AN EXTREMAL PROBLEM FOR POSITIVE DEFINITE MATRICES

BY

T. W. ANDERSON and I. OLKIN

TECHNICAL REPORT NO. 34
JULY 1978

PREPARED UNDER CONTRACT N00014-75-C-0442
(NR-042-034)
OFFICE OF NAVAL RESEARCH
THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
AN EXTREMAL PROBLEM FOR POSITIVE DEFINITE MATRICES

by

T. W. ANDERSON and I. OLKIN

TECHNICAL REPORT NO. 34
JULY 1978

PREPARED UNDER CONTRACT N00014-75-C-0442
(NR-042-034)
OFFICE OF NAVAL RESEARCH

Theodore W. Anderson, Project Director

Reproduction in Whole or in Part is Permitted for
any Purpose of the United States Government.
Approved for public release; distribution unlimited.

Also issued as Technical Report No. 130 under National Science Foundation
Grant MPS 75-09450 - Department of Statistics, Stanford University.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
AN EXTREMAL PROBLEM FOR POSITIVE DEFINITE MATRICES

T. W. Anderson and I. Olkin
Stanford University

Abstract

A problem studied by Flanders (1975) is to minimize the function
\[ f(R) = \text{tr}(SR + TR^{-1}) \]
over the set of positive definite matrices \( R \), where \( S \) and \( T \) are positive semi-definite matrices. Alternative proofs that may have some intrinsic interest are provided. The proofs explicitly yield the infimum of \( f(R) \). One proof is based on a convexity argument and the other on a sequence of reductions to a univariate problem.

1. Introduction.

Flanders (1975) studied a matrix problem that arose in electric circuit theory. Let \( z_1, \ldots, z_m, w_1, \ldots, w_m \) be complex column vectors of length \( n \) and consider the real-valued function

\[ f(R) = z_1^* R z_1 + \cdots + z_m^* R z_m + w_1^* R^{-1} w_1 + \cdots + w_m^* R^{-1} w_m, \]

where \( R > 0 \) denotes an \( n \times n \) positive definite Hermitian matrix. The problem is to minimize \( f(R) \) over the set of positive definite \( R \).

If we set
\[
Z = (z_1, \ldots, z_m), \quad W = (w_1, \ldots, w_m),
\]
\[
S = ZZ^*, \quad T = WW^*,
\]
then (1) becomes

\[
(2) \quad f(R) = \text{tr} R(z_1 z_1^* + \cdots + z_m z_m^*) + \text{tr} R^{-1}(w_1 w_1^* + \cdots + w_m w_m^*) \\
= \text{tr} R z z^* + \text{tr} R^{-1} w w^* \\
= \text{tr} R S + \text{tr} R^{-1} T,
\]

where \( S \) and \( T \) are positive semi-definite matrices of order \( n \). We shall write \( A \geq 0 \) and \( A > 0 \) to denote that \( A \) is positive semi-definite and positive definite, respectively.

The result obtained by Flanders (1975) is as follows:

**Theorem 1.** If \( f(R) \) is defined on the set of positive definite matrices by (1) or (2), and if \( A = z^* w \), then

\[
\begin{align*}
(i) \quad \inf_{R > 0} f(R) &= 2 \text{tr} (A A^*)^{1/2} = 2 \text{tr}(A^* A)^{1/2}, \\
(ii) \quad f(R_0) &= 2 \text{tr}(A A^*)^{1/2} \quad \text{for some } R_0 > 0 \quad \text{if and only if} \\
\text{rank } (z) &= \text{rank } (w) = \text{rank } (A).
\end{align*}
\]

Flanders first proves that

\[
f(R) \geq 2 \text{tr}(A A^*)^{1/2},
\]

and then discusses the approach to equality. However, the matrix \( R \) that achieves equality (when the condition of (ii) is satisfied) is not exhibited in any simple manner. We now provide alternative somewhat simpler proofs of (i) that may have some intrinsic interest.
2. Matrices are of full rank.

First Alternative Proof. This proof is based on the fact that \( f(R) \) is a convex function of \( R \). The function \( \text{tr} RS \) is linear in \( R \), and \( \text{tr} R^{-1}T \) is convex in \( R \), i.e.,

\[
\text{tr}(\alpha R_1 + (1-\alpha)R_2)^{-1} T \leq \alpha \text{tr} R_1^{-1}T + (1-\alpha) \text{tr} R_2^{-1}T, \quad 0 \leq \alpha \leq 1,
\]

for \( R_1 \) and \( R_2 \) positive definite. The inequality is strict unless \( R_1 = R_2 \) or \( \alpha = 0 \) or \( 1 \). Consequently, \( f(R) \) is (strictly) convex, and we need to minimize a convex function over a convex set. Since \( f(R) \to \infty \) as \( R \to 0 \) or as \( R \to \infty \), the minimum is achieved at an interior point, namely where \( df(R)/dR = 0 \). But

