THE SUM OF THE ELEMENTS OF THE POWERS OF A MATRIX

MARVIN MARCUS AND MORRIS NEWMAN

1. Introduction and results. In the first two sections of this paper A will be assumed to be an irreducible nonnegative n-square matrix; $A \geqq 0$. Let $s_k = s_k(A)$ denote the sum of the entries in the matrix A^k, where k is a positive integer. The problem considered in the first section is the convergence of the ratio s_k/s_{k-1} as $k \to \infty$. In § 3 we obtain an inequality relating the s_k for various k in the case A is a Hermitian matrix and in § 4 we discuss convexity properties of s_2/s_1.

Let λ_1 be the dominant positive characteristic root of A which can be taken as 1 for the purposes of our subsequent arguments. If h is the number of charcteristic roots of A of modulus 1, then they are the roots of $\lambda^h - 1 = 0$ and are all simple [3]. Let $\varepsilon = e^{2\pi i/h}$ so that $1, \varepsilon, \varepsilon^2, \cdots, \varepsilon^{h-1}$ are the roots of modulus 1. Choose permutation matrices P and Q so that

$$PAP^T = \begin{bmatrix} 0 & A_1 & & & 0 \\ & 0 & A_2 & & \\ & & \ddots & \ddots & \\ & & & & A_{h-1} \\ A_h & & & & 0 \end{bmatrix} \tag{1}$$

and

$$QA^TQ^T = \begin{bmatrix} 0 & B_1 & & & 0 \\ & 0 & B_2 & & \\ & & \ddots & \ddots & \\ & & & & B_{h-1} \\ B_h & & & & 0 \end{bmatrix} \tag{2}$$

where the zero blocks down the main diagonal in both (1) and (2) are square. We shall asume henceforth that A is in this *Frobenius normal form*. In other words we assume A is already in the form given on the right in (1). Let $u_1, \cdots, u_h$ and $v_1, \cdots, v_h$ be the characteristic vectors of A and A^T corresponding to $1, \varepsilon, \cdots, \varepsilon^{h-1}$ respectively. We write for the *maximal characteristic vector*

$$u_1 = z_1 \dot{+} \cdots \dot{+} z_h \,, \tag{3}$$

Received September 27, 1961. This work was supported in part by the Office of Naval Research. The authors wish to express their thanks to K. Goldberg for many valuable suggestions.

where $\dot{+}$ indicates direct sum and the partitioning in (3) is conformal to the partitioning into main diagonal blocks in (1). That is, if the square blocks down the main diagonal in (1) are of sizes $n_1, \cdots, n_h$ then z_t has coordinates numbered $n_{t-1}+1, \cdots, n_t (n_0 = 0)$ all positive, the rest zero. Thus no two of the z_t have positive coordinates in the same position and the direct sum notation is appropriate.

Since Qv_1 is the maximal characteristic vector of QA^TQ^T we can partition Qv_1 into a direct sum exactly as was done with u_1: $Qv_1 = m_1 \dot{+} \cdots \dot{+} m_h$. Then $v_1 = Q^Tm_1 \dot{+} \cdots \dot{+} Q^Tm_h$ and we set $Q^Tm_t = w_t$, $t = 1, \cdots, h$, to obtain

$$v_1 = w_1 \dot{+} \cdots \dot{+} w_h \,. \tag{4}$$

Let $r = r(A)$ and $c = c(A)$ be the n-tuples of row and column sums of A respectively. Partition r and c conformally with v_1 and u_1 respectively as in (4) and (3):

$$r = r_1 \dot{+} \cdots \dot{+} r_h$$
$$c = c_1 \dot{+} \cdots \dot{+} c_h \,.$$

The notations (z_i, c_i) and (w_i, r_i) will be used for the ordinary euclidean inner product.

Our main result is in terms of these inner products.

THEOREM 1. *$\lim_{k\to\infty} s_k/s_{k-1}$ exists and is equal to the dominant characteristic root if and only if the numbers*

$$\sum_{i=1}^{h} (z_i, c_i)(w_{\alpha-i+1}, r_{\alpha-i+1}) \tag{5}$$

are all equal for $\alpha = 1, \cdots, h$.

We remark that the indices in (5) are to be reduced modulo h.

