JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 65B, No. 2, April–June 1961

Comparison Theorems for Symmetric Functions of Characteristic Roots[1]

Marvin Marcus

(February 10, 1961)

Several theorems are proved that give necessary and sufficient conditions for $A-B$ to be positive semidefinite Hermitian. The conditions are in terms of the comparison of elementary symmetric functions of the characteristic roots of $A+X$ and $B+X$ as X varies over positive definite Hermitian matrices.

1. Introduction

In a recent paper [4][2] M. Stone obtained the following result in reproving certain theorems of Ehrenfeld [1]: *if F and G are positive definite n-square Hermitian matrices, then*

$$d(I+AF)\geq d(I+AG) \tag{1}$$

for all positive definite n-square Hermitian matrices A if and only if $F-G$ is positive semidefinite Hermitian. Here d denotes determinant and henceforth $A>0$ (≥ 0) will signify that A is positive (positive semidefinite) Hermitian. $F\geq G$ ($F>G$) will mean $F-G\geq 0$ ($F-G>0$). Note that since $A>0$ if and only if $A^{-1}>0$ the condition (1) is the same as saying

$$d(A+F)\geq d(A+G) \tag{2}$$

for all $A>0$. It is in the form (2) that we investigate what happens when we replace d by some other elementary symmetric function of the characteristic roots. To be specific, suppose E_r $(x_1, \ldots, x_n)$ denotes the r^{th} elementary symmetric function of the indicated variables:

$$E_r(x_1, \cdots, x_n)=\sum_{1\leq i_1<\ldots<i_r\leq n}\prod_{t=1}^{r} x_{i_t}.$$

For a fixed r, $1\leq r\leq n$, let $E_r(A)$ denote $E_r(\lambda_1(A), \ldots, \lambda_n(A))$ where $\lambda_i(A)$, $i=1, \ldots, n$, are the characteristic roots of A. If the $\lambda_i(A)$ are real we will choose our notation so that $\lambda_1(A)\geq \ldots \geq \lambda_n(A)$. Then the problem we pose is to find conditions on the characteristic roots of F and G such that

$$E_r(A+F)\geq E_r(A+G) \text{ for all } A\geq 0.$$

We have

THEOREM 1. *Assume $F\geq 0$, $G\geq 0$, and $1\leq r\leq n$. If*

$$E_{r-p}(\lambda_n(F), \ldots, \lambda_{p+1}(F))\geq E_{r-p}(\lambda_1(G), \ldots, \lambda_{n-p}(G)), \tag{3}$$

$$p=0, \ldots, r-1$$

it follows that

$$E_r(A+F)\geq E_r(A+G) \text{ for all } A\geq 0. \tag{4}$$

In case $G=I$, (3) *becomes*

$$E_{r-p}(\lambda_n(F), \cdots, \lambda_{p+1}(F))\geq \binom{n-p}{r-p}, \quad p=0, \cdots, r-1, \tag{5}$$

which is both necessary and sufficient for

$$E_r(A+F)\geq E_r(A+I) \quad \text{for all } A\geq 0. \tag{6}$$

We remark that for $r=n$ the second part of theorem 1 simply becomes: $\lambda_n(F)\geq 1$ *if and only if* $d(A+F)\geq d(A+I)$ *for all* $A\geq 0$. Now in (1) multiply both sides by $d(A^{-1}R^*R)$ where R is a nonsingular matrix satisfying $R^*GR=I$, $R^*FR=K$ and we find that

$$d(R^*A^{-1}R+K)\geq d(R^*A^{-1}R+I).$$

Now as A runs over all positive definite matrices so does $R^*A^{-1}R$ and from the above remark we conclude that $\lambda_n(K)\geq 1$. The characteristic roots of K are just the characeristic roots of $G^{-1}F$ and thus

$$G^{-1/2}FG^{-1/2}\geq I, \qquad F\geq G,$$

and the result in [4] follows. A somewhat stronger multiplicative analogue of theorem 1 is available.

