ELSEVIER Linear Algebra and its Applications 321 (2000) 285–293

LINEAR ALGEBRA AND ITS APPLICATIONS

www.elsevier.com/locate/laa

More on restricted canonical correlations

Matjaž Omladič [a], Vesna Omladič [b]

[a] Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia
[b] Faculty of Social Sciences, University of Ljubljana, Kardeljeva pl. 5, 1000 Ljubljana, Slovenia

Received 18 March 1999; accepted 14 June 2000

Submitted by G.P.H. Styan

Abstract

The problem of the first canonical correlation between two random vectors subject to some natural constraints is treated in the paper. The problem is usually referred to as restricted canonical correlation. A new approach to solving the problem is given by translating it into a generalized eigenvalue problem with an $n \times n$ real symmetric matrix A and a positive definite matrix B of the same size.

1. Introduction

The problem of restricted canonical correlation was introduced in [1]. The problem of finding the first canonical correlation between two random vectors is extended to the case, where some (or all) of the canonical variate coefficients are subject to some constraints. The most natural possible restriction might be to have these coefficients non-negative. The reader is referred to [1] for a set of motivating examples as well as a discussion how some other possible restrictions may be reduced to the case of non-negativity. Das and Sen [1] are solving the problem using the Kuhn–Tucker Lagrangian theory. We propose here a somewhat different approach. Namely, a useful approach to the solution of the standard canonical correlation problem translates the problem to a generalized eigenvalue problem with an $n \times n$ real symmetric matrix A and a positive definite matrix B of the same size. We show that this approach can be extended to the case of restricted canonical correlations as well.

E-mail addresses: matjaz.omladic@fmf.uni-lj.si (M. Omladič), vesna.omladic@uni-lj.si (V. Omladič).

PII: S 0 0 2 4 - 3 7 9 5 (0 0) 0 0 1 9 8 - 1

Our approach to the problem is based on the fact given in Proposition 3.1 that the restricted canonical correlation equals the maximal eigenvalue of the generalized eigenproblem for matrices A_K and B_K obtained, respectively, from the starting matrices A and B by crossing out all rows and columns with indices not in K, where K is the set of indices such that the canonical weights for the variables indexed by this set are strictly positive. This fact reduces the problem to finding the set K of variable indices so that the eigenvector corresponding to the maximal generalized eigenvalue for matrices A_K and B_K has positive entries.

The problem therefore becomes a search problem over $2^m - 1$ sets of indices, where m is the total number of variables on which the non-negativity restriction is imposed. In the search process for any of the $2^m - 1$ sets of indices K the corresponding generalized eigenproblem for the symmetric matrix A_K and the positive definite matrix B_K is solved and the solution corresponding to the maximal eigenvalue is tried out to see whether

(a) the corresponding eigenvector has positive entries on K, and

(b) it is a local maximum.

At the end of the process, the global solution is given by the one of the local solutions satisfying both (a) and (b) that has the maximal generalized eigenvalue. In this respect our main results are Propositions 3.1 and 3.2 (where a necessary condition for (b) is given) for the case that restriction is imposed on all variables, and Theorem 3.3, where these results are given for the general case. Namely, if restrictions are not imposed for all n variables in a data set, we denote by L the set of indices of variables on which no non-negativity restrictions are imposed, and by L' the set of indices of variables on which they are imposed. In this case the search process goes only through the subsets K of the set L'.

The paper is organized as follows: the problem of restricted canonical correlation is presented in Section 2 together with a reduction of this problem to an optimization problem. Our main results are given in Section 3 as described above. Section 4 gives examples pointing out to some problems in applications of these methods.

