PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 25, Fasc. 2 (2005), pp. 223–229

GENERALIZED SEPARATION THEOREMS FOR SINGULAR VALUES OF A MATRIX AND THEIR APPLICATIONS IN CANONICAL CORRELATION ANALYSIS

BY

C. XIE (Hamilton, Ontario), R. S. SINGH (Guelph, Ontario),
B. SMITH (Halifax, Nova Scotia) and X. LU* (Calgary, Alberta)

Abstract. Rao (1979) established separation theorems for singular values of a matrix and showed their applications in multivariate analysis. In this paper, we provide generalized separation theorems for singular values of a matrix and use them to find some interesting relations between canonical correlations and conditional canonical correlations.

2000 AMS Subject Classification: 62H20, 15A18.

Key words and phrases: Canonical correlations; conditional canonical correlations; singular values.

1. INTRODUCTION

Suppose $(X', Y', Z')'$ has a joint multinormal distribution, where X, Y and Z are $p\times 1$, $q\times 1$ and $s\times 1$ random vectors, respectively. The canonical correlations between Y and Z might be different from the canonical correlations between Y and Z when X is given. The latter will be called *conditional canonical correlations.* In this paper, we generalize the separation theorems given by Rao [2], and then use them to establish the relations between canonical correlations and conditional canonical correlations.

The following notation is used throughout the paper. The singular values of a matrix A are denoted by $\sigma_1(A)\geqslant\sigma_2(A)\geqslant\ldots$, and the eigenvalues of A which is Hermitian are denoted by $\lambda_1(A)\geqslant\lambda_2(A)\geqslant\ldots$ We use A', $r(A)$ and A^* to denote the transpose, the rank and the complex conjugate of A, respectively. A real-valued function $\|\cdot\|$, defined on the space $S_{m\times n}$ of $m\times n$ complex

* Xuewen Lu's research was supported in part by the NSERC Discovery Grant 73-1069 of Canada.

matrices, is called a *unitarily invariant norm* if it satisfies the following conditions:

(i) $\|X\| > 0$ if $X \neq 0$;

(ii) $\|cX\| = |c| \cdot \|X\|$;

(iii) $\|X+Y\| \leqslant \|X\| + \|Y\|$;

(iv) $\|VXU\| = \|X\|$ for any unitary matrices V and U of orders m and n, respectively.

We present the main results in Section 2. Applications of these results in canonical correlation analysis are considered in Section 3. The proofs of all theorems are given in Section 4.

2. MAIN RESULTS

First, we state the Separation Theorem for Singular Values (STSV) of a matrix by Rao [2]. STSV has been used successfully to solve some problems in multivariate analysis.

THEOREM 1 (see Rao [2]). *Let A be $m\times n$, B be $m\times r$ and C be $n\times k$ matrices such that $B^*B = I_r$ and $C^*C = I_k$. Then*

$$\sigma_{t+i}(A) \leqslant \sigma_i(B^*AC) \leqslant \sigma_i(A), \tag{1}$$

where $i = 1, \ldots, \min(r, k)$ and $t = m+n-r-k$.

Now we state the Generalized Separation Theorem for Singular Values (GSTSV) of a matrix. GSTSV can be considered as a generalization of STSV if we only take into account the right inequality of (1), since the right inequality of STSV is a special case of GSTSV, when $\sigma_1(B) = 1$ and $\sigma_1(C) = 1$.

THEOREM 2. *Let A be $m\times n$, B be $m\times r$ and C be $n\times k$ matrices such that $\sigma_1(B) \leqslant 1$ and $\sigma_1(C) \leqslant 1$. Then*

$$\sigma_i(B^*AC) \leqslant \sigma_i(A),$$

where $i = 1, 2, \ldots, h$ and $h = \min(r, k, m, n)$.

COROLLARY 2.1. *Let A be $m\times n$, B be $m\times m$ and C be $n\times n$ matrices such that $\sigma_1(B) \leqslant 1$ and $\sigma_1(C) \leqslant 1$. Then*

$$\|B^*AC\| \leqslant \|A\|$$

for any unitarily invariant norm.

The result follows from Theorem 2 and Lemma 4.1 in Section 4.

