SC

SYMMETRIZING MATRICES

A thesis submitted

by

E. J. Desautels

to

the Faculty of Pure and Applied Science

of the University of Ottawa

in partial fulfillment of the requirements

for the degree of

Master of Science

in the subject of

Mathematics

ABSTRACT

SYMMETRIZING MATRICES

A symmetrizing matrix of an arbitrary n-square matrix M is defined as an n-square symmetric matrix B such that BM = M′B. Elementary properties of symmetrizing matrices are established, and an interpretation of a symmetrizing matrix B of M is given with B as the representation of a scalar-product, not necessarily positive definite, with respect to which the arbitrary matrix M, symmetrized by B, represents a self-adjoint operator.

Some basic concepts of linear algebra are discussed, leading to a complete derivation of the Jordan canonical form theorem. By considering an arbitrary square matrix in its Jordan canonical form, a complete solution of the symmetrization problem is given, arriving at the results of M. Marcus and N. A. Khan [Pacific J. Math., 10 (1960) 1337-1346]. For the class of nonderogatory matrices it is shown that the symmetrizing matrices are congruent to direct sums of persymmetric matrices.

Examples are given of symmetrizing matrices for a companion matrix. References are provided for the construction of symmetrizing matrices of arbitrary matrices.

ACKNOWLEDGEMENT

The author is deeply indebted to Professor J. L. Howland for initially proposing the topic for this thesis, and for his unfailing assistance and encouragement. The author wishes to thank the University of Ottawa for financial support in the preparation of this thesis, and also the IBM Systems Research Institute, New York City, for their cooperation in the typing of the manuscript.

INTRODUCTION

The concept of symmetrization of linear operators appears to have been first formulated by Marty [18] in the case of linear integral operators; the text by Zaanen [21] provides references and a detailed development. Later authors used this concept in the study of the eigenfunctions and eigenvalues of the integral equations of classical potential theory [3] and of certain generalizations of these equations [10].

It is reasonable to conjecture that the symmetrization of finite-dimensional operators may lead to new results or methods in the spectral theory of matrices, or to new procedures for the numerical evaluation of matrix eigenvalues and eigenvectors. In particular, two methods for the evaluation of the zeros of a polynomial as eigenvalues of its symmetrized companion matrix have been reported ([11] and [12]). Procedures for the evaluation of the eigenvalues of an arbitrary matrix by symmetrization techniques are currently under development ([5] and [14]).

This thesis is a self-contained account of the fundamental theory of symmetrizing matrices, giving the elementary properties and simple geometric interpretations of such matrices, and a detailed accounting of the number and nature of such matrices which exist for a given matrix. Problems relating to the practical construction of symmetrizing matrices, or to the signatures of the related quadratic forms, etc., are not discussed.

TABLE OF CONTENTS

CHAPTER I

ELEMENTARY PROPERTIES OF SYMMETRIZING MATRICES

This chapter will provide an informal development of the elementary properties of symmetrizing matrices. The following chapters will attempt a more rigorous treatment.

1. Symmetrizing Matrices

We shall use A' to denote the transpose of the matrix A and say that B is a symmetrizing matrix of M if B and M are related as follows:

Definition: Given an n-square matrix M over a field F, a symmetrizing matrix of M is an n-square symmetric matrix B over F such that

(1) ... $BM = M'B$.

If B is a symmetrizing matrix of M, we shall say that B symmetrizes M. Two lemmas follow immediately from the definition, the first of which may be used to motivate calling B a "symmetrizing" matrix.

Lemma 1: If B is a symmetrizing matrix of M, then BM is a symmetric matrix.

Proof: We have $B' = B$ and $BM = M'B$.

Thus $(BM)' = M'B' = M'B = BM$.

q.e.d.

Lemma 2: If B is a symmetrizing matrix of M, then BM is also a symmetrizing matrix of M.

Proof: $(BM)M = (M'B)M = M'(BM)$

q.e.d.

Corollary: If B is a symmetrizing matrix of M, then $BM, BM^2, \ldots BM^k \ldots$ are also symmetrizing matrices of M.

Certain classes of matrices are trivially symmetrizable. For instance, if M is a symmetric matrix, then it is symmetrized by the identity matrix I, and also by M itself. In particular if M is a diagonal matrix, every diagonal matrix symmetrizes M. Examples of symmetrizing matrices for a matrix which is not symmetric will be given in the final chapter.

In the event M is symmetrized by a nonsingular matrix B, we immediately obtain from $BM = M'B$ the relation $MB^{-1} = B^{-1}M'$. Thus we have

Lemma 3: If M has a nonsingular symmetrizing matrix B, then M' has a symmetrizing matrix B^{-1}.
Obviously the converse is also true.

2. The Space of Symmetrizing Matrices

If we denote the set of all symmetrizing matrices of M by S_M, and if A,B are any elements of S_M, then for any scalar c, $(cA + B)M = c(AM) + BM = c(M'A) + (M'B) = M'(cA + B)$. Thus linear combinations of elements of S_M belong to S_M. Since the zero matrix is a symmetrizing matrix of M, it is clear that S_M is a vector space over F, in particular, a subspace of the space of all $n \times n$ symmetric matrices over F. At this point we can only say of the dimension of S_M that $\dim S_M \leq n(n + 1)/2$.

3. The Case of Matrices with Simple Characteristic Roots

We will establish a lemma relating the characteristic vector of a matrix with a symmetrizing matrix.

Lemma: If M is symmetrized by B and if M has a characteristic value c and vector v, then M' has a characteristic value c and vector Bv.

Proof: $Mv = cv$

$M'(Bv) = BMv = c(Bv)$

q.e.d.

It is clear that if a matrix M has simple characteristic values, then its transpose M' has the same simple characteristic values. If these are $c_1 \ldots c_n$ and the associated characteristic vectors are $v_1 \ldots v_n$, and, if C is the diagonal matrix with diagonal elements $c_1 \ldots c_n$ and V the matrix with columns $v_1 \ldots v_n$, then we have the identity $M = VCV^{-1}$, as V is necessarily invertible.

Theorem 1: If M has simple characteristic values, then $B = (V^{-1})'CV^{-1}$ is a (nonsingular) symmetrizing matrix of M.

Proof: Clearly B is symmetric, as B is congruent to the diagonal matrix C. From the above remark, $M = VCV^{-1}$. Then we have $BM = (V^{-1})'C^2V^{-1}$, and $M'B = (V^{-1})'C^2V^{-1}$.

Then $BM = M'B$, along with B symmetric, establishes the theorem.

q.e.d.

Corollary 1: $(V^{-1})'CV^{-1}$, $(V^{-1})'C^2V^{-1} \ldots (V^{-1})'C^kV^{-1} \ldots$ all symmetrize M.

Corollary 2: The matrix M and its symmetrizing matrix B are simultaneously diagonable, M by similarity and B by congruence.

Corollary 3: The matrix M is symmetrized by $(V^{-1})'Q^kV^{-1}$ for k = 1, 2... where Q is any diagonal matrix.

Corollary 4: The matrix M' is symmetrized by VQ^kV' for k = 1, 2... where Q is any diagonal matrix.

Corollary 5: The vectors Bv_i are characteristic vectors of M belonging to the characteristic values c_i.

We will reconsider some of the above results in a different notation. The Kronecker product of an m × m matrix A and an n × n matrix B is an m·n × m·n matrix $A \times B = (a_{ij}B)$. Elementary properties of Kronecker products are given in [16] p. 5. If v is an n × 1 matrix and u a 1 × n matrix, the expression v × u will be the n × n matrix $v \times u = (v_i u)$.

Let M be the matrix of Theorem 1, then $M = VCV^{-1}$ where the columns of V are the characteristic vectors v_i associated with the characteristic values c_i of M, the rows of V^{-1} are the characteristic vectors u_i of M' and C is the diagonal matrix $(c_1 \ldots c_n)$. Then the matrix $B = (V^{-1})'QV^{-1}$ symmetrizes M, where Q is any diagonal matrix.

In order to obtain our result, it will be sufficient to consider a 2 × 2 matrix M,

$$M = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}$$

$$= \begin{bmatrix} c_1v_{11} & c_2v_{12} \\ c_1v_{21} & c_2v_{22} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}$$

$$= \begin{bmatrix} c_1v_{11}u_{11} + c_2v_{12}u_{21} & c_1v_{11}u_{12} + c_2v_{12}u_{22} \\ c_1v_{21}u_{11} + c_2v_{22}u_{21} & c_1v_{21}u_{12} + c_2v_{22}u_{22} \end{bmatrix}$$

Defining the vectors $u_i = (u_{i1}, u_{i2})$, $v_i = \begin{bmatrix} v_{1i} \\ v_{2i} \end{bmatrix}$, i = 1, 2

$$M = \begin{bmatrix} c_1 v_{11}(u_1) + c_2 v_{12}(u_2) \\ c_1 v_{21}(u_1) + c_2 v_{22}(u_2) \end{bmatrix} = c_1 v_1 \times u_1 + c_2 v_2 \times u_2.$$

Thus $M = \sum_{i=1,2} c_i v_i \times u_i$. For an n-square matrix M, $M = \sum_{i=1\ldots n} c_i v_i \times u_i$, and similarly $M' = \sum_{i=1\ldots n} c_i u_i' \times v_i'$. For a symmetrizing matrix $B = (V^{-1})'QV^{-1}$ of M, where Q is an arbitrary diagonal matrix with elements $q_1 \ldots q_n$, the expression is $B = \sum_{i=1\ldots n} q_i u_i' \times u_i$. Finally we obtain the result

$$(2) \quad \ldots BM = (V^{-1})'QCV^{-1} = \sum_{i=1\ldots n} q_i c_i u_i' \times u_i.$$

The preceding formulae, along with (2) show that B maps characteristic vectors of M into arbitrary multiples of characteristic vectors of M', the arbitrariness being introduced by the matrix Q.

The system (1) may be considered as a matrix equation $BM - M'B = 0$ in the unknown matrix B; using this approach, we prove

Theorem 2: An n-square matrix M with simple characteristic values has an n-parameter family of symmetrizing matrices.

Proof: Let B be an n-square matrix such that $BM - M'B = 0$. The coefficient matrix of this linear system is

$$(3) \quad I \times M - M \times I$$

and it has n^2 characteristic values $c_i - c_j$ $i, j = 1 \ldots n$ where $c_i \ldots c_n$ are the n distinct characteristic values of M ([1] p. 231). Thus precisely n of the n^2 characteristic values of (3) vanish, and (3) has rank $n^2 - n$ ([16] p. 13, note 3.11-4).