\[
df = \text{tr} S(dR) - \text{tr} R^{-1} (dR)R^{-1}T = \text{tr} dR(S - R^{-1}TR^{-1});
\]

so that there is an interior point \( \tilde{R} \) satisfying \( S - \tilde{R}^{-1}TR^{-1} = 0 \), or equivalently,

\[
(3) \quad \tilde{R}S\tilde{R} = T,
\]

that is the minimizer of \( f(R) \). Note that \( \text{tr} \tilde{R}S = \text{tr} \tilde{R}^{-1}T \), so that

\[
f(\tilde{R}) = 2 \text{tr} \tilde{R}S.
\]

Furthermore, from (3),
\[ S^{1/2} R S^{1/2} = (S^{1/2} R S^{1/2})^2 = S^{1/2} T S^{1/2} ; \]

so that

\[ S^{1/2} R S^{1/2} = (S^{1/2} T S^{-1/2})^{1/2} , \]

and

\[ f(\tilde{R}) = 2 \text{tr} \tilde{R} S = 2 \text{tr}(S^{1/2} T S^{-1/2})^{1/2} . \]

Here we have used the positive definite square root. However, any square root, e.g., \( S = LL^* \), \( T = MM^* \) can be used, in which case the result is

\[ f(\tilde{R}) = 2 \text{tr}(L^* M M^* L)^{1/2} = 2 \text{tr}(M^* L L^* M)^{1/2} . \]

Let \( \lambda_i(A) \) denote the characteristic roots of the matrix \( A \). Because

\[
\text{tr}(L^* M M^* L)^{1/2} = \sum \lambda_i [(L^* M M^* L)^{1/2}] = \sum \lambda_i [L^* (L^* M M^* L)]^{1/2} \\
= \sum \lambda_i (M M^* L L^* M)^{1/2} = \sum \lambda_i (T S)^{1/2} , \\
= \text{tr}(T^{1/2} S T^{-1/2})^{1/2} = \text{tr}(S^{1/2} T S^{-1/2})^{1/2} ,
\]

we see that all square roots yield the same result.
Second Alternative Proof. This proof is based on a sequence of reductions until finally we obtain an extremal problem of distinct variables. There exists a nonsingular matrix $H$ such that

$$T^{-1} = HH^*, \quad S = HD_dH^*, \quad $$(1)

where $D_d$ is a diagonal matrix with real elements $(d_1, \ldots, d_n)$ ordered $0 < d_1 \leq \cdots \leq d_n$, which are the roots of

$$0 = \|S^{-1}T^{-1}\| = |T^{-1}| \cdot |ST^{-1}Y| = |T^{-1}| \cdot |S^{1/2}T^{1/2}S^{-1/2}Y| \ .$$

Then

$$f(R) = \text{tr} RH_dH_d^* + \text{tr} R^{-1}H^*H^{-1},$$

$$= \text{tr} H_d^*RHD_d + \text{tr}(H^*RH)^{-1} \equiv \text{tr} GD_d + \text{tr} G^{-1},$$

where $G = H^*RH$. Let $G = QD_\lambda Q^*$, where $Q$ is unitary and $D_\lambda = \text{diag} (\lambda_1, \ldots, \lambda_n)$ is diagonal with positive diagonal elements ordered $0 < \lambda_n \leq \cdots \leq \lambda_1$. The function to be minimized (with respect to $Q$ and $D_\lambda$) is

$$(4) \quad \text{tr} QD_\lambda Q^*D_d + \text{tr} D_\lambda^{-1} = \sum_{i,j=1}^n d_i \lambda_j |q_{ij}|^2 + \sum_{j=1}^n \lambda_j^{-1} \ .$$

By a theorem of von Neumann (1937) [see also Fan (1951)], the minimum of the first sum in $(4)$ with respect to $Q$ is $\sum_{j=1}^n d_j \lambda_j$, and
the minimizing $Q$ is $I$. Then the minimum of $\sum_{j=1}^{n} d_j \lambda_j + \sum_{j=1}^{n} \lambda_j^{-1}$ with respect to $\lambda_1, \ldots, \lambda_n$, is attained at $\lambda_j = d_j^{-1/2}$ and the minimized value of (4) is $2 \sum_{j=1}^{n} d_j^{1/2}$.

Remark. If one or more of the $d_j$ are 0, then the infimum cannot be attained.

The minimizing matrix $R$ is

$$R = (H^*)^{-1} G H^{-1} = (H^*)^{-1} D^{-1/2} d H^{-1} = (H D^{-1/2} H^*)^{-1}.$$  

This matrix satisfies (3).


In the case when rank($S$) and/or rank($T$) is less than $n$ we show how to reduce the problem to a canonical form from which we may then invoke the result for full rank.

