SIAM J. MATRIX ANAL. APPL.
Vol. 15, No. 3, pp. 903–908, July 1994

A NOTE ON EXTREME CORRELATION MATRICES*

CHI-KWONG LI† AND BIT-SHUN TAM‡

Abstract. An $n \times n$ complex Hermitian or real symmetric matrix is a correlation matrix if it is positive semidefinite and all its diagonal entries equal one. The collection of all $n \times n$ correlation matrices forms a compact convex set. The extreme points of this convex set are called extreme correlation matrices. In this note, elementary techniques are used to obtain a characterization of extreme correlation matrices and a canonical form for correlation matrices. Using these results, the authors deduce most of the existing results on this topic, simplify a construction of extreme correlation matrices proposed by Grone, Pierce, and Watkins, and derive an efficient algorithm for checking extreme correlation matrices.

Key words. correlation matrix, extreme point, perturbation, rank, linear span

AMS subject classification. 15A48

Let $\mathcal{M} = \mathcal{H}_n$ or $\mathcal{S}_n$, the real linear space of all $n \times n$ Hermitian matrices and the real linear space of all $n \times n$ symmetric matrices, respectively. A positive semidefinite matrix $A = (a_{ij}) \in \mathcal{M}$ with $a_{11} = \cdots = a_{nn} = 1$ is called a *correlation matrix*. The term correlation matrix comes from statistics, where the entries of a real correlation matrix occur as correlations between pairs of random variables. It is easy to see that the collection of $n \times n$ correlation matrices forms a compact convex set, and we are interested in its extreme points. Recall that an element x in a convex set S is an *extreme point* if $x = ty + (1-t)z$ for $y, z \in S$ and $0 < t < 1$ implies $y = z = x$, that is, if x can be a convex combination of points of S in only trivial ways.

We shall call an extreme point of the set of correlation matrices an *extreme correlation matrix*. This concept has been studied in [1], [4], and [2]. In those papers, different approaches were used to determine all possible ranks of extreme correlation matrices, to construct extreme correlation matrices of different ranks, and to give simple characterizations of extreme correlation matrices in low dimensional cases. In this note, we use an elementary approach to prove several results on the subject. Using our results, one can deduce easily all of the main results in the three papers mentioned above. Moreover, we simplify a construction of extreme correlation matrices proposed in [2], derive an efficient algorithm for checking extreme correlation matrices, and compare our condition with the one given in [4]. A question posed in [2] is also discussed.

In the following we shall concentrate mainly on the Hermitian case, the slightly more difficult case. For the real case, we also give some results that have no analogs in the Hermitian case.

1. Basic results.

Given an $n \times n$ correlation matrix A, a nonzero Hermitian matrix B is said to be a *perturbation* of A if $A \pm tB$ are correlation matrices for some

*Received by the editors March 22, 1992; accepted for publication (in revised form) November 20, 1992.

†Department of Mathematics, The College of William and Mary, Williamsburg, Virginia 23187 (ckli@cs.wm.edu). The work of this author was supported in part by National Science Foundation grant DMS 91-00344.

‡Department of Mathematics, Tamkang University, Tamsui, Taiwan 25137, Republic of China (bsm01@twntku10.bitnet). The work of this author was supported by the National Science Council of the Republic of China.

(and hence for all sufficiently small) $t > 0$. Clearly, A is not extreme if and only if A has a perturbation. In fact, if $A = (A_1 + A_2)/2$ for two distinct correlation matrices A_1 and A_2, then $B := (A_1 - A_2)/2$ is such that $A \pm B$ are correlation matrices. We give a characterization of perturbations of a given correlation matrix and a characterization of extreme correlation matrices in the following theorem.

THEOREM 1. *Let $A \in \mathcal{H}_n$ be an $n \times n$ correlation matrix of rank r. Suppose that $A = XQX^*$, where $X \in \mathbb{C}^{n \times r}$ and $Q \in \mathcal{H}_r$. Then*

(a) *$B \in \mathcal{H}_n$ is a perturbation of A if and only if all diagonal entries of B equal zero and $B = XRX^*$ for some nonzero $R \in \mathcal{H}_r$, and*

(b) *A is extreme if and only if*

$$\operatorname{span}\{x_j x_j^* : 1 \leq j \leq n\} = \mathcal{H}_r,$$

where x_j is the jth column of X^*.

Proof. With the notation of the theorem, one sees easily that $\operatorname{rank} X = r$ and Q is positive definite.