Thus there are n linearly independent solutions B ([19] p. 54).

It remains to show that these solutions are symmetric. We have from Corollary 1 of Theorem 1 a set of symmetrizing matrices $B_1, B_2, \ldots B_k \ldots$ for M, where $B_k = (V^{-1})'C^kV^{-1}$. Consider a linear combination of these matrices

$$a_1B_1 + \ldots + a_nB_n = (V^{-1})'C(a_1I + \ldots + a_nC^{n-1})V^{-1}$$

for any n scalars a_i, not all zero. This expression cannot vanish, for otherwise the minimal polynomial of C would have degree $\leq n - 1$, yet it must have degree n, as C is similar to M and M has simple characteristic values. Thus $B_1 \ldots B_n$ are linearly independent and they must form a basis for the solution space which has dimension n. Thus every solution B of $BM - M'B = 0$ is a linear combination of $B_1 \ldots B_n$, hence every such B is symmetric and we have proven the theorem.

Corollary: The dimension of the space of symmetrizing matrices of M, S_M, is n.

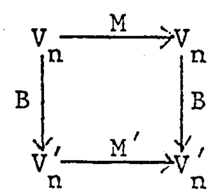
A study of the symmetrizing matrices of companion matrices has been published by Lancaster [15] in which results similar to the above ones are established. Far more general results will be established in Chapter IV, such that many of the results in this chapter will follow as corollaries.

4. Symmetrizing Matrices and Scalar Products

Up to this point we have considered the matrix M and any symmetrizing matrix B of M merely as n-square arrays of elements of a field F. However, n-square matrices may be interpreted as something more, namely as representations of linear operators in a given basis for an n-dimensional vector space. Relations among linear operators give rise to matrix equations in every basis in the space, equations which may be transformed into one another by the application of the formulae for change of basis. Symmetrizing matrices may be given a natural operator - theoretic or invariant interpretation, and the corresponding rules of coordinate transformation derived.

Given a basis $e_1 \ldots e_n$ for V_n, there exists a basis $f^1 \ldots f^n$ for the dual space V_n' of V_n which has the property $[e_j, f^i] = 0 \quad i \neq j$, $[e_i, f^i] = 1$ $i, j = 1 \ldots n$ ([9] p. 92). Corresponding to the transpose matrix M' of the matrix M, and to the basis f^i, is the linear operator M' on V_n which we shall refer to as the transpose operator, in accordance with [9] pp. 103-105. The operator M' is also called the adjoint of M, ([8] p. 79), but we shall reserve this term for later use in this section. The transpose operator M' is defined by the condition $[Mx, f] = [x, M'f]$ for every x in V_n and f in V_n' ([8] p. 78).

Let B be a symmetrizing matrix of the matrix M. Then B is symmetric and $BM = M'B$. A natural operator interpretation of this matrix equation is as follows. With respect to a basis $e_1 \ldots e_n$ of V_n and its dual $f^1 \ldots f^n$ of V_n', B represents a linear transformation from V_n to V_n' (we speak of linear operators when mappings of a space into itself are involved, and linear transformations when mappings of a space into another space are involved). This interpretation of B is a valid one, for if B does map V_n into V_n', then for any x in V_n, Mx is in V_n and Bx is in V_n' and the equation $BM = M'B$ means the elements Mx of V_n and Bx of V_n' are mapped into the same element of V_n', by B and M' respectively; i.e., $B(Mx) = M'(Bx)$. Schematically,



Given a symmetrizing matrix B of M, then with respect to the basis and its dual as given above, the linear transformation B of V_n into V_n' is defined by

(4) $Be_i = b_{ji} f^j$

Then if $x = x^k e_k$ and $y = y^i e_i$ are any elements of V_n, $By = By^i e_i = y^i Be_i = y^i b_{ji} f^j$ is in V_n' and thus the bilinear form $[x, By] = y^i b_{ji} f^j x^k e_k = y^i b_{ji} x^j$ is a scalar in F. We will denote $[x, By]$ by $(x, y)_B$, and examine it more closely after the following lemma.

It is well known that the matrix representation of a linear operator M transforms by similarity when the basis of the space V_n is changed ([8] pp. 84-85). If the linear operator P maps the basis $e_1 \ldots e_n$ of V_n into the basis $u_1 \ldots u_n$ of V_n, then $u_i = Pe_i = p_i^j e_j$ and $P^{-1}[M]_e P = [M]_u$ where the subscript indicates the basis involved in the representation. If B is a symmetrizing matrix of M, then we claim B will transform by congruence.

<u>Lemma</u>: A similarity transformation of M induces a congruence transformation on any symmetrizing matrix B of M.

Proof: Let $x = x^i u_i = x^i p_i^j e_j$ and $y = y^k u_k = y^k p_k^l e_l$ be any elements of V_n.

Then $(x, y)_B = x^i p_i^j e_j B y^k p_k^l e_l = (x^i p_i^j y^k p_k^l) e_j b_{ml} f^m = y^k (p_k^l b_{jl} p_i^j) x^i$.

But $(x, y)_B = y^k b_{ik} x^i$ in the basis u_i. Thus the representation of B under change of basis changes to PBP'; that is, it transforms by congruence, whereas M transforms by similarity.

q.e.d.

Note that this lemma is true under more general conditions for the matrix B; B need not be a symmetrizing matrix of M (see, for example, [9] p. 288).

We shall introduce the concept of a scalar product on a real or complex finite dimensional space V_n. The expression $(—, —)$ is said to be a scalar product on V_n ([8] p. 121) if it maps ordered pairs x, y of V_n into the field F; (x, y) is a scalar in F such that

(i) $(x, y) = \overline{(y, x)}$ (the bar denotes complex conjugation)

(ii) $(ax + by, z) = a(x, z) + b(y, z)$

(iii) $(x, x) \geq 0$; $(x, x) = 0$ if and only if $x = 0$.

We claim that a symmetrizing matrix B "almost" defines a scalar product on V_n. Consider the expression $(x, y)_B = [x, By]$. Condition (ii) is certainly satisfied and if V_n is a real vector space, then (i) is also satisfied, as B is a symmetric matrix. The third condition will be satisfied if $x^i b_{ji} x^j \geq 0$ for every $x = (x^1 \ldots x^n)$ in V_n; if such is the case, B is a positive definite matrix. However, a symmetrizing matrix B need not be positive definite; hence, we speak of $(x, y)_B$ as a pseudo-scalar product.

In a vector space V_n with a scalar product $(—, —)$ or a pseudo-scalar product $(—, —)_B$, for every linear operator on V_n a unique linear operator M^*, the adjoint of M, may be defined on V_n; $(M^*y, x)_B = (y, Mx)_B$ ([8] p. 132). Note that the adjoint is necessarily associated with a given scalar product. We may now inquire under what conditions M is self-adjoint with respect to the pseudo-scalar product $(—, —)_B$.

Theorem: A linear operator M on a real vector space is self-adjoint with respect to a pseudo-scalar product if and only if the pseudo-scalar product is defined by a symmetrizing matrix B of M.

Proof: Let M be self-adjoint with respect to the scalar product defined by B. Since the adjoint is defined by $(M^*y, x)_B = (y, Mx)_B$, then $(M^*y, x)_B = [M^*y, Bx] = [y, (M^*)'Bx]$ and $(y, Mx)_B = [y, BMx]$ give $[y, (M^*)'Bx] = [y, BMx]$. As this is to be true for all vectors x and y, then $BM = (M^*)'B$. If M and

M^* are to be the same linear operator, they must have the same transpose M'. Thus $(M^*)' = M'$ and so $BM = M'B$. Since B represents a scalar product in a real vector space, B is symmetric. Thus B is a symmetrizing matrix of M.

Conversely, if B symmetrizes M, it may be used to define a pseudo-scalar product $(—, —)_B$, with respect to which the adjoint M^* is given by $(M^*y, x)_B = (y, Mx)_B$. As above, we obtain $[y, (M^*)'Bx] = [y, BMx]$ or equivalently $(M^*)'B = BM$. Since B symmetrizes M, $BM = M'B = (M^*)'B$. Thus $M' = (M^*)'$, and so $M = M^*$, i.e., M is self-adjoint with respect to the scalar product defined by the symmetrizing matrix B.

CHAPTER II

SOME FUNDAMENTAL CONCEPTS OF LINEAR ALGEBRA

In order that symmetrizing matrices may be treated in a more rigorous fashion and that such a treatment be self-contained, this chapter will deal with those concepts of Linear Algebra that will be fundamental to the work that follows. The definitions, axioms and theorems that will be taken for granted may be found in [2], [8], or [9].

1. Basis of a Vector Space

Let V be a vector space over the field F, and let S denote the set of s elements $x_1 \ldots x_s$ of V. A linear combination of the elements of S is a sum $a^i x_i$ where $a^1 \ldots a^s$ are arbitrary scalars in F; thus $a^i x_i$ is in V. If there are scalars $a^1 \ldots a^s$ not all zero such that $a^i x_i = 0$, then S is said to be a linearly dependent set; otherwise S is linearly independent. The set of all possible linear combinations of S over F forms a subspace of V known as the space spanned by S. If V contains a linearly independent set S which spans V, then S is called a basis of V. Note that a basis as defined here is a finite set of elements of V; we will emphasize this by speaking of finite bases of V, if any exist.

2. Uniqueness of Dimension

A vector space V is said to be finite-dimensional if it has a finite basis; that is, if it contains a finite set of elements which are linearly independent, and which span V.

<u>Theorem</u>: Every finite basis of a vector space V has the same number of elements.

<u>Proof</u>: Let $x_1 \ldots x_n$ and $y_1 \ldots y_m$ be two finite sets of elements of V, where V has a finite basis. Let $x_1 \ldots x_n$ span V, and $y_1 \ldots y_m$ be a linearly independent set in V. Thus the first set has one defining property of a basis, while the second set has the other defining property of a basis.

Since $x_1 \ldots x_n$ span V

(1) $y_m x_1 \ldots x_n$

is a linearly dependent set which spans V. Thus some element of the spanning set $x_1 \ldots x_n$ is a linear combination of the remainder of the set (1). This can be chosen to be x_n, if necessary by a change of indices. Removing x_n from the set (1),

$y_m x_1 \ldots x_{n-1}$ still span V.