2. Applications to statistics

Denote the joint variance–covariance matrix of two random vectors $\boldsymbol{Y}^1$ and $\boldsymbol{Y}^2$ of respective sizes l and m, by

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

Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be two columns of constants of respective sizes l and m, and let the random variables X_1 and X_2 be defined as the respective linear combinations of the random variables from these two sets with these constants as coefficients, i.e.,

$X_1 = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{Y}^1$ and $X_2 = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{Y}^2$. Then, the covariance of these two variables is given as $\operatorname{cov}(X_1, X_2) = \boldsymbol{a}^{\mathrm{T}}\Sigma_{12}\boldsymbol{b}$, while their correlation coefficient is

$$\rho\,(X_1, X_2) \;=\; \frac{\boldsymbol{a}^{\mathrm{T}}\Sigma_{12}\boldsymbol{b}}{\sqrt{\boldsymbol{a}^{\mathrm{T}}\Sigma_{11}\boldsymbol{a}}\,\sqrt{\boldsymbol{b}^{\mathrm{T}}\Sigma_{22}\boldsymbol{b}}}.$$

The classical method of canonical correlations searches for the maximal among these correlations when columns $\boldsymbol{a}$ and $\boldsymbol{b}$ run over all possible choices, while the method of *restricted canonical correlations* (as introduced in [1]) searches for the maximal among the correlations when columns $\boldsymbol{a}$ and $\boldsymbol{b}$ run over all possible non-negative choices. Of course, only for some of the coefficients the non-negativity restrictions may be imposed.

Let us now extend the standard techniques to translate this problem into an optimization problem to be considered in Section 3. Let L be a subset of the set $\{1, 2, \dots, l\}$, and let M be a subset of the set $\{1, 2, \dots, m\}$. Denote by L' the complement of the set L in $\{1, 2, \dots, l\}$ and by M' the complement of the set M in $\{1, 2, \dots, m\}$. Moreover, we introduce regions R and S in the l- and m-dimensional, real vector space, respectively, defined by

$$R = \left\{\boldsymbol{c} = (c_i)_{i=1}^{l};\; c_i \geqslant 0 \text{ for } i \in L'\right\}$$

and

$$S = \left\{\boldsymbol{d} = (d_i)_{i=1}^{m};\; d_i \geqslant 0 \text{ for } i \in M'\right\}.$$

We want to find a pair of vectors $\boldsymbol{a} \in R$ and $\boldsymbol{b} \in S$ satisfying the condition

$$\frac{\boldsymbol{a}^{\mathrm{T}}\Sigma_{12}\boldsymbol{b}}{\sqrt{\boldsymbol{a}^{\mathrm{T}}\Sigma_{11}\boldsymbol{a}}\sqrt{\boldsymbol{b}^{\mathrm{T}}\Sigma_{22}\boldsymbol{b}}} = \max_{\boldsymbol{c}\in R,\ \boldsymbol{d}\in S} \frac{\boldsymbol{c}^{\mathrm{T}}\Sigma_{12}\boldsymbol{d}}{\sqrt{\boldsymbol{c}^{\mathrm{T}}\Sigma_{11}\boldsymbol{c}}\sqrt{\boldsymbol{d}^{\mathrm{T}}\Sigma_{22}\boldsymbol{d}}}. \tag{1}$$

It is not difficult to see that solutions of this equation exist. Namely, observe that quotient (1) does not change if we multiply either $\boldsymbol{a}$ or $\boldsymbol{b}$ by a positive constant. So let us restrict ourselves with no loss of generality to vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ satisfying an additional condition $\sqrt{\boldsymbol{a}^{\mathrm{T}}\Sigma_{11}\boldsymbol{a}} = \sqrt{\boldsymbol{b}^{\mathrm{T}}\Sigma_{22}\boldsymbol{b}} = 1$. Since the set of vectors satisfying all these conditions is compact, and since a continuous function always attains its supremum on a compact set, the conclusion follows.

