PRINCIPAL SUBMATRICES OF NORMAL AND HERMITIAN MATRICES

BY
R. C. THOMPSON

1. Introduction

In this paper we obtain inequalities and location theorems linking all the eigenvalues of all of the principal $k \times k$ submatrices of a normal or Hermitian $n \times n$ matrix A to the eigenvalues of A. We also obtain inequalities for certain expressions involving $k \times k$ subdeterminants of A. In addition we examine the possible occurrences of a multiple eigenvalue of A among the eigenvalues of the principal $k \times k$ submatrices of A. Certain of our theorems for normal matrices hold only when $k = n - 1$. It is an interesting and open question to find analogues of these theorems for $k \times k$ principal submatrices. For Hermitian matrices we obtain stronger theorems than are possible for arbitrary normal matrices. In one of our theorems (Theorem 3) we only require that A be diagonable.

2. Notation

In this paper $A = (A_{ij})$ denotes an $n \times n$ diagonable matrix with eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$. Usually A will be normal. In general the eigenvalues are not all distinct so let $\mu_1, \mu_2, \cdots, \mu_s$ denote the distinct eigenvalues, where the multiplicity of μ_i is e_i for $1 \leqq i \leqq s$; $e_1 + \cdots + e_s = n$.

We arrange the notation so that

$$(\lambda_1, \lambda_2, \cdots, \lambda_n) = (\mu_1, \cdots, \mu_1, \mu_2, \cdots, \mu_2, \cdots, \mu_s, \cdots, \mu_s).$$

When A is Hermitian we assume $\mu_1 < \mu_2 < \cdots < \mu_s$.

For fixed integers n and k, $1 \leqq k < n$, Q_{nk} denotes the set of all sequences $\omega = \{i_1, i_2, \cdots, i_k\}$ of integers such that $1 \leqq i_1 < i_2 < \cdots < i_k \leqq n$. We always let

$$\omega = \{i_1, i_2, \cdots, i_k\} \quad \text{and} \quad \tau = \{j_1, j_2, \cdots, j_k\}$$

be two typical elements of Q_{nk}. The $k \times k$ matrix B defined by

$$B_{\alpha\beta} = A_{i_\alpha j_\beta}, \qquad 1 \leqq \alpha, \beta \leqq k,$$

is denoted by $A[\omega \mid \tau]$. The $(n - 1) \times (n - 1)$ matrix obtained by deleting row i and column j from A is denoted by $A(i \mid j)$. We let $f(\lambda)$, $f_{[\omega]}(\lambda)$, $f_{(i)}(\lambda)$ stand for the characteristic polynomials of A, $A[\omega \mid \omega]$, $A(i \mid i)$, respectively. We let

$$f_{[\omega]}(\lambda) = \lambda^k - c_{\omega 1}\lambda^{k-1} + c_{\omega 2}\lambda^{k-2} - \cdots + (-1)^k c_{\omega k}.$$

Here, of course, $c_{\omega j}$ is the sum of the principal $(k - j) \times (k - j)$ subdetermi-

Received January 29, 1965.

nants of $A[\omega \mid \omega]$. The roots of $f_{[\omega]}(\lambda)$ are denoted by $\eta_{\omega 1}, \eta_{\omega 2}, \cdots, \eta_{\omega k}$. When A is Hermitian we arrange the numbering so that $\eta_{\omega 1} \leqq \eta_{\omega 2} \leqq \cdots \leqq \eta_{\omega k}$.

For integers $k \geqq 1$ and r, $0 \leqq r \leqq k$, we define $E_r(a_1, a_2, \cdots, a_k)$ by the polynomial identity

$$\prod_{i=1}^{k} (\lambda + a_i) = \sum_{r=0}^{k} E_r(a_1, a_2, \cdots, a_k)\lambda^{k-r}.$$

We shall always let $h = h(a_1, \cdots, a_k)$ be an arbitrary linear function of k variables. We set

$$E_r(\Lambda_\omega) = E_r(\lambda_{i_1}, \lambda_{i_2}, \cdots, \lambda_{i_k})$$

and, for reasons of compactness, we define

$$h(\Lambda_\omega) = h(E_1(\Lambda_\omega), E_2(\Lambda_\omega), \cdots, E_k(\Lambda_\omega)),$$

and

$$h(A[\omega \mid \omega]) = h(c_{\omega 1}, c_{\omega 2}, \cdots, c_{\omega k}).$$

We let G_α denote the geometric mean of the positive real numbers

$$| \mu_\beta - \mu_\alpha |, \quad \beta = 1, 2, \cdots, \alpha - 1, \alpha + 1, \cdots, s.$$

We set $\rho_\alpha = (e_\alpha/n)^{1/(s-1)}G_\alpha$, $\rho = \{\sum_{\alpha=1}^{s} \prod_{\beta=1,\beta\neq\alpha}^{s} | \mu_\alpha - \mu_\beta |^{-1}\}^{-1/(s-1)}$.

The circles with center μ_α and radii ρ_α, ρ, G_α, $(\Omega e_\alpha)^{1/(s-1)}G_\alpha$ are denoted by C_α, C^α, ${}_\alpha C$, ${}^\alpha C$, respectively. Here $\Omega = 4n^{-1}(n + 2)^{-1}$ if n is even and $\Omega = 4(n + 1)^{-2}$ if n is odd.

As is usual, the transpose and complex conjugate transpose of A are indicated with A^T, A^*, respectively. The k^{th} compound of A is $C_k(A)$. The identity matrix is denoted by I.