THEOREM 2. *Assume $F\geq 0$, $G\geq 0$, and $1\leq r\leq n$. Then*

$$E_r(AF)\geq E_r(AG) \text{ for all } A\geq 0 \tag{7}$$

[1] This research was supported in part by the Office of Naval Research.
[2] Figures in brackets indicate the literature references at the end for this paper.

if and only if $\prod_{j=1}^{r} \lambda_j(F^{-1}G) \leq 1.$ (8)

In order to prove these results we use some results concerning compound matrices and Grassmann products [2]. We give the coordinate definition of these items and list several of their properties. If $x_1, \ldots, x_r$ are vectors in the unitary space of n-tuples, $x_i=(x_{i1}, \ldots, x_{in})$, $1 \leq r \leq n$, then $x_1 \wedge \ldots \wedge x_r$ is the $\binom{n}{r}$-tuple whose coordinates are the r-square subdeterminants of the r$\times n$ matrix (x_{ij}), $i=1, \ldots, r$, $j=1, \ldots, n$ arranged in lexicographic order. If A is an n-square matrix, then $C_r(A)$ is the $\binom{n}{r}$-square matrix whose entries are the r-square subdeterminants of A arranged in doubly lexicographic order according to the row and column indices of A. That is, if $1 \leq i_1 < \ldots < i_r \leq n$ and $1 \leq j_1 < \ldots < j_r \leq n$ are two increasing sets of r integers then the $(i_1, \ldots, i_r)$, $(j_1, \ldots, j_r)$ element of $C_r(A)$ is the determinant of the matrix $A[i_1, \ldots, i_r | j_1, \ldots, j_r]$ whose (s, t) entry is $a_{i_s j_t}$. If $i_s=j_s$, $s=1, \ldots, r$ we denote the corresponding principal submatrix by $A[i_1, \ldots, i_r]$. If $A \geq 0$ then $C_r(A) \geq 0$; $C_r(A)x_1 \wedge \ldots \wedge x_r = Ax \wedge \ldots \wedge Ax_r$; if (,) is the usual unitary inner product in the space of m-tuples then $(x_1 \wedge \ldots \wedge x_r, y_1 \wedge \ldots \wedge y_r) = d\{(x_i, y_j)\}$, $i, j=1, \ldots, r$.

We remark that the condition (3) of theorem 1 will not imply that $F \geq G$ in the case $r < n$. For even if $G=I$, we can easily construct F so that $\lambda_n(F) < 1$ and yet (3) holds for $p=0, \ldots, r-1$. However, there is a direct generalization of Stone's result to the compound matrix.

THEOREM 3. *If* $F > 0$, $G > 0$ *and* $1 \leq r \leq n$, *then*

$$C_r(I+AF) \geq C_r(I+AG) \text{ for all } A>0$$

$$\text{if and only if } F \geq G. \quad (9)$$

Once again, $r=n$ is precisely the result in [4].

2. Proofs

We prove theorem 1. Let $A=UXU^*$, U unitary, $X=\text{diag}(x_1, \ldots, x_n)$. Then $E_r(A+F)-E_r(A+G) = E_r(U^*AU+U^*FU)-E_r(U^*AU+U^*GU) = E_r(X+K)-E_r(X+H)$ where $K=U^*FU \geq 0$, $H=U^*GU \geq 0$. Let

$$\varphi(x_1, \ldots, x_n: U)=E_r(X+K)-E_r(X+H).$$

Then

$$\varphi(x_1, \ldots, x_n: U)=\sum_{1\leq\alpha_1<\ldots<\alpha_r\leq n} \{d[\text{diag}(x_{\alpha_1}, \ldots, x_{\alpha_r}) + K[\alpha_1, \ldots, \alpha_r]]-d[\text{diag}(x_{\alpha_1}, \ldots, x_{\alpha_r}) + H[\alpha_1, \ldots, \alpha_r]]\}. \quad (10)$$

