J Glob Optim (2011) 49:91–107
DOI 10.1007/s10898-010-9536-6

Towards the global solution of the maximal correlation problem

Lei-Hong Zhang · Li-Zhi Liao · Li-Ming Sun

Received: 31 August 2009 / Accepted: 11 February 2010 / Published online: 26 February 2010
© Springer Science+Business Media, LLC. 2010

Abstract The maximal correlation problem (MCP) aiming at optimizing correlation between sets of variables plays a very important role in many areas of statistical applications. Currently, algorithms for the general MCP stop at solutions of the multivariate eigenvalue problem for a related matrix A, which serves as a necessary condition for the global solutions of the MCP. However, the reliability of the statistical prediction in applications relies greatly on the global maximizer of the MCP, and would be significantly impacted if the solution found is a local maximizer. Towards the global solution of the MCP, we have obtained four results in the present paper. First, the sufficient and necessary condition for global optimality of the MCP when A is a positive matrix is extended to the nonnegative case. Secondly, the uniqueness of the multivariate eigenvalues in the global maxima of the MCP is proved either when there are only two sets of variables involved, or when A is nonnegative. The uniqueness of the global maximizer of the MCP for the nonnegative irreducible case is also proved. These theoretical achievements lead to our third result that if A is a nonnegative irreducible matrix, both the Horst-Jacobi algorithm and the Gauss-Seidel algorithm converge *globally* to the *global* maximizer of the MCP. Lastly, some new estimates of the multivariate eigenvalues related to the global maxima are obtained.

This research was supported in part by grants from Hong Kong Baptist University (FRG) and the Research Grant Council of Hong Kong.

L.-H. Zhang
Department of Applied Mathematics, Shanghai University of Finance and Economics,
777 Guoding Road, 200433 Shanghai, People's Republic of China
e-mail: longzlh@gmail.com

L.-Z. Liao (✉) · L.-M. Sun
Department of Mathematics, Hong Kong Baptist University, Kowloon Tong,
Kowloon, Hong Kong, People's Republic of China
e-mail: liliao@hkbu.edu.hk

L.-M. Sun
e-mail: lmsun@nau.edu.cn

Keywords Multivariate statistics · Canonical correlation · Maximal correlation problem · Multivariate eigenvalue problem · Power method · Gauss-Seidel method · Global maximizer · Nonnegative irreducible matrix

1 Introduction

Assessing the relationship between sets of random variables arises from lots of applications, among which are cluster analysis, data classification, pattern recognition, principal component analysis, and bioinformatics. Canonical correlation analysis (CCA) is an important and effective tool for that purpose; see for example, [5–7,10,15,17]. As early as 1935, based on the observation of many real world examples and phenomena, Hotelling studied how to find the linear combination of one set of variables that correlates maximally with the linear combination of another set of variables, which later is known as the maximal correlation problem (MCP) [8,9]. If the optimal linear combination is successfully found, we then have the advantage of using one set of variables to predict the other. Optimizing correlation between $m > 2$ sets of variables can be easily generalized from the case of two sets of variables, and we recommend [2, Sect. 2] for a brief but comprehensive background information about the MCP. For the purpose of discussion in the present paper, it is enough to only introduce the notation and the optimization problem that is related to the MCP.

Let $A \in \mathbb{R}^{n\times n}$ be a symmetric and positive definite matrix, and

$$\mathcal{P} = \{n_1, n_2, \ldots, n_m\} \tag{1.1}$$

be a set of positive integers with $\sum_{i=1}^m n_i = n$. Partition A and a vector $\mathbf{x} \in \mathbb{R}^n$ into block forms according to $\mathcal{P}$ as follows,

$$A = \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1m} \\ A_{21} & A_{22} & \ldots & A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \ldots & A_{mm} \end{pmatrix} \in \mathbb{R}^{n\times n}, \tag{1.2}$$

$$\mathbf{x} = (\mathbf{x}_1^\top, \ldots, \mathbf{x}_m^\top)^\top \in \mathbb{R}^n, \tag{1.3}$$

with $A_{ij} \in \mathbb{R}^{n_i\times n_j}$ and $\mathbf{x}_i \in \mathbb{R}^{n_i}$, respectively. The optimal solution to the MCP then corresponds to the *global* maximizer of the following equality constrained optimization problem:

$$\begin{cases} \text{Maximize } r(\mathbf{x}) := \mathbf{x}^\top A\mathbf{x} \\ \text{subject to } \ \|\mathbf{x}_i\|_2 = 1, \quad i = 1, 2, \ldots, m. \end{cases} \tag{1.4}$$

Denote the constraint set as

$$\mathcal{M} := \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}_i\|_2 = 1,\ \mathbf{x}_i \in \mathbb{R}^{n_i},\ \text{for } i = 1, 2, \ldots, m.\}. \tag{1.5}$$

It is clear that maximizing $r(\mathbf{x})$ on $\mathcal{M}$ is equivalent to maximizing $\mathbf{x}^\top(A + cI^{[n]})\mathbf{x}$ on $\mathcal{M}$ for any $c \in \mathbb{R}$, where $I^{[n]}$ stands for the identity matrix of order n. Therefore, no generality is lost when A is assumed to be positive definite.

