A MAXIMIZATION PROBLEM AND ITS APPLICATION TO CANONICAL CORRELATION

by

Morris L. Eaton

Technical Report No. 233

September, 1974

Department of Statistics
University of Minnesota

Abstract

Let Σ be an $n \times n$ positive definite matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0$ and let $M = \{x,y | x \in R^n, y \in R^n, x \neq 0, y \neq 0, x'y = 0\}$. Then $\sup_{x,y \in M} \frac{x'\Sigma y}{\sqrt{x'\Sigma x \, y'\Sigma y}} = \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}$.

If $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ is a partitioning of Σ, let θ_1 be the largest canonical correlation associated with the above partitioning. The above result yields $\theta_1 \leq \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}$.

Let Σ be an $n \times n$ positive definite matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n > 0$ and associated eigenvectors $x_1, \dots, x_n$, $\|x_i\| = 1$, $i = 1, \dots, n$, $x_i' x_j = 0$ if $i \neq j$. The main result of this note 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

$$(1) \qquad \sup_{x,y \in M} \frac{x' \Sigma y}{\sqrt{x' \Sigma x \; y' \Sigma y}} = \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n} .$$

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

The proof of Theorem 1 is based on the following two lemmas.

Lemma 1: Suppose $w_o \in R^n$ and $v_o \in R^n$ are both non-zero. Then

$$(2) \qquad \sup_{\substack{\|u\| = 1 \\ u'w_o = 0}} (u' v_o)^2 = \|v_o\|^2 - \frac{(w_o' v_o)^2}{\|w_o\|^2} .$$

Proof: Let $P = I_n - \frac{w_o w_o'}{\|w_o\|^2}$ so P is the orthogonal projection onto the orthogonal complement of w_o. Then

$$(3) \qquad \sup_{\substack{\|u\|=1 \\ u'w_o=0}} (u'v_o)^2 = \sup_{\substack{\|u\|=1 \\ u'w_o=0}} ((Pu)'v_o)^2$$

$$= \sup_{\substack{\|u\|=1 \\ u'w_o=0}} (u'Pv_o)^2 \leq \|Pv_o\|^2$$

by the Cauchy-Schwartz inequality. But equality in the above inequality is attained by setting $u = \frac{Pv_o}{\|Pv_o\|}$ when $Pv_o \neq 0$ (the case of $Pv_o = 0$ is obvious). Noting that $\|Pv_o\|^2 = \|v_o\|^2 - \frac{(w_o' v_o)^2}{\|w_o\|^2}$, the proof is complete.

Lemma 2: Suppose X is a random variable with $\Pr\{m \le X \le M\} = 1$ where $m > 0$. Then

$$(4) \qquad \frac{1}{\mathcal{E}X\,\mathcal{E}X^{-1}} \ge \frac{4mM}{(m+M)^2} .$$

Proof: This follows immediately from Lemma 2.2 in Marshall and Olkin (1964) by setting (in the Olkin-Marshall notation) $Z \equiv 1$, $s = 1$ and $r = -1$. Of course, (4) is just the Kantorovich Inequality.

Proof of Theorem 1: Fix $x \neq 0$, $x \in R^n$. Then

$$(5) \qquad \sup_{\substack{y \neq 0 \\ x'y=0}} \left(\frac{x' \Sigma y}{\sqrt{y' \Sigma y}} \right)^2 = \sup_{\substack{y \neq 0 \\ x' \Sigma^{-\frac{1}{2}} y=0}} \left(\frac{x' \Sigma^{\frac{1}{2}} y}{\sqrt{y' y}} \right)^2 =$$

$$\sup_{\substack{\|y\|=1 \\ (\Sigma^{-\frac{1}{2}}x)' y=0}} [(\Sigma^{\frac{1}{2}} x)' y]^2 = \|\Sigma^{\frac{1}{2}} x\|^2 - \frac{((\Sigma^{-\frac{1}{2}}x)' \Sigma^{\frac{1}{2}} x)^2}{\|\Sigma^{-\frac{1}{2}} x\|^2} =$$

$$x' \Sigma x - \frac{(x'x)^2}{x' \Sigma^{-1} x}$$

where the next to the last equality follows from Lemma 1 with $v_o = \Sigma^{\frac{1}{2}} x$ and $w_o = \Sigma^{-\frac{1}{2}} x$. Here, $\Sigma^{\frac{1}{2}}$ is the unique positive definite square root of Σ and $\Sigma^{-\frac{1}{2}} \equiv (\Sigma^{\frac{1}{2}})^{-1}$.

