JOURNAL OF MULTIVARIATE ANALYSIS 6, 422–425 (1976)

# A Maximization Problem and its Application to Canonical Correlation

MORRIS L. EATON*

*Department of Theoretical Statistics, University of Minnesota, Minneapolis, Minnesota 55455*

*Communicated by J. N. Srivastava*

Let $\Sigma$ be an $n \times n$ positive definite matrix with eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n > 0$ and let $M = \{x, y \mid x \in R^n, y \in R^n, x \neq 0, y \neq 0, x'y = 0\}$. Then for $x, y$ in $M$, we have that $x'\Sigma y/(x'\Sigma x\, y'\Sigma y)^{1/2} \leqslant (\lambda_1 - \lambda_n)/(\lambda_1 + \lambda_n)$ and the inequality is sharp. If

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

is a partitioning of $\Sigma$, let $\theta_1$ be the largest canonical correlation coefficient. The above result yields $\theta_1 \leqslant (\lambda_1 - \lambda_n)/(\lambda_1 + \lambda_n)$.

Let $\Sigma$ be an $n \times n$ positive definite matrix with eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n > 0$ and associated eigenvectors $x_1, \ldots, x_n$, $\| x_i \| = 1$, $i = 1, \ldots, n$, $x_i' x_j = 0$ if $i \neq j$. The main result of this paper is

THEOREM 1. *Let $M = \{x, y \mid x \in R^n, y \in R^n, x \neq 0, y \neq 0, x'y = 0\}$. Then*

$$\sup_{x,y \in M} \frac{x'\Sigma y}{(x'\Sigma x y'\Sigma y)^{1/2}} = \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}. \tag{1}$$

*Equality in* (1) *is achieved for $x = x_1 + x_n$ and $y = x_1 - x_n$.*

Received February 5, 1975; revised January 20, 1976.

AMS 1970 subject classification: Primary 15A42, secondary 62H20.

Key words and phrases: Matrix inequality, eigenvalue, canonical correlation.

* This research was supported in part by the National Science Foundation Grant, NSF-GP-34482.

*Proof.* Let $\delta = 2\lambda_1\lambda_n/(\lambda_1 + \lambda_n)$. For $y$ and $x$ in $M$,

$$\begin{aligned} x'\Sigma y &= x'\Sigma y - \delta x'y \\ &= x'\Sigma^{1/2}(I - \delta\Sigma^{-1})\,\Sigma^{1/2}y \\ &\leqslant (x'\Sigma xy'\Sigma y)^{1/2} \max_{1\leqslant i\leqslant n} |\mu_i| \end{aligned} \tag{2}$$

where $\mu_i = 1 - (\delta/\lambda_i)$, $i = 1,\ldots, n$ are the eigenvalues of $I - \delta\Sigma^{-1}$. But

$$\max_{1\leqslant i\leqslant n} |\mu_i| = 1 - \frac{\delta}{\lambda_1} = \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}$$

so

$$\sup_{x,y\in M} \frac{x'\Sigma y}{(x'\Sigma xy'\Sigma y)^{1/2}} \leqslant \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}. \tag{3}$$

However, equality is clearly achieved for $x = x_1 + x_n$ and $y = x_1 - x_n$. This completes the proof.

*Remark.* We now outline a longer, but somewhat more informative, proof of (3). Using the Cauchy–Schwartz inequality, one can show that for fixed $x \neq 0$,

$$\sup_{\substack{y\neq 0\\ yx'=0}} \frac{(x'\Sigma y)^2}{x'\Sigma xy'\Sigma y} = 1 - \frac{(x'x)^2}{x'\Sigma xx'\Sigma^{-1}x}. \tag{4}$$

Then, using the Kantorovich inequality (see Marshall and Olkin [4]), it follows that

$$\sup_{x\neq 0} \left[1 - \frac{(x'x)^2}{x'\Sigma xx'\Sigma^{-1}x}\right] \leqslant 1 - \frac{4\lambda_1\lambda_n}{(\lambda_1 + \lambda_n)^2} = \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right)^2 \tag{5}$$

which yields (3) and (1) follows by setting $x = x_1 + x_n$ and $y = x_1 - x_n$.