3. Preliminary calculations

Let $D = \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ and let $A = SDS^{-1}$ for some nonsingular S. Then $\lambda I - A = S(\lambda I - D)S^{-1}$. Hence

$$C_k(\lambda I - A) = C_k(S)C_k(\lambda I - D)C_k(S)^{-1}.$$

The diagonal elements of $C_k(\lambda I - A)$ are the $f_{[\omega]}(\lambda)$ for $\omega \in Q_{nk}$. The diagonal elements of the diagonal matrix $C_k(\lambda I - D)$ are the polynomials

$$\prod_{\beta\in\tau} (\lambda - \lambda_\beta), \qquad \tau \in Q_{nk}. \tag{1}$$

Hence

$$f_{[\omega]}(\lambda) = \sum_{\tau\in Q_{nk}} \det S[\omega \mid \tau] \det S^{-1}[\tau \mid \omega] \prod_{\beta\in\tau} (\lambda - \lambda_\beta). \tag{2'}$$

When $S = U$ is unitary and A is normal, (2′) becomes

$$f_{[\omega]}(\lambda) = \sum_{\tau\in Q_{nk}} | \det U[\omega \mid \tau] |^2 \prod_{\beta\in\tau} (\lambda - \lambda_\beta). \tag{2}$$

We rewrite (2) in vector matrix language as

$$\begin{bmatrix} \cdots \\ f_{[\omega]}(\lambda) \\ \cdots \end{bmatrix} = W \begin{bmatrix} \cdots \\ \prod_{\beta\in\tau}(\lambda - \lambda_\beta) \\ \cdots \end{bmatrix}. \tag{3}$$

Here, in (3), the column vector on the left has as components the $f_{[\omega]}(\lambda)$, ordered lexicographically, and the column vector on the right has as components the polynomials (1), ordered lexicographically. The matrix W is non-negative and doubly stochastic; its entries are the $|\det U[\omega \mid \tau]|^2$, in doubly lexicographic order. We compare coefficients of the same power of λ on each side of (3). As an easy consequence we get

$$\begin{bmatrix} \cdots \\ h(A[\omega \mid \omega]) \\ \cdots \end{bmatrix} = W \begin{bmatrix} \cdots \\ h(\Lambda_\tau) \\ \cdots \end{bmatrix}. \tag{4}$$

The column vector on the left side of (4) has as components the numbers $h(A[\omega \mid \omega])$, ordered lexicographically, whereas the column vector on the right side of (4) has as components the numbers $h(\Lambda_\tau)$, ordered lexicographically. From (4) we get on taking real parts (indicated by R) and absolute values:

$$\begin{bmatrix} \cdots \\ R(h(A[\omega \mid \omega])) \\ \cdots \end{bmatrix} = W \begin{bmatrix} \cdots \\ Rh((\Lambda_\tau)) \\ \cdots \end{bmatrix}, \tag{5}$$

$$\begin{bmatrix} \cdots \\ |h(A[\omega \mid \omega])| \\ \cdots \end{bmatrix} \leqq W \begin{bmatrix} \cdots \\ |h(\Lambda_\tau)| \\ \cdots \end{bmatrix}. \tag{6}$$

The inequality in (6) is componentwise.

Now let $k = n - 1$. Then, given $\omega, \tau \in Q_{n,n-1}$, there exist unique integers i, j for which $1 \leqq i, j \leqq n$, $i \notin \omega$, $j \notin \tau$. Since U is unitary, $U^{-1} = U^*$. Consequently $(\det U)^{-1} \det U[\omega \mid \tau](-1)^{i+j} = \bar{U}_{ij}$; hence

$$|\det U[\omega \mid \tau]|^2 = |U_{ij}|^2.$$

Moreover,

$$\textstyle\prod_{\beta \in \tau} (\lambda - \lambda_\beta) = f(\lambda)/(\lambda - \lambda_j),$$

and $f_{[\omega]}(\lambda) = f_{(i)}(\lambda)$. So (3) may be rewritten as

$$\begin{bmatrix} \cdots \\ f_{(i)}(\lambda) \\ \cdots \end{bmatrix} = W \begin{bmatrix} \cdots \\ f(\lambda)(\lambda - \lambda_j)^{-1} \\ \cdots \end{bmatrix} \tag{7}$$

and (2) becomes

$$f_{(i)}(\lambda) = \textstyle\sum_{j=1}^{n} |U_{ij}|^2 f(\lambda)(\lambda - \lambda_j)^{-1}. \tag{8}$$

All our results will follow from these formulas.

4. Normal matrices

Except in Theorem 3, A is always a normal matrix in §4.

THEOREM 1. *For given $\omega \in Q_{nk}$, $h(A[\omega \mid \omega])$ lies in the convex hull of the complex numbers $h(\Lambda_\tau)$ as τ runs over Q_{nk}.*

Proof. This is immediate from (4) since W is nonnegative and doubly stochastic. This is a generalization of a result in [5] which had also been proved independently by M. Marcus.

THEOREM 2. *For fixed* $\omega \in Q_{nk}$,
(i) $\max_U |h(A[\omega \mid \omega])| = \max_{\tau\in Q_{nk}} |h(\Lambda_\tau)|$,
(ii) $\max_U R(h(A[\omega \mid \omega]) = \max_{\tau\in Q_{nk}} R(h(\Lambda_\tau))$,
(iii) $\min_U R(h(A[\omega \mid \omega]) = \min_{\tau\in Q_{nk}} R(h(\Lambda_\tau))$.

Remark. $\max_U$, $\max_\tau$ denote, respectively, the maximum of the quantity in question as U varies over all unitary matrices or as τ varies over all sequences of Q_{nk} . Similarly for the min.

Proof. That the left members of (i), (ii) are always $\leqq$ the right members follows from (6), (5) since W is doubly stochastic. Equality is achieved by taking U to be a permutation matrix such that UDU^* has

$$\lambda_{j_\alpha}$$

at the (i_α , i_α) position, $1 \leqq \alpha \leqq k$. Then $f_{[\omega]}(\lambda)$ is the polynomial (1), so that $h(A[\omega \mid \omega]) = h(\Lambda_\tau)$.