Thus,

$$(6)\qquad \sup_{x,y\,\epsilon\,M} \frac{(x'\Sigma y)^2}{x'\Sigma x\; y'\Sigma y} = \sup_{x\neq 0} \frac{1}{x'\Sigma x}\left(x'\Sigma x - \frac{(x'x)^2}{x'\Sigma^{-1}x}\right) =$$

$$\sup_{x\neq 0}\; 1 - \frac{(x'x)^2}{x'\Sigma x\; x'\Sigma^{-1}x} = 1 - \inf_{\|x\|=1} \frac{1}{x'\Sigma x\; x'\Sigma^{-1}x}$$

Now, write $\Sigma = \Gamma D \Gamma'$ where Γ is an $n \times n$ orthogonal matrix and D is a diagonal matrix with diagonal elements $\lambda_1,\ldots,\lambda_n$ - the eigenvalues of Σ. Then, with $w = \Gamma x$,

$$(7)\qquad \inf_{\|x\|=1} \frac{1}{x'\Sigma x\; x'\Sigma^{-1}x} = \inf_{\|w\|=1} \frac{1}{w'\,Dw\; w'\,D^{-1}w} =$$

$$\inf_{\|w\|=1} \frac{1}{\left(\sum_1^n w_i^2\lambda_i\right)\left(\sum_1^n w_i^2\lambda_i^{-1}\right)}\,.$$

Since $\|w\| = 1$, $\sum_1^n w_i^2 = 1$. Denoting by X the random variable which takes on the value λ_i with probability w_i^2, we see that for all w, $\|w\| = 1$,

$$(8)\qquad \frac{1}{\left(\sum_1^n \lambda_i w_i^2\right)\left(\sum_1^n w_i^2\lambda_i^{-1}\right)} = \frac{1}{\mathcal{E}X\;\mathcal{E}X^{-1}} \geq \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2}$$

by Lemma 2 with $m = \lambda_n$ and $M = \lambda_1$.

Thus, combining (6), (7) and (8),

$$(9) \qquad \sup_{x,y \in M} \frac{(x'\Sigma y)^2}{x'\Sigma x \; y'\Sigma y} = 1 - \inf_{\|x\|=1} \frac{1}{x'\Sigma x \; x'\Sigma^{-1}x} \leq$$

$$1 - \frac{4\lambda_1 \lambda_n}{(\lambda_1 + \lambda_n)^2} = \frac{(\lambda_1 - \lambda_n)^2}{(\lambda_1 + \lambda_n)^2} .$$

Taking square roots, we now have the inequality

$$(10) \qquad \sup_{x,y \in M} \frac{x'\Sigma y}{\sqrt{x'\Sigma x \; y'\Sigma y}} \leq \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n} .$$

However, setting $x = x_1 + x_n$ and $y = x_1 - x_n$, we see that in (10), we actually achieve equality. This completes the proof of Theorem 1.

<u>Corollary 1</u>: Let Σ 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

$$(11) \qquad \sup_{x,y \in M_A} \frac{x'\Sigma y}{\sqrt{x'\Sigma x \; y'\Sigma y}} = \frac{\mu_1 - \mu_n}{\mu_1 + \mu_n}$$

where μ_1 is the largest eigenvalue of $A^{-1}\Sigma$ and μ_n is the smallest eigenvalue of $A^{-1}\Sigma$.

<u>Proof</u>: This follows immediately from Theorem 1.

Consider $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ where Σ_{11} is $p \times p$ and Σ_{22} is $q \times q$ with $p + q = n$. As is well known, (Anderson (1958), p 289 - or Eaton (1972) Chapter 10) the largest canonical correlation coefficient, say θ_1, is given by

$$(12) \qquad \theta_1 = \sup_{\substack{0 \neq a \in R^p \\ 0 \neq b \in R^q}} \frac{a' \Sigma_{12} b}{\sqrt{a' \Sigma_{11} a \, b' \Sigma_{22} b}}$$

<u>Theorem 2</u>: For any partitioning of Σ,

$$(13) \qquad \theta_1 \leq \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}$$

where $\lambda_1 \geq \ldots \geq \lambda_n > 0$ are the eigenvalues of Σ.

<u>Proof</u>: For $a \in R^p$ and $b \in R^q$, set $a^* = \binom{a}{o} \in R^n$ and $b^* = \binom{o}{b} \in R^n$. Then we have

$$(14) \qquad \theta_1 = \sup_{\substack{0 \neq a^* \in R^n \\ 0 \neq b^* \in R^n}} \frac{a^{*\prime} \Sigma b^*}{\sqrt{a^{*\prime} \Sigma a^* \, b^{*\prime} \Sigma b^*}}$$

$$\leq \sup_{x,y \in M} \frac{x' \Sigma y}{\sqrt{x' \Sigma x \, y' \Sigma y}} = \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}$$

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 (13) was also established by Haberman (1974) using a different method.

To show that the inequality (13) is sharp, consider $p \le q$ and

$$(15) \qquad \Sigma = \begin{pmatrix} I_p & (D_\theta \ 0) \\ \begin{pmatrix} D_\theta \\ 0 \end{pmatrix} & I_q \end{pmatrix}$$

where D_θ: $p \times p$ is diagonal with diagonal entries $1 \ge \theta_1 \ge \theta_2 \ge \dots \ge \theta_p \ge 0$. For Σ partitioned as in (15), θ_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 (13) is sharp. One can also show that when $p \ge 2$ and for Σ given in (15), we have $\theta_2 = (\lambda_2 - \lambda_{n-1})/(\lambda_2 + \lambda_{n-1})$. This might lead one to conjecture that for general Σ and $p \ge 2$, $q \ge 2$, the inequality $\theta_2 \le (\lambda_2 - \lambda_{n-1})/(\lambda_2 + \lambda_{n-1})$ holds. However, it is possible to construct a 4×4 Σ where the inequality does not hold.

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. of Math. Anal. and Appl. 8, p. 503-14.