(a) Let B be a matrix with all diagonal entries equal to zero and of the form XRX^* for some nonzero $R \in \mathcal{H}_r$. Since X is a matrix of full column rank, we have $\operatorname{rank} XRX^* = \operatorname{rank} R > 0$. Thus B is nonzero. Since Q is positive definite, $Q \pm tR$ is positive semidefinite for all sufficiently small $t > 0$. It follows that B is a perturbation of A. Conversely, suppose B is a perturbation of A. Evidently, all diagonal entries of B equal zero. Append to X an $n \times (n-r)$ matrix Y such that the $n \times n$ matrix $V = (X|Y)$ is nonsingular. Clearly A can be expressed as $V(Q \oplus 0_{n-r})V^*$. Write B as VCV^*, where $C \in \mathcal{H}_n$. Partition C in the same way as $Q \oplus O_{n-r}$. Since B is a perturbation of A, $(Q \oplus O_{n-r}) \pm tC$ is positive semidefinite for some $t > 0$. It follows that except for its $(1,1)$ block, which we denote by R, the blocks of C are all zero. Hence B is of the form XRX^*, with $0 \neq R \in \mathcal{H}_r$.

(b) Since A is not extreme if and only if it has a perturbation, by part (a) one sees that A is extreme if and only if for any $R \in \mathcal{H}_r$, $R = 0$ whenever all diagonal entries of XRX^* equal zero. In terms of the usual inner product on $\mathcal{H}_r$ defined by $\langle X, Y \rangle = tr(XY)$, we can reformulate the last condition as: $R = 0$ whenever $\langle R, x_j x_j^* \rangle = 0$ for all j, $1 \leq j \leq n$; or equivalently, $\operatorname{span}\{x_j x_j^* : 1 \leq j \leq n\} = \mathcal{H}_r$. Thus our result follows. □

Notice that for a given rank r positive semidefinite matrix A, there are two standard ways to decompose it as XQX^*. One way is to take Q to be a diagonal matrix whose diagonal entries are all the nonzero eigenvalues of A and form the matrix X whose columns are the corresponding eigenvectors. Another way is to find X such that $A = XX^*$, i.e., to take $Q = I_r$. In both cases, there are standard algorithms and computer programs to do the decomposition.

From Theorem 1 and its real analog, one easily deduces the following result proved in [1] (for the Hermitian case) and [2] (for the real case).

COROLLARY 2. *If A is an $n \times n$ extreme Hermitian (respectively, real symmetric) correlation matrix of rank r, then $r^2 \leq n$ (respectively, $r(r+1) \leq 2n$).*

In [4] and [2] it is shown that if r satisfies the inequality in the corollary, then there exists a rank r extreme correlation matrix. One can verify the constructions of extreme correlation matrices in those papers using our Theorem 1. We shall suggest a construction after proving Theorem 3. To state the result, we need the following definition and notation. A matrix is a (real) generalized permutation matrix if it is a unitary (respectively, real orthogonal) matrix with exactly one nonzero entry in each row and each column. Denote by $J_{r,s}$ the $r \times s$ matrix all of whose entries equal 1. For simplicity we use J_r to represent $J_{r,r}$.

THEOREM 3. *Let $A \in \mathcal{H}_n$. Then A is a correlation matrix if and only if there exists a generalized permutation matrix P such that $PAP^* = (B_{st})$, a $p \times p$ block matrix with $B_{st} = b_{st}J_{k(s),k(t)}$, $(k(1)+\cdots+k(p) = n)$, where $(b_{st}) \in \mathcal{H}_p$ is a correlation matrix all of whose off-diagonal entries have moduli less than one. Furthermore, we have*

(a) $\operatorname{rank} A = \operatorname{rank}(b_{st})$;

(b) *A is extreme if and only if (b_{st}) is extreme.*

Proof. To prove the "only if" part, express A in the form XX^*, where $X \in \mathbb{C}^{n\times r}$ and $r = \operatorname{rank} A$. Denote the jth column of X^* by x_j. Since each diagonal entry of A is equal to 1, each x_j is a vector of unit length. Now permute the rows of X and then multiply each row with a suitable scalar of absolute value one so that rows of X that differ by unit multiples are grouped together and become equal. The resulting effect on X is equivalent to applying a generalized permutation similarity to A. Since the inner product between any two linearly independent unit vectors always has modulus less than one, A is transformed to the required form. That (b_{st}) is a correlation matrix follows from the observation that it is a principal submatrix of A of the required form.

