The Annals of Mathematical Statistics
1968, Vol. 39, No. 3, 859–863

MINIMIZATION OF EIGENVALUES OF A MATRIX AND OPTIMALITY OF PRINCIPAL COMPONENTS

BY MASASHI OKAMOTO[1] AND MITSUYO KANAZAWA

Osaka University

1. Introduction. Let $x' = (x_1, x_2, \dots, x_p)$ be a random vector with mean vector $E(x) = 0$ and variance matrix $E(xx') = \Sigma$. Let $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_p \geqq 0$ be the eigenvalues of Σ in order of decreasing magnitude, and $v_1, v_2, \cdots, v_p$ be the corresponding orthonormal eigenvectors.

The principal components of x, namely $v_1'x, v_2'x, \cdots, v_p'x$ were introduced by Hotelling [3], and since then characterized by various optimal properties. Almost all of these optimal properties, however, are stated in terms of linear functions of $x_1, x_2, \cdots, x_p$. For example, Rao [4] characterizes the first $k(\leqq p)$ principal components as a linear form $y = T'x$ with a $p \times k$ matrix T which minimizes the trace or the Euclidean norm of the residual variance matrix of x after subtracting its best linear predictor based on y. The unique exception is Darroch [2] who deals with the optimality within the class of all random variables with at most k dimensions.

The purpose of this paper is to characterize the first k principal components by a more general optimal property containing those due to Rao or Darroch as special cases. Lemma 3 in Section 2 dealing with simultaneous minimization of the eigenvalues of a non-negative definite matrix is of an algebraic character, and may be interesting by itself.

2. Notation and lemmas. Let $\mathfrak{A} = \mathfrak{A}_p$ be the set of all real non-negative definite matrices of order p. A partial order in the set $\mathfrak{A}$ is defined as usual; $A \geqq B$ if and only if $A - B \in \mathfrak{A}$. For any $A \in \mathfrak{A}$ let $\lambda_1(A) \geqq \lambda_2(A) \geqq \cdots \geqq \lambda_p(A)$ be the eigenvalues of A in order of decreasing magnitude. The following two Lemmas will be stated without proof.

LEMMA 1. *A necessary and sufficient condition for a real-valued function $f(A)$ defined on $\mathfrak{A}$ to be*

(i) *strictly increasing, that is, $f(A) \geqq f(B)$ if $A \geqq B$, and $f(A) > f(B)$ if moreover $A \neq B$, and*

(ii) *invariant under orthogonal transformation, that is, $f(P'AP) = f(A)$ for any orthogonal matrix P,*

is that $f(A)$ is identical to some function $g(\lambda_1(A), \cdots, \lambda_p(A))$ of the eigenvalues of A which is strictly increasing in each argument.

It is noted that the trace as well as the Euclidean norm of a matrix enjoys this property of a function f.

Now we denote by $M_k(A)$ $(k = 1, 2, \cdots, p)$ the linear subspace spanned by

Received April 28, 1967.

[1] Now visiting professor at the Statistical Laboratory and Department of Statistics, Iowa State University.

the eigenvectors corresponding to the eigenvalues larger than $\lambda_k(A)$. This does not necessarily involve the eigenvalues $\lambda_1(A), \cdots, \lambda_{k-1}(A)$ because of the possibility that $\lambda_{k-1}(A) = \lambda_k(A)$. Furthermore, for any matrix L we denote by $M(L)$ the linear subspace spanned by the column vectors of L.

LEMMA 2. *For any $A \varepsilon \mathfrak{A}$, any positive integer k and any matrix L with p rows and with rank $<k$ it holds that*

$$\sup_{L'x=0} x'Ax/x'x \geqq \lambda_k(A),$$

sup *denoting the supremum for p-vectors x satisfying $L'x = 0$. A sufficient condition that the equality sign holds is that*

$$M(L) \supset M_k(A),$$

which is also a necessary condition in case $k = p$ and implies that $M(L)$ is orthogonal to some eigenvector of the matrix A corresponding to the smallest eigenvalue $\lambda_p(A)$.

This result is referred in Bellman ([1], p. 113) to the Courant-Fischer min-max theorem without the statement about L attaining the equality.

LEMMA 3. *If A, B and $A - B$ all belong to $\mathfrak{A}$ and if B is at most of rank k, then for each $i = 1, 2, \cdots, p$ it holds that*

$$\lambda_i(A - B) \geqq \lambda_{k+i}(A), \tag{1}$$

where $\lambda_j = \lambda_j(A)$ is defined to be zero for $j > p$. A necessary and sufficient condition that the equality sign holds in (1) for every i simultaneously is that

$$B = \lambda_1 v_1 v_1' + \lambda_2 v_2 v_2' + \cdots + \lambda_k v_k v_k', \tag{2}$$

where $v_1, v_2, \cdots, v_k$ are orthonormal eigenvectors of A corresponding to $\lambda_1, \lambda_2, \cdots, \lambda_k$.