In case A is symmetric then the roots of modulus 1 can be only 1 or -1. Thus $h = 1$ or 2. In case $h = 1$ (A is *primitive*) then Theorem 1 automatically holds since there is only one item (5). In case $h = 2$ we have.

THEOREM 2. *If*

$$A = \begin{bmatrix} O_p & C \\ C^T & O_q \end{bmatrix}$$

is irreducible and has maximal characteristic vector

$$u_1 = z_1 \dot{+} z_2 = (a_1, \cdots, a_p) \dot{+} (b_1, \cdots, b_q)$$

then a necessary and sufficient condition that $\lim_{k\to\infty} s_k/s_{k-1}$ *exist and equal the dominant characteristic root is that*

$$\sum_{i=1}^{p} a_i = \sum_{i=1}^{q} b_i \ . \tag{6}$$

In certain cases more can be said about the convergence of s_k/s_{k-1}.

THEOREM 3. *If* A *is a positive semi-definite symmetric irreducible nonnegative matrix then* s_k/s_{k-1} *approaches the dominant positive characteristic root of* A. *Moreover,*

$$s_k/s_{k-1} \leqq s_{k+1}/s_k \tag{7}$$

with equality if and only if $r(A^{(k-1)/2})$ *and* $r(A^{(k+1)/2})$ *are linearly dependent.*

An immediate consequence of the above results is the

COROLLARY. *If* A *is a nonnegative symmetric matrix and* A^2 *is irreducible then* $\sqrt{s_{2k}/s_{2k-2}}$ *converges to the dominant characteristic root monotonically.*

In general $\lim_{k\to\infty} s_k/s_{k-1}$ may exist without the convergence being monotone, e.g. $A = \begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix}$. Then $A^2 = 2A + 3I$,

$$s_{k+1} = 2s_k + 3s_{k-1}$$

and hence $s_k = (9/4)3^k - (1/4)(-1)^k$,

$$s_k/s_{k-1} = (3^{k+2} + (-1)^{k+1})/(3^{k+1} + (-1)^k) \ ,$$

which converges to 3 but not monotonically.

2. Proofs. Choose S to bring A to Jordan normal form

$$A = S \begin{bmatrix} 1 & 0 & & & & \\ & \varepsilon & & & & \\ & & \ddots & & & \\ 0 & & & \varepsilon^{h-1} & & \\ & & & & \lambda_{h+1} & \\ & & & & & \ddots \\ & & & & & & \lambda_n \end{bmatrix} S^{-1} \ . \tag{8}$$

Let E_{ij} be the n-square matrix whose single nonzero entry is a 1 in position (i, j) and set $J = \sum_{i,j} E_{ij}$. Then

$$s_k = tr(JA^k) = tr\Big(JS\Big(\sum_{t=1}^{h}\varepsilon^{(t-1)k}E_{tt}\Big)S^{-1}\Big) + \delta_k \,. \tag{9}$$

Since $|\lambda_t| < 1$ for $t \geqq h+1$ we conclude that $\lim_{k\to\infty}\delta_k = 0$. If x is an n-vector it will be convenient to denote by $\sigma(x)$ the sum of the coordinates of x. Moreover, S_t and S^t will designate the tth row and column of the n-square matrix S respectively. Let $\mu_t = (S^{-1}JS)_{tt}$, $t = 1, \cdots, h$ and note that

$$\begin{aligned} s_k - \delta_k &= \sum_{t=1}^{h} tr(E_{tt}S^{-1}JS)\varepsilon^{(t-1)k} \\ &= \sum_{t=1}^{h}(S^{-1}JS)_{tt}\varepsilon^{(t-1)k} \\ &= \sum_{t=1}^{h}\mu_t\varepsilon^{(t-1)k} \,, \end{aligned} \tag{10}$$

and

$$\begin{aligned} \mu_t &= \sum_{\alpha,\beta=1}^{n}(S^{-1})_{t\alpha}J_{\alpha\beta}S_{\beta t} \\ &= \sum_{\alpha,\beta}(S^{-1})_{t\alpha}S_{\beta t} \\ &= \sigma(S_t^{-1})\sigma(S^t) \,, \qquad t = 1, \cdots, h \,. \end{aligned} \tag{11}$$