Define $n = l + m$ and introduce a real symmetric $n \times n$ matrix A and a positive definite $n \times n$ matrix B by

$$A = \begin{bmatrix} 0 & \Sigma_{12} \\ \Sigma_{21} & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}.$$

Take arbitrary columns $\boldsymbol{a}$ and $\boldsymbol{b}$ and define

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix}.$$

Furthermore, denote by T the set of all $\boldsymbol{x}$, with the property that with respect to this block partition vector $\boldsymbol{a}$ belongs to R and vector $\boldsymbol{b}$ belongs to S. Then, we have:

Lemma 2.1. *A column $\boldsymbol{x} \in T$ is a solution of*

$$\frac{\boldsymbol{x}^{\mathrm{T}} A\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}} B\boldsymbol{x}} = \max_{y\in T} \frac{\boldsymbol{y}^{\mathrm{T}} A\boldsymbol{y}}{\boldsymbol{y}^{\mathrm{T}} B\boldsymbol{y}}. \tag{2}$$

if and only if the columns $\boldsymbol{a}$ and $\boldsymbol{b}$ are the solutions of (1).

Proof. Let us first rewrite (2) into

$$\frac{\boldsymbol{a}^{\mathrm{T}}\Sigma_{12}\boldsymbol{b} + \boldsymbol{b}^{\mathrm{T}}\Sigma_{21}\boldsymbol{a}}{\boldsymbol{a}^{\mathrm{T}}\Sigma_{11}\boldsymbol{a} + \boldsymbol{b}^{\mathrm{T}}\Sigma_{22}\boldsymbol{b}} = \max_{\boldsymbol{c}\in R,\ \boldsymbol{d}\in S} \frac{\boldsymbol{c}^{\mathrm{T}}\Sigma_{12}\boldsymbol{d} + \boldsymbol{d}^{\mathrm{T}}\Sigma_{21}\boldsymbol{c}}{\boldsymbol{c}^{\mathrm{T}}\Sigma_{11}\boldsymbol{c} + \boldsymbol{d}^{\mathrm{T}}\Sigma_{22}\boldsymbol{d}}, \tag{2'}$$

using the above definitions. Denote by λ the solution of problem (1) and by μ the solution of problem (2) which is also equal to the solution of problem $(2')$. Let a pair of vectors $\boldsymbol{a} \in R$ and $\boldsymbol{b} \in S$ satisfy (1) and observe that by multiplying either of the two vectors by a positive constant, the quotient on the left-hand side of (1) does not change. So we may assume with no loss of generality that $\sqrt{\boldsymbol{a}^{\mathrm{T}}\Sigma_{11}\boldsymbol{a}} = \sqrt{\boldsymbol{b}^{\mathrm{T}}\Sigma_{22}\boldsymbol{b}}$. Denote this constant by α to get

$$\frac{\boldsymbol{a}^{\mathrm{T}}\Sigma_{12}\boldsymbol{b} + \boldsymbol{b}^{\mathrm{T}}\Sigma_{21}\boldsymbol{a}}{\boldsymbol{a}^{\mathrm{T}}\Sigma_{11}\boldsymbol{a} + \boldsymbol{b}^{\mathrm{T}}\Sigma_{22}\boldsymbol{b}} = \frac{2\boldsymbol{a}^{\mathrm{T}}\Sigma_{12}\boldsymbol{b}}{2\alpha^2} = \frac{\boldsymbol{a}^{\mathrm{T}}\Sigma_{12}\boldsymbol{b}}{\sqrt{\boldsymbol{a}^{\mathrm{T}}\Sigma_{11}\boldsymbol{a}}\sqrt{\boldsymbol{b}^{\mathrm{T}}\Sigma_{22}\boldsymbol{b}}}.$$