Then the set $y_{m-1} y_m x_1 \ldots x_{n-1}$ is a linearly dependent set. That is, there are scalars a, b, $c_1 \ldots c_{n-1}$, not all zero, such that $ay_{m-1} + by_m + c_1 x_1 + \ldots + c_{n-1} x_{n-1} = 0$. If a = 0 and/or b = 0, then some x_i, i = 1...n-1 may be removed from the set without affecting its spanning property. If $a \neq 0$ and $b \neq 0$, then of necessity at least one $c_i \neq 0$, i = 1...n-1; otherwise the pair y_{m-1}, y_m would be linearly dependent. Thus some x_i, say x_{n-1}, may be removed from the set, leaving $y_{m-1} y_m x_1 \ldots x_{n-2}$ as a set which spans V.

Continuing in this manner, at each step creating a linearly dependent set by the addition of a new y_i, and retaining a spanning set after the deletion of an x_i, if $m > n$, we obtain a set of n elements $y_k \ldots y_m$, k = m - n + 1, which spans V. Yet this is a proper subset of the set of m elements $y_1 \ldots y_m$, which was assumed linearly independent. Consequently $m \leq n$. If both $x_1 \ldots x_n$ and $y_1 \ldots y_m$ are finite bases of V, then they each have both the above spanning and linear independence properties.

Then $m \leq n$ and $n \leq m$, thus m = n, which is to say every finite basis of V has the same number of elements.

q.e.d.

The above theorem, whose proof is due to Halmos ([8] pp. 13-14), also ([2] p. 169) allows one to associate a unique positive integer to each finite-dimensional vector space V, the dimension of V, namely the number of elements of any of the bases of V. The concept of dimension is very important in the study of vector spaces, if only because it is invariant under changes of basis.

Corollary: If V is a finite-dimensional space, every linearly independent set in V may be extended to a basis.

Proof: Let $x_1 \ldots x_n$ be a basis for V, and $y_1 \ldots y_m$, $m \leq n$ be a linearly independent set in V. Then proceeding as in the above proof of the theorem, we obtain a set $y_1 \ldots y_m \; x_1 \ldots x_{n-m}$ of n elements which spans V. This set must be linearly independent, for otherwise, as in the above proof, the set

$y_1 \ldots y_m \; x_1 \ldots x_{n-m-1}$ would span V. But n-1 elements cannot span a space of dimension n, for every spanning set must have at least n elements.

Thus the set $y_1 \ldots y_m \; x_1 \ldots x_{n-m}$ spans V and is linearly independent: i.e., we have extended the linearly independent set $y_1 \ldots y_m$ into a basis for V.

q.e.d.

3. Direct Sums

We shall be using the concept of direct sum extensively in the next chapter, as this is an effective tool in the study of a finite-dimensional space V.

Let W_1 and W_2 be subspaces of V. Then by $W_1 + W_2$ is meant the set of all elements $w_1 + w_2$ with w_i in W_i, i = 1, 2. It is not very difficult to see that $W_1 + W_2$ is a subspace of V. If the elements of $W_1 + W_2$ can be expressed as a sum of elements of W_1 and W_2 in only one way, then W_1 and W_2 are disjoint, and the sum $W_1 + W_2$ is said to be direct, and it is written $W_1 \oplus W_2$.

In extending the concept of direct sum to more than two subspaces, we must first introduce the notion of independence among subspaces. Let $W_1 \ldots W_k$ be subspaces of V; then they are independent if $x_1 + \ldots + x_k = 0$, x_i in W_i, implies each $x_i = 0$, $i = 1 \ldots k$. Equivalently, $W_1 \ldots W_k$ are independent if each element x in the sum $W_1 + \ldots + W_k$ may be written as a sum of elements $x = x_1 + \ldots + x_k$, x_i in W_i, $i = 1 \ldots k$, in only one way. We observe that whereas disjointness is a property of a pair of subspaces, independence is a property of a set of subspaces. The sum of k independent subspaces $W_1 \ldots W_k$ is then called the direct sum of these subspaces, and is written $W_1 \oplus \ldots \oplus W_k$.

<u>Theorem</u>: If V is a finite-dimensional vector space over the field F, and if $W_1 \ldots W_k$ are subspaces of V, then $V = W_1 \oplus \ldots \oplus W_k$ if and only if the union of the bases for each W_i is a basis for V.

Proof: Let $V = W_1 \oplus \ldots \oplus W_k$, and let $S_i : x_1^i \ldots x_{d_i}^i$, ($d_i$ the dimension of W_i)

be a basis for W_i, $i = 1\ldots k$. If S is the set consisting of all the elements of each S_i, $i = 1\ldots k$, then a linear combination of elements of S has the form $a^{j}_{1}x^{1}_{j} + a^{j}_{2}x^{2}_{j} + \ldots + a^{j}_{k}x^{k}_{j}$, where for instance $a^{j}_{1}x^{1}_{j} = a^{1}_{1}x^{1}_{1} + a^{2}_{1}x^{1}_{2} + \ldots + a^{d_1}_{1}x^{1}_{d_1}$. That is, a linear combination of elements of S has the form $y_1 + y_2 + \ldots + y_k$ where for instance $y_1 = a^{j}_{1}x^{1}_{j}$. Thus every vector in V can be expressed in terms of a linear combination of elements of S, and S spans V. The set S is clearly linearly independent, for if $y_1 + \ldots + y_k = 0$, then the independence of each W_i implies $y_i = 0$, $i = 1\ldots k$. Thus S is a basis for V.

The converse follows in much the same manner.

q.e.d.

<u>Corollary</u>: If $V = W_1 \oplus \ldots \oplus W_k$, then $\dim V = \dim W_1 + \ldots + \dim W_k$.

4. Linear Transformations, Operators and Matrices

Given two vector spaces V and W over the same field F, a rule T which assigns to each element x in V an element Tx in W such that $T(ax + by) = aTx + bTy$ for all elements x, y of V and a, b of F, is called a linear transformation from V into W. It is not very difficult to see that the set of all linear transformations from V into W itself forms a vector space over F.

We are particularly interested in that set of linear transformations which maps V into V. Such linear transformations are distinguished by calling them linear operators on V. Thus if T is a linear operator on V, then Tx is in V for every x in V. In particular, if V is an n-dimensional space with a basis $e_1\ldots e_n$, then $Te_i = t^{j}_{i}e_j$ $i = 1\ldots n$. That is, with respect to the basis $e_1\ldots e_n$, T determines n^2 scalars t^{j}_{i}, which is called the matrix representation of T with respect to the basis $e_1\ldots e_n$. Conversely, every n-square matrix defines a linear operator, given some basis for the n-dimensional vector space involved.

A linear operator T does not in any manner depend upon the basis under consideration; its matrix representation certainly does. If P is the linear operator which maps the basis $e_1\ldots e_n$ of V into another basis $v_1\ldots v_n$ of V,

$Pe_i = v_i$, then P is known as a nonsingular operator, and the inverse operator P^{-1} exists and has the property that $P^{-1}v_i = e_i$ and $PP^{-1} = P^{-1}P = I$, the identity operator. Then $(P^{-1}TP)e_i = (P^{-1}T)v_i = (P^{-1})b_i^jv_j = b_i^je_j$; the operator $P^{-1}TP$ then has in the basis $e_1 \ldots e_n$ the same matrix representation as the operator T has in the basis $v_1 \ldots v_n$. Conversely, if the matrix of T in the $e_1 \ldots e_n$ basis is (t_i^j), then its matrix in the $v_1 \ldots v_n$ basis is $P(t_i^j)P^{-1} = (b_i^j)$. The matrix of a linear operator is said to undergo a similarity transformation whenever the basis is changed. Thus similar matrices represent the same linear operator.

We are now in a position to go on to a rigorous development of the Jordan canonical form for linear operators on a finite-dimensional vector space, which will provide the key to a complete solution of the matrix symmetrization problem.

CHAPTER III

THE JORDAN CANONICAL FORM

Given an arbitrary linear operator A on a finite dimensional space V it will be established that V is the direct sum of subspaces on which A has a particularly simple form. Thus the study of a general linear operator A will be reduced to the study of the restrictions of A to the component subspaces of V. The proofs in this chapter are due to Cater ([4] pp. 391, 393); the assumption of an algebraically closed field at the outset yields proofs which are more elegant than the standard textbook proofs.

1. An Elementary Development

Let A be a linear operator on the vector space V of dimension n, over the algebraically closed field F. For each scalar c of F let V_c be the set of all vectors x of V which are annihilated by some power of the operator (A - c). That is, the vector x is in V_c if and only if, for some positive integer k, $(A - c)^k x = 0$. If x, y are any vectors of V_c then for any scalars a, b, ax + by is also in V_c. Thus each V_c is a subspace of V, in fact, an invariant subspace of V under A. For if x is any vector in V_c, then for some k, $(A - c)^k x = 0$. But $(A - c)^k (Ax) = ((A - c)^k A)x = A(A - c)^k x = A0 = 0$. Thus Ax is also in V_c.

<u>Lemma 1</u>: The subspaces V_c span V.

For any vector x in V there are n + 1 scalars $a_0 \ldots a_n$ not all zero such that $(a_i A^i)x = 0$, for any set of n + 1 vectors in an n-dimensional vector space V is linearly dependent. Thus for each x in V, there is at least one polynomial p with coefficients in F, of degree $\leq n$, such that $p(A)x = 0$. The polynomial p is determined jointly by the vector x in question, and the linear operator A. Let m(x) be the minimal degree of such polynomials for the vector x, given a linear operator A. The proof of this lemma will be an inductive argument on the degree m(x).

Proof of Lemma 1: We wish to show that the subspaces V_c, c ranging over F, span V. Let them span $W \leq V$. Then, if for some x in V, $m(x) = 1$, there are scalars a, b in F such that $(bA + a)x = 0$, with $b \neq 0$. Thus $(A - (-a/b))x = 0$ and x is in V_c, where $c = -a/b$, thus in W.

Assume that $m(u) \leq k - 1$ implies that u is in W, and consider an x in V for which $m(x) = k$. Then there is a polynomial p of degree k such that $p(A)x = 0$. Because F is an algebraically closed field, the polynomial p may be completely factored into linear factors; that is, there exist k + 1 scalars c, $c_1 \ldots c_k$ in F such that $p(A) = c(A - c_1)(A - c_2)\ldots(A - c_k)$. In case all the c_i are equal, say $c_i = b$, $i = 1,\ldots, k$ $p(A)x = c(A - b)^k x = 0$ and x belongs to V_b; therefore x belongs to W. Otherwise suppose the c_i are not equal. Then for some i and j between 1 and k, $i \neq j$, $c_i \neq c_j$. Consider the nonzero vectors $(A - c_i)x$ and $(A - c_j)x$. These are annihilated by the operators $(A - c_1)\ldots(A - c_{i-1})$ $(A - c_{i+1})\ldots(A - c_k)$ and $(A - c_1)\ldots$ $(A - c_{j-1})$ $(A - c_{j+1})\ldots(A - c_k)$ respectively. Thus in accordance with the induction hypothesis, the vectors $(A - c_i)x$ and $(A - c_j)x$ belong to W, as polynomial operators of degree k - 1 have been found which annihilate them. Since W is a vector space, and it contains $(A - c_i)x$ and $(A - c_j)x$, it must also contain $(c_j - c_i)^{-1}((A - c_i)x - (A - c_j))x = x$. Thus $m(u) = k$ implies that u is in W, and this completes the induction argument.