From (8) we have

$$AS^t = \varepsilon^{t-1}S^t \,, \qquad t = 1, \cdots, h$$

and since the dimension of the null space of $A - \varepsilon^{t-1}I$ is 1 for $t = 1, \cdots, h$ we conclude that

$$S^t = c_t u_t \,, \qquad t = 1, \cdots, h \tag{12}$$

for appropriate nonzero scalars c_t. Similarly

$$S_t^{-1} = d_t v_t \,, \qquad t = 1, \cdots, h \,. \tag{13}$$

From (11), (12) and (13) we have, for e the n-tuple all of whose coordinates is 1,

$$\begin{aligned} \mu_t &= \sigma(S_t^{-1})\sigma(S_t) = d_t c_t \sigma(v_t)\sigma(u_t) \\ &= d_t c_t (v_t, e)(u_t, e) \\ &= d_t c_t (A^T(v_t/\varepsilon^{t-1}), e)(A(u_t/\varepsilon^{t-1}), e) \\ &= d_t c_t \varepsilon^{2(1-t)}(v_t, Ae)(u_t, A^T e) \\ &= d_t c_t \varepsilon^{2(1-t)}(v_t, r)(u_t, c) \,. \end{aligned} \tag{14}$$

The vectors u_t and v_t, $t = 2, \cdots, h$, have explicit representations in terms of u_1, v_1 and ε as follows:

$$u_t = z_1 \dot{+} \varepsilon^{t-1} z_2 \dot{+} \cdots \dot{+} \varepsilon^{(t-1)(h-1)} z_h$$

$$v_t = w_1 \dot{+} \varepsilon^{(t-1)} w_2 \dot{+} \cdots \dot{+} \varepsilon^{(t-1)(h-1)} w_h \qquad t = 2, \cdots, h .$$

Let $k_t = d_t c_t \varepsilon^{2(1-t)} \neq 0$, $\zeta_i = (z_i, c_i)$, $\eta_i = (w_i, r_i)$, $i = \cdots, h$ and we compute from (14) and the fact that $\varepsilon^{(t-1)h} = 1$ that

$$\begin{aligned} \mu_t &= k_t(v_t, r)(u_t, c) \\ &= k_t \sum_{i-1}^{h} (w_i, r_i)\varepsilon^{(t-1)(i-1)} \sum_{i=1}^{h} (z_i, c_i)\varepsilon^{(t-1)(i-1)} \\ &= k_t \sum_{\alpha=1}^{h} \left(\sum_{i=1}^{h} \zeta_i \eta_{\alpha-i+1} \right) \varepsilon^{(t-1)(\alpha-1)} , \end{aligned} \tag{15}$$

where the subscripts are always reduced modulo h.

LEMMA 1. *$\mu_2 = \cdots = \mu_h = 0$ if and only if the sums $\sum_{i=1}^{h} \zeta_i \eta_{\alpha-i+1}$ are all equal for $\alpha = 1, \cdots, h$.*

Proof. Set $f_\alpha = \sum_{i=1}^{h} \zeta_i \eta_{\alpha-i+1}$ and from (15) the conditions $\mu_t = 0$, $t = 2, \cdots, h$ are equivalent to the system of linear equations

$$\sum_{\alpha=1}^{h} f_\alpha \varepsilon^{(\alpha-1)(t-1)} = 0 \qquad t = 2, \cdots, h . \tag{16}$$

Since $\sum_{\alpha=1}^{h} \varepsilon^{(\alpha-1)(t-1)} = 0$ $(1 \leqq t - 1 < h)$, each of the equations (16) has the solution $f_1 = \cdots = f_h$. On the other hand, the $(h-1)$-square submatrix of coefficients in (16) obtained by deleting the first column in the coefficient matrix has as its determinant the Vandermonde of $\varepsilon, \varepsilon^2, \cdots, \varepsilon^{h-2}$ to within a nonzero constant multiple. Thus the system (16) has rank $h - 1$ and $f_1 = \cdots = f_h$ is the *only* solution of (16). The proof of Theorem 1 will then be complete if we establish

LEMMA 2. *$\lim_{k\to\infty} s_k/s_{k-1}$ exists if and only if $\mu_2 = \cdots = \mu_h = 0$. If it exists it has value* 1.