So λ is no greater than μ. Now, let $\boldsymbol{x}$ be a solution of problem (2) and let the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ be its parts with respect to the above block partition. If we introduce $\boldsymbol{c} = s\boldsymbol{a}$ and $\boldsymbol{d} = t\boldsymbol{b}$ for some positive constants s and t, we get

$$\frac{\boldsymbol{c}^{\mathrm{T}}\Sigma_{12}\boldsymbol{d} + \boldsymbol{d}^{\mathrm{T}}\Sigma_{21}\boldsymbol{c}}{\boldsymbol{c}^{\mathrm{T}}\Sigma_{11}\boldsymbol{c} + \boldsymbol{d}^{\mathrm{T}}\Sigma_{22}\boldsymbol{d}} = \frac{st\left[\boldsymbol{a}^{\mathrm{T}}\Sigma_{12}\boldsymbol{b} + \boldsymbol{b}^{\mathrm{T}}\Sigma_{21}\boldsymbol{a}\right]}{s^2\left[\boldsymbol{a}^{\mathrm{T}}\Sigma_{11}\boldsymbol{a}\right] + t^2\left[\boldsymbol{b}^{\mathrm{T}}\Sigma_{22}\boldsymbol{b}\right]}.$$

The quotient on the right-hand side is a function of s and t. A standard computation reveals that it has its maximum at

$$\frac{s}{t} = \frac{\sqrt{\boldsymbol{b}^{\mathrm{T}}\Sigma_{22}\boldsymbol{b}}}{\sqrt{\boldsymbol{a}^{\mathrm{T}}\Sigma_{11}\boldsymbol{a}}},$$

so that

$$\mu = \frac{\boldsymbol{a}^{\mathrm{T}}\Sigma_{12}\boldsymbol{b}}{\sqrt{\boldsymbol{a}^{\mathrm{T}}\Sigma_{11}\boldsymbol{a}}\,\sqrt{\boldsymbol{b}^{\mathrm{T}}\Sigma_{22}\boldsymbol{b}}}$$

is no greater than λ, and consequently, they are equal. $\square$

3. Solving the optimization problem

For any $n \times n$ real symmetric matrix $A = (a_{ij})$ and any subset K of the set $\{1, 2, \ldots, n\}$, let A_K be the principal submatrix of matrix A made of components a_{ij} with both indices i and j in K. Similarly, for any n-tuple $\boldsymbol{x} = (x_i)$ denote by $\boldsymbol{x}_K$ the subvector made of components x_i such that i belongs to K.

Fix now a real symmetric $n \times n$ matrix A and a positive definite matrix B. We would like to find a vector $\boldsymbol{x} \geqslant 0$ satisfying the condition

$$\frac{\boldsymbol{x}^{\mathrm{T}} A\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}} B\boldsymbol{x}} = \max_{y \geqslant 0} \frac{\boldsymbol{y}^{\mathrm{T}} A\boldsymbol{y}}{\boldsymbol{y}^{\mathrm{T}} B\boldsymbol{y}}. \tag{3}$$

We can see that this equation has a solution $\boldsymbol{x}$ such that $\boldsymbol{x} \geqslant 0$. In the proof of this fact we may restrict ourselves with no loss of generality to vectors $\boldsymbol{x}$ satisfying an additional condition $\boldsymbol{x}^{\mathrm{T}} B\boldsymbol{x} = 1$. Since the set of vectors satisfying both $\boldsymbol{x} \geqslant 0$ and $\boldsymbol{x}^{\mathrm{T}} B\boldsymbol{x} = 1$ is compact, and since a continuous function always attains its supremum on a compact set, the conclusion follows. In the following proposition, let $\boldsymbol{x} \geqslant 0$ be a solution of (3) and let K be a subset of $\{1, 2, \ldots, n\}$ made of indices i such that $x_i > 0$. It is clear that in this case $\boldsymbol{x}_K$ solves the problem

$$\frac{\boldsymbol{x}_K^{\mathrm{T}} A_K \boldsymbol{x}_K}{\boldsymbol{x}_K^{\mathrm{T}} B_K \boldsymbol{x}_K} = \max_{y_K \geqslant 0} \frac{\boldsymbol{y}_K^{\mathrm{T}} A_K \boldsymbol{y}_K}{\boldsymbol{y}_K^{\mathrm{T}} B_K \boldsymbol{y}_K}$$

with $\boldsymbol{x}_K \geqslant 0$.