We have shown that an arbitrary x in V belongs to W. Hence W includes V. But W is a subspace of V. Therefore W = V and the subspaces V_c span V.

q.e.d.

Lemma 2: The subspaces V_c are linearly independent, i.e., if $x_i \in V_{c_i}$ for $1 \leq i \leq m$ and $c_i \neq c_j$ for all i, j then $x_1 + x_2 + \ldots + x_m = 0$ implies $x_1 = x_2 = x_3 = \ldots = x_m = 0$.

Proof: There exist positive integers m_i, $1 \leq i \leq m$, such that $(A - c_i)^{m_i} x_i = 0$, but $(A - c_i)^{m_i - 1} x_i \neq 0$. For any $1 \leq j \leq m$ and $1 \leq k \leq m$ consider

$$(A - c_k)^{m_k-1}(A - c_j)x_k = (A - c_k)^{m_k-1}[(A - c_k)x_k + (c_k - c_j)x_k]$$

$$= (A - c_k)^{m_k}x_k + (A - c_k)^{m_k-1}(c_k - c_j)x_k$$

$$= (A - c_k)^{m_k-1}(c_k - c_j)x_k$$

Similarly $(A - c_k)^{m_k-1}(A - c_j)^{m_j}x_k = (A - c_k)^{m_k-1}(c_k - c_j)^{m_j}x_k$

and $\quad (A - c_k)^{m_k-1}\prod_{\substack{j=1\\j\neq k}}^{m}(A - c_j)^{m_j}x_k = (A - c_k)^{m_k-1}\prod_{\substack{j=1\\j\neq k}}^{m}(c_k - c_j)^{m_j}x_k$

Thus $(A - c_k)^{m_k-1}\prod_{\substack{j=1\\j\neq k}}^{m}(A - c_j)^{m_j}[x_1 + x_2 +\ldots+ x_m] = (A - c_k)^{m_k-1}\prod_{\substack{j=1\\j\neq k}}^{m}(A - c_j)^{m_j}x_k$

$$= (A - c_k)^{m_k-1}\prod_{\substack{j=1\\j\neq k}}^{m}(c_k - c_j)^{m_j}x_k$$

and, if $x_1 + x_2 +\ldots+ x_m = 0$, then $(A - c_k)^{m_k-1}\prod_{\substack{j=1\\j\neq k}}^{m}(c_k - c_j)^{m_j}x_k = 0$

But $\prod_{\substack{j=1\\j\neq k}}^{m}(c_k - c_j)^{m_j} \neq 0$, and $x_k \neq 0$ implies that $(A - c_k)^{m_k-1}x_k = 0$.

Thus $x_k = 0$, and this is true for all $1 \leq k \leq m$.

q.e.d.

Corollary: The subspaces V_c are disjoint.

Corollary: There are at most $r \le n$ scalars $c_1 \ldots c_r$ for which V_{c_i} is nonempty.

These lemmas with their corollaries together give the theorem:

Theorem 1: The finite dimensional vector space V is the direct sum of all the nonempty subspaces V_{c_i}.

Corollary: The linear operator A is reduced by the nonempty subspaces V_{c_i}.

Proof: This corollary is evident if one recalls that a linear operator A on a vector space V is said to be reduced by a set of subspaces if each subspace is invariant under A, and if V is the direct sum of these subspaces ([8] p. 72).

2. The Jordan Canonical Form

The Jordan Canonical Form Theorem may be derived from Theorem 1 by a further subdivision of the nonempty subspaces V_c, and the choice of appropriate bases in the subspaces thus defined. In this derivation, the concepts of nilpotent operator and quotient space occur.

A linear operator A is said to be nilpotent when, for some positive integer r, $A^r \equiv 0$. The least such integer r is the index of nilpotence of A. Thus, the operator A - c, restricted to the nonempty subspace V_c is nilpotent.

If W is a subspace of V, elements x, y of V are said to be congruent modulo W in case x - y is an element of W. It is easily seen that congruence is an equivalence relation on V; the associated equivalence classes are called cosets of W. The collection V/W of all cosets of W is called the quotient space of V with respect to W and with the obvious definitions of the linear operations, may be shown to be a linear vector space over the field F of V.

The immediate objective is the following result:

Theorem 2: If A is nilpotent operator on V, then there exist vectors $x_1 \ldots x_r$ in V such that the nonzero vectors in the collection $x_i, Ax_i, A^2x_i, \ldots$. $i = 1, \ldots r$ constitute a basis for V.

This theorem may be established by induction on the index q of nilpotence of A. The following result gives an upper bound for q.

Lemma 1: If A is nilpotent of index q on a space V of dimension n, then $q \le n$.

Proof: Since A is nilpotent on V, then $V = V_c$, with $c = 0$, as each element of V is annihilated by some power of A. Now $A^n = 0$. For if $A^n \neq 0$, consider for any x in V the set of $n + 1$ elements of V; $x, Ax, A^2x, \ldots A^nx$. This set must be linearly dependent, since V has dimension n. Thus there are $n + 1$ scalars $c_0 \ldots c_n$ not all zero such that $c_0x + c_1Ax + \ldots + c_nA^nx = 0$. Let j be the least index for which $c_j \neq 0$. Then

$$(1) \quad c_jA^jx + \ldots c_nA^nx = 0$$

Let x be annihilated by A^k; that is, $A^{k-1}x \neq 0$ and $A^kx = 0$. Apply the linear operator A^{k-j-1} to the vector in (1), obtaining $c_jA^{k-1}x + (c_{j+1}A^kx + \ldots + c_nA^{n+k-j-1}x) = 0$, or $c_jA^{k-1}x = 0$. This is contrary to the definition of k. Thus it must be that $A^n = 0$. Since q is the least index for which $A^q \equiv 0$, it follows that $q \le n$.

q.e.d.

If A is nilpotent of index q, the null space W of A has nonzero elements. For, by the minimal property of q, there exists at least one x in V such that $A^{q-1}x \neq 0$, and the element $A^{q-1}x$ is clearly an element of W. Thus, the quotient space V/W is different from V, and a linear transformation B may be defined on V/W by observing that A takes the same value throughout a coset of W. The inductive proof of Theorem 2 is based upon the following result.

Lemma 2: If A is nilpotent of index q, then B is nilpotent of index $q - 1$.

Proof: For every x in V, A^px is in W for some $p \le q - 1$. Thus, for every X in V/W, $B^pX = W$ for some $p \le q - 1$. Since there exist x in V for which $p = q - 1$, correspondingly there exist X in V/W for which $p = q - 1$. Thus B is nilpotent of index $q - 1$.

q.e.d.

Finally it is necessary to relate bases in V/W to bases in V. To this end, it may be noted that a subset Z of V consisting of just one element chosen arbitrarily from each of the cosets $\neq W$ of V/W, together with the zero vector from W, is a vector subspace of V. For, Z is trivially isomorphic with V/W. Thus, a basis in V/W may be used to construct a basis in Z by selecting an element x_i of V from each of the basis elements in V/W and defining Z to be the subspace of V spanned by the x_i. Then, we may show:

Lemma 3: $V = W \oplus Z$.

Proof: Let x be an element of V, and let X be the coset containing x. Let x' be the element of Z selected from X. Then x and x' are congruent modulo W - i.e., $x - x' = y$ is an element of W or $x = x' + y$. Thus every $x \in V$ may be written uniquely as the sum of $x' \in Z$ and $y \in W$.

Suppose that x is in both W and Z. Then, since x is in W, X = W and $x' = 0$. Also, since x is in Z, $x = x'$ and $y = 0$. Thus $x = x' + y = 0$ so that W, Z are disjoint.

q.e.d.

Thus, a basis in V/W defines a basis in Z and this may be extended to a basis in V by adding to it a basis for W.

Proof of Theorem 2: Suppose that A is nilpotent of index q on V.

If q = 1, then $A^1 = A = 0$ annihilates every vector in V, and any basis of V will suffice.

Assume that the conclusion is valid for $q \leq m - 1$, and let the index of nilpotence of A be m. Then B is nilpotent of index m - 1 on V/W, and, by the inductive hypothesis, there exists a basis of the prescribed form for V/W. This gives rise immediately to a basis of the prescribed form, say

$$\begin{array}{ll} x_1 & Ax_1 \ldots A^{r_1}x_1 \\ x_2 & Ax_2 \ldots A^{r_2}x_2 \\ \cdot & \\ \cdot & \\ \cdot & \\ x_s & Ax_x \ldots A^{r_s}x_s \end{array}$$

for a subspace Z of V such that $V = W \oplus Z$.

Now, the vectors $A^{r_1+1}x_1, A^{r_2+1}x_2 \ldots A^{r_s+1}x_s$ are elements of W - i.e., are congruent to 0 modulo W. The nonzero elements among these are linearly independent, for if, for some scalars c_i

$$c_1A^{r_1+1}x_1 + c_2A^{r_2}x_2 + \ldots + c_sA^{r_s+1}x_s = 0$$

then $A[c_1A^{r_1}x_1 + c_2A^{r_2}x_2 + \ldots + c_sA^{r_s}x_s] = 0$

whence $c_1A^{r_1}x_1 + c_2A^{r_2}x_2 + \ldots + c_sA^{r_s}x_s$ is in W.

But this is impossible, since the cosets of these elements are elements of a basis in V/W. Thus, the nonzero elements among $A^{r_1+1}x_1, A^{r_2+1}x_2 \ldots A^{r_s+1}$ form at least a partial basis for W. Upon completing this basis by adding elements $x_{s+1}, x_{s+2}, \ldots x_r$ of W to it, a complete basis for W, whence a complete basis for V is determined.

