In}^*(R(B);-A^+) + (0,0,\ n-p)\end{aligned}$$

and, as in equation (3.7),

$$\begin{aligned}\operatorname{In}^*(R(B);-A^+) = \operatorname{In}^*(A^+N(B^T);-A) &= \operatorname{In}^*(N(B^T)^A;-A)\\ &\quad - (0,0,\ \dim\{N(A)\cap N(B^T)\}).\end{aligned}$$

Substitution into (6.25) then yields

Lemma 6.3. Let the $(m+n)\times(m+n)$ symmetric matrix M be partitioned as

$$M = \begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix} \tag{6.26}$$

where B is $m\times n$ of rank p. Let

$$T = [A \quad B].$$

Then

$$\operatorname{In}^*(N(T);M) = \operatorname{In}^*(N(B^T)^A;-A) + (0,0,\ n-p) \tag{6.27}$$

Lemma 6.3 was first obtained by Han (1986a, Theorem 2.2) in his investigations concerning the Wolfe dual that arises in nonlinear programming.

When M is of the form (6.26) we have obtained three expressions for $\operatorname{In} M$, namely (5.3), (6.23) and (6.24). It has already been remarked that equations (6.23) and (5.3) are, in a certain sense, dual. If $\operatorname{In} M$ is eliminated between (5.3) and (6.24), rather than between (6.23) and (6.24) we obtain, after some calculation,

$$\operatorname{In}^*(N(T);M) = \operatorname{In}(-A) - \operatorname{In}^*(N(B^T);-A) + (-d^\circ,-d^\circ,d^\circ + d + n-p)$$

where

$$d = \dim\{N(B^T)\cap N(B^T)^A\} = \dim\{N(A)\cap N(B^T)\} + d^\circ(N(B^T))$$

and

$$d^\circ(N(B^T)) = d^\circ = \dim\{A\, N(B^T)\cap R(B)\}.$$

As was to be expected this last result can be obtained directly from (6.27) by use of Corollary 2.7 to eliminate the term $\operatorname{In}^*(N(B^T)^A;-A)$.

As a final comment it is instructive to notice that the proofs of (6.12) and (6.24) both start with an application of Corollary 2.7 to the matrix M. However, the choice of subspace S is different. When $S = (R(A),0)$ the analysis leads to (6.12). When $S = (\mathbb{R}^m,0)$, (6.24) is obtained.

§7. Tests for Positivity of Restricted Quadratic Forms

The preceding development has exploited the equivalence of symmetric matrices and quadratic forms, but the main focus has been on inertia theorems for matrices. Emphasis is now switched to consideration of quadratic forms. In particular, we shall consider tests that guarantee either

$$x^T Ax > 0 \qquad \forall x \in S\backslash\{0\} \tag{7.1}$$

or

$$x^T Ax \geq 0 \qquad \forall x \in S. \tag{7.2}$$

Here, as before, A is an $m \times m$ symmetric matrix and S is a subspace of $\mathbb{R}^m$. It is further assumed that

$$S = N(B^T) \tag{7.3}$$

where B is an $m \times n$ matrix of rank p.

One motivation for the study of conditions such as (7.1) and (7.2) arises in constrained nonlinear programming. Then A is to be interpreted as the Hessian of the Lagrangian, and B is the transpose of the matrix that is the gradient of the constraints. Details can be found in many references, for example, Hestenes (1966).

In the notation of this development, condition (7.1) is equivalent to

$$\mathrm{In}^*(S;A) = (m-p,0,0), \tag{7.4}$$

and (7.2) is equivalent to

$$\nu^*(S;A) = 0. \tag{7.5}$$

It is apparent that the preceding theory is relevant, but that theory merely translates conditions such as (7.4) or (7.5) into different criteria. The immediate question that arises is whether any new criterion is actually easier to verify than the original condition. Concomitantly, before a proper discussion of the problem can be undertaken, some assumptions must be made as to what quantities are comparitively easy to compute. The assumptions that will be made here are:

(i) If a concrete expression for a matrix is known, its inertia can be calculated.