Proposition 3.1. *Under the above assumptions $\lambda = (\boldsymbol{x}^{\mathrm{T}} A\boldsymbol{x})/(\boldsymbol{x}^{\mathrm{T}} B\boldsymbol{x})$ is the maximal generalized eigenvalue of the generalized eigenvalue problem $A_K \boldsymbol{x}_K = \lambda B_K \boldsymbol{x}_K$ with $\boldsymbol{x}_K$ equal to a corresponding eigenvector.*

Proof. Introduce the spectral decomposition of the symmetric matrix

$$B_K^{-1/2} A_K B_K^{-1/2} = \Sigma_r \lambda_r P_r,$$

where λ_r are its eigenvalues and P_r are the corresponding (necessarily symmetric) spectral idempotents, whose total sum equals the identity matrix I. This implies that

$$\lambda = \frac{\boldsymbol{x}_K^{\mathrm{T}} A_K \boldsymbol{x}_K}{\boldsymbol{x}_K^{\mathrm{T}} B_K \boldsymbol{x}_K} = \sum_r \lambda_r \frac{\boldsymbol{x}_K^{\mathrm{T}} B_K^{1/2} P_r B_K^{1/2} \boldsymbol{x}_K}{\boldsymbol{x}_K^{\mathrm{T}} B_K \boldsymbol{x}_K}$$

is a convex combination of the eigenvalues of $B_K^{-1/2} A_K B_K^{-1/2}$. It follows that λ is no greater than the maximal of these eigenvalues. Assume that the eigenvalues are indexed in decreasing order so that λ_0 is the maximal one. Let $\boldsymbol{y}_K$ be an eigenvector corresponding to this eigenvalue and assume with no loss of generality $\boldsymbol{y}_K^{\mathrm{T}} \boldsymbol{y}_K = 1$. Denote $\gamma = \boldsymbol{x}_K^{\mathrm{T}} B_K \boldsymbol{x}_K$ and introduce

$$\boldsymbol{z}_K = \boldsymbol{x}_K \cos\varphi + B_K^{-1/2} \boldsymbol{y}_K \sin\varphi.$$

Observe that $\boldsymbol{z}_K$ is as close to $\boldsymbol{x}_K$ as we want, if φ is close enough to 0. It is also clear that $\boldsymbol{z}_K$ has strictly positive entries on K for φ close enough to 0, because $\boldsymbol{x}_K$ satisfies this condition. So

$$\boldsymbol{z}_K^{\mathrm{T}} B_K \boldsymbol{z}_K = \gamma \cos^2\varphi + 2\boldsymbol{x}_K^{\mathrm{T}} B_K^{1/2} \boldsymbol{y}_K \cos\varphi \sin\varphi + \sin^2\varphi,$$

and

$$\begin{aligned}
\boldsymbol{z}_K^{\mathrm{T}} A_K \boldsymbol{z}_K = & \lambda\gamma \cos^2\varphi \\
& + 2\boldsymbol{x}_K^{\mathrm{T}} B_K^{1/2} B_K^{-1/2} A_K B_K^{-1/2} \boldsymbol{y}_K \cos\varphi \sin\varphi \\
& + \boldsymbol{y}_K^{\mathrm{T}} B_K^{-1/2} A_K B_K^{-1/2} \boldsymbol{y}_K \sin^2\varphi \\
= & \lambda\gamma \cos^2\varphi \\
& + 2\lambda_0 \boldsymbol{x}_K^{\mathrm{T}} B_K^{1/2} \boldsymbol{y}_K \cos\varphi \sin\varphi \\
& + \lambda_0 \sin^2\varphi.
\end{aligned}$$