Thus the basis consisting of the nonzero vectors among

$$\begin{array}{lll}
x_1 & Ax_1 \ldots A^{r_1}x_1 & A^{r_1+1}x_1 \\
x_2 & Ax_2 \ldots A^{r_2}x_2 & A^{r_2+1}x_2 \\
\vdots & & \vdots \\
x_s & Ax_s \ldots A^{r_s}x_s & A^{r_s+1}x_s \\
x_s+1 & & \\
\vdots & & \\
x_r & &
\end{array}$$

constitutes the required basis for V.

q.e.d.

An arbitrary linear operator A on a finite-dimensional vector space V is not necessarily nilpotent. However by Theorem 1, if V is a vector space over an algebraically closed field F, then there is a positive integer r such that V is the direct sum of r subspaces

$$V = V_{c_1} \oplus \ldots \oplus V_{c_r}$$

Let c be any one of the $c_1 \ldots c_r$; then the operator (A - c) is clearly nilpotent on the subspace V_c, and its index of nilpotence cannot exceed the dimension of V_c. Theorem 2 then provides vectors $x_1 \ldots x_k$ in V_c such that the nonzero vectors of the collection x_i, $(A - c)x_i$, $(A - c)^2 x_i$, ... i = 1, ..., k constitute a basis of V_c. Thus there are determined positive integers q_i, i = 1, ..., k such that $(A - c)^{q_i} x_i = 0$ i = 1, ..., k and the following vectors are a basis of V_c:

$$x_1,\ (A - c)x_1,\ (A - c)^2 x_1,\ \ldots\ (A - c)^{q_1 - 1} x_1$$

$$x_2,\ (A - c)x_2,\ (A - c)^2 x_2,\ \ldots\ (A - c)^{q_2 - 1} x_2$$

....

$$x_k,\ (A - c)x_k,\ (A - c)^2 x_k,\ \ldots\ (A - c)^{q_k - 1} x_k$$

If the basis vectors are labelled $y_1, y_2, \ldots y_p$

$$y_1 = x_1,\ y_2 = (A - c)x_1,\ y_3 = (A - c)^2 x_1,\ \ldots,\ y_{q_1} = (A - c)^{q_1 - 1} x_1$$

$$y_{q_1+1} = x_2,\ \ldots \qquad ,\ y_{q_1+q_2} = (A - c)^{q_2 - 1} x_2$$

....

$$y_{p-q_k+1} = x_k,\ \ldots \qquad ,\ y_p = (A - c)^{q_k - 1} x_k$$

where $p = q_1 + q_2 + \ldots + q_k$; then we may obtain the matrix representation of the linear operator A_c induced by the restriction of A to the subspace V_c, in terms of the basis $y_1 \ldots y_p$. For we have

$$(A - c)y_1 = y_2 \qquad Ay_1 = cy_1 + y_2$$

$$(A - c)y_2 = y_3 \qquad Ay_2 = cy_2 + y_3$$

$$\cdots \qquad \cdots$$

$$(A - c)y_{q_1-1} = y_{q_1} \qquad Ay_{q_1-1} = cy_{q_1-1} + y_{q_1}$$

$$(A - c)y_{q_1} = 0 \qquad Ay_{q_1} = cy_{q_1}$$

$$(A - c)y_{q_1+1} = y_{q_1+2} \qquad Ay_{q_1+1} = cy_{q_1+1} + y_{q_1+2}$$

$$\cdots \qquad \cdots$$

$$\cdots \qquad \cdots$$

$$(A - c)y_{p-1} = y_p \qquad Ay_{p-1} = cy_{p-1} + y_p$$

$$(A - c)y_p = 0 \qquad Ay_p = cy_p$$

Let us suppose for a moment that $p = q_1$. Then we have the vectors $y_1 \ldots y_p$ forming a basis of V_c, with $(A - c)y_i = y_{i+1}$, $i = 1, \ldots, p - 1$ and $(A - c)y_p = 0$. Then the matrix in this basis of the restriction A_c of A to the subspace V_c is a p-square matrix

$$[A_c] = \begin{bmatrix} c & 0 & 0 & & & & \\ 1 & c & 0 & & & & \\ 0 & 1 & c & & & & \\ 0 & 0 & 1 & \ddots & & & \\ & & & \ddots & c & 0 & 0 \\ & & & & 1 & c & 0 \\ & & & & 0 & 1 & c \end{bmatrix}$$

where all the missing entries are zero. Such a matrix can be written as the sum of two matrices (matrix sum, not a direct sum), $[A_c] = cI + U$, where I is a p-square identity matrix, and U is a p-square matrix composed of zeros except for its subdiagonal, which is all ones; U is known as an auxilliary unit matrix. Any matrix which can be written as a sum $cI + U$ is called an elementary Jordan matrix.

Returning to the general case in which $p \neq q$, it is clear that the matrix $[A_c]$ will be a direct sum of k elementary Jordan matrices,

$$[A_c] = (cI + U)_1 \oplus \ldots \oplus (cI + U)_k$$

where $(cI + U)_i$ indicates that the enclosed matrices are both q_i-square. Such a direct sum of elementary Jordan matrices is called a Jordan canonical form. Thus we may say that the operator A has been reduced to its Jordan canonical form on the subspace V_c.

Theorem 1 provides us with a decomposition of the linear operator A into a direct sum of the operators $A_{c_1}, \ldots A_{c_r}$ induced on each subspace $V_{c_1} \ldots V_{c_r}$ by A. It follows that the matrix representation of A, in the basis consisting of the union of the bases of the subspaces V_{c_i} as they have been labelled above, is the direct sum of r Jordan canonical forms $[A] = [A_{c_1}] \oplus \ldots \oplus [A_{c_r}]$, where $[A_{c_i}] = (cI + U)_1 \oplus \ldots \oplus (cI + U)_{q_i}$. The matrix $[A]$ is known as the Jordan canonical form of the linear operator A.

Since similar matrices represent the same linear operator in different bases, and we have shown that for any linear operator a basis can be found in which the representation of the operator is a Jordan canonical form, we can now state a very important theorem:

Theorem 3: Every square matrix over an algebraically closed field is similar to a direct sum of Jordan canonical matrices.

No mention has been made of the uniqueness of the Jordan canonical form of the matrix representation of an operator. Suffice it to say that if the integers q_i provided by Theorem 2 are ordered so that $q_1 \leq q_2 \leq \ldots \leq q_n$, the Jordan canonical form will then be uniquely determined ([9] p. 207).

In the choice of the basis $y_1 \ldots y_p$ for the subspace V_c, we could just as well have reversed the order of the basis vectors so that $y_1 = (A - c)^{q_1 - 1} x_1$, $y_2 = (A - c)^{q_1 - 2} x_1, \ldots, y_{q_1 - 1} = (A - c)x_1$, $y_q = x_1$, $y_{q_1 + 1} = (A - c)^{q_2 - 1} x_2$, and so on. Then we would obtain the relations $(A - c)y_1 = 0$ $(A - c)y_2 = y_1, \ldots, (A - c)y_{q_1} = y_{q_1 - 1}$, $(A - c)y_{q_1 + 1} = 0$, and so on. Then in this basis, the restriction of A to the subspace V_c has the representation $[A_c] = (cI + U')_1 + \ldots + (cI + U')_{q_1}$, where U' is the transpose of the auxiliary unit matrix U. We may now speak of lower Jordan forms or upper Jordan forms, according as the auxiliary unit matrices involved have their nonzero elements below or above the main diagonal.

In connection with the auxiliary unit matrices, we wish to introduce the concept of nilpotent matrix. As for operators, a matrix A is said to be nilpotent if for some positive integer k, $A^k = 0$. Clearly any auxiliary unit matrix U is nilpotent; and if it is m-square, then $U^m = 0$. Thus we can say that in the above basis, the matrix representation of the restriction of A to the subspace V_c is the sum of a nonsingular matrix cI and a nilpotent matrix U.

A matrix may always be partitioned into blocks. In the event that it may be partitioned so that the blocks which do not include diagonal elements are all zero, the matrix is called quasi-diagonal. We observe that a Jordan matrix is a quasi-diagonal matrix, and that every matrix is similar to a quasi-diagonal matrix, according to Theorem 3.

3. Characteristic Values and Vectors

The Jordan canonical form of a linear operator is usually arrived at by considerations involving characteristic values and vectors; see for instance [7] pp. 132-163. We will reformulate the results of the preceding section in terms of these new concepts.

A scalar c is said to be a characteristic value of a linear operator A if there is a nonzero vector x such that $(A - c)x = 0$. The vector x is then called a characteristic vector associated with the characteristic value c of A. If there is a nonzero vector y such that $(A - c)^k y = 0$ for some positive integer k, then y is a principal vector associated with the characteristic value c of A. Then a characteristic vector x is simply a principal vector for which k = 1.

In the discussion leading to Theorem 3, we had a basis $y_1 \ldots y_p$ of V_c such that $(A - c)y_i = y_{i+1}$ for $i = 1 \ldots p - 1$, $i \neq q_1, q_1 + q_2, \ldots, p - q_k$ and $(A - c)y_j = 0$ for $j = q_1, q_1 + q_2, \ldots, p - q_k, p$. Thus the basis vectors are all principal vectors associated with the characteristic value c of A, as some positive integer k_i may be found for each i such that $(A - c)^{k_i} y_i = 0$. The subset $y_{q_1}, y_{q_1 + q_2}, \ldots, y_p$ of the basis vectors are all characteristic of c, as for these vectors $k_i = 1$ for $i = q_1, q_1 + q_2, \ldots p$.

We may speak of the characteristic values and vectors of a square matrix just as we have done for a linear operator. In that case, the characteristic values of the matrix A are the solutions of the determinantal equation $|A - cI| = 0$. A polynomial of degree n, f(c), is defined by $f(c) = |A - cI|$. It may be shown that the zeros of the polynomial f(c) are simply the characteristic values of A, and that similar matrices have the same characteristic polynomial ([9] p. 165). It is clear then that the characteristic values of A are the same set of scalars, whether A is being considered as a linear operator, or any matrix representation of the linear operator A. The characteristic vectors of the matrix A may be similarly defined as the nontrivial solutions of the matrix equation $(A - cI)x = 0$, where x is an $n \times 1$ column matrix. For $|A - cI|$ is zero only when c is a characteristic value of A. Then the only nontrivial solutions of $(A - cI)x = 0$ are the characteristic vectors associated with the characteristic values of A ([9] p. 23).

In general the n zeros of a polynomial of degree n are not all distinct. If the r distinct zeros of the characteristic polynomial of A are $c_1 \ldots c_r$, then each zero c_i has a multiplicity m_i, such that $f(c) = \prod_{i=1\ldots r} (c_i - c)^{m_i}$ and $\sum_{i=1\ldots r} m_i = n$. The m_i are often referred to as the algebraic multiplicities of the c_i to prevent their confusion with the geometric multiplicities of c_i, which refer to the dimensions of the subspace spanned by all the characteristic vectors of c_i ([8] pp. 104-105).