COROLLARY 2.2. *Let M be $m\times m$ and N be $n\times n$ nonsingular matrices. Let A be $m\times n$, B be $m\times r$ and C be $n\times k$ matrices such that $\sigma_1(B^*M) \leqslant 1$ and*

$\sigma_1(NC) \leqslant 1$. *Then*

$$\sigma_i(B^* AC) \leqslant \sigma_i(M^{-1} AN^{-1}),$$

where $i = 1, 2, \ldots, h$ *and* $h = \min(m, n, r, k)$.

Proof. Using Theorem 2, we have

$$\sigma_i(B^* AC) = \sigma_i(B^* MM^{-1} AN^{-1} NC) \leqslant \sigma_i(M^{-1} AN^{-1}),$$

where $i = 1, 2, \ldots, h$.

COROLLARY 2.3. *Let* M *be* $m \times m$ *and* N *be* $n \times n$ *nonsingular matrices. Let* A *be* $m \times n$, B *be* $m \times m$ *and* C *be* $n \times n$ *matrices such that* $\sigma_1(B^* M) \leqslant 1$ *and* $\sigma_1(NC) \leqslant 1$. *Then*

$$\|B^* AC\| \leqslant \|M^{-1} AN^{-1}\|$$

for any unitarily invariant norm.

This result follows from Corollary 2.2 and Lemma 4.1 in Section 4.

3. APPLICATIONS IN CANONICAL CORRELATION ANALYSIS

In this section, we will discuss some applications of GSTSV in canonical correlation analysis. In these applications, the canonical correlations turn out to be singular values of some matrices.

Let X be $p \times 1$, Y be $q \times 1$ and Z be $s \times 1$ random vectors. Now, we suppose $(X', Y', Z')'$ has the normal distribution and its dispersion matrix is

$$D\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{pmatrix},$$

where $D(X) = \Sigma_{11}$, $D(Y) = \Sigma_{22}$, $D(Z) = \Sigma_{33}$, $\mathrm{COV}(X, Y) = \Sigma_{12}$, $\mathrm{COV}(X, Z) = \Sigma_{13}$, and $\mathrm{COV}(Y, Z) = \Sigma_{23}$.

Canonical correlation analysis is a method of summarizing relationships between two sets of variables. The objective is to find linear combinations of one set of variables which are most highly correlated with linear combinations of a second set of variables. Here we consider the relations between two sets of canonical correlations: one between Y and Z, the other between $Y \mid X = x$ and $Z \mid X = x$.

THEOREM 3. *Let* $(X', Y', Z')'$ *be as defined before, and we assume*

$$\min\big(r(\Sigma_{21}), r(\Sigma_{13})\big) \leqslant k \quad or \quad r(\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{13}) \leqslant k.$$

Then

$$\varrho_{Y,Z|X=x}(i) \geqslant \varrho_{Y,Z}(i+k), \tag{2}$$

where $\varrho_{Y,Z}(i)$ *and* $\varrho_{Y,Z|X=x}(i)$ *denote the ith canonical correlation between Y and Z and the ith canonical correlation between* $Y\,|\,X = x$ *and* $Z\,|\,X = x$, *respectively.*

THEOREM 4. *Let* $(X', Y', Z')'$ *be as defined before and X and Z be independent random variables. Then*

$$\varrho_{Y,Z|X=x}(i) \geqslant \varrho_{Y,Z}(i), \tag{3}$$

where $\varrho_{Y,Z}(i)$ *and* $\varrho_{Y,Z|X=x}(i)$ *are defned in Theorem* 3.

4. PROOFS OF THEOREMS

The following two lemmas will be used in the proofs of main theorems in this paper.

LEMMA 4.1 (see Rao [2]). *Let* X_i *be a matrix with singular values* $\sigma_{1i} \geqslant \ldots \geqslant \sigma_{ri}$ *for* $i = 1, 2$. *Then* $\|X_1\| \geqslant \|X_2\|$ *for any unitarily invariant norm if and only if*

$$\sigma_{11} + \ldots + \sigma_{k1} \geqslant \sigma_{12} + \ldots + \sigma_{k2}, \quad k = 1, 2, \ldots, r.$$