Now, if λ_0 were strictly greater than λ, it is clear from these expressions that φ can be chosen on one hand small enough to make the entries of $\boldsymbol{z}_K$ on K strictly positive and on the other hand such that the quotient

$$\frac{\boldsymbol{z}_K^{\mathrm{T}} A_K \boldsymbol{z}_K}{\boldsymbol{z}_K^{\mathrm{T}} B_K \boldsymbol{z}_K}$$

is strictly greater than λ contradicting its (local) maximality. This proves that λ_0 is no greater than λ; so they are equal. Finally, it is easy to see that the eigenvalues of the symmetric matrix $B_K^{-1/2} A_K B_K^{-1/2}$ are the same as the generalized eigenvalues of the generalized eigenvalue problem $A_K \boldsymbol{x}_K = \lambda B_K \boldsymbol{x}_K$, where the corresponding eigenvector $\boldsymbol{x}_K$ is in the relation $\boldsymbol{x}_K = B_K^{-1/2} \boldsymbol{y}_K$ with the eigenvector $\boldsymbol{y}_K$ of the matrix $B_K^{-1/2} A_K B_K^{-1/2}$. $\square$

Proposition 3.2. *Under the assumptions of Proposition* 3.1 *it holds that* $(A - \lambda B)\boldsymbol{x} \leqslant 0$ *and the set of indices where this vector is strictly negative is disjoint with* K.

Proof. In the proof of necessity, it suffices to consider the case when K contains all but one index. So let $K = \{1, 2, \ldots, n-1\}$ and write A, B, and $\boldsymbol{x}$ in a block form with respect to this set K:

$$A = \begin{bmatrix} A_K & \boldsymbol{b} \\ \boldsymbol{b}^{\mathrm{T}} & \gamma \end{bmatrix}, \qquad B = \begin{bmatrix} B_K & \boldsymbol{c} \\ \boldsymbol{c}^{\mathrm{T}} & \delta \end{bmatrix}, \quad \text{and} \quad \boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_K \\ 0 \end{bmatrix}.$$

By Proposition 3.1 we have

$$(A - \lambda B)\boldsymbol{x} = \begin{bmatrix} 0 \\ (\boldsymbol{b} - \lambda \boldsymbol{c})^{\mathrm{T}} \boldsymbol{x}_K \end{bmatrix},$$

so that the set of indices where the vector $(A - \lambda B)\boldsymbol{x}$ is non-zero is disjoint with K. It remains to show that $(\boldsymbol{b} - \lambda \boldsymbol{c})^{\mathrm{T}} \boldsymbol{x}_K$ is non-positive. To this end assume with no loss of generality that $\boldsymbol{x}^{\mathrm{T}} B \boldsymbol{x} = 1$ and define for any $\varphi \in [0, \pi/2]$ the vector

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{x}_K \cos\varphi \\ \sin\varphi \end{bmatrix}.$$

It is clear that $\boldsymbol{y} \geqslant 0$. Also we have that

$$\boldsymbol{y}^{\mathrm{T}} B \boldsymbol{y} = \cos^2 \varphi + 2\boldsymbol{c}^{\mathrm{T}} \boldsymbol{x}_K \cos \varphi \sin \varphi + \delta \sin^2 \varphi,$$

and that

$$\boldsymbol{y}^{\mathrm{T}} A \boldsymbol{y} = \lambda \cos^2 \varphi + 2\boldsymbol{b}^{\mathrm{T}} \boldsymbol{x}_K \cos \varphi \sin \varphi + \gamma \sin^2 \varphi.$$

An elementary computation reveals that the first derivative of the quotient $(\boldsymbol{y}^{\mathrm{T}} A\boldsymbol{y})/(\boldsymbol{y}^{\mathrm{T}} B\boldsymbol{y})$ as a function of φ at $\varphi = 0$ equals $2\,(\boldsymbol{b} - \lambda\boldsymbol{c})^{\mathrm{T}}\boldsymbol{x}_K$. Now, if $(\boldsymbol{b} - \lambda\boldsymbol{c})^{\mathrm{T}}\boldsymbol{x}_K$ were strictly positive, the quotient $(\boldsymbol{y}^{\mathrm{T}} A\boldsymbol{y})/(\boldsymbol{y}^{\mathrm{T}} B\boldsymbol{y})$, which equals λ at $\varphi = 0$, would be strictly increasing as a function of φ, contradicting the maximality of λ. □