Remark. The theory of Schur convex and concave functions [4] in combination with (5) or (6) yields many inequalities linking symmetric functions of the real numbers $R(h(A[\omega \mid \omega]))$ (or of $|h(A[\omega \mid \omega])|$) as ω varies over Q_{nk} for fixed k to the same symmetric functions of real numbers $R(h(\Lambda_\tau))$ (or of $|h(\Lambda_\tau)|$, respectively) as τ varies over Q_{nk} .

When A is merely diagonable it follows from (2′) that

$$(9) \quad f_{[\omega]}(\lambda) = \sum\nolimits_{\tau\in Q_{nk}} \det S[\omega \mid \tau] \det S^{-1}[\tau \mid \omega] f(\lambda) \prod\nolimits_{\beta\notin\tau} (\lambda - \lambda_\beta)^{-1}.$$

If $e_\alpha - (n - k) \geqq 1$, then

$$(10) \qquad (\lambda - \mu_\alpha)^{e_\alpha - n + k}$$

is a divisor of the right side of (9), hence of the left also. Thus μ_α is a root of $f_{[\omega]}(\lambda)$ with multiplicity at least $e_\alpha - n + k$. If may happen that μ_α is a root of $f_{[\omega]}(\lambda)$ with multiplicity $> e_\alpha - n + k$. However we have

$$(11) \qquad \sum\nolimits_{\omega\in Q_{nk}} f_{[\omega]}(\lambda) = ((n - k)!)^{-1} f^{(n-k)}(\lambda).$$

Here $f^{(n-k)}(\lambda)$ denotes the derivative of $f(\lambda)$ of order $n - k$. Formula (11) follows by summing (2′) over $\omega \in Q_{nk}$ and using

$$\sum\nolimits_{\omega\in Q_{nk}} \det S[\omega \mid \tau] \det S^{-1}[\tau \mid \omega] = 1.$$

(This follows from $C_k(S^{-1})C_k(S) = I$.) In fact, however, (11) holds for all matrices (not just diagonable ones) and can be proved in general by considering the the determinant $\det(tI - (\lambda I - A))$ and using Taylor's theorem.

In any event it follows from (11) that

$$(\lambda - \mu_\alpha)^{e_\alpha - n+k+1}$$

cannot be a factor of every $f_{[\omega]}(\lambda)$. This completes the proof of Theorem 3.

THEOREM 3. *Let k be fixed and let A be an $n \times n$ matrix over a field K for which μ_α is an eigenvalue with multiplicity e_α.*

(i) *Suppose A is diagonable and $e_\alpha - (n-k) \geqq 1$. Then each $A[\omega \mid \omega]$, $\omega \in Q_{nk}$, has μ_α as an eigenvalue with multiplicity at least $e_\alpha - (n-k)$.*

(ii) *Suppose A is arbitrary and K has characteristic zero or larger than n. Then not every $A[\omega \mid \omega]$ can have μ_α as an eigenvalue with multiplicity at least the larger of $\{e_\alpha - (n-k) + 1, 1\}$.*

Theorem 3(i) is false when A is not diagonable. A counterexample is

$$\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}.$$

For the rest of §4 we suppose A is normal and $k = n-1$. To avoid trivial situations we assume $s \geqq 2$ so that A is not scalar. We know that $A(i \mid i)$ has μ_α as eigenvalue with multiplicity $e_\alpha - 1$ or larger. Thus μ_α with multiplicity $e_\alpha - 1$ is always a root of $A(i \mid i)$, $1 \leqq \alpha \leqq s$. We call these the *trivial* eigenvalues of $A(i \mid i)$. In addition there exist $s-1$ additional eigenvalues of $A(i \mid i)$, denoted by $\xi_{i1}, \xi_{i2}, \cdots, \xi_{i,s-1}$. We call these the *nontrivial* eigenvalues of $A(i \mid i)$. It may happen that the nontrivial eigenvalues of $A(i \mid i)$ are not all distinct and that some of the nontrivial eigenvalues of $A(i \mid i)$ equal some of the trivial eigenvalues. So we now have

$$f_{(i)}(\lambda) = \prod_{j=1}^{s} (\lambda - \mu_j)^{e_j - 1} \prod_{j=1}^{s-1} (\lambda - \xi_{ij}). \tag{12}$$

From (8) we get

$$f_{(i)}(\lambda) = \sum_{\beta=1}^{s} \theta_{i\beta} f(\lambda)(\lambda - \mu_\beta)^{-1}, \tag{13}$$

where

$$\theta_{i\beta} = \sum_{j : \lambda_j = \mu_\beta} | U_{ij} |^2. \tag{14}$$

The sum in (14) is over all integers j for which $\lambda_j = \mu_\beta$. Now substitute (12) and

$$f(\lambda) = \prod_{j=1}^{s} (\lambda - \mu_j)^{e_j}$$

into (13), cancel the common factor and then set $\lambda = \mu_\alpha$. The result is

$$\theta_{i\alpha} = \prod_{j=1}^{s-1} (\mu_\alpha - \xi_{ij}) \prod_{j=1, j \neq \alpha}^{s} (\mu_\alpha - \mu_j)^{-1}, \quad 1 \leqq \alpha \leqq s, 1 \leqq i \leqq n. \tag{15}$$

It follows from (14) that $\theta_{i\alpha} \geqq 0$, and that

$$\sum_{i=1}^{n} \theta_{i\alpha} = e_\alpha, \qquad 1 \leqq \alpha \leqq s, \tag{16}$$

$$\sum_{\alpha=1}^{s} \theta_{i\alpha} = 1, \qquad 1 \leqq i \leqq n. \tag{17}$$

Moreover we have

LEMMA 1. *The $n \times n$ matrix in which the column vectors*

$$e_\alpha^{-1}(\theta_{1\alpha}, \theta_{2\alpha}, \cdots, \theta_{n\alpha})^T$$

appear exactly e_α times, $1 \leqq \alpha \leqq s$, is nonnegative and doubly stochastic.