COROLLARY 1. *Let $\Sigma$ and $A$ be two $n \times n$ positive definite matrices and let $M_A = \{x, y \mid x \in R^n, y \in R^n, x \neq 0, y \neq 0, x'Ay = 0\}$. Then*

$$\sup_{x,y\in M_A} \frac{x'\Sigma y}{(x'\Sigma xy'\Sigma y)^{1/2}} = \frac{\mu_1 - \mu_n}{\mu_1 + \mu_n} \tag{6}$$

*where $\mu_1$ is the largest eigenvalue of $A^{-1}\Sigma$ and $\mu_n$ is the smallest eigenvalue of $A^{-1}\Sigma$.*

*Proof.* This follows immediately from Theorem 1.

Consider

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where $\Sigma_{11}$ is $p \times p$ and $\Sigma_{22}$ is $q \times q$ with $p + q = n$. As is well known, (Anderson [1, p. 289] or Eaton [2, Chap. 10]) the largest canonical correlation coefficient, say $\theta_1$, is given by

$$\theta_1 = \sup_{\substack{0 \neq a \in R^p \\ 0 \neq b \in R^q}} \frac{a' \Sigma_{12} b}{(a' \Sigma_{11} a b' \Sigma_{22} b)^{1/2}} \cdot \tag{7}$$

THEOREM 2. *For any partitioning of* $\Sigma$,

$$\theta_1 \leqslant \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n} \tag{8}$$

*where* $\lambda_1 \geqslant \cdots \geqslant \lambda_n > 0$ *are the eigenvalues of* $\Sigma$.

*Proof.* For $a \in R^p$ and $b \in R^q$, set $a^* = \binom{a}{0} \in R^n$ and $b^* = \binom{0}{b} \in R^n$. Then we have

$$\begin{aligned} \theta_1 &= \sup_{\substack{0 \neq a^* \in R^n \\ 0 \neq b^* \in R^n}} \frac{a^{*\prime} \Sigma b^*}{(a^{*\prime} \Sigma a^* b^{*\prime} \Sigma b^*)^{1/2}} \\ &\leqslant \sup_{x, y \in M} \frac{x' \Sigma y}{(x' \Sigma x y' \Sigma y)^{1/2}} = \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n} \end{aligned} \tag{9}$$

by Theorem 1. The inequality holds because $a^{*\prime} b^* = 0$ so the second sup is over a larger set of vectors than is the first sup. The proof is complete.

The inequality in (9) was also established by Haberman [3] using a different method.

To show that the inequality (9) is sharp, consider $p \leqslant q$ and

$$\Sigma = \begin{pmatrix} I_p & (D_\theta 0) \\ \begin{pmatrix} D_\theta \\ 0 \end{pmatrix} & I_q \end{pmatrix} \tag{10}$$

where $D_\theta$: $p \times p$ is diagonal with diagonal entries $1 \geqslant \theta_1 \geqslant \theta_2 \geqslant \cdots \geqslant \theta_p \geqslant 0$. For $\Sigma$ partitioned as in (10), $\theta_1$ is the largest canonical correlation and it is not hard to show that $\lambda_1 = 1 + \theta_1$ and $\lambda_n = 1 - \theta_1$. Hence $\theta_1 = (\lambda_1 - \lambda_n)/(\lambda_1 + \lambda_n)$ so (9) is sharp. One can also show that when $p \geqslant 2$ and for $\Sigma$ given in (10), we have $\theta_2 = (\lambda_2 - \lambda_{n-1})/(\lambda_2 + \lambda_{n-1})$. This might lead one to conjecture that for general $\Sigma$ and $p \geqslant 2$, $q \geqslant 2$, the inequality $\theta_2 \leqslant (\lambda_2 - \lambda_{n-1})/(\lambda_2 + \lambda_{n-1})$

holds. However, it is possible to construct a $4 \times 4$ matrix $\Sigma$ where the inequality does not hold.

## Acknowledgment

I would like to thank the referee for providing the shortened proof of Theorem 1.

## References

[1] Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

[2] Eaton, M. L. (1972). *Multivariate Statistical Analysis*. University of Copenhagen, Copenhagen. Denmark.

[3] Haberman, S. (1974). *How much do Gauss–Markov and Least Squares Estimates Differ? A Coordinate-free Approach*. University of Chicago preprint.

[4] Marshall, A. W. and Olkin, I. (1964). Reversal of the Lyaponov, Hölder, and Minkowski inequalities and other extensions of the Kantorovich inequality. *J. Math. Anal. Appl.* 8 503–514.