We may also consider a slightly more general optimization problem. Let $A = (a_{ij})$ be an $n \times n$ real symmetric matrix and let L be a subset of the set $\{1, 2, \ldots, n\}$. Denote by L' the complement of the set L in $\{1, 2, \ldots, n\}$. Introduce a region R in the n-dimensional real vector space, defined by

$$R = \left\{\boldsymbol{x} = (x_i)_{i=1}^n;\ x_i \geqslant 0 \text{ for } i \in L'\right\}.$$

We would like to find a vector $\boldsymbol{x} \in R$ satisfying the condition

$$\frac{\boldsymbol{x}^{\mathrm{T}} A\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}} B\boldsymbol{x}} = \max_{y \in R} \frac{\boldsymbol{y}^{\mathrm{T}} A\boldsymbol{y}}{\boldsymbol{y}^{\mathrm{T}} B\boldsymbol{y}}. \tag{4}$$

It is not difficult to see that this equation has a solution $\boldsymbol{x} \in R$. This can be done similarly as above with optimization problem (3). Using similar ideas we can prove even more, namely:

Theorem 3.3. *Let $\boldsymbol{x} \in R$ be any solution of* (4), *let K be a subset of L' made of indices i with $x_i > 0$, and let $\lambda = (\boldsymbol{x}^{\mathrm{T}} A\boldsymbol{x})/(\boldsymbol{x}^{\mathrm{T}} B\boldsymbol{x})$. Then, the following are true*:

1. *The value λ equals the maximal eigenvalue of the generalized eigenvalue problem for matrices $A_{K\cup L}$ and $B_{K\cup L}$, while $\boldsymbol{x}_{K\cup L}$ equals a corresponding eigenvector.*
2. *The set of indices where the vector $(A - \lambda B)\boldsymbol{x} \leqslant 0$ is strictly negative is disjoint with $K \cup L$.*

Proof. The proof of this theorem follows exactly the same steps as the proof of Propositions 3.1 and 3.2. The key observation is the following. It is clear that $\boldsymbol{x}_{K\cup L}$ solves the problem

$$\frac{\boldsymbol{x}_{K\cup L}^{\mathrm{T}}\ A_{K\cup L}\ \boldsymbol{x}_{K\cup L}}{\boldsymbol{x}_{K\cup L}^{\mathrm{T}}\ B_{K\cup L}\ \boldsymbol{x}_{K\cup L}} = \max_{y_{K\cup L} \geqslant 0} \frac{\boldsymbol{y}_{K\cup L}^{\mathrm{T}}\ A_{K\cup L}\ \boldsymbol{y}_{K\cup L}}{\boldsymbol{y}_{K\cup L}^{\mathrm{T}}\ B_{K\cup L}\ \boldsymbol{y}_{K\cup L}}.$$

To get assertion 1, use the spectral decomposition of the symmetric matrix $B_{K\cup L}^{-1/2} A_{K\cup L} B_{K\cup L}^{-1/2}$ similarly as in the proof of Proposition 3.1. This observation also suffices to get assertion 2. Namely, by assertion 1 the vector $(A - \lambda B)\boldsymbol{x}$ has all the components with indices i belonging to the set $K \cup L$ equal to 0. In the proof of the fact that the rest of the components are non-positive we use similar arguments as in the proof of Proposition 3.2. □

4. Counterexamples

The results of Section 3 suggest a simple algorithm for solving the statistical problem of restricted canonical correlations via the proposed optimization problem. Namely, for any choice of subset K of indices L' one could simply compute the maximal generalized eigenvalue λ and corresponding eigenvector of the generalized eigenproblem for matrices $A_{K\cup L}$ and $B_{K\cup L}$. We take under consideration the solutions $\boldsymbol{x}$ such that

(A) $\boldsymbol{x}_K$ has strictly positive entries, and

(B) $(A-\lambda B)\boldsymbol{x} \leqslant 0$.