We now can improve Theorem 3 somewhat, when $k = n - 1$.

THEOREM 4. *Let α be fixed. The number of integers i, $1 \leqq i \leqq n$, for which $A(i \mid i)$ has μ_α as a nontrivial eigenvalue is at most $n - e_\alpha$. When this bound is attained then for each of the remaining e_α integers i, the nontrivial eigenvalues of $A(i \mid i)$ are $\mu_1, \cdots, \mu_{\alpha-1}, \mu_{\alpha+1}, \cdots, \mu_s$. Conversely, the number of integers i, $1 \leqq i \leqq n$, for which $A(i \mid i)$ has $\mu_1, \cdots, \mu_{\alpha-1}, \mu_{\alpha+1}, \cdots, \mu_s$ as the nontrivial eigenvalues is at most e_α. When this bound is attained then for each of the remaining $n - e_\alpha$ integers i, $A(i \mid i)$ has μ_α as a nontrivial eigenvalue.*

Remark. The bounds are attained when A is diagonal. However they can be attained when A is nondiagonal. An example is

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

Proof. By (14), (17) each of the terms in the sum (16) is between 0 and 1. So there must be at least e_α integers i for which $\theta_{i\alpha} \neq 0$. By (15), $\theta_{i\alpha} = 0$ if and only if $A(i \mid i)$ has μ_α as a nontrivial eigenvalue. Hence μ_α is a nontrivial eigenvalue of $A(i \mid i)$ for at most $n - e_\alpha$ integers i. When this bound is achieved, $\theta_{i\alpha} = 0$ for $n - e_\alpha$ values of i, and hence $\theta_{i\alpha} = 1$ for e_α values of i. But, by (17), $\theta_{i\alpha} = 1$ implies $\theta_{i\beta} = 0$ for all $\beta \neq \alpha$, and by (15) this can happen only if $\mu_1, \cdots, \mu_{\alpha-1}, \mu_{\alpha+1}, \cdots, \mu_s$ are all nontrivial eigenvalues of $A(i \mid i)$. The converse follows by reversing these steps.

THEOREM 5. *A necessary and sufficient condition that an $n \times n$ normal matrix A be diagonal is that each $(n - 1) \times (n - 1)$ principal submatrix of A has as its eigenvalues an $(n - 1)$-subset of the eigenvalues of A.*

Proof. When A is diagonal the condition is obvious. Suppose the condition is satisfied. Then the nontrivial eigenvalues of $A(i \mid i)$ are $\mu_1, \cdots, \mu_s$, omitting $\mu_{t(i)}$. Then, by (15), $\theta_{i\alpha} = 0$ except when $\alpha = t(i)$, and then $\theta_{i,t(i)} = 1$. So any $\theta_{\gamma\delta}$ is 0 or 1. Because of (16), there exist exactly e_α integers i for which $t(i) = \alpha$. When $t(i) = \alpha$, $\theta_{i\beta} = 0$ for all $\beta \neq \alpha$, so by (14), $U_{ij} = 0$ for all j for which $\lambda_j \neq \mu_\alpha$. The number of j for which $\lambda_j = \mu_\alpha$ is exactly e_α. When $t(i) \neq \alpha$, $\theta_{i\alpha} = 0$ and (14) then forces $U_{ij} = 0$ for all j for which $\lambda_j = e_\alpha$. Thus U is 0 except for blocks U_α lying at the intersection of rows numbered i for which $t(i) = \alpha$ and columns numbered j for which $\lambda_j = \mu_\alpha$. These columns j are exactly the columns j for which

$$e_1 + \cdots + e_{\alpha-1} + 1 \leqq j \leqq e_1 + \cdots + e_\alpha .$$

(See §2.) We may find a permutation matrix P such that

$$PU = \operatorname{diag}(U_1, U_2, \cdots, U_s).$$

Now PAP^T is diagonal if and only if A is. Moreover $PAP^T = (PU)D(PU)^* = D$ since $U_1, \cdots, U_s$ are each unitary and the main diagonal of D partitions into scalar segments. Hence A is diagonal.

THEOREM 6. *For an appropriate unitary U, $A_{ii} = (\text{trace } A)/n$ and $f_{(i)}(\lambda) = f'(\lambda)n^{-1}$, for all i, $1 \leqq i \leqq n$.*

Proof. Take $U_{ij} = \zeta^{(i-1)(j-1)}n^{-1/2}$, $1 \leqq i, j \leqq n$, where ζ is a primitive root of unity of order n. Then use (2) with $\lambda = 0$ and $k = 1$, and (8).

THEOREM 7. *Let α be fixed. Then either:* (i) *for at least one i, $A(i \mid i)$ has a nontrivial eigenvalue inside C_α, and for at least one i, $A(i \mid i)$ has a nontrivial eigenvalue outside C_α; or* (ii) *for every i, $A(i \mid i)$ has all its nontrivial eigenvalues on the boundary of C_α.*

Proof. We use the fact that always $\theta_{i\alpha} = \mid \theta_{i\alpha} \mid$. Suppose all the nontrivial eigenvalues of all $A(i \mid i)$ lie on the boundary of or outside of C_α, and at least one $A(i \mid i)$ has a nontrivial eigenvalue outside C_α. Then $\mid \mu_\alpha - \xi_{ij} \mid \geqq \rho_\alpha$ for all i, j, with strict inequality at least once. Then (16) becomes

$$\textstyle\sum_{i=1}^{n} (\rho_\alpha/G_\alpha)^{s-1} < e_\alpha ;$$

hence

$$\rho_\alpha < (e_\alpha/n)^{1/(s-1)}G_\alpha = \rho_\alpha .$$

This is a contradiction. Similarly we show that it cannot happen that all $A(i \mid i)$ have all their nontrivial eigenvalues on the boundary of or inside of C_α, with strictly inside at least once.