If we examine the Jordan canonical form of a linear operator A and restrict our attention to the i-th block of the direct sum decomposition, denoted by A_i, then the diagonal elements of A_i are the characteristic value c_i, and c_i appears only in the block A_i. If the algebraic multiplicity of c_i is m_i, then A_i is a m_i-square matrix. The square matrix may in turn be decomposed into a direct sum. If we required that the direct sum decomposition of A_i be in terms of elementary Jordan matrices, then clearly the decomposition is unique. Then A_i is the direct sum of r_i elementary Jordan matrices, each $q_{i,j}$-square, such that $\sum_{j=1\ldots r_i} q_{i,j} = m_i$, and r_i is the geometric multiplicity of c_i.

As the n-dimensional vector space V is the direct sum of r subspaces V_{c_i}, we may then reformulate Theorem 3 as

<u>Theorem 4</u>: If A is a linear operator on a finite dimensional vector space V over an algebraically closed field F, then there exist positive integers r, $m_1 \ldots m_r$ and scalars $c_1 \ldots c_r$ in F such that $c_1 \ldots c_r$ are the r distinct characteristic values of A, respectively of multiplicities $m_1 \ldots m_r$; V is the direct sum of r subspaces $V_{c_1} \ldots V_{c_r}$, each subspace V_{c_i} reduces A and the operator $(A - c_i)$ is nilpotent on V_{c_i}, $i = 1 \ldots r$. Furthermore, the integers r, $m_1 \ldots m_r$ and scalars $c_1 \ldots c_r$ are uniquely determined by A, V and F.

We will be referring to nonderogatory matrices in the next chapter. In this connection, the polynomial $e_{j,i}(c) = (c - c_j)^{q_{j,i}}$ is called an elementary divisor of multiplicity $q_{j,i}$ associated with the characteristic value c_j of A; one such elementary divisor is associated with each elementary Jordan matrix in the Jordan canonical form of A. If the multiplicity of an elementary divisor is one, it is said to be simple (or linear); if all the elementary divisors of A are simple, A is said to be simple. The characteristic polynomial of A can easily be seen to be the product of all the elementary divisors of A, and according to the Cayley-Hamilton Theorem ([9] p. 166), A satisfies its own characteristic equation. The polynomial which is the product of the elementary divisors of maximal multiplicity for each c_i is called the minimal polynomial of A, for it is the polynomial in A of minimal degree which vanishes identically ([9] p. 168). This of course implies that the minimal polynomial of A always divides the characteristic polynomial of A. In the event that the minimal and characteristic polynomials of A are identical, A is said to be nonderogatory.

CHAPTER IV
SYMMETRIZING MATRICES

In order to determine how many independent symmetrizing matrices X a given matrix M may have, we will first find matrices X, not necessarily symmetric, which satisfy the matrix equation

(1) $XM = M'X$.

This equation has been studied by Marcus and Khan [17] in great detail and a complete solution has been found. However their basic approach involves the use of Kronecker products and their properties; a much simpler development will be given, adapted from [6] pp. 215-220.

1. Basic Results

Applying Theorem 4 of the preceding chapter, M may be transformed to its Jordan canonical form M_1; thus there is a nonsingular matrix P such that $M = PM_1P^{-1}$. Using this expression in (1) we obtain $(P'XP)M_1 = M_1'(P'XP)$. We observe once again that a similarity transformation of M induces a congruence transformation on any symmetrizing matrix of M. Since congruence transformations preserve symmetry, let us consider (1) as the equation relating a Jordan canonical form M with an unknown matrix X. It will then be possible, if a symmetric solution X is found, to generate, for all matrices similar to M, symmetrizing matrices congruent to X.

We may write

$$(2) \quad \begin{aligned} M &= (c_1I + U)_{p_1} \oplus \dots \oplus (c_kI + U)_{p_k} \\ M' &= (c_1I + U')_{p_1} \oplus \dots \oplus (c_kI + U')_{p_k} \end{aligned}$$

where the subscript p_i indicates the matrices contained in the parentheses are p_i-square, and U_{p_i} are p_i-square lower auxiliary unit matrices, $i = 1\dots k$. Let X be partitioned conformally with M, $X = (X_{ij})$, X_{ij} a $p_i \times p_j$ submatrix of X, $i, j = 1\dots k$. For example, if

$$M = \left[\begin{array}{cc|c|c} a & 0 & 0 & 0 \\ 1 & a & 0 & 0 \\ \hline 0 & 0 & a & 0 \\ \hline 0 & 0 & 0 & b \end{array}\right], \text{ then } X = \left[\begin{array}{cc|c|c} * & * & * & * \\ * & * & * & * \\ \hline * & * & * & * \\ \hline * & * & * & * \end{array}\right], \text{ since } p_1 = 2,\ p_2 = p_3 = 1.$$

The Jordan canonical form being a quasi-diagonal matrix, we recall the rules for multiplication of partitioned and quasi-diagonal matrices; when a partitioned matrix is multiplied on the left (on the right) by a quasi-diagonal matrix, then the rows (the columns) are multiplied on the left (on the right) by the corresponding diagonal blocks of the quasi-diagonal matrix ([6] pp. 42-43). The equation $XM = M'X$ then breaks up into k^2 equations $(c_i I + U)_{p_i} X_{ij} = X_{ij}(c_j I + U')_{p_j}$ $i, j = 1 \ldots k$, or equivalently,

$$(3) \quad (c_i - c_j)X_{ij} = X_{ij}U'_{p_j} - U_{p_i}X_{ij} \quad i, j = 1 \ldots k.$$

Let us fix i and j, and set $c = c_i - c_j$, $V = U_{p_i}$, $W = U'_{p_j}$, and $Y = X_{ij}$;

Y is then a $p_i \times p_j$ rectangular matrix. Then (3) becomes $cY = YW - VY$, which will be the subject of the following two lemmas.

Lemma 1: The only solution of $cY = YW - VY$, $c \neq 0$, is the zero matrix Y.

Proof: We have $cY = YW - VY$. Multiply both sides by $c \neq 0$, $c^2Y = cYW - cVY = (YW - VY)W + V(YW - VY) = YW^2 + 2VYW + V^2Y$. Repeating this process $r - 2$ times we obtain

$$(4) \quad c^rY = \sum_{l+m=r} (-1)^l \binom{r}{l} V^mYW^l.$$

However the matrices V and W are both nilpotent, V of index p_i and W of index p_j; i.e., $V^{p_i} = W^{p_j} = 0$. If in (4) we set $r \geq p_i + p_j - 1$, then each term V^mYW^l with $l + m = r$ vanishes. For if $m < p_i$, then $l = r - m \geq p_j$ and $W^l = 0$, or if $l < p_j$, then $m = r - l \geq p_i$ and $V^m = 0$. Thus each term in the expansion of the righthand side of (4) vanishes, leaving $c^rY = 0$. Since $c \neq 0$, necessarily $Y = 0$.

q.e.d.

If for a fixed i and j, $c_i = c_j$, then $c = 0$ and we are lead to consider the equation $YW = VY$; let us set $p_i = m$ and $p_j = n$. Thus V is m-square, W is n-square.

The proofs which follow are from [17] pp. 1340-1344.

<u>Lemma 2</u>: The solution of $YW = VY$ is an $m \times n$ matrix Y having $\min(m, n)$ arbitrary parameters.

Proof: Let $y_i = \mathrm{col}(y_{1i}, y_{2i} \cdots y_{mi})$, $i = 1 \ldots n$, be the columns of Y. Since V is an m-square lower auxiliary unit matrix, it may easily be verified that $Vy_i = \mathrm{col}(0, y_{1i}, \ldots, y_{m-1,i})$ $i = 1 \ldots n - 1$. Similarly, since W is an upper auxiliary unit matrix $YW = [0, y_1, y_2 \cdots y_{n-1}]$. Selecting and equating the element (r, j) of VY and YW, we obtain $Y_{r,j-1} = Y_{r-1,j}$ $r = 2 \ldots m$, $j = 2 \ldots n$ and $y_{1,i} = y_{k,1} = 0$, $i = 1 \ldots n - 1$, $k = 1 \ldots m - 1$. Suppose $n \le m$ and consider the elements $y_{i,j}$ for which $i + j \le n + 1$; i.e., $y_{n-k,k+1} = y_{n-k-1,k+2} = \cdots = y_{1,n} = 0$ for $k = 1 \ldots n - 1$. The j-th column of Y may then be written $\mathrm{col}(y_{1,j}\ y_{2,j}\ y_{3,j} \cdots y_{m,j}) = \mathrm{col}(y_{1,j}\ y_{1,j+1}\ y_{1,j+2} \cdots y_{1,j+m-1}) = \mathrm{col}(0 \ldots 0\ y_{1,n+1}, y_{1,n+2}, \cdots y_{1,j+m-1}) = \mathrm{col}(0 \ldots 0\ x_{n+1}, x_{n+2}, \cdots x_{j+m-1})$ where $x_{i+j-2} = y_{i,j-1} = y_{i-1,j}$.

The $m \times n$ matrix Y may thus be partitioned, for $n \le m$, into an $n \times n$ symmetric matrix Y_1 and an $(m - n) \times (m - n)$ zero matrix 0_1, such that Y is the direct sum of Y_1 and 0_1. Besides being symmetric, Y_1 is also persymmetric, as all the elements in the j-th perpendicular to the main diagonal of Y, are equal to x_j. In the case $n \le m$, Y_1 and consequently Y has n arbitrary parameters. Similarly it may be shown that in the other case, $m \le n$, Y_1 has m arbitrary parameters, and is persymmetric, thus Y has m arbitrary parameters.

In any case, Y has $\min(m, n)$ arbitrary parameters.

q.e.d.

As an illustration of Lemma 2, consider

$$V = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad W = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Then $VY - YW = \begin{bmatrix} 0 & 0 & 0 & 0 \\ y_{11} & y_{12} & y_{13} & y_{14} \\ y_{21} & y_{22} & y_{23} & y_{24} \end{bmatrix} - \begin{bmatrix} 0 & y_{11} & y_{12} & y_{13} \\ 0 & y_{21} & y_{22} & y_{23} \\ 0 & y_{31} & y_{32} & y_{33} \end{bmatrix} = 0$

The solution is $Y = \begin{bmatrix} 0 & 0 & 0 & x_4 \\ 0 & 0 & x_4 & x_5 \\ 0 & x_4 & x_5 & x_6 \end{bmatrix}$ where $y_{14} = y_{23} = y_{32} = x_4$, $y_{24} = y_{33} = x_5$, $y_{34} = x_6$.