(ii) Calculations of inverses, or generalized inverses, are to be avoided.

According to these maxims, calculation of $\operatorname{In} B^T AB$, which involves knowledge of $\operatorname{In}^* B^T AB$ and the rank of B, would be preferred to calculation of $\operatorname{In}^*(N(B^T);A)$. This is because the concrete matrix whose inertia is associated with $\operatorname{In}^*(N(B^T);A)$ is $(I-BB^+)A(I-BB^+)$, which matrix must first be found.

The theory described in this work offers two indirect approaches to the verification of conditions such as (7.1) or (7.2). The first indirect approach is offered by Corollary 2.7, or one of its mutations involving A^+. For example, equations (3.1), (4.1) and Corollary (4.1) provide the equation

$$\operatorname{In}^*(N(B^T);A) = \operatorname{In} A - \operatorname{In} B^T A^+ B + (d,d,\ n-p-2d) \tag{7.6}$$

where

$$d = \dim\{R(AB) \cap N(B^T)\}. \tag{7.7}$$

Thus, conditions (7.4) and (7.5) on $\operatorname{In}^*(N(B^T);A)$ can be translated into conditions on $\operatorname{In} A$, $\operatorname{In}^*(R(B);A^+)$ and d. According to assumption (i) above, it is easier to calculate $\operatorname{In} A$ than to calculate $\operatorname{In}^*(N(B^T);A)$. However, two other quantities remain to be determined, and the generalized inverse A^+ appears. Nevertheless, this first indirect approach was successfully employed in an analogous infinite-dimensional problem considered by Maddocks (1985). The crucial feature of that example was that while $S = N(B^T)$

was of infinite dimension, S^A was of low dimension. Consequently, calculation of $\mathrm{In}^*(R(B);A^+)$ was reduced to the solution of a small number of inhomogeneous equations of the form

$$Ax_i = b_i,$$

where b_i is a basis for $R(B)$. The analysis of the problem of Morse that was presented at the end of §6 provides another example of this first, indirect approach to the determination of $\mathrm{In}^*(N(B^T);A)$.

The second indirect approach is provided by Lemma 5.2, which may be rewritten

$$\mathrm{In}^*(N(B^T);-A) = (m-d,\ m-d,\ n-p+2d) - \mathrm{In}\,M. \tag{7.8}$$

where

$$d = \dim\{N(B^T)\cap N(B^T)^A\} = \delta^*(N(B^T);-A) \tag{7.9}$$

and M is the *bordered matrix* (cf. Cottle, 1974, Chabrillac & Crouzeix, 1984)

$$M = \begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix} \tag{7.10}$$

It should here be remarked that

$$\nu^*(N(B^T);-A) = \pi^*(N(B^T);A),\ \pi^*(N(B^T);-A) = \nu^*(N(B^T);A),$$
$$\text{and}\qquad \delta^*(N(B^T);-A) = \delta^*(N(B^T);A),$$

so that (7.8) does provide an expression for $\mathrm{In}^*(N(B^T);A)$. Actually, from the point of view of this development it would be more natural to introduce a bordered matrix with upper left element $-A$, but that approach is not in accord with previous works.

According to our assumptions, calculation of $\operatorname{In} M$ is straightforward, albeit that M is an $(m+n)\times(m+n)$ matrix rather than a $m\times m$ matrix. The degree of complexity involved in calculation of d is not immediately apparent. However, the particular structure of (7.8) makes this issue mute for we have

<u>Theorem 7.1</u>. Let S and M be defined as in (7.3) and (7.10). Then M has at least p positive, p negative and $(n-p)$ zero eigenvalues. Moreover,

(i) property (7.1) holds if and only if $\pi(M) = m$;

and

(ii) property (7.2) holds if and only if $\nu(M) = p$.

This result is due to Chabrillac & Crouzeix (1984, Theorem 1). It also follows from (2.2), (7.4), (7.5), (7.8) and (7.9).