We see from (10) that φ is a polynomial in $x_1, \ldots, x_n$ of degree r in which the nonconstant terms form a multilinear polynomial. Let

$$\varphi_{i_1, \ldots i_p}(x_1, \ldots, x_n: U)=\frac{\partial^p \varphi}{\partial x_{i_1} \ldots \partial x_{i_p}}(x_1, \ldots, x_n: U)$$

and

$$\overset{\circ}{\varphi}_{i_1 \ldots i_p}=\varphi_{i_1 \ldots i_p}(0, \ldots, 0: U).$$

Then

$$\varphi(x_1, \ldots, x_n: U)=\sum_{p=0}^{r} \sum_{1\leq i_1<\ldots<i_p\leq n} x_{i_1} \ldots x_{i_p} \overset{\circ}{\varphi}_{i_1 \ldots i_p}, \quad (11)$$

where the term corresponding to $p=0$ is just $\varphi(0, \ldots, 0: U)$. Let w denote the increasing sequence $i_1 < \ldots < i_p$. Unless w is a subset of $\alpha_1 < \ldots < \alpha_r$ the partial derivative satisfies

$$\frac{\partial^p}{\partial x_{i_p} \ldots \partial x_{j_p}} \{d[\text{diag}(x_{\alpha_1}, \ldots, x_{\alpha_r})+K[\alpha_1 \ldots \alpha_r]] - d[\text{diag}(x_{\alpha_1}, \ldots, x_{\alpha_r})+H[\alpha_1, \ldots, \alpha_r]]\}=0. \quad (12)$$

If w is a subset of $\alpha_1 < \ldots < \alpha_r$ let $\alpha_1^w < \ldots < \alpha_{r-p}^w$ be the ordered complementary set of w in $\alpha_1 < \ldots < \alpha_r$. In this case the derivative in (12) has the value

$$d[\text{diag}(x_{\alpha_1^w}, \ldots, x_{\alpha_{r-p}^w})+K[\alpha_1^w, \ldots, \alpha_{r-p}^w]] - d[\text{diag}(x_{\alpha_1^w}, \ldots, x_{\alpha_{r-p}^w})+H[\alpha_1^w, \ldots, \alpha_{r-p}^w]]. \quad (13)$$

Setting $x_1= \ldots =x_n=0$ in (13) we have

$$\phi^0_{i_1 \ldots i_p}=\sum_{1\leq\alpha_1<\ldots<\alpha_r\leq n}' \{d(K[\alpha_1^w, \ldots, \alpha_{r-p}^w] - d(H[\alpha_1^w, \ldots, \alpha_{r-p}^w])\}, \quad (14)$$

where $\sum'$ indicates that the summation is taken only over those $\alpha_1 < \ldots < \alpha_r$ containing $w=(i_1 < \ldots < i_p)$ as a subset. Returning to (11) we set $x_{i_1}= \ldots =x_{i_p}=t$ and all other $x_j=0$. Then $\phi(0, \ldots, t, \ldots, t, \ldots 0: U) = t^p \phi^0_{i_1 \ldots i_p}+L_{p-1}(t)$ where L_{p-1} is a polynomial of degree at most $p-1$ in t. Since $\phi(x_1, \ldots, x_n: U) \geq 0$ we conclude, by letting t increase, that $\phi^0_{i_1 \ldots i_p} \geq 0$. Conversely, if $\phi^0_{i_1 \ldots i_p} \geq 0$ and $\phi(0, \ldots, 0, U) \geq 0$, then $\phi(x_1, \ldots, x_n: U) \geq 0$ for all non-negative $x_1, \ldots, x_n$. Let $j_i < \ldots < j_{n-p}$ denote the set complementary to $i_1 < \ldots < i_p$ in $1, \ldots, n$. Let e_j, $j=1, \ldots, n$, be the unit n-tuple with 1 in position j, 0 elsewhere. We may rewrite (14) as