Proof. From (10) we have

$$s_k/s_{k-1} = \left(\sum_{t=1}^{h} \mu_t \varepsilon^{(t-1)k} + \delta_k \right) \Big/ \left(\sum_{t=1}^{h} \mu_t \varepsilon^{(t-1)(k-1)} + \delta_{k-1} \right) ,$$

and since $\lim_{k\to\infty} \delta_k = 0$, $\lim_{k\to\infty} s_k/s_{k-1}$ exists if and only if

$$g_k = m_k/m_{k-1} = \sum_{t=1}^{h} \mu_t \varepsilon^{(t-1)k} \Big/ \sum_{t=1}^{h} \mu_t \varepsilon^{(t-1)(k-1)}$$

approaches a limit. Note that m_k is periodic of period h. Also $\mu_1 = c_1 d_1 \sigma(u_1)\sigma(v_1) \neq 0$ follows from (3) and (4) and so the condition is clearly sufficient. Since g_k takes on only a finite number of values it follows

that if g_k approaches a limit l then $g_k = l$ for all k. Moreover if g_k exists,

$$\begin{aligned} g_k &= m_k/m_{k-1} = m_k/m_{k-h-1} \\ &= \prod_{\alpha=1}^{h+1}(m_{k-\alpha+1}/m_{k-\alpha}) \\ &= g_k^{h+1} , \end{aligned}$$

and thus $l = 1$. But then $m_k = m_{k-1}$ and we conclude that

$$\sum_{t=2}^{h}\mu_t(1 - \varepsilon^{(1-t)})\varepsilon^{(t-1)k} = 0 . \tag{17}$$

Letting $k = 0, \cdots, h-2$ successively in (17) and noting that $\prod_{0\leqq i<j\leqq h-2}(\varepsilon^i - \varepsilon^j) \neq 0$ we conclude that $\mu_t(1 - \varepsilon^{1-t}) = 0$, $t = 2, \cdots, h$, and hence that $\mu_2 = \cdots = \mu_h = 0$.

To proceed to the proof of Theorem 2 note that the maximal characteristic vectors of A and $A^T = A$ are given by

$$\begin{aligned} u_1 = v_1 &= z_1 \dot{+} z_2 = (a_1, \cdots, a_p) \dot{+} (b_1, \cdots, b_q) . \\ c_1 &= (\sigma(C_1), \cdots, \sigma(C_p)) \\ c_2 &= (\sigma(C^1), \cdots, \sigma(C^q)), \text{ and} \\ r_1 &= c_1, \ r_2 = c_2 . \end{aligned}$$

The condition that the items (5) be equal for $\alpha = 1, 2$ becomes, in succession,

$$\begin{aligned} (z_1, c_1)(z_2, c_2) + (z_2, c_2)(z_1, c_1) &= (z_1, c_1)^2 + (z_2, c_2)^2 , \\ (z_1, c_1) &= (z_2, c_2) , \\ \sum_{i=1}^{p} a_i\sigma(c_i) &= \sum_{i=1}^{q} b_i\sigma(C^i) , \\ \sigma\Big(\sum_{i=1}^{p} a_i C_i\Big) &= \sigma\Big(\sum_{i=1}^{q} b_i C^i\Big) . \end{aligned} \tag{18}$$

Now $Cb = a$, $C^T a = b$ and hence $a = \sum_{i=1}^{q} b_i C^i$, $b = \sum_{i=1}^{p} a_i C_i$. We then have from (18) that $\sigma(a) = \sigma(b)$ is equivalent to (5) in the case A symmetric and $h = 2$.

The convergence of s_k/s_{k-1} in Theorem 3 is clear since $h = 1$. If A is positive semi-definite and $\alpha \geqq 0$ let A^α be the unique positive semi-definite determination. Then if p and q are nonnegative,

$$\begin{aligned} s_{(p+q)/2}^2 &= (A^{p+q/2}e, e)^2 \\ &= (A^{p/2}e, A^{q/2}e)^2 \\ &\leqq (A^p e, e)(A^q e, e) \\ &= s_p s_q \end{aligned} \tag{19}$$

with equality if and only if $A^{p/2}e$ and $A^{q/2}e$ are linearly dependent. Set $p = k - 1$ and $q = k + 1$ to finish the argument.