It should be remarked that while the two lines of attack suggested here are in some respects similar, they also possess intrinsic differences. The first method is based upon A-orthogonality in $\mathbb{R}^m$: the second method is based upon M-orthogonality in $\mathbb{R}^{m+n}$.

The theory of the bordered matrix has further ramifications. For example, Han (1986b) discusses its role in the theory of Wolfe duality in nonlinear programming. Another potentially interesting observation is that in problems where the subspace S of (7.1) or (7.2) is defined in terms of the range of B, equation (6.23) can be used as the basis of a result involving a bordered matrix with top left entry A^+.

Finally, it should be stressed that there are approaches to conditions (7.1) and (7.2) other than the inertia, or index, theory described here. For example, there are tests involving determinants, and tests based upon penalization or augmentability methods. Chabrillac & Crouzeix (1984) survey several of these techniques.

§8. Summary & Discussion

The first part of this work introduces an inertia $\mathrm{In}^*(S;A)$ that is associated with a pair comprising a subspace S and a symmetric matrix A. The main result is Theorem 2.6 which states relations between $\mathrm{In}^*(S;A)$ and $\mathrm{In}^*(S^A;A)$, where S^A is the subspace that is the A-orthogonal complement of S. This theory essentially comprises a specialization to finite-dimensions, and an extension of, a general Hilbert space theory developed by Hestenes (1951) and Maddocks (1985).

The definition and properties of $\mathrm{In}^*(S;A)$ are nontrivial precisely because the matrix A need not map the subspace S to itself. Indeed the quantity

$$d^\circ(S;A) = \dim\{S \cap S^A \cap R(A)\},$$

known as the relative nullity of A on S, plays a central role in this analysis. Han & Fujiwara (1985) and Han (1986b) developed a comparable finite-dimensional theory under additional hypotheses guaranteeing that the relative nullity $d^\circ(S;A)$ vanishes. A corollary of the analysis given here is that $d^\circ(S;A)$ vanishes precisely when $\mathbb{R}^m$ can be written as the sum $S+S^A$.

The analysis of §§2 and 3 is applied in §§4,5 and 6 to obtain results concerning the (standard) inertias of symmetric matrices with particular structures. The first results (Corollaries 4.1 - 4.3) concern matrices of the form $B^T AB$. The statement of Corollary 4.1 is rather complicated, but it is the direct generalization to matrices B not of full rank, of Sylvester's Law of Inertia.

Sections 5 and 6 treat symmetric partitioned matrices M of the form

$$M = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}. \tag{8.1}$$

The main result for such matrices is Theorem 6.1, which relates the inertias of the matrices M, A and M/A, the generalized Schur complement of A in M. The statement of Theorem 6.1 is again somewhat complicated, but it subsumes inequalities obtained by Carlson et al (1974), and, in the case of A being invertible, an inertia theorem of Haynsworth (1968) is recovered.

Several of the results on partitioned matrices concern the case $C = 0$, i.e.

$$M = \begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix}. \tag{8.2}$$

For such matrices three different expressions for the inertia of M are obtained, namely equations (5.3), (6.23) and (6.24). Equation (5.3) includes a widely known result, here stated as Corollary 5.4, in which B has full rank. Chabrillac & Crouzeix (1984) obtained a result closely related to equation (5.3), but they state the result in terms of properties of quadratic forms (vide infra). Han & Fujiwara (1985) obtain an expression equivalent to (5.3) in the special case that the quantity d, defined in (5.4), vanishes.

Equation (5.3) is here derived directly, whereas equation (6.23) is obtained as a special case of Theorem 6.1. However, equations (5.3) and (6.23) are also shown to be dual in a sense associated with the A-orthogonality theory developed in §§2 and 3.

Equations (5.3) and (6.23) involve various restricted inertias of A, but only the unrestricted inertia of M. However, equation (6.24) involves the inertia of M restricted to the nullspace of the $m \times (m+n)$ matrix

$$T = [A \quad B].$$

When the triple $\operatorname{In} M$ is eliminated between equations (6.23) and (6.24) further information is obtained, which is here formalized as Lemma 6.3. This result is originally due to Han (1986a), who exploited it in the context of the Wolfe dual that arises in nonlinear programming.