THEOREM 8. *Let i be fixed. Then either:* (i) *$A(i \mid i)$ has at least one nontrivial eigenvalue inside one of $C^1, \cdots, C^s$ and at least one nontrivial eigenvalue outside one of $C^1, \cdots, C^s$; or* (ii) *each nontrivial eigenvalue of $A(i \mid i)$ lies on the boundary of every one of $C^1, \cdots, C^s$.*

Proof. Suppose each nontrivial eigenvalue of $A(i \mid i)$ is on the boundary of, or outside of, every one of $C^1, \cdots, C^s$, with strictly outside at least once. Then $\mid \mu_\alpha - \xi_{ij} \mid \geqq \rho$ for all α, j, with strict inequality at least once. Then (15) and (17) produce

$$\textstyle\sum_{\alpha=1}^{s} \rho^{s-1} \prod_{\beta=1,\beta\neq\alpha}^{s} \mid \mu_\alpha - \mu_\beta \mid^{-1} < 1,$$

hence $\rho < \rho$. Similarly we cannot have each nontrivial eigenvalue of $A(i \mid i)$ inside of or on the boundary of each of $C^1, \cdots, C^s$, with strictly inside at least once.

The exceptional cases in Theorems 8 and 9 can happen; for an example, consider the matrix

$$\frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

THEOREM 9. *Let i, α be fixed. Then either*: (i) *$A(i \mid i)$ has a nontrivial eigenvalue inside ${}_\alpha C$; or* (ii) *the nontrivial eigenvalues of $A(i \mid i)$ are $\mu_1, \cdots, \mu_{\alpha-1}, \mu_{\alpha+1}, \cdots, \mu_s$ and each of these numbers lies on the boundary of ${}_\alpha C$.*

Proof. We know $\theta_{i\alpha} \leqq 1$. If all nontrivial eigenvalues of $A(i \mid i)$ are on the boundary of or outside of ${}_\alpha C$, and at least one nontrivial eigenvalue is outside ${}_\alpha C$, then $\mid \mu_\alpha - \xi_{ij} \mid \geqq G_\alpha$ for all j with strict inequality at least once. Then

$$G_\alpha^{s-1}/G_\alpha^{s-1} < \theta_{i\alpha} \leqq 1.$$

This is a contradiction. So all nontrivial eigenvalues of $A(i \mid i)$ are on the boundary of ${}_\alpha C$ or else at least one is inside ${}_\alpha C$. If all nontrivial eigenvalues are on the boundary of ${}_\alpha C$ then $\mid \mu_\alpha - \xi_{ij} \mid = G_\alpha$ for all j; hence $\theta_{i\alpha} = 1$. Then (17) forces $\theta_{i\beta} = 0$ for all $\beta \neq \alpha$, so that by (15), the nontrivial eigenvalues of $A(i \mid i)$ are $\mu_1, \cdots, \mu_{\alpha-1}, \mu_{\alpha+1}, \cdots, \mu_s$.

The exceptional circumstance can happen. An example is diag$(1, -1, 0)$.

THEOREM 10. *There always exists a permutation σ of $1, 2, \cdots, n$ such that $A(\sigma(i) \mid \sigma(i))$ has a nontrivial eigenvalue on the boundary of or outside of ${}^\alpha C$, for all i such that $e_1 + \cdots + e_{\alpha-1} + 1 \leqq i \leqq e_1 + \cdots + e_\alpha$, and all α, $1 \leqq \alpha \leqq s$.*

Proof. This follows from the known [2] fact that a doubly stochastic matrix contains a diagonal every element of which is $\geqq \Omega$. The result now follows by combining Lemma 1 with (15).

THEOREM 11. *Let G_{ij} denote the geometric mean of the distances from μ_j to the nontrivial eigenvalues of $A(i \mid i)$. Among the G_{ij} for fixed j and variable i, certain G_{ij} will be zero but at least e_j are not zero. Suppose* (*for notational simplicity*) *that $G_{ij} \neq 0$ for $1 \leqq i \leqq m$ and $G_{ij} = 0$ for $i > m$. Then*

$$e_j/nG_j \leqq e_j/mG_j \leqq \left(\textstyle\sum_{i=1}^m G_{ij}\right)/m \leqq (e_j/m)^{1/(s-1)}G_j \leqq G_j.$$

Proof. We may write (16) as

$$\textstyle\sum_{i=1}^m (G_{ij}/G_j)^{s-1} = e_j.$$

Because $0 < G_{ij}/G_j \leqq 1$, the left side of the sum is increased by removing the exponent $s - 1$. This gives the lower bound. The upper bound is obtained by using the fact that the function x^{s-1} is concave up. Many other inequalities of this nature can be proved. We do not pursue the matter further, however.

5. Hermitian matrices

In §5, A is assumed to be Hermitian. Recall that $\mu_1 < \mu_2 < \cdots < \mu_s$ so that $\lambda_1 \leqq \lambda_2 \leqq \lambda_3 \leqq \cdots \leqq \lambda_{n-1} \leqq \lambda_n$.

Let h be a real linear function. Then by Theorem 2(ii), for fixed $\omega \in Q_{nk}$, $\max_U h(A[\omega \mid \omega]) = \max_\tau h(\Lambda_\tau)$. It is possible to say a little more about the sequence $\tau \in Q_{nk}$ for which $h(\Lambda_\tau)$ is maximal.

THEOREM 12. *Let h be a real linear function of k variables, let $\omega \in Q_{nk}$ be fixed. Then*

$$\max_U h(A[\omega \mid \omega]) = \max_{0 \leqq t \leqq k} h(\Lambda_{\delta(t)})$$

where

$$\delta(t) = \{1, 2, \cdots, t, n - k + t + 1, n - k + t + 2, \cdots, n\} \in Q_{nk}.$$

A similar result holds for the min.