$$\begin{aligned}\phi^0_{i_1 \ldots i_p} &= \sum{}' \{dK[\sigma_1, \ldots, \sigma_{r-p}]-dH[\sigma_1, \ldots, \sigma_{r-p}]\} \\ &= \sum{}' \{(C_{r-p}(K)e_{\sigma_1} \wedge \ldots \wedge e_{\sigma_{r-p}}, e_{\sigma_1} \wedge \ldots \wedge e_{\sigma_{r-p}}) \\ &\quad -(C_{r-p}(H)e_{\sigma_1} \wedge \ldots \wedge e_{\sigma_{r-p}}, e_{\sigma_1} \wedge \ldots \wedge e_{\sigma_{r-p}})\} \\ &= \sum{}' (C_{r-p}(F)u_{\sigma_1} \wedge \ldots \wedge u_{\sigma_{r-p}}, u_{\sigma_1} \wedge \ldots \wedge u_{\sigma_{r-p}}) \\ &\quad - \sum{}' (C_{r-p}(G)u_{\sigma_1} \wedge \ldots \wedge u_{\sigma_{r-p}}, \\ &\qquad u_{\sigma_1} \wedge \ldots \wedge u_{\sigma_{r-p}}),\end{aligned} \quad (15)$$

where $\sum'$ indicates that the summation extends over precisely those increasing sequence $\sigma_1 < \ldots < \sigma_{r-p}$ which are subsets of $j_1 < \ldots < j_{n-p}$. Also $u_j = Ue_j$, $j=1, \ldots, n$ is an orthonormal set of vectors (recall that U is unitary). We use an extremal result in [3: theorem 1, p. 525] to conclude from (15) that

$$\phi^0_{i_1 \ldots i_p} \geq E_{r-p}(\lambda_{p+1}(F), \ldots, \lambda_n(F)) - E_{r-p}(\lambda_1(G), \ldots, \lambda_{n-p}(G)) \geq 0.$$

In case $G=I$ we have from (15) again that

$$\phi^0_{i_1 \ldots i_p} = \sum{}'(C_{r-p}(F)u_{\sigma_1} \wedge \ldots \wedge u_{\sigma_{r-p}}, u_{\sigma_1} \wedge \ldots \wedge u_{\sigma_{r-p}}) - \binom{n-p}{r-p}$$

for all choices of sets of $n-p$ orthonormal vectors $u_1, \ldots, u_{n-p}$, and another application of the above cited extremal result completes the proof.

To proceed to the proof of theorem 2, choose a nonsingular R such that $F=RR^*$, $G=RDR^*$, $D=\text{diag}\ (\lambda_1(F^{-1}G), \ldots, \lambda_n(F^{-1}G))$. Let P be an arbitrary nonsingular matrix and since any $A>0$ is of the form $(PR^{-1})^*(PR^{-1})$ we have that (7) is equivalent to

$$E_r((PR^{-1})^*(PR^{-1})RR^*) \geq E_r((PR^{-1})^*(PR^{-1})RDR^*),$$

$$E_r(P^*P) \geq E_r(P^*PD),$$

or

$$E_r(A) \geq E_r(AD) \text{ for all } A>0. \tag{16}$$

In (16) replace A by VXV^*, V unitary, $X=\text{diag}\ (x_1, \ldots, x_n) \geq 0$ to obtain

$$E_r(X) \geq E_r(XH), \qquad H=V^*DV$$

or

$$tr[C_r(X)(I-C_r(H))]=0, \tag{17}$$

where I is the $\binom{n}{r}$-square identity matrix. It is not difficult to check that (17) holds for all non-negative diagonal X if and only if every diagonal element of $I-C_r(H)$ is non-negative. That is,

$$1-(C_r(V^*DV)e_{\alpha_1} \wedge \ldots \wedge e_{\alpha_r}, e_{\alpha_1} \wedge \ldots \wedge e_{\alpha_r}) \geq 0$$

must hold for all V and all $1 \leq \alpha_1 < \ldots < \alpha_r \leq n$. But this is precisely equivalent to

$$(C_r(D)u_{\alpha_1} \wedge \ldots \wedge u_{\alpha_r}, u_{\alpha_1} \wedge \ldots \wedge u_{\alpha_r}) \leq 1$$

for all orthonormal $u_{\alpha_1}, \ldots, u_{\alpha_r}$. As in the proof of theorem 1 we have finally that (7) holds if and only if (8) does.