To prove the "If" part, suppose $A = (B_{st})$ as described in the theorem. It is clear that $A \in \mathcal{H}_n$ and its main diagonal entries are all equal to one. Note that we can write $J_{k(s),k(t)}$ as $\tilde{e}_{k(s)}\tilde{e}^t_{k(t)}$, where $\tilde{e}_j$ denotes the $j \times 1$ vector of all 1's. If $x_j \in \mathbb{C}^{k(j)}$ for $j = 1, \ldots, p$, then by direct calculations, the value of the quadratic form of A at x with $x^* = (x_1^*, \ldots, x_p^*)$ is equal to the value of the quadratic form of $(b_{st}) \in \mathcal{H}_p$ at $\tilde{x} = (\tilde{e}^t_{k(1)}x_1, \ldots, \tilde{e}^t_{k(p)}x_p)^t$ and so is nonnegative.

It is not difficult to show that $\operatorname{rank} A = \operatorname{rank}(b_{st})$. That A is extreme if and only if (b_{st}) is extreme follows readily from Theorem 1. □

Notice that the real analog of Theorem 3 also holds. To obtain the statement and the proof for the real case, one only needs to replace $\mathcal{H}_n$ by $\mathcal{S}_n$, generalized permutation matrices by real generalized permutation matrices, complex scalars by real scalars, etc.

Notice that the matrix (B_{st}) in Theorem 3 is a *block Kronecker product* of the matrices (b_{st}) and $(J_{k(s),k(t)})$. We refer the readers to [3] and its references for the definition and properties of this product.

2. A construction and an algorithm. There are at least two ways that Theorem 3 can help to study extreme correlation matrices. First, it helps to reduce the dimension of a problem under consideration. Second, if one can find an $n \times n$ rank r extreme correlation matrix, then one can use Theorem 3 to construct an $m \times m$ rank r extreme correlation matrices for any $m \geq n$. We illustrate the latter idea by describing a construction of extreme correlation matrices. (Note that this construction is a modification of the one given in [2].)

2.1. Construction of extreme correlation matrices. By the preceding discussion, for a given r it suffices to construct an $n \times n$ rank r extreme correlation matrix for $n = r^2$ in the Hermitian case, and for $n = r(r+1)/2$ in the real case. Then one can get an $m \times m$ rank r extreme correlation matrix for any $m \geq n$. We shall again use e_j to denote the jth column of I_r.

For the Hermitian case, assume $n = r^2$. Set $A = XX^*$ with $X \in \mathbb{C}^{n\times r}$ such that the first r columns of X^* form I_r, the next $r(r-1)/2$ columns consist of vectors of the form $(e_s + e_t)/\sqrt{2}$ with $1 \leq s < t \leq r$, and the rest of the $r(r-1)/2$ columns consist of vectors of the form $(e_s + ie_t)/\sqrt{2}$. Using Theorem 1, one verifies readily that $A \in \mathcal{H}_n$ is an extreme correlation matrix.

For the real case, assume $n = r(r+1)/2$. Let $\tilde{A} = \tilde{X}\tilde{X}^t$ where $\tilde{X}$ is obtained from X constructed in the Hermitian case by deleting the last $r(r-1)/2$ rows. Again by Theorem 1, one can show easily that $\tilde{A} \in \mathcal{S}_n$ is an extreme correlation matrix.

2.2. An algorithm for checking extreme correlation matrices. By Theorems 1, 3 (and its proof), and Corollary 2, one derives readily the following algorithm to determine whether a given Hermitian correlation matrix is extreme. A similar algorithm also holds for the real case.

Step 1. Express A as XX^*, where $X \in \mathbb{C}^{n\times r}$, $r = \operatorname{rank} A$.

Step 2. Form a matrix Y from the distinct (up to unit multiples) rows of X. Say $Y \in \mathbb{C}^{p\times r}$. (Then YY^* is equal to the matrix (b_{st}) as given in Theorem 3.)

Step 3. Determine $\operatorname{rank} Y$. If $\operatorname{rank} Y = r$ satisfies $r^2 > p$, then A is not extreme. Otherwise, proceed to Step 4.

Step 4. Determine the dimension of $\operatorname{span}\{y_j y_j^* : 1 \le j \le p\}$, where y_j is the jth column of Y^*. It is r^2 if and only if A is extreme.

An efficient way to perform Step 4 is to construct a $p \times r^2$ matrix F as follows. For each j between 1 and p, the first r entries of the jth row of F are $|y_{j1}|^2, |y_{j2}|^2, \ldots, |y_{jr}|^2$, arranged in the natural order, and its remaining $r^2 - r$ entries $y_{jk}\overline{y}_{jl}, \overline{y}_{jk}y_{jl}$, $1 \le k < l \le r$ (indexed by ordered pairs (k,l), and with conjugate entries adjacent) are arranged in the usual lexicographic order. Then $\operatorname{rank} F = \dim \operatorname{span}\{y_j y_j^* : 1 \le j \le p\}$.