That is, the maximizing element $\tau \in Q_{nk}$ consists of the t smallest and $k - t$ largest integers between 1 and n for some t. This is an extension of a result in [3]. (The initial, or terminal, segments of $\delta(t)$ are absent if $t = 0$, or k.)

Proof. For fixed i and fixed $x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_k$,

$$h(E_1(x_1, \cdots, x_k), \cdots, E_k(x_1, \cdots, x_k))$$

is a linear function of x_i. Suppose

$$\delta = \{1, \cdots, j - 1, \cdots, p + 1, \cdots, n\}$$

is the element of Q_{nk} for which $h(\Lambda_\delta) = \max_\tau h(\Lambda_\tau)$. In δ we suppose j, p are respectively the smallest, largest integers for which $j, p \notin \delta$. We show that if an integer $g \in \delta$ exists for which $j < g < p$ then we may increase the length of either the initial or terminal segment in δ, without decreasing the length of the other segment, and retaining the maximal property. A finite number of repetitions of this produces the result. Now $h(\Lambda_\delta) = \lambda_g \alpha + \beta$ where α, β are real numbers not depending on λ_g, λ_j, or λ_p. If $\alpha = 0$ or if $\alpha \neq 0$ but $\lambda_g = \lambda_j$ then we keep the maximal property if, in δ, we delete g and insert j. If $\alpha \neq 0$ but $\lambda_g = \lambda_p$ then we keep the maximal property if, in δ, we delete g and insert p. If $\alpha \neq 0$ but $\lambda_j < \lambda_g < \lambda_p$ then deleting g from δ and inserting either j or p increases $h(\Lambda_\delta)$. This contradicts the maximal property of δ.

We arrange the nontrivial eigenvalues of $A(i \mid i)$ in increasing order. Then the well known [1] fact that

$$(18) \qquad \mu_1 \leqq \xi_{i1} \leqq \mu_2 \leqq \xi_{i2} \leqq \cdots \leqq \mu_{s-1} \leqq \xi_{i,\, s-1} \leqq \mu_s$$

follows from (13) by a simple graphical argument. Conversely, for fixed i, given arbitrary real numbers $\xi_{i1}, \cdots, \xi_{i,\, s-1}$ satisfying (18) we can find unitary U such that the nontrivial eigenvalues of $A(i \mid i)$ are $\xi_{i1}, \cdots, \xi_{i,\, s-1}$. This follows from the observation that if

$$\varphi_1(\lambda) = \prod_{j=1}^{s-1} (\lambda - \xi_{ij}), \qquad \varphi_2(\lambda) = \prod_{j=1}^{s} (\lambda - \mu_j),$$

then [6],

$$\sum_{\alpha=1}^{s} \varphi_1(\mu_\alpha)\varphi_2'(\mu_\alpha)^{-1} = 1.$$

Moreover $\varphi_1(\mu_\alpha)\varphi_2'(\mu_\alpha)^{-1} \geqq 0$. If we put $\theta_{i\alpha} = \varphi_1(\mu_\alpha)\varphi_2'(\mu_\alpha)^{-1}$ for $1 \leqq \alpha \leqq s$ then (17) holds. Now use (14) to construct row i of a unitary matrix. For any unitary U with this row i, (13) is valid with $\theta_{i\alpha}$ as just defined. Let

$$\varphi(\lambda) = f_{(i)}(\lambda)\prod_{\alpha=1}^{s}(\lambda - \mu_\alpha)^{-(e_\alpha-1)} = \sum_{\beta=1}^{s}\theta_{i\beta}\prod_{\alpha=1,\ \alpha\neq\beta}^{s}(\lambda - \mu_\alpha).$$

We have to prove that $\varphi(\lambda) = \varphi_1(\lambda)$. But $\varphi(\mu_\alpha) = \varphi_1(\mu_\alpha)$ for $1 \leqq \alpha \leqq s$ and

$$\text{degree}\ \varphi(\lambda) = \text{degree}\ \varphi_1(\lambda) = s - 1.$$

Thus we have given a new proof of the following well known theorem [1].

THEOREM 13. *Let i be fixed. The inequalities* (18) *are necessary and sufficient for the existence of a unitary U such that $A(i \mid i)$ has $\xi_{i1}, \cdots, \xi_{i,\,s-1}$ as its nontrivial eigenvalues.*

It is well known [1] (and easily follows) that for a given $\omega \in Q_{nk}$, $\lambda_j \leqq \eta_{\omega j} \leqq \lambda_{n-k+j}$ for $1 \leqq j \leqq k$.

LEMMA 2.

$$\theta_{i\alpha} \leqq (\mu_\alpha - \xi_{i\beta})(\mu_\alpha - \mu_\beta)^{-1} \quad \text{if} \quad 1 \leqq \beta \leqq \alpha - 1; \tag{19}$$

$$\theta_{i\alpha} \leqq (\xi_{i\beta} - \mu_\alpha)(\mu_{\beta+1} - \mu_\alpha)^{-1} \quad \text{if} \quad \alpha \leqq \beta \leqq s - 1; \tag{20}$$

$$\begin{aligned} &\{(\mu_\alpha - \xi_{i,\,\alpha-1})(\mu_\alpha - \mu_1)^{-1}\}\{(\xi_{i\alpha} - \mu_\alpha)(\mu_s - \mu_\alpha)^{-1}\} \\ &\qquad \leqq \theta_{i\alpha} \\ &\qquad \leqq \{(\mu_\alpha - \xi_{i,\,\alpha-1})(\mu_\alpha - \mu_{\alpha-1})^{-1}\}\{(\xi_{i\alpha} - \mu_\alpha)(\mu_{\alpha+1} - \mu_\alpha)^{-1}\} \\ &\qquad\qquad\qquad\qquad \text{if} \quad \alpha \neq 1, s. \end{aligned} \tag{21}$$