To prove theorem 3 it will be convenient to let $\stackrel{c}{=}$ denote the relation of Hermitian congruence. Then if $A>0$,

$$C_r(I+AF)-C_r(I+AG)=C_r(A)\{C_r(A^{-1}+F)-C_r(A^{-1}+G)\} \stackrel{c}{=} (C_r(A))^{1/2}[C_r(A^{-1}+F)-C_r(A^{-1}+G)](C_r(A))^{-1/2}.$$

This last matrix has the same roots as the Hermitian matrix $C_r(A^{-1}+F)-C_r(A^{-1}+G)$ and hence (9) is equivalent to

$$C_r(A+F) \geq C_r(A+G) \qquad \text{for all } A>0.$$

Now

$$C_r(A+F)-C_r(A+G)=C_r(A+F)-(C_r(G))^{1/2}C_r(G^{-1/2}AG^{-1/2}+I)(C_r(G))^{1/2} \stackrel{c}{=} C_r(G^{-1/2}AG^{-1/2}+G^{-1/2}FG^{-1/2})-C_r(G^{-1/2}AG^{-1/2}+I).$$

Thus (9) is equivalent to

$$C_r(A+H)-C_r(A+I) \geq 0 \quad \text{for all } A>0 \tag{18}$$

where $H=G^{-1/2}HG^{-1/2}$.

By a unitary congruence we may assume $H=\text{diag}\ (h_1, \ldots, h_n)$ and by setting $A=\text{diag}\ (x_1, \ldots, x_n) \geq 0$ we see that

$$\prod_{t=1}^{r}(x_{i_t}+h_{i_t}) \geq \prod_{t=1}^{r}(x_{i_t}+1)$$

for any non-negative numbers $x_{i_1}, \ldots, x_{i_r}$. This clearly implies that each $h_t \geq 1$, $t=1, \ldots, n$. Thus, $0 \geq H-I=G^{-1/2}FG^{-1/2}-I \stackrel{c}{=} F-G$.

Conversely suppose $F-G \geq 0$. Then $H \geq I$ and if we set $B=A+I>0$ we would like to conclude that

$$C_r(A+H)=C_r(B+H-I) \geq C_r(B)=C_r(A+I). \tag{19}$$

But (19) is equivalent to

$$C_r(I+K) \geq C_r(I), \quad K=B^{-1/2}(H-I)B^{-1/2}. \tag{20}$$

After a unitary congruence (20) reduces to

$$C_r(I+\text{diag}\ (k_1, \ldots, k_n)) \geq C_r(I),$$

where $k_\alpha \geq 0$ are the characteristic roots of K. The proof is complete.

2. References

[1] S. Ehrenfeld, Complete class theorems in experimental design, Proc. of the Third Berkeley Symposium on Mathematical Statistics and Probability **1,** 57 (1954–1955), Univ. of Calif. Press, Berkeley, 1956; Math. Rev. **18,** 946.

[2] W. Graeub, Lineare Algebra, ch. 5 (Springer-Verlag Berlin, 1958).

[3] M. Marcus and J. L. McGregor, Extremal properties of Hermitian matrices. Can. J. Math. **8,** 524 (1956).

[4] M. Stone, Application of a measure of information to the design and comparison of regression experiments. Ann. Math. Statist. **30,** 55 (1959); Math. Rev. **21,** 980.

(Paper 65B2–49)