Note that $f(R) = tr RS + tr R^{-1} T$ is invariant with respect to the transformation

$$(R, S, T) \rightarrow (QRQ^*, Q^*^{-1} SQ^{-1}, QTQ^*)$$

for any nonsingular matrix $Q$. Furthermore, the ranks of $S$, $T$ and $ST$ are invariant under this transformation. By judiciously choosing a matrix $Q$ we can effect a simplification of the problem.

Theorem 2. If $S \succeq 0$, $T \succeq 0$, then there exists a nonsingular matrix $V$ such that
\begin{equation}
\mathcal{T} = VTV^* = \begin{pmatrix} I_{\tau} & 0 \\ 0 & 0 \end{pmatrix}, \quad S = V^{*}S^{-1}V^{-1} = \begin{pmatrix} D & 0 \\ 0 & M \end{pmatrix}.
\end{equation}

where $\tau = \text{rank}(T)$, $D_d = \text{diag}(d_1, \ldots, d_\tau)$, $\text{rank}(M) = \text{rank}(S) - \text{rank}(ST)$, and $\text{rank}(D_d) = \text{rank}(ST)$.

Suppose for the moment that Theorem 2 holds. Then

\[
f(R) = \text{tr} \, RV^* \left( \begin{array}{cc} D_d & 0 \\ 0 & M \end{array} \right) V + \text{tr} \, R^{-1}V^{-1} \left( \begin{array}{cc} I_{\tau} & 0 \\ 0 & 0 \end{array} \right) V^{*} - 1.
\]

\[
= \text{tr} \, \tilde{R} \left( \begin{array}{cc} D_d & 0 \\ 0 & M \end{array} \right) + \text{tr} \, \tilde{R}^{-1} \left( \begin{array}{cc} I_{\tau} & 0 \\ 0 & 0 \end{array} \right),
\]

where $\tilde{R} = VRV^*$. Minimization of $f(R)$ over $R > 0$ is equivalent to minimization over $\tilde{R} > 0$. Consequently,

\[
f(R) = \text{tr} \, \tilde{R}_{11}D_d + \text{tr} \, \tilde{R}_{22}M + \text{tr} \, (\tilde{R}^{-1})_{11}.
\]

Since $\text{tr} \, (\tilde{R}_{11} - \tilde{R}_{12} \tilde{R}_{22}^{-1} \tilde{R}_{21}^{-1})^{-1} \geq \text{tr} \, \tilde{R}_{11}^{-1}$, $f(R)$ is minimized by taking $\tilde{R}_{12} = 0$. Then

\[
\min f(R) = \min \left[ \text{tr} \, \tilde{R}_{11}D_d + \text{tr} \, \tilde{R}^{-1}_{11} + \text{tr} \, \tilde{R}_{22}M \right].
\]

Three rank cases need to be considered.

(1) Rank($T$) > rank($ST$). Then one or more of the $d_j$ are 0, in which case the infimum is not attained. (See discussion leading to the remark in the second alternative proof.)
(ii) Rank(S) > rank(ST). Then M ≠ 0 and the infimum of 0 is not attained.

(iii) Rank(T) = rank(S) = rank(ST). Then M = 0 and rank(D_d) = τ, so that

\[
\inf_{R > 0} f(R) = \inf_{R_{11} > 0} (\text{tr} \tilde{R}_{11} D_d + \text{tr} \tilde{R}_{11}^{-1})
\]

and the problem has been reduced to the case of the one with full rank.

This completes the proof of Theorem 1. ||

To prove Theorem 2 we use the following two lemmas.

Lemma 1. A nonsingular matrix

\[
G = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}, \quad G_{11} : \tau \times \tau
\]

satisfies

\[
(6) \quad G \begin{pmatrix} I_\tau & 0 \\ 0 & 0 \end{pmatrix} G^* = \begin{pmatrix} I_\tau & 0 \\ 0 & 0 \end{pmatrix}
\]

if and only if G_{21} = 0, G_{11} is unitary, and G_{22} is nonsingular.

Proof. Multiply in (6) to obtain \( G_{11} G_{11}^* = I_\tau \), \( G_{21} G_{21}^* = 0 \), from which the conclusion follows. ||

Lemma 2. Given \( Q > 0 \) there exists a \( G_{12} \) and a nonsingular \( G_{22} \) such that

\[
Q_{11} G_{12} + Q_{12} G_{22} = 0.
\]
Proof. $Q$ has the triangular decomposition

$$Q = \begin{pmatrix} T_{11} & 0 \\ T_{12} & T_{22} \end{pmatrix} \begin{pmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{pmatrix} = \begin{pmatrix} T_{11}T_{11} & T_{11}T_{12} \\ T_{12}T_{11} & T_{12}T_{12} + T_{12}T_{22} \end{pmatrix}.$$ 

The rows of $(T_{11}, T_{12})$ span a space of dimension less than or equal to $r$. The $n-r$ columns of $(\hat{G}_{12})$ can be chosen to be orthogonal to the rows of $(T_{11}, T_{12})$ and linearly independent (so that $G_{22}$ is nonsingular).