The Corollary follows from Theorem 3.

3. The Hermitian case. In this section A is assumed to be an n-square Hermitian matrix with characteristic roots $\lambda_1, \cdots, \lambda_n$. We have

THEOREM 4. *Let p, q, m, t be nonnegative integers and assume that $t = \min(p, q, m, t)$ is even and $p + q = m + t$ is even. Then*

$$s_p s_q \leqq s_m s_t \ . \tag{20}$$

Proof. Let $D = \operatorname{diag}(\lambda_1, \cdots, \lambda_n)$ and choose a unitary matrix U such that $U^*AU = D$. Then

$$s(A) = (1/n)tr(JAJ) = (1/n)tr(JUDU^*J) \ .$$

It is not difficult to compute that each main diagonal element of $JUDU^*J$ is $\sum_{i=1}^n \lambda_i |\sigma(U^i)|^2$. Let $w_i = |\sigma(U^i)|^2$ and then

$$s(A) = \sum_{i=1}^n \lambda_i w_i \ . \tag{21}$$

Replacing A by A^p in (21) yields

$$s_p = \sum_{i=1}^n \lambda_i^p w_i$$

and (20) is equivalent to

$$\sum_{i=1}^n \lambda_i^m w_i \sum_{i=1}^n \lambda_i^t w_i - \sum_{i=1}^n \lambda_i^p w_i \sum_{i=1}^n \lambda_i^q w_i \geqq 0 \ . \tag{22}$$

The left side of (22) becomes, after symmetrizing the sums,

$$\begin{aligned} &(1/2)\sum_{i,j} w_i w_j (\lambda_i \lambda_j)^t [\lambda_j^{m-t} + \lambda_i^{m-t} - \lambda_i^{p-t}\lambda_j^{q-t} - \lambda_i^{q-t}\lambda_j^{p-t}] \\ &= (1|2)\sum_{i,j} w_i w_j (\lambda_i \lambda_j)^t (\lambda_j^{p-t} - \lambda_i^{p-t})(\lambda_j^{q-t} - \lambda_i^{q-t}) \\ &= (1/2)\sum_{i,j} w_i w_j (\lambda_i \lambda_j)^t \lambda_i^{p+q-2t} ((\lambda_j/\lambda_i)^{p-t} - 1)((\lambda_j/\lambda_i)^{q-t} - 1) \ . \end{aligned}$$

Since t and $p + q - 2t$ are even and moreover $p - t \geqq 0$, $q - t \geqq 0$ it follows that $(\lambda_i\lambda_j)^t \lambda_i^{p+q-2t}((\lambda_j/\lambda_i)^{p-t} - 1)((\lambda_j/\lambda_i)^{q-t} - 1) \geqq 0$ for all i, j.

Setting $t = 0$ in Theorem 4 yields the

COROLLARY. *If p and q are nonnegative integers and $p + q$ is even then*

$$s_p s_q \leqq n s_{p+q} \,. \tag{23}$$

we remark that formula (21) doesn't yield $s_0 = \sum_{i=1}^n w_i = n$ unless A is nonsingular. But the singular case follows from the nonsingular one by the standard continuity argument.

In case A has nonnegative entries then a specialization of an inequality in [1] implies that $n^2 s_3 \geqq s_1$. We conjecture that $n s_3 \geqq s_1 s_2$ in analogy with (23).

4. Some remarks on s_2/s_1.

Let $f(t) = f(t_1, \cdots, t_n) = \sum_{1 \leqq i < j \leqq n} t_i t_j / \sum_{i=1}^n t_i$ and note that

$$f(a+b) - f(a) - f(b) = \sum_{i=1}^n \Big(a_i \sum_{j=1}^n b_j - b_i \sum_{j=1}^n a_j \Big)^2 \Big/ 2 \sum_{i=1}^n (a_i + b_i) \sum_{i=1}^n a_i \sum_{i=1}^n b_i \,.$$