Choosing the maximal among the so obtained eigenvalues λ brings us to the global solution of the optimization problem, and consequently, with the solution of the restricted canonical correlation problem. If the sum n of the cardinalities of two sets for which the canonical correlation is to be computed is not too big, the method should give a result in real time.

It is plausible that a more sophisticated algorithm exists. However, let us point out some of the problems that occur when searching for a better algorithm. One of the questions is: if for a choice of K the eigenvector $\boldsymbol{x}_{K\cup L}$ of the generalized eigenproblem for matrices $A_{K\cup L}$ and $B_{K\cup L}$ does not satisfy conditions (A) and (B), could we dismiss all sets greater than K from further investigation? The answer to this question is negative as the following example shows.

Define the matrices

$$A = \begin{bmatrix} 6 & 6 & -1 \\ 6 & 6 & 6 \\ -1 & 6 & 6 \end{bmatrix}$$

and $B = 14I$ and consider the problem for the set L being empty. Then, the partial solutions of the generalized eigenvalue problem for singleton sets K are all equal to $\lambda = 3/7$. If we take the sets $\{1\}$, $\{2\}$, and $\{3\}$, respectively, for the set K, we get corresponding eigenvector equal to $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$, respectively, satisfying condition (A), but not conditon (B). For the set K equal to $\{1,3\}$ the maximal eigenvalue is 1/2 with $(1,0,-1)$ as the eigenvector, so that condition (A) cannot be satisfied. For the choices $\{1,2\}$ and $\{2,3\}$ for K we do get an eigenvector with positive entries corresponding to the eigenvalue 6/7, namely, $(1,1,0)$ and $(0,1,1)$, respectively. Observe that condition (A) is satisfied in these two cases, while condition (B) is not satisfied. Notice now that the "greatest" set $\{1,2,3\}$ yields the global maximum 1 with the eigenvector $(2,3,2)$. Observe that in this example the point of global maximum is the only one satisfying both conditions (A) and (B).

So a natural question is whether the two conditions are also sufficient. A possible positive answer to this question would help us improve our method substantially. However, let us give an example showing that this question has a negative answer as well. Consider the matrices

$$A = \begin{bmatrix} 3 & -1 & -1 \\ -1 & 2 & 2 \\ -1 & 2 & 2 \end{bmatrix}$$

and $B = 5I$, and assume again that the set L is empty. Then, the best among the partial solutions of the generalized eigenvalue problem for singleton sets equals $\lambda = 3/5$ on $\{1\}$. This point clearly satisfies the two conditions. The maximal eigenvalue corresponding to the sets $\{1, 2\}$ and $\{1, 3\}$ has eigenvector whose entries cannot be all made strictly positive, while the set $\{2, 3\}$ has eigenvalue 4/5 with corresponding eigenvector $(0, 1, 1)$. This point also satisfies the two conditions. The greatest set $\{1, 2, 3\}$ yields eigenvalue 1 with the eigenvector $(1, -1, -1)$. So the global maximum is attained at $(0, 1, 1)$ with eigenvalue 4/5. Notice that the set $\{2, 3\}$, where we have found the global maximum, is disjoint with the set $\{1\}$, where a local maximum has been found.

Question 1. Is there an easily verifyable set of conditions, necessary and sufficient for a global maximum or at least a local maximum of problem (3) to occur?

Question 2. Is it possible to reduce the complexity of the proposed search process using one of the standard techniques of optimization, such as active constraint strategy [2]?

Acknowledgment

The authors are indebted to the anonymous referee for the suggestions that improved this paper and gave them some ideas for further work on the subject.

References

[1] S. Das, P.K. Sen, Restricted canonical correlations, Linear Algebra Appl. 210 (1994) 29–47.
[2] R. Fletcher, Practical Methods of Optimization. Vol. 2. Constrained Optimization, Wiley, Chichester, 1981.
[3] J.R. Schott, Matrix Analysis for Statistics, Wiley, NewYork, 1997.