The first half of the lemma is a corollary of the Weyl inequality (see Bellman [1], p. 119), but the second half is new as far as the authors are aware.

PROOF. Applying Lemma 2 to the matrix $A - B$ and choosing L of rank $<i$ for which the equality holds, we have for each i

$$\begin{aligned} \lambda_i(A - B) &= \sup_{L'x=0} x'(A - B)x/x'x \\ &\geqq \sup_{L'x=0,B'x=0} x'Ax/x'x \geqq \lambda_{k+i}(A) = \lambda_{k+i}. \end{aligned} \tag{3}$$

Now suppose that

$$\lambda_i(A - B) = \lambda_{k+i} \quad \text{for every} \quad i, \tag{4}$$

and we shall prove (2), the converse being obvious. Without loss of generality we can assume the matrix A to be diagonal, that is, $A = \operatorname{diag}(\lambda_1, \cdots, \lambda_p)$ with diagonal elements $\lambda_1, \cdots, \lambda_p$. Combining (3), Lemma 2, and the relation (4) for $i = p - k$, we conclude that $M(B, L)$, and *a fortiori*, $M(B)$ is orthogonal to some eigenvector of A corresponding to the smallest eigenvalue λ_p. This implies that there exists an orthogonal matrix P_p leaving the eigenspaces of A invariant, or $P_p'AP_p = A$, and satisfying

$$P_p'BP_p = \begin{pmatrix} B_{p-1} & 0 \\ 0 & 0 \end{pmatrix}$$

for some $B_{p-1} \varepsilon \mathcal{A}_{p-1}$. Define $A_{p-1} = \text{diag}(\lambda_1, \cdots, \lambda_{p-1})$, then the set of the eigenvalues of $A - B$ is identical with that of

$$P_p'(A - B)P_p = \begin{pmatrix} A_{p-1} - B_{p-1} & 0 \\ 0 & \lambda_p \end{pmatrix}.$$

Therefore it follows from (4) that

$$\begin{aligned} \lambda_i(A_{p-1} - B_{p-1}) &= \lambda_{k+i} \quad \text{for} \quad i = 1, \cdots, p - k - 1, \\ &= 0 \quad \text{for} \quad i = p - k, \cdots, p - 1. \end{aligned}$$

By a mathematical induction we have an orthogonal matrix P_j of order j $(j = k + 1, \cdots, p)$ and a $B_j \varepsilon \mathcal{A}_j$ $(j = k, \cdots, p)$ such that

$$P_j'A_jP_j = A_j \quad \text{and} \quad P_j'B_jP_j = \begin{pmatrix} B_{j-1} & 0 \\ 0 & 0 \end{pmatrix} \tag{5}$$

for any $j = k + 1, \cdots, p$ and

$$\begin{aligned} \lambda_i(A_j - B_j) &= \lambda_{k+i} \quad \text{for} \quad i = 1, \cdots, j - k, \\ &= 0 \quad \text{for} \quad i = j - k + 1, \cdots, j \end{aligned} \tag{6}$$

for any $j = k, \cdots, p$, where $A_j = \text{diag}(\lambda_1, \cdots, \lambda_j)$.

From (6) for $j = k$ we obtain that $B_k = A_k$. Define

$$V = P_p \begin{pmatrix} P_{p-1} & 0 \\ 0 & I_1 \end{pmatrix} \cdots \begin{pmatrix} P_{k+1} & 0 \\ 0 & I_{p-k-1} \end{pmatrix},$$

where I_j stands for a unit matrix of order j. Then, V is an orthogonal matrix of order p and the relations (5) for $j = k + 1, \cdots, p$ imply

$$V'AV = A, \tag{7}$$

$$V'BV = \begin{pmatrix} A_k & 0 \\ 0 & 0 \end{pmatrix}. \tag{8}$$

It is readily seen that by (7) the columns $v_1, v_2, \cdots, v_p$ of V give a set of orthonormal eigenvectors of A and that (8) is equivalent to (2), which completes the proof.

3. An optimal property of principal components.

THEOREM. *Let A be any $p \times k$ matrix and $y' = (y_1, y_2, \cdots, y_k)$ be any random vector. Let f be a real-valued function defined on $\mathcal{A}$ which is strictly increasing and invariant under any orthogonal transformation in the sense in Lemma 1. Then*

$$F_1 = f(E(x - Ay)(x - Ay)')$$

is minimized with respect to A and y when and only when

$$Ay = v_1v_1'x + v_2v_2'x + \cdots + v_kv_k'x$$

and the minimum value of F_1 is $g(\lambda_{k+1}, \cdots, \lambda_p, 0, \cdots, 0)$, where v_j $(j = 1, \cdots, k)$ are orthonormal eigenvectors of Σ corresponding to λ_j's and g is the function introduced in Lemma 1.