Proof. We write

$$\theta_{i\alpha} = \prod_{\beta=1}^{\alpha-1}\{(\mu_\alpha - \xi_{i\beta})(\mu_\alpha - \mu_\beta)^{-1}\}\prod_{\beta=\alpha}^{s-1}\{(\xi_{i\beta} - \mu_\alpha)(\mu_{\beta+1} - \mu_\alpha)^{-1}\}. \tag{22}$$

Because of (18), each of the bracketed fractions in (22) is between 0 and 1. Hence dropping some of the fractions increases the value of the expression. This proves (19), (20), and half of (21). We now write

$$\begin{aligned} \theta_{i\alpha} = {} & [(\mu_\alpha - \xi_{i,\,\alpha-1})(\xi_{i\alpha} - \mu_\alpha)(\mu_\alpha - \mu_1)^{-1}(\mu_s - \mu_\alpha)^{-1}] \\ & \cdot \prod_{\beta=1}^{\alpha-2}\{(\mu_\alpha - \xi_{i\beta})(\mu_\alpha - \mu_{\beta+1})^{-1}\} \\ & \cdot \prod_{\beta=\alpha+1}^{s-1}\{(\xi_{i\beta} - \mu_\alpha)(\mu_\beta - \mu_\alpha)^{-1}\}. \end{aligned} \tag{23}$$

By (18), each of the fractions in $\{\ \ \}$ braces in (23) is $\geqq 1$. This proves the other half of (21).

Notation.

$$_{\alpha}A_{\alpha+1} = n^{-1}\sum_{i=1}^{n}\xi_{i\alpha}, \qquad 1 \leqq \alpha < s.$$

That is, $_{\alpha}A_{\alpha+1}$ is the arithmetic mean of the nontrivial eigenvalues of the $A(i \mid i)$ belonging to the interval $[\mu_\alpha, \mu_{\alpha+1}]$.

THEOREM 14. *For* $1 \leqq \beta < s$,

$$
\begin{aligned}
&(n-1)n^{-1}\mu_\beta + n^{-1}\mu_{\beta+1} \\
&\qquad \leqq \mu_{\beta+1} - \min_{\alpha:\alpha\leqq\beta}(n-e_\alpha)n^{-1}(\mu_{\beta+1}-\mu_\alpha) \\
&\qquad \leqq {}_\beta A_{\beta+1} \\
&\qquad \leqq \mu_\beta + \min_{\alpha:\alpha>\beta}(n-e_\alpha)n^{-1}(\mu_\alpha-\mu_\beta) \\
&\qquad \leqq n^{-1}\mu_\beta + (n-1)n^{-1}\mu_{\beta+1}\,.
\end{aligned}
\tag{24}
$$

If $\alpha \neq 1, s$,

$$
\begin{aligned}
e_\alpha(\mu_\alpha-\mu_{\alpha-1})(\mu_{\alpha+1}-\mu_\alpha) &\leqq \textstyle\sum_{i=1}^n(\mu_\alpha-\xi_{i,\,\alpha-1})(\xi_{i\alpha}-\mu_\alpha) \\
&\leqq e_\alpha(u_\alpha-\mu_1)(\mu_s-\mu_\alpha).
\end{aligned}
\tag{25}
$$

For any α,

$$
\begin{aligned}
&\textstyle\sum_{\beta=1}^{\alpha-1}({}_\beta A_{\beta+1}-\mu_\beta)(\mu_{\beta+1}-\mu_\beta)^{-1} + \sum_{\beta=\alpha+1}^{s}(\mu_\beta - {}_{\beta-1}A_\beta)(\mu_\beta-\mu_{\beta-1})^{-1} \\
&\qquad\qquad \geqq (n-e_\alpha)n^{-1}.
\end{aligned}
\tag{26}
$$

(*Empty sums are defined to be zero.*)

Proof. (24) follows from (19), (20), (16), using

$$(\mu_\alpha-\xi_{i\beta})(\mu_\alpha-\mu_\beta)^{-1} = 1-(\xi_{i\beta}-\mu_\beta)(\mu_\alpha-\mu_\beta)^{-1}$$

and

$$(\xi_{i\beta}-\mu_\alpha)(\mu_{\beta+1}-\mu_\alpha)^{-1} = 1-(\mu_{\beta+1}-\xi_{i\beta})(\mu_{\beta+1}-\mu_\alpha)^{-1}.$$

(25) follows immediately from (21). To get (26) use

$$\textstyle\sum_{i=1}^n(\theta_{i1}+\cdots+\theta_{i,\,\alpha-1}+\theta_{i,\,\alpha+1}+\cdots+\theta_{is}) = n-e_\alpha$$

in combination with (19) and (20).

THEOREM 15. *Given* α, *there exist integers* i, j ($i \neq j$), *depending on* α *such that* (27), (28), (29) *all hold.*

$$\xi_{i,\,\alpha-1} \leqq e_\alpha n^{-1}\mu_{\alpha-1} + (n-e_\alpha)n^{-1}\mu_\alpha \quad \textit{if} \quad \alpha\neq 1; \tag{27}$$

$$\xi_{i\alpha} \geqq e_\alpha n^{-1}\mu_{\alpha+1} + (n-e_\alpha)n^{-1}\mu_\alpha\,, \quad \textit{if} \quad \alpha\neq s; \tag{28}$$

$$
\begin{aligned}
(u_\alpha-\xi_{j,\,\alpha-1})(\xi_{j\alpha}-\mu_\alpha) &\leqq e_\alpha n^{-1}(\mu_\alpha-\mu_1)(\mu_s-\mu_\alpha) \\
&\qquad \textit{if} \quad \alpha \neq 1, s.
\end{aligned}
\tag{29}
$$

Proof. From (16), $\theta_{i\alpha} \geqq e_\alpha n^{-1}$ for at least one i. Then from (16) again,

$$\theta_{j\alpha} \leqq (n-1)^{-1}(e_\alpha-\theta_{i\alpha}) \leqq e_\alpha n^{-1}$$

for at least one $j \neq i$. The proof is now completed by use of (19), (20), (21).