Proof of Theorem 2. Let $U$ be any nonsingular matrix such that

$$T = U \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} U^*,$$

and define

$$\hat{S} = U^* S U.$$

Let $G$ be any nonsingular matrix satisfying Lemma 2 and define

$$\tilde{S} = G^* \hat{G}$$

$$= \begin{pmatrix} G_{11}^* \hat{S}_{11}G_{11} & G_{12}^* (\hat{s}_{11}G_{12} + \hat{s}_{12}G_{22}) \\ (G_{12}^* \hat{S}_{12} + G_{22}^* \hat{S}_{22})G_{11} & G_{12}^* (\hat{s}_{11}G_{12} + \hat{s}_{12}G_{22}) + G_{22}^* (\hat{s}_{11}G_{12} + \hat{s}_{22}G_{22}) \end{pmatrix}.$$

Using Lemma 2, the matrix $G_{12}$ can be chosen so that $\hat{s}_{11}G_{12} + \hat{s}_{12}G_{22} = 0$; so that
\[ S = \begin{pmatrix}
    G_{11}^* \hat{S}_{11} G_{11} & 0 \\
    0 & G_{22}^* (\hat{S}_{21} G_{12} + \hat{S}_{22} G_{22}^{-1}) G_{22}
\end{pmatrix} \]

Since \( G_{11} \) is unitary, we may choose it to diagonalize \( \hat{S}_{11} \). This completes the proof of Theorem 2. \( \| \)

Remark. In the second alternative proof we used the fact that if \( A \) and \( B \) are positive definite, then both can be diagonalized simultaneously by a nonsingular matrix \( W \), i.e.,

\( (7) \quad A = W W^*, \quad B = W D \theta W^* \),

where \( D \theta = \text{diag}(\theta_1, \ldots, \theta_n) \), and \( \theta_1 \geq \cdots \geq \theta_n \geq 0 \) are the characteristic roots of \( A^{-1} B \). When the hypotheses of positive definiteness are removed the simultaneous decomposition may no longer be accomplished in general. Theorem 2 complements the above result by providing a simultaneous decomposition. This can be stated in a form parallel to \( (7) \).

Theorem 3. If \( A \geq 0, B \geq 0 \), then there exists a nonsingular matrix \( W \) such that

\[
W A W^* = \begin{pmatrix}
    I_r & 0 & 0 & 0 \\
    0 & I & 0 & 0 \\
    0 & 0 & 0 & 0 \\
    0 & 0 & 0 & 0
\end{pmatrix},
\]

\[
W B W^* = \begin{pmatrix}
    D \theta & 0 & 0 & 0 \\
    0 & 0 & 0 & 0 \\
    0 & 0 & I & 0 \\
    0 & 0 & 0 & 0
\end{pmatrix},
\]

10
where \( D_\Theta = \text{diag}(\theta_1, \ldots, \theta_\tau) \) are the nonzero characteristic roots of

\[ |B - \Theta A| = 0. \]

**Proof.** The result follows from Theorem 2 by noting that if \( V \) is a matrix satisfying (5), then for any nonsingular matrix \( C \),

\[
W = V \begin{pmatrix}
I & 0 \\
0 & C
\end{pmatrix}
\]

also satisfies (5). Since \( M \geq 0 \), there exists a \( C \) such that \( CMC^* = \text{diag}(I, 0) \).

An equivalent form of Theorem 3 was obtained (but not published) by Olkin (1951). We suspect that this result is even older, but have not been able to locate a reference.
References


AN EXTREMAL PROBLEM FOR POSITIVE DEFINITE MATRICES

T. W. ANDERSON and I. OLKIN

Department of Statistics
Stanford University
Stanford, California

Office of Naval Research
Statistics & Probability Program Code 436
Arlington, Virginia 22217

Office of Naval Research
Statistics & Probability Program Code 436
Arlington, Virginia 22217

Approved for public release; distribution unlimited.

Extremal problem, positive definite matrices, electric circuit theory, diagonalization of matrices

A problem studied by Flanders (1975) is to minimize the function
\[ f(R) = \text{tr}(SR + TR^{-1}) \]
over the set of positive definite matrices \( R \), where \( S \) and \( T \) are positive semi-definite matrices of rank \( m \).

Alternative proofs that may have some intrinsic interest are provided. The proofs explicitly yield the infimum to \( f(R) \). One proof is based on a convexity argument and the other on a sequence of reductions to a univariate problem.