REMARK. When $f(A) = \text{tr}(A)$, the optimal property above reduces to that due to Darroch [2], while it affords a generalization of Rao's result [4] when $f(A) = \|A\|$. It is noted that this theorem does not necessarily hold when f is assumed only to be strictly increasing and also that the Ay which attains the minimum of F_1 is uniquely determined when and only when $\lambda_k \neq \lambda_{k+1}$.

PROOF. Let r be the rank of the matrix Σ. If $r \leqq k$ the problem is trivial; F_1 is minimized uniquely by taking $Ay = x$. Suppose, therefore, $r > k$.

Without loss of generality we can assume $E(yy') = I_k$, and let $E(xy') = B$. Since $\begin{pmatrix} \Sigma & B \\ B' & I_k \end{pmatrix}$ is the variance matrix of a joint random vector (x', y'), it is non-negative definite and hence the matrix $\Sigma - BB'$, too. Since

$$E(x - Ay)(x - Ay)' = \Sigma - BB' + (A - B)(A - B)'$$
$$\geqq \Sigma - BB',$$

we have

$$f(E(x - Ay)(x - Ay)') \geqq f(\Sigma - BB') = F_2, \quad \text{say},$$

with the equality sign if and only if $A = B$.

By Lemma 1 F_2 is an increasing function of each eigenvalue of $\Sigma - BB'$, which we shall try to minimize. Lemma 3 implies that

$$\lambda_i(\Sigma - BB') \geqq \lambda_{k+i}(\Sigma)$$

for each i, where the equality holds for all i simultaneously if and only if

$$BB' = \lambda_1v_1v_1' + \lambda_2v_2v_2' + \cdots + \lambda_kv_kv_k', \tag{9}$$

v_j's $(j = 1, \cdots, p)$ being orthonormal eigenvectors of Σ corresponding to λ_j's. Let $V = (v_1, v_2, \cdots, v_p)$ and $\Lambda = \text{diag}(\lambda_1, \cdots, \lambda_p)$, then (9) is written as

$$BB' = V\Lambda^{\frac{1}{2}} \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} \Lambda^{\frac{1}{2}}V'. \tag{10}$$

We denote by Λ^* the diagonal matrix obtained by substituting ones for zeroes in the diagonal of Λ and define

$$Q = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} = \Lambda^{*-\frac{1}{2}}V'B, \tag{11}$$

where Q_1 is a $k \times k$ and Q_2 is a $(p - k) \times k$ matrix. From (10) and (11) it follows that

$$\begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} = QQ' = \begin{pmatrix} Q_1Q_1' & Q_1Q_2' \\ Q_2Q_1' & Q_2Q_2' \end{pmatrix},$$

and hence $Q_1Q_1' = I_k$ and $Q_2 = 0$. Therefore, F_2 is minimized by taking

$$B = V\Lambda^{*\frac{1}{2}}\begin{pmatrix} Q_1 \\ 0 \end{pmatrix} = G, \quad \text{say,}$$

where Q_1 is any orthogonal $k \times k$ matrix, and the minimum value of F_2 is

$$f(\lambda_{k+1}v_{k+1}v_{k+1}' + \cdots + \lambda_p v_p v_p') = g(\lambda_{k+1}, \cdots, \lambda_p, 0, \cdots, 0),$$

where g is the function defined in Lemma 1.

Though the argument from now on is the same as that in Darroch [2] except that we here admit Σ to be singular, we shall give it for the sake of completeness of the proof. Define

$$H = V\Lambda^{*-\frac{1}{2}}\begin{pmatrix} Q_1 \\ 0 \end{pmatrix} \quad \text{and} \quad v = H'x.$$

Then it is easily seen that

$$\Sigma H = G \quad \text{and} \quad H'G = I_k.$$

Now we shall show that there exists uniquely a random vector y satisfying the conditions $E(yy') = I_k$ and $E(xy') = G$. In fact v is the solution, for

$$E(xv') = E(xx')H = \Sigma H = G$$

and

$$E(vv') = H'E(xv') = H'G = I_k.$$

Uniqueness follows from the fact

$$E(v - y)(v - y)' = E(vv') - E(vy') - E(y'v) + E(yy')$$
$$= I_k - I_k - I_k + I_k = 0,$$

since $E(vy') = H'E(xy') = H'G$. Thus F_1 is minimized by taking

$$Ay = Gv = GH'x = V\begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}V'x$$
$$= v_1v_1'x + v_2v_2'x + \cdots + v_kv_k'x.$$

REFERENCES

[1] BELLMAN, R. (1960). *Introduction to Matrix Analysis*. McGraw-Hill, New York.

[2] DARROCH, J. N. (1965). An optimal property of principal components. *Ann. Math. Statist.* **36** 1579–1582.

[3] HOTELLING, H. (1933). Analysis of a complex of statistical variables into principal components. *J. Educ. Psych.* **24** 417–441; 498–520.

[4] RAO, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyā Ser. A.* **26** 329–358.