Explanation. Consider the following real subspace of $\mathbb{C}^{r^2}$:

$$W = \{(t_1, \ldots, t_{r^2})^t : t_j \in \mathbb{R}, \quad j = 1, \ldots, r; \qquad \text{and}$$
$$t_{r+2m-1} = \overline{t}_{r+2m}, \quad m = 1, \ldots, (r^2 - r)/2\}.$$

Note that the real span of the row vectors of F is included in W, and is isomorphic with the subspace of $\mathcal{H}_r$ spanned by $\{y_j y_j^* : 1 \le j \le p\}$. But any set of vectors in W that is linearly independent over $\mathbb{R}$ is also linearly independent over $\mathbb{C}$, so $\operatorname{rank} F = \dim \operatorname{span}\{y_j y_j^* : 1 \le j \le p\}$.

Notice that the equivalent condition in [4] for an extreme correlation matrix can also be deduced readily as follows. Denote by f_j^t the jth row of F. Note that the vectors $f_1, \ldots, f_p$ all lie in the (real) hyperplane $\{f = (f_1, \ldots, f_{r^2})^t \in W : \sum_{i=1}^r f_j = 1\}$ of W (since the row vectors of Y are of unit length, as YY^* is a correlation matrix). But this hyperplane does not contain the zero vector, so we have

$$\begin{aligned}\dim \operatorname{span}\{f_j : 1 \le j \le p\} &= 1 + \dim \operatorname{span}\{f_j - f_p : 1 \le j \le p-1\} \\ &= 1 + \dim \operatorname{span}\{f_j - f_{j+1} : 1 \le j \le p-1\}.\end{aligned}$$

Denote by D_A the $(p-1) \times r^2$ matrix whose jth row is $(f_j - f_{j+1})^t$. Then A is extreme if and only if $r^2 = \dim \operatorname{span}\{f_j : 1 \le j \le p\} (= \operatorname{rank} F)$ if and only if $\operatorname{rank} D_A = r^2 - 1$, which is the condition given in [4]. (In [4] the matrix D_A is obtained from the matrix X instead of from Y. But this does not affect our argument.)

3. Further results. We first consider two results that are valid only for the real case.

COROLLARY 4. *Let $A \in \mathcal{S}_n$ be a rank two correlation matrix. Suppose P is a real generalized permutation matrix such that PAP^t is equal to (B_{st}), a $p \times p$ block matrix that satisfies the conditions as given in Theorem* 3. *Then A is extreme if and only if $p \ge 3$.*

Proof. "Only if" part. Since $\operatorname{rank} A = \operatorname{rank}(b_{st})$, p cannot be 1. If $p = 2$, then (b_{st}) is not extreme since it is nonsingular, and hence A is also not extreme according to Theorem 3.

"If" part. Suppose that A is not extreme. Since $\operatorname{rank} A = 2$, each (relative) boundary point of the face (of the set of $n \times n$ correlation matrices) generated by A is a matrix of rank one. So there exist two rank one correlation matrices A_1, A_2 such that $A = \lambda A_1 + (1-\lambda)A_2$ for some λ with $0 < \lambda < 1$. By applying a generalized permutation similarity to A, we may assume that $A_1 = J_n$ and

$$A_2 = \begin{pmatrix} J_k & -J_{k,n-k} \\ -J_{k,n-k} & J_{n-k} \end{pmatrix}$$

for some k between 1 and $n-1$. But then we have

$$A = \begin{pmatrix} J_k & \alpha J_{k,n-k} \\ \alpha J_{n-k,k} & J_{n-k} \end{pmatrix},$$

where $\alpha = \lambda + (-1)(1-\lambda)$ is of absolute value less than one. So in this case, $p = 2$. □

COROLLARY 5. *A 3×3 real symmetric correlation matrix of rank two is extreme if and only if its off-diagonal entries all have absolute values less than one.*

Two remarks are in order. First, by Corollaries 4 and 5, one sees that a rank two correlation matrix $A \in \mathcal{S}_n$ is extreme if and only if A has a principal submatrix that is an extreme 3×3 correlation matrix.