We now obtain estimates for the average value of the $\eta_{\omega j}$ as ω runs over Q_{nk}, j fixed.

THEOREM 16. *For fixed j and k, $1 \leqq k \leqq n-1$, $1 \leqq j \leqq k$,*

$$\sum_{r=0}^{n-k} \Psi_r \lambda_{n-k+j-r} \leqq \binom{n}{k}^{-1} \sum_{\omega \in Q_{nk}} \eta_{\omega j} \leqq \sum_{r=0}^{n-k} \Psi_r \lambda_{j+r}, \tag{30}$$

where

$$\Psi_r = E_r(n-1, n-2, \cdots, k)\{\textstyle\prod_{i=k+1}^{n} i\}^{-1}, \quad 0 \leqq r \leqq n-k, \tag{31}$$

$$\textstyle\sum_{r=0}^{n-k} \Psi_r = 1. \tag{32}$$

Remark. We remind the reader that the $\eta_{\omega j}$ are in increasing order for fixed $\omega \in Q_{nk}$, so that $\lambda_j \leqq \eta_{\omega j} \leqq \lambda_{n-k+j}$. Hence the average of the $\eta_{\omega j}$ for fixed j as ω runs over Q_{nk} lies between λ_j and λ_{n-k+j}. In Theorem 16 we obtain convex combinations of $\lambda_j, \lambda_{j+1}, \cdots, \lambda_{n-k+j}$ which are upper and lower bounds for this average.

Proof. We may express (24) in the form

$$n^{-1}((n-1)\lambda_j + \lambda_{j+1}) \leqq n^{-1} \textstyle\sum_{\omega \in Q_{n,n-1}} \eta_{\omega j} \leqq n^{-1}(\lambda_j + (n-1)\lambda_{j+1}). \tag{33}$$

Hence (30) is true when $k = n-1$. Suppose the result established for $k+1$. Then we have

$$\sum_{r=0}^{n-(k+1)} \phi_r \lambda_{n-(k+1)+j-r} \leqq \binom{n}{k+1}^{-1} \sum_{\tau \in Q_{n,k+1}} \eta_{\tau j} \leqq \textstyle\sum_{r=0}^{n-(k+1)} \phi_r \lambda_{j+r}, \quad 1 \leq j \leq k+1, \tag{34}$$

with

$$\phi_r = E_r(n-1, \cdots, k+1)\{\textstyle\prod_{i=k+2}^{n} i\}^{-1}, \quad 0 \leqq r \leqq n-k-1. \tag{35}$$

Now, for a given $\tau \in Q_{n,\, k+1}$ there exist exactly $k+1$ sequences $\omega \in Q_{nk}$ for which $\omega \subset \tau$. So by using (33) for $(k+1) \times (k+1)$ Hermitian matrices,

$$(k+1)^{-1}(k\eta_{\tau j} + \eta_{\tau,\, j+1}) \leqq (k+1)^{-1} \textstyle\sum_{\omega \in Q_{nk},\, \omega \subset \tau} \eta_{\omega j} \leqq (k+1)^{-1}(\eta_{\tau j} + k\eta_{\tau,\, j+1}). \tag{36}$$

We sum (36) over all sequences $\tau \in Q_{n,\, k+1}$, and then divide by $\binom{n}{k+1}$. The number of times a given $\eta_{\omega j}$ will appear in the central member of the resulting inequality (call it $*$) is just the number of $\tau \in Q_{n,\, k+1}$ for which $\omega \subset \tau$; that is, exactly $(n-k)$ times. Now

$$\binom{n}{k+1}(k+1)(n-k)^{-1} = \binom{n}{k}.$$

We use (34) for j and $j+1$ on the left and right sums in our inequality *.

We then obtain (30) on recognizing that

$$\Psi_0 = \phi_0(k+1)^{-1}, \quad \Psi_{n-k} = k\phi_{n-k-1}(k+1)^{-1}, \quad \Psi_r = (k\phi_{r-1} + \phi_r)(k+1)^{-1}$$

for $1 \leqq r < n - k$. That (32) holds follows immediately by setting $\lambda = 1$ in the polynomial identity

$$\prod_{j=k}^{n-1} (\lambda + j) = \sum_{r=0}^{n-k} E_r(n-1, n-2, \cdots, k)\lambda^{n-k-r}.$$

The proof is complete.

The author wishes to thank the referee for pointing out that Theorem 3 holds for diagonable as well as normal matrices.

References

1. L. Mirsky, *Matrices with prescribed characteristic roots and diagonal elements*, J. London Math. Soc., vol. 33 (1958), pp. 14–21.
2. M. Marcus and R. Ree, *Diagonals of doubly stochastic matrices*, Quart. J. Math. (2), vol. 10 (1959), pp. 296–302.
3. M. Marcus, B. N. Moyls, and R. Westwick, *Extremal properties of Hermitian matrices II*, Canad. J. Math., vol. II (1959), pp. 379–382.
4. A. Ostrowski, *Sur quelques applications des fonctions convexes et concaves au sens de I. Schur*, J. Math. Pures Appl. (9), vol. 31 (1952), pp. 253–292.
5. R. C. Thompson, *A note on normal matrices*, Canad. J. Math., vol. 15 (1963), pp. 220–225.
6. Problem E 1623, Amer. Math. Monthly, vol. 71 (1964), pp. 682–683.

The University of California
Santa Barbara, California