LEMMA 4.2 (see Gel'fand and Naimark [1]). *Let A and B be* $n \times n$ *complex matrices. Then*

$$\prod_{s=1}^{k} \sigma_{i_s}(AB) \leqslant \prod_{s=1}^{k} \sigma_{i_s}(A)\,\sigma_s(B), \quad 1 \leqslant i_1 \leqslant \ldots \leqslant i_k \leqslant n, \text{ and } k = 1, 2, \ldots, n,$$

with equality for $k = n$. *Especially, for* $k = 1$, *we have*

$$\sigma_i(AB) \leqslant \sigma_i(A)\,\sigma_1(B), \quad i = 1, 2, \ldots, n.$$

Proof of Theorem 2. Let $p = \max(r, k, m, n)$,

$$D = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}_{p \times p}, \quad E^* = \begin{pmatrix} B^* & 0 \\ 0 & 0 \end{pmatrix}_{p \times p} \quad \text{and} \quad F = \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix}_{p \times p},$$

where 0's are matrices of appropriate order and consist of zeroes only. Then, for $i = 1, 2, \ldots, h$, we obtain

$$E^* DF = \begin{pmatrix} B^* AC & 0 \\ 0 & 0 \end{pmatrix}_{p \times p},$$

$$\sigma_i(D) = \sigma_i(A), \quad \sigma_i(E^*) = \sigma_i(B),$$

$$\sigma_i(F) = \sigma_i(C) \quad \text{and} \quad \sigma_i(E^* DF) = \sigma_i(B^* AC).$$

Using Lemma 4.2, we have

$$\sigma_i(E^*DF) \leqslant \sigma_1(E^*)\,\sigma_i(DF) \leqslant \sigma_1(E^*)\,\sigma_i(D)\,\sigma_1(F),$$

where $i = 1, 2, \ldots, p$. Therefore

$$\sigma_i(B^* AC) = \sigma_i(E^* DF) \leqslant \sigma_1(E^*)\sigma_i(D)\sigma_1(F)$$
$$= \sigma_1(B^*)\sigma_i(A)\sigma_1(C) \leqslant \sigma_i(A),$$

where $i = 1, 2, \ldots, h$.

In order to prove Theorems 3 and 4, we quote two lemmas:

LEMMA 4.3 (see Rao [2]). *Let A be an $m \times n$ matrix of rank r and B be an $m \times n$ matrix of rank $\leqslant k$. Then*

$$\sigma_i(A-B) \geqslant \sigma_{k+i}(A) \tag{4}$$

for any i, where $\sigma_{k+i}(A)$ is defined to be zero for $i+k > r$. The equality of (4) *is attained for all i if and only if $k \leqslant r$ and*

$$B = \sigma_1 P_1 Q_1^* + \ldots + \sigma_k P_k Q_k^*,$$

while the singular value decomposition of A is

$$A = \sigma_1 P_1 Q_1^* + \ldots + \sigma_r P_r Q_r^*.$$

LEMMA 4.4 (see Srivastava and Carter [3]). *Let*

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_p\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right) \quad \text{and} \quad \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} > 0,$$

where X_1, X_2, μ_1, μ_2, Σ_{11} and Σ_{22} are $r \times 1$, $s \times 1$, $r \times 1$, $s \times 1$, $r \times r$ and $s \times s$ matrices, respectively. Then the conditional distribution of X_1, given X_2, is

$$N_r(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2), \Sigma_{1.2}), \quad \text{where } \Sigma_{1.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} > 0.$$

Proof of Theorem 3. Using Lemma 4.4, we obtain the dispersion matrix of $(Y', Z')'$, given $X = x$, in the form

$$D\begin{pmatrix} Y \mid X = x \\ Z \mid X = x \end{pmatrix} = \begin{pmatrix} \Sigma_{22} & \Sigma_{23} \\ \Sigma_{32} & \Sigma_{33} \end{pmatrix} - \begin{pmatrix} \Sigma_{21} \\ \Sigma_{31} \end{pmatrix} \Sigma_{11}^{-1} (\Sigma_{12} \quad \Sigma_{13})$$
$$= \begin{pmatrix} \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} & \Sigma_{23} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{13} \\ \Sigma_{32} - \Sigma_{31}\Sigma_{11}^{-1}\Sigma_{12} & \Sigma_{33} - \Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13} \end{pmatrix} \geqslant 0.$$