2. Explicit Determination of the Symmetrizing Matrices

Lemma 3: Let M be an n-square matrix with the single characteristic value c, and let $(x - c)^{n_i}$ be an elementary divisor of M of multiplicity r_i, $i = 1...p$ and $n_1 > n_2 >...> n_p$. Then the most general matrix X satisfying $XM = M'X$ has

$$\sum_{i=1...p} (r_i^2 n_i + 2r_i \sum_{j+i+1...p} r_j n_j)$$ arbitrary parameters.

Proof: Without any loss of generality we can assume that $M = W_1 +...+ W_p$ where $W_i = U_{n_i} \oplus ... \oplus U_{n_i}$ (direct sum of r_i terms), and U_{n_i} is an n_i-square auxiliary unit matrix, for in the discussion leading to Lemma 1, we observed that in the solution of $XM = M'X$ the fundamental equation (3) reduced to one involving only the nilpotent part of M.

Partitioning X conformally with M, the equation $U_{n_i} X_{ij} = X_{ij} U'_{n_j}$ determines the structure of any block X_{ij} in the partitioning of X. For a block X_{ii} associated with a pair U_{n_i}, U'_{n_i}, Lemma 2 states X_{ii} will have n_i arbitrary parameters. There are r_i blocks U_{n_i}, and r_i blocks U'_{n_i}, thus these blocks will form r_i^2 pairs, each with n_i arbitrary parameters. This accounts for $r_i^2 n_i$ parameters.

Any block X_{ij} corresponding to a pair U_{n_i}, U'_{n_j} with $n_i \neq n_j$ will have $\min(n_i, n_j)$ parameters. Assuming $i < j$, then $n_i > n_j$ and for one block U_{n_i}

there are r_j blocks U'_{n_j} each contributing n_j parameters, and this will be true for $j = i + 1, i + 2 \ldots p$. Thus the one block U_{n_i} is associated with $\sum_{j=i+1\ldots p} r_j n_j$ arbitrary parameters. Since each block U_{n_i} is repeated r_i times, the number of parameters associated with U_{n_i} becomes $r_i \sum r_j n_j$. Each time the equation $U_{n_i} X_{ij} = X_{ij} U'_{n_i}$ occurs with $i < j$, it also occurs with $i > j$, which in effect doubles the number of parameters in the last sum given.

There are p blocks U_{n_i}, thus the matrix X will have

$$\sum_{i=1\ldots p} (r_i^2 n_i + 2r_i \sum_{j=i+1\ldots p} r_j n_j) \quad \text{arbitrary parameters.}$$

q.e.d.

An example may provide a better understanding of Lemma 3. Let M be a matrix as described in Lemma 3, with $p = 2$, $r_1 = n_1 = 2$, $r_2 = n_2 = 1$. For instance,

$$M = \begin{bmatrix} c & 0 & 0 & 0 & 0 \\ 1 & c & 0 & 0 & 0 \\ 0 & 0 & c & 0 & 0 \\ 0 & 0 & 1 & c & 0 \\ 0 & 0 & 0 & 0 & c \end{bmatrix} ; \text{ then most general solution X of } XM = M'X$$

is

$$X = \left[\begin{array}{cc:cc:c} x_1 & x_2 & x_3 & x_4 & x_5 \\ x_2 & 0 & x_4 & 0 & 0 \\ \hdashline x_6 & x_7 & x_8 & x_9 & x_{10} \\ x_7 & 0 & x_9 & 0 & 0 \\ \hdashline x_{11} & 0 & x_{12} & 0 & x_{13} \end{array}\right]$$

This X has 13 arbitrary parameters $x_1 \ldots x_{13}$, which agrees with $\sum_{i=1,2}(r_i^2 n_i + 2r_i \sum_{j=i+1,2} r_j n_j)$. Note that each block in the partitioning of X is persymmetric.

Corollary: Let X be a symmetrizing matrix of the matrix M of Lemma 3. Then X has

$$1/2 \sum_{i=1\ldots p} \left[r_i(r_i + 1)n_i + 2r_i \sum_{j=i+1\ldots p} r_j n_j \right] \text{ parameters.}$$

Proof: In the partitioning of X in Lemma 3, there are r_i diagonal blocks, each having n_i parameters, and each block is persymmetric, hence symmetric. The remaining $r_i^2 - r_i$ off-diagonal blocks each have n_i parameters. For X to be symmetric, these blocks will be equal in pairs; thus the number of parameters is halved. Thus the square blocks of order n_i contribute $r_i n_i + (r_i^2 - r_i)n_i/2$, or $r_i(r_i + 1)n_i/2$ parameters.

The remaining off-diagonal rectangular blocks of dimension $n_i \times n_j$ contribute $r_i \sum_{j=i+1\ldots p} r_j n_j$ parameters to a symmetric matrix X.

Thus taking into account the square and nonsquare blocks in the partitioning of X for $i = 1\ldots p$ we obtain the result stated in this Corollary.

q.e.d.

In the example preceding this Corollary, if we require that X be symmetric, then the number of arbitrary parameters is reduced from 13 to 9.

We may now establish a theorem for an arbitrary square matrix M.

Theorem 1: Let M be an n-square matrix with distinct characteristic values $c_1\ldots c_p$, $p \le n$, and let $(x - c_i)^{e_{ij}}$, $j = 1\ldots n_i$, $e_{i1} > \ldots > e_{in_i}$ be the elementary divisors of M, where each $(x - c_i)^{e_{ij}}$ has multiplicity r_{ij}. Then the most general matrix X satisfying

(5) $XM = M'X$

has $\sum_{i=1\ldots p} \left[\sum_{j=1\ldots n_i} (r_{ij}^2 e_{ij} + 2r_{ij} \sum_{k=j+1\ldots n_i} r_{ik} e_{ik}) \right]$ arbitrary parameters and the most general symmetrizing matrix of M has

$$1/2 \sum_{i=1\ldots p} \left[\sum_{j=1\ldots n_i} (r_{ij}(r_{ij} + 1)e_{ij} + 2r_{ij} \sum_{k=j+1\ldots n_i} r_{ik} e_{ik}) \right] \text{ arbitrary parameters.}$$

Proof: Let X and M be partitioned according to the reduction of the n-dimensional space $V = V_{c_1} \oplus \dots \oplus V_{c_p}$ of Theorem 1, Chapter III. Then by Lemma 1 above, the nondiagonal blocks X_{ij} in the partitioning of X are zero, as they are associated with pairs of blocks in the partitioning of M which belong to pairs c_i, c_j with $c_i \neq c_j$.

Upon partitioning the i-th diagonal block X_{ii} of X conformally with the partitioning of M in accordance with the Jordan canonical form theorem, and applying Lemma 3, the number of parameters in the diagonal block X_{ii} is

$\sum_{j=1\dots n_i} (r_{ij}^2 e_{ij} + 2r_{ij} \sum_{k=j+1\dots n_i} r_{ik}e_{ik})$. By summing this expression for $i = 1\dots p$ we obtain the first formula.

For a symmetric X, the Corollary to Lemma 3 provides, for each diagonal block in the (second) partitioning of X,

$$1/2 \sum_{j=1\dots n_i} (r_{ij}(r_{ij} + 1)e_{ij} + 2r_{ij} \sum_{k=j+1\dots n_i} r_{ik}e_{ik})$$

parameters. Summing for $i = 1\dots p$, the second formula is obtained.

q.e.d.

3. Symmetrizing Matrices for a Nonderogatory Matrix

In Chapter I some results were derived concerning the symmetrizing matrices of a matrix with simple characteristic values. Such a matrix is simply a nonderogatory matrix, as its minimal and characteristic polynomials are equal. We have established a much stronger result for a nonderogatory matrix M, (see [17] p. 1338).

Theorem 2: Every matrix X satisfying $XM = M'X$ is symmetric if and only if M is nonderogatory.

Proof: If M is nonderogatory, $n_i = r_{ij} = 1$ for $i = 1\dots n_i$ and $j = 1\dots p$, in which case the number of parameters in a general matrix X satisfying the above relation is $\sum_{i=1\dots p} e_{i1}$ and for a symmetric X the number of parameters is also the same. Thus every matrix X satisfying $XM = M'X$ is symmetric.

Conversely, if every matrix satisfying $XM = M'X$ is symmetric, then equating the formulae of Theorem 1 for the number of parameters in a general and a symmetric X we obtain

$$\sum_{i=1\ldots p}\left[\sum_{j=1\ldots n_i}(r_{ij}(r_{ij}-1)e_{ij}/2 + r_{ij}\sum_{k=j+1\ldots n_i} r_{ik}e_{ik})\right] = 0.$$

Since n_i, r_{ij} and e_{ij} are all positive integers, necessarily $r_{ij} = 1$ for $j = 1\ldots n_i$ and $n_i = 1$; i.e., $n_i = 1$ and $r_{ij} = 1$ for $j = 1$, both for all $i = 1\ldots p$. Then there is one and only one elementary divisor corresponding to each characteristic value of M, and since the minimal polynomial is the product of the elementary divisors of maximal multiplicity, clearly M is nonderogatory.

q.e.d.

Theorem 2 has also been proven in a different fashion in [20].

Corollary: The symmetrizing matrices of a nonderogatory matrix M are congruent to direct sums of persymmetric matrices.

Proof: This follows from the fact that the symmetrizing matrix of a nonderogatory matrix in Jordan canonical form is a direct sum of persymmetric matrices.

q.e.d.

The following example is an illustration of Theorem 2 and its Corollary. Let M be

$$M = \begin{bmatrix} a & 0 & 0 & 0 & 0 \\ 1 & a & 0 & 0 & 0 \\ 0 & 0 & b & 0 & 0 \\ 0 & 0 & 1 & b & 0 \\ 0 & 0 & 0 & 0 & c \end{bmatrix}$$

Then M is a nonderogatory and $p = 3$, $e_1 = e_2 = 2$, $e_3 = 1$, $r_1 = r_2 = r_3 = 1$ and the most general symmetrizing matrix X for M should have $\sum_{i=1\ldots p} e_i = 5$ arbitrary parameters. By a straightforward computation we find

$$X = \begin{bmatrix} A & B & 0 & 0 & 0 \\ B & 0 & 0 & 0 & 0 \\ 0 & 0 & C & D & 0 \\ 0 & 0 & D & 0 & 0 \\ 0 & 0 & 0 & 0 & E \end{bmatrix}$$

as the most general symmetric matrix satisfying $XM = M'X$.