Second, the "if" parts of Corollaries 4 and 5 are both invalid in the Hermitian case. Indeed, for any $n \geq 2$, if we take A_n to be the matrix $(J_n + uu^*)/2$, where $u = (1, \mu, \ldots, \mu^{n-1})^t$, μ a primitive nth root of unity, then A_n is a nonextreme Hermitian correlation matrix of rank two, all of whose off-diagonal entries have moduli less than one.

By Theorem 1, we have the following observation for rank two correlation matrices in $\mathcal{H}_n$.

OBSERVATION. *Suppose $A = XQX^* \in \mathcal{H}_n$ with $X \in \mathbb{C}^{n\times 2}$ and $Q \in \mathcal{H}_2$ is a rank two correlation matrix. Let S be a 2×2 nonsingular submatrix of X^*. Then A is extreme if and only if there are two column vectors $u = (u_1, u_2)^t$ and $v = (v_1, v_2)^t$ of the matrix $S^{-1}X^*$, such that $\overline{u}_1 u_2$ and $\overline{v}_1 v_2$ are complex numbers that are not nonzero real multiples of each other.*

In the lemma in [2], it was shown that an equivalent condition for a real symmetric correlation matrix to be extreme is that its nullspace is maximal among the nullspaces of all correlation matrices. (The corresponding result for the Hermitian case also holds.) Clearly another equivalent condition is that the range space of the matrix is minimal among the range spaces of all correlation matrices. In [2] the authors also posed the question of determining the structure of the nullspace of a correlation matrix. Below we give an answer to the dual question of characterizing the linear subspaces of $\mathbb{C}^n$ (also $\mathbb{R}^n$) that can be the range space of a correlation matrix.

THEOREM 6. *A subspace of $\mathbb{C}^n$ (or $\mathbb{R}^n$) is the range space of a correlation matrix if and only if it has a basis (or a spanning set) $\{v_1, \ldots, v_r\}$ such that $\sum_{j=1}^{r} v_j \circ \overline{v}_j = (1, \ldots, 1)^t \in \mathbb{R}^n$, where $\overline{x}$ denotes the complex conjugate of the vector x, and $x \circ y$ denotes the Schur (Hadamard/entrywise) product of x and y.*

Proof. Suppose W is the range space of the correlation matrix A. Let $A = XX^*$ with $X \in \mathbb{C}^{n\times r}$, where $r = \operatorname{rank} A$. Then the columns of X form a basis for W that satisfies the required properties.

Conversely, if W is a subspace that has a spanning set as described in the theorem, then $A = XX^*$, where the columns of $X \in \mathbb{C}^{n\times r}$ are the vectors from the spanning set, is the required correlation matrix. □

COROLLARY 7. *A subspace of $\mathbb{C}^n$ (or $\mathbb{R}^n$) is the nullspace of a correlation matrix if and only if its orthogonal complement has a spanning set satisfying the condition in Theorem 6.*

Note added in proof. After the paper had been accepted for publication, the authors found that a slight modification of the proof of Theorem 1 yields the following result that covers [1, Thm. 3].

THEOREM 8. *Under the hypotheses and notation of Theorem 1, the face of the convex set of $n \times n$ correlation matrices generated by A is of dimension*

$$r^2 - \dim \operatorname{span} \{x_j x_j^* : 1 \leq j \leq n\}.$$

Proof. It is clear that the dimension of the face generated by A is equal to the dimension of the space generated by the perturbations of A. According to Theorem 1(a) (or its proof), a nonzero matrix is a perturbation of A if and only if it is of the form XRX^*, where X is $n \times r$ and R is $r \times r$ lying in the orthogonal complement of $\operatorname{span}\{x_j x_j^* : 1 \leq j \leq n\}$. Since X has full column rank, the mapping $R \mapsto XRX^*$ is a linear isomorphism. □

Acknowledgment. The authors wish to thank Dr. H.J. Woerdeman for drawing their attention to this subject. Thanks are also due to Dr. R. Horn and the referees for their helpful comments that improved our exposition.

REFERENCES

[1] J. P. R. CHRISTENSEN AND J. VESTERSTRØM, *A note on extreme positive definite matrices*, Math. Ann., 244 (1979), pp. 65–68.

[2] R. GRONE, S. PIERCE, AND W. WATKINS, *Extremal correlation matrices*, Linear Algebra Appl., 134 (1990), pp. 63–70.

[3] R.A. HORN AND R. MATHIAS, *Block-matrix generalizations of Schur's basic theorems on Hadamard products*, Linear Algebra Appl., 172 (1992), pp. 337–346.

[4] R. LOEWY, *Extreme points of a convex subset of the cone of positive semidefinite matrices*, Math. Ann., 253 (1980), pp. 227–232.