Equation (6.24) remains valid in the case $C \neq 0$. It then encompasses results of Morse (1971) and Cottle (1974) that apply in either of the cases of A or M being nonsingular.

The properties of restricted quadratic forms are considered in §7. Two approaches to tests for positivity of a quadratic form $x^T Ax$ on a subspace $S = N(B^T) \subset \mathbb{R}^m$ are described. The first test exploits A-orthogonality in $\mathbb{R}^m$. The second test, originally due to Chabrillac & Crouzeix (1984), involves the $(m+n)\times(m+n)$ bordered matrix M of the form (8.2). This last test is here shown to be a consequence of M-orthogonality in $\mathbb{R}^{m+n}$.

It has already been remarked that the essence of the theory developed in §§2 and 3 can be obtained by specialization to finite-dimensions of a more general Hilbert space theory. As is detailed in §§4 to 7, this descent from infinite-dimensions allows both an extension and a unification of known finite-dimensional results. Contrariwise, few of the proofs utilized here are intrinsically finite-dimensional, and some of the known matrix results now indicate potentially viable routes to new theorems valid in Hilbert space. For example, the bordered matrix (8.2) is the direct analogue of certain systems of linear differential equations that arise in the study of the second-variation in the isoperimetric calculus of variations. (see, for example, Maddocks, 1985,§3).

Acknowledgements: I should like to thank H. Wolkowicz, who apprised me of much of the literature cited here, and the referees, who provided additional references.

References

D. Carlson, E. Haynsworth and T. Markham (1974); A generalization of the Schur complement by means of the Moore-Penrose inverse, SIAM J.Applied Math. 26, pp.169-175.

Y. Chabrillac and J.-P. Crouzeix (1984); Definiteness and semi-definiteness of quadratic forms revisited, Linear Algebra and Its Applications 63, pp.283-292.

R. W. Cottle (1974); Manifestations of the Schur complement, Linear Algebra and Its Applications 8, pp.189-211.

J. Dancis (1986); A qualitative formulation of Sylvester's Law of inertia III, Linear Algebra and Its Applications 80, pp. 141-158.

J. Dancis (1987); On the inertias of symmetric matrices and bounded self-adjoint operators, submitted to Linear Algebra and Its Applications.

J. Gregory (1980); Quadratic Form Theory and Differential Equations, Academic Press, New York.

S.-P. Han (1986a); On the Hessian of Lagrangian and second-order optimality conditions, SIAM J.Control and Optimization 24, pp.339-345.

S.-P. Han (1986b); Relative eigenvalues of Hermitian matrices, Linear Algebra and Its Applications 81, pp.75-88.

S.-P. Han and O. Fujiwara (1985); An inertia theorem for symmetric matrices and its application to nonlinear programming, Linear Algebra and Its Applications 72, pp.47-58.

E. V. Haynsworth (1968); Determination of the inertia of a partitioned Hermitian matrix, Linear Algebra and Its Applications 1, pp.73-81.

M. R. Hestenes (1951); Applications of the theory of quadratic forms in Hilbert space to the calculus of variations, Pacific Journal of Mathematics 1, pp. 525-581.

M. R. Hestenes (1966); Calculus of Variations and Optimal Control Theory, John Wiley, New York.

J. H. Maddocks (1985); Restricted quadratic forms and their application to bifurcation and stability in constrained variational principles, SIAM J.Applied Math. 16, pp.47-68.

M. Morse (1971); Subordinate quadratic forms and their complementary forms, Revue Roumaine de Math Pures et Appliques 16, pp.558-569.

D. V. Ouellette (1981); Schur complements and statistics, Linear Algebra and Its Applications 36, pp.187-295.

G. P. H. Styan (1985); Schur Complements and Linear Statistical Models, Proc. First Tampere Seminar on Linear Models (1983), published by Dept. of Math. Sci. Univ. Tampere, Finland.

H. F. Weinberger (1974); Variational methods for eigenvalue approximation, C.B.M.S. Conference Series #15. SIAM.