Note that X is the direct sum of $\begin{bmatrix} A & B \\ B & 0 \end{bmatrix}$, $\begin{bmatrix} C & D \\ D & 0 \end{bmatrix}$, $[E]$, all three being persymmetric.

CHAPTER V

THE SYMMETRIZING MATRICES OF A COMPANION MATRIX

With any polynomial $f(c) = -c^n + a_1c^{n-1} + \ldots + a_n$ we can associate a matrix M, known as the companion matrix of f(c).

$$M = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ 1 & 0 & & 0 \\ 0 & 1 & & 0 \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ 0 & 0 & & 0 \end{bmatrix}$$

The companion matrix M of f(c) has f(c) as its characteristic polynomial, and its minimal polynomial is also f(c) ([2] p. 318). Thus a companion matrix is a nonderogatory matrix. The following sections will be concerned with obtaining symmetrizing matrices for a companion matrix.

1. A Specific Example

We have established for a nonderogatory n-square matrix M that for any symmetrizing matrix B,

(i) B has n arbitrary parameters

(ii) B is congruent to the direct sum of persymmetric matrices.

Using these facts, let us attempt to find a symmetrizing matrix for the 4×4 companion matrix M of the polynomial $f(c) = -c^4 + a_1c^3 + a_2c^2 + a_3c + a_4$,

$$M = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

As a first attempt at obtaining a symmetrizing matrix B for M, let B be a persymmetric matrix

$$B = \begin{bmatrix} a & b & c & d \\ b & c & d & e \\ c & d & e & f \\ d & e & f & g \end{bmatrix}$$

The matrix equation $BM - M'B = 0$ yields the set of equations

$$\begin{aligned} &ba_1 + c = aa_2 + c \qquad && ca_2 + f = ba_3 + f \\ (1)\quad &ca_1 + d = aa_3 + d && da_2 + g = ba_4 \\ &da_1 + f = aa_4 && da_3 + h = ca_4 \end{aligned}$$

Of the 7 parameters involved, 4 may be set arbitrarily. The system (1) is greatly simplified if we set $a = b = c = 0$. Then (1) reduces to $da_1 + f = 0$, $da_2 + g = 0$, $da_3 + h = 0$. Clearly the best strategy is to choose d as the last parameter at our disposal, and set $d = -1$. Then $f = a_1$, $g = a_2$ and $h = a_3$. The symmetrizing matrix B is then

$$B = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & a_1 \\ 0 & -1 & a_1 & a_2 \\ -1 & a_1 & a_2 & a_3 \end{bmatrix}$$

We know that $BM, BM^2, \ldots BM^k, \ldots$ are also symmetrizing matrices of M, and that any symmetrizing matrix of M is a linear combination of B, BM, BM^2 and BM^3.

$$BM = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & -1 & a_1 & 0 \\ -1 & a_1 & a_2 & 0 \\ 0 & 0 & 0 & -a_4 \end{bmatrix}, \quad BM^2 = \begin{bmatrix} 0 & -1 & 0 & 0 \\ -1 & a_1 & 0 & 0 \\ 0 & 0 & -a_3 & -a_4 \\ 0 & 0 & -a_4 & 0 \end{bmatrix}, \quad BM^3 = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -a_2 & -a_3 & -a_4 \\ 0 & -a_3 & -a_4 & 0 \\ 0 & -a_4 & 0 & 0 \end{bmatrix}$$

Then it is clear that for BM^4,

$$-BM^4 = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ a_2 & a_3 & a_4 & 0 \\ a_3 & a_4 & 0 & 0 \\ a_4 & 0 & 0 & 0 \end{bmatrix} = a_4B + a_3BM + a_2BM^2 + a_1BM^3.$$

The matrix BM^5 no longer has the simple structure of the preceding matrices, i.e., it is not a direct sum of persymmetric matrices,

$$-BM^5 = \begin{bmatrix} a_1^2 + a_2 & a_1a_2 + a_3 & a_1a_3 + a_4 & a_1a_4 \\ a_1a_2 + a_3 & a_2^2 + a_4 & a_2a_3 & a_2a_4 \\ a_1a_3 + a_4 & a_2a_3 & a_3^2 & a_3a_4 \\ a_1a_4 & a_2a_4 & a_3a_4 & a_4^2 \end{bmatrix}$$

The preceding examples have been published by Lancaster [15] who, however, does not attempt a generalization of the symmetrizing matrix for other than companion matrices.

2. A General Solution

If the polynomial $f(c) = -c^4 + a_1c^3 + a_2c^2 + a_3c + a_4$ has the distinct roots p, q, r, and s, then the companion matrix M of f(c) is diagonalised by the Vandermonde matrix V,

$$V = \begin{bmatrix} p^3 & q^3 & r^3 & s^3 \\ p^2 & q^2 & r^2 & s^2 \\ p & q & r & s \\ 1 & 1 & 1 & 1 \end{bmatrix}, \text{ for } MV = \begin{bmatrix} p^4 & q^4 & r^4 & s^4 \\ p^3 & q^3 & r^3 & s^3 \\ p^2 & q^2 & r^2 & s^2 \\ p & q & r & s \end{bmatrix} = VC$$

where C is the diagonal matrix with the diagonal entries p, q, r and s. The Vandermonde matrix V is nonsingular whenever the entries p, q, r, and s are distinct ([1] p. 186). Thus $M = VCV^{-1}$ is similar to a diagonal matrix.

We can say of any symmetrizing matrix A of M

(i) A is a linear combination of B, BM, BM^2, BM^3.

(ii) $A = (V^{-1})'QV^{-1}$ where Q is some diagonal matrix.

and these statements are equivalent. Similar remarks apply to the symmetrizing matrices of the transpose M' of the companion matrix M. In particular, symmetrizing matrices of M' have the form VQV', which involves the Vandermonde matrix V itself, rather than its inverse, and thus lends itself more readily to a practical method for the construction of symmetrizing matrices in terms of symmetric functions of the roots.

We have avoided discussing the construction of symmetrizing matrices of arbitrary matrices because our development was based upon knowing the Jordan canonical form of a given matrix. In practice the Jordan canonical form is the least and last known property of a given matrix. The interested reader may wish to refer to a paper entitled "The Construction of Matrices Symmetrizing an Arbitrary Matrix," [13], and also "Matrix Symmetrization Methods for the Algebraic Eigenproblem" [14].

REFERENCES

1. R. Bellman, Introduction to Matrix Analysis, McGraw-Hill, New York, 1960.

2. G. Birkhoff and S. MacLane, A Survey of Modern Algebra, Macmillan, New York, 1957.

3. J. Blumenfeld and W. Mayer, Über Poincarésche Fundamentalfunktionen Sitz. d. Akad. der Wiss. Wien 123-2a-2 (1914) 2011-2047.

4. S. Cater, An Elementary Development of the Jordan Canonical Form, American Math. Monthly, 69 (1962) 391-393.

5. F. Farrell, Thesis, University of Ottawa, In preparation.

6. F. R. Gantmacher, Matrix Theory, Volume 1, Chelsea, New York, 1956.

7. I. M. Gelfand, Lectures on Linear Algebra, Interscience, New York, 1961.

8. P. R. Halmos, Finite-Dimensional Vector Spaces, 2nd ed., Van Nostrand, Princeton, 1958.

9. K. Hoffman and R. Kunze, Linear Algebra, Prentice-Hall, Englewood Cliffs, N.J., 1961.

10. J. L. Howland, Induced Potentials, Ph.D. Thesis, Harvard, 1955.

11. J. L. Howland, A Method for Computing the Real Roots of a Determinental Equation, American Math. Monthly, 68 (1961) 235-239.

12. J. L. Howland, On Some Methods for Computing the Roots of Polynomials, Proc. IFIP Congress 62, Munich (1962) p. 116.

13. J. L. Howland, The Construction of Matrices Symmetrizing an Arbitrary Matrix, A paper prepared for the Can. Math. Congress, Quebec City, P.Q., June 4, 1963.

14. J. L. Howland and F. Farrell, Matrix Symmetrization Methods for the Algebraic Eigenproblem, A.C.M. National Conference, August 1963, To be published.

15. P. Lancaster, Symmetric Transformations of the Companion Matrix, NABLA, Bull. of the Malayan Math. Soc., (Sept. 1961) 146-148.

16. M. Marcus, Basic Theorems in Matrix Theory, U.S. Dept. of Commerce, National Bureau of Standards, Applied Math Series 57, Washington, 1960.

17. M. Marcus and N. A. Khan, On a Commutator Result of Taussky and Zassenhaus, Pacific J. Math., 10 (1960) 1337-1346.

18. J. Marty, Existence de Solutions Singulières pour Certaines Equations de Fredholm, C. R. 150 (1910) 1031-1033.

19. S. Perlis, Theory of Matrices, Addison-Wesley, Reading, Mass., 1952.

20. O. Taussky and H. Zassenhaus, On the Similarity Transformation Between a Matrix and Its Transpose, Pacific J. Math., 9 (1959) 893-896.

21. A. C. Zaanen, Linear Analysis, Interscience, New York, 1953.

BIBLIOGRAPHY

R. Baer, Linear Algebra and Projective Geometry, Academic Press, New York, 1952.

W. S. Burnside and A. W. Panton, The Theory of Equations, Volume 2, Dover, New York, 1960.

D. T. Finkbeiner II, Introduction to Matrices and Linear Transformations, Freeman, New York, 1960.

W. F. Freiberger, The International Dictionary of Applied Mathematics, Van Nostrand, Princeton, 1960.

H. L. Hamburger and M. E. Grimshaw, Linear Transformations in N-Dimensional Vector Space, Cambridge University Press, 1951.

F. B. Hildebrand, Methods of Applied Mathematics, Prentice-Hall, Englewood Cliffs, N.J., 1952.

N. H. Kuiper, Linear Algebra and Geometry, North-Holland Publishing Co., 1962.

M. Marcus, Linear Operations on Matrices, American Math. Monthly, 69 (1962) 837-847.

T. Muir, The Theory of Determinants, Dover, New York, 1960.

S. Narayan, A Textbook of Matrices, Stevens Co., New York, 1957.

H. A. Osborn, A Class of Bilinear Forms, Trans. American Math. Soc., 90 (1959) 485-498.

R. R. Stoll, Linear Algebra and Matrix Theory, McGraw-Hill, New York, 1952.