Then the canonical correlations between Y and Z are the singular values of $\Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33}^{-1/2}$ and the canonical correlations between $Y \mid X = x$ and $Z \mid X = x$ are singular values of

$$(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1/2}(\Sigma_{23} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{13})(\Sigma_{33} - \Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13})^{-1/2}.$$

It is easy to see that

$$\sigma_1(\Sigma_{22}^{-1/2}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{1/2}) = \sqrt{\lambda_1(\Sigma_{22}^{-1/2}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})\Sigma_{22}^{-1/2})}$$
$$= \sqrt{\lambda_1(I - \Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1/2})} \leqslant 1.$$

In the same way we can get

$$\sigma_1\left((\Sigma_{33}-\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13})^{1/2}\Sigma_{33}^{-1/2}\right) \leqslant 1.$$

By Theorem 2, we have

$$\begin{aligned}(5)\quad & \sigma_i(\Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33}^{-1/2}-\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33}^{-1/2})\\ &= \sigma_i\big(\Sigma_{22}^{-1/2}(\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{1/2}(\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1/2}\\ &\quad \times(\Sigma_{23}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{13})(\Sigma_{33}-\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13})^{-1/2}(\Sigma_{33}-\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13})^{1/2}\Sigma_{33}^{-1/2}\big)\\ &\leqslant \sigma_i\big((\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1/2}(\Sigma_{23}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{13})(\Sigma_{33}-\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13})^{-1/2}\big)\\ &= \varrho_{Y,Z|X=x}(i).\end{aligned}$$

Since

$$\begin{aligned} r(\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33}^{-1/2}) &= r(\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{13})\\ &\leqslant \min\big(r(\Sigma_{21}),\, r(\Sigma_{13})\big) \leqslant k,\end{aligned}$$

by Lemma 4.3 we can get

$$\begin{aligned}(6)\quad & \sigma_i(\Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33}^{-1/2}-\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33}^{-1/2})\\ &\geqslant \sigma_{i+k}(\Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33}^{-1/2}) = \varrho_{Y,Z}(i+k).\end{aligned}$$

Then (5) and (6) together implies

$$\varrho_{Y,Z|X=x}(i) \geqslant \varrho_{Y,Z}(i+k).$$

Proof of Theorem 4. Since X and Z are independent random variables, we have

$$\Sigma_{13}=0, \qquad \Sigma_{31}=0.$$

By Lemma 4.2, we obtain

$$\begin{aligned}\varrho_{Y,Z|X=x}(i) &= \sigma_i\big((\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1/2}\Sigma_{23}\Sigma_{33}^{-1/2}\big)\\ &\geqslant \sigma_i\big(\Sigma_{22}^{-1/2}(\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{1/2}(\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1/2}\Sigma_{23}\Sigma_{33}^{-1/2}\big)\\ &= \sigma_i(\Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33}^{-1/2}) = \varrho_{Y,Z}(i).\end{aligned}$$

REFERENCES

[1] I. M. Gel'fand and M. A. Naimark, *The relation between the unitary representations of the complex unimodular group and its unitary subgroup*, Izv. Akad. Nauk SSSR Ser. Mat. 14 (1950), pp. 239–260.

[2] C. R. Rao, *Separation theorems for singular values of matrices and their applications in multivariate analysis*, J. Multivariate Anal. 9 (1979), pp. 362–377.

[3] M. S. Srivastava and E. M. Carter, *An Introduction to Applied Multivariate Statistics*, North-Holland, New York 1983.

Corresponding author:
Changchun Xie
PHRI, Department of Medicine
McMaster University
Hamilton, Ontario, L8L 2X2
Canada
E-mail: xiech@mcmaster.ca

R. S. Singh
Department of Mathematics & Statistics
University of Guelph
Guelph, Ontario, N1G 2W1
Canada
E-mail: rssingh@uoguelph.ca

Bruce Smith
Department of Mathematics & Statistics
Dalhousie University
Halifax, Nova Scotia, B3H 3J5
Canada
E-mail: bsmith@mathstat.dal.ca

Xuewen Lu
Department of Mathematics & Statistics
University of Calgary
Calgary, Alberta, T2N 1N4
Canada
E-mail: lux@math.ucalgary.ca

Received on 7.2.2005;
revised version on 20.7.2005