It follows that if $\sum_{i=1}^n a_i > 0$ and $\sum_{i=1}^n b_i > 0$ then $f(a+b) \geqq f(a) + f(b)$ with equality if and only if the sets $(a_1, \cdots, a_n)$ and $(b_1, \cdots, b_n)$ are proportional. Define the functions

$$g(t) = \sum_{i=1}^n t_i^2 \Big/ \sum_{i=1}^n t_i \,,$$

$$h(t) = \Big(\sum_{i=1}^n t_i^2 + \sum_{1 \leqq i < j \leqq n} t_i t_j \Big) \Big/ \sum_{i=1}^n t_i \,,$$

and observe that

$$g(t) = \sum_{i=1}^n t_i - 2f(t) \,,$$

$$h(t) = \sum_{i=1}^n t_i - f(t) \,.$$

Then if $\sum_{i=1}^n a_i > 0$, $\sum_{i=1}^n b_i > 0$,

$$\begin{aligned} g(a+b) &= \sum_{i=1}^n a_i + \sum_{i=1}^n b_i - 2f(a+b) \\ &\leqq \sum_{i=1}^n a_i - 2f(a) + \sum_{i=1}^n b_i - 2f(b) \\ &= g(a) + g(b) \,, \end{aligned} \tag{24}$$

and similarly $h(a+b) \leqq h(a) + h(b)$.
Equality holds in the preceding two inequalities if and only if the sets a and b are proportional. From the inequality (24) we can then prove

THEOREM 5. *If A and B are symmetric n-square matrices satisfying $\sum_{i,j=1}^n a_{ij} > 0$, $\sum_{i,j=1}^n b_{ij} > 0$ then*

$$s_2(A+B)/s_1(A+B) \leqq s_2(A)/s_1(A) + s_2(B)/s_1(B) \tag{25}$$

with equality if and only if $r(A)$ and $r(B)$ are proportional.

Proof. From the formula

$$s_2(A)/s_1(A) = \sum_{i=1}^{n}\sigma(A_i)^2 \Big/ \sum_{i=1}^{n}\sigma(A_i) = g(r(A))$$

we compute that

$$\begin{aligned} s_2(A+B)/s_1(A+B) &= g(r(A+B)) = g(r(A)+r(B)) \\ &\leqq g(r(A)) + g(r(B)) \\ &= s_2(A)/s_1(A) + s_2(B)/s_1(B)\ . \end{aligned}$$

A similar result can be formulated for the function h. It might be conjectured that a convexity result like (25) is true for the functions $s_r(A)/s_{r-1}(A)$, $r > 2$. This is not the case: take

$a_1 = \frac{1}{3}$, $a_2 = \cdots = a_{n-1} = 0$, $a_n = 1$, $b_1 = \cdots = b_{n-1} = 0$, $b_n = 1$ and observe that

$$\sum_{i=1}^{n}(a_i+b_i)^r \Big/ \sum_{i=1}^{n}(a_i+b_i)^{r-1} = (3^{-r}+2^r)/(3^{-(r-1)}+2^{r-1})$$

whereas

$$\sum_{i=1}^{n}a_i^r \Big/ \sum_{i=1}^{n}a_i^{r-1} + \sum_{i=1}^{n}b_i^r \Big/ \sum_{i=1}^{n}b_i^{r-1} = (3^{-r}+1)(3^{-(r-1)}+1)+1$$

and it is simple to check that

$$(3^{-r}+2^r)/(3^{-(r-1)}+2^{r-1}) > (3^{-r}+1)/(3^{-(r-1)}+1)+1$$

for $r \geqq 3$.

The referee suggests that the arguments of the paper could be rephrased in terms of the vector e, where $e^T = (1, 1, \cdots, 1)$. Thus $s_k = e^T A^R e$, $J = ee^T$, $\sigma(x) = e^T x$, etc. He also notes that e could be replaced by any other positive vector with conclusions similar to those obtained in the paper. We have not thought it advisable to pursue the matter further.

REFERENCES

1. F. V. Atkinson, G. A. Watterson, P. A. P. Moran, *A matrix inequality*, Quart J. Math., Oxford (2), **11**, (1960), 137-140.
2. G. H. Hardy, J. E. Littlewood, G. Pólya, *Inequalities* (Cambridge, Second Edition, 1952).
3. H. Wielandt, *Unzerlegbare, nicht negative Matrizen*, Math Zeit. **52** (1950), 642-648.

THE UNIVERSITY OF CALIFORNIA, SANTA BARBARA AND
THE UNITED STATES NATIONAL BUREAU OF STANDARDS