The first order necessary optimal condition for (1.4) is the existence of real scalars $\lambda_1, \ldots, \lambda_m$ (the *Lagrange multipliers*) and a vector $\mathbf{x} \in \mathbb{R}^n$ such that the system of equations

$$\begin{cases} A\mathbf{x} = \Lambda\mathbf{x}, \\ \|\mathbf{x}_i\|_2 = 1, \quad i = 1, 2, \ldots, m, \end{cases} \tag{1.6}$$

is satisfied [2,17], where

$$\Lambda := \text{diag}\{\lambda_1 I^{[n_1]}, \lambda_2 I^{[n_2]}, \ldots, \lambda_m I^{[n_m]}\}. \tag{1.7}$$

The multivariate eigenvalue problem (MEP) is cast exactly as the system of Eq. (1.6), and because of the analogy of the role of Λ in the MEP with the classical eigenvalue, $\lambda_1, \ldots, \lambda_m$ are usually called the *multivariate eigenvalues*. Besides of being a necessary condition for the MCP, recently, from an entirely different and non-statistical setting, the MEP itself finds applications in the perturbation analysis of linear dynamical systems subject to additive bounded noises [18].

Thinking the pair $(\mathbf{x}, \Lambda)$ that solves (1.6) as the solution of the MEP, Chu and Watterson [2] discovered that there are precisely $\prod_{i=1}^{m}(2n_i)$ solutions for the *generic* matrix A whose n eigenvalues are distinct. It can also be verified from (1.6) that the objective function value $r(\mathbf{x})$ at the solution pair $(\mathbf{x}, \Lambda)$ of the MEP is $r(\mathbf{x}) = \sum_{i=1}^{m} \lambda_i$, and the global optimal solution of the MCP corresponds to the one with the largest $\sum_{i=1}^{m} \lambda_i$. Moreover, it is interesting to note that when reducing the MEP to the classical eigenvalue problem by setting $m = 1$, λ_1 becomes the objective function value and is exactly the largest eigenvalue of A.

However, it must be pointed out, on the one hand, that the case $m > 1$ has essential differences from the classical eigenvalue problem. First, the MCP with $m = 1$ is to maximize the Rayleigh quotient on the sphere, which has no other local maximizer except for the global one [3], that is, any local maximizer is a global maximizer. When $m > 1$, different situation occurs. It has been observed that both global and local maxima exist (see [2] and [18]). Secondly, even though the global objective function value r^* is unique in both $m = 1$ and $m > 1$, there may exist different combinations of $\lambda_1^*, \ldots, \lambda_m^*$ with the same sum $r^* = \sum_{i=1}^{m} \lambda_i^*$ when $m > 1$ (see Example 1 in Sect. 3). Therefore, a very interesting and important question arising here is how many combinations of $\lambda_1^*, \ldots, \lambda_m^*$, or simply Λ^*, exist in the global solutions pair $(\mathbf{x}^*, \Lambda^*)$. One of our four main contributions in this paper is to show, based on a sufficient condition we established, that Λ^* is unique either when $m = 2$, or when A is a *nonnegative matrix*. Moreover, it is further shown that the global maximizer of the MCP with a nonnegative *irreducible* matrix A is also essentially unique. This achievement is a step towards better understanding the MCP and the MEP, and may guide the design of the numerical algorithm for solving the MCP successfully.

During our investigation, on the other hand, we also see many analogies between the classical eigenvalue problem with the MEP, especially when A is a nonnegative matrix. A corresponding Collatz-Wielandt formula for the MEP is established and many nice analogies are summarized in Sect. 5.

In fact, the MCP with a nonnegative matrix A or with a nonnegative solution also arises frequently in practical applications; some real world examples can be found, e.g., in [6,14,16]. In the MCP (1.4), the matrix A is closely related to the sample correlation matrix for the m sets of variables which would be nonnegative if the variables involved are of positive associations ([2,6,16]). Furthermore, in some practical applications, such as in audio and video signals [14], the nonnegative linear combinations of variables are required. In Sect. 4, we will also provide a real world example with a nonnegative irreducible matrix A to which our results can be applied.

When turning to the computational issue, we are disappointed to note that up to date, no algorithm is able to solve the MCP for the general case. All iterative algorithms available are mainly centered around satisfying the necessary condition (1.6), the MEP. The earliest iteration, the Horst-Jacobi algorithm [6], which is summarized as Algorithm 1 below, is of Jacobi-type recurrence structure and is proved to converge monotonically to a solution of the MEP in [2]. An improvement, the Gauss-Seidel algorithm [2] summarized as Algorithm 2, is developed by adopting the Gauss-Seidel-type iteration. Following the similar arguments in proving the convergence of the Horst-Jacobi algorithm [2], the Gauss-Seidel algorithm is shown to converge monotonically to a solution of the MEP in [19] recently. However, numerical experiments show that for either algorithm, the computed solution depends closely on the starting point, and if the starting point is not specially selected, there is a high probability that the solutions found are local maxima [19]. Without the global maximizer, the maximal correlation would not be established, making the statistical prediction less reliable.

Algorithm 1 The Horst-Jacobi algorithm for the MEP [6].

Given $\mathbf{x}^{(0)} \in \mathbb{R}^n$,
for $k = 0, 1, \ldots,$ **do**
  for $i = 1, 2, \ldots, m$ **do**
    $\mathbf{y}_i^{(k)} := \sum_{j=1}^{m} A_{ij}\mathbf{x}_j^{(k)}$
    $\lambda_i^{(k)} := \|\mathbf{y}_i^{(k)}\|_2$
    $\mathbf{x}_i^{(k+1)} := \frac{\mathbf{y}_i^{(k)}}{\lambda_i^{(k)}}$
  end for
end for

Algorithm 2 The Gauss-Seidel algorithm for the MEP [2].

Given $\mathbf{x}^{(0)} \in \mathbb{R}^n$,
for $k = 0, 1, \ldots,$ **do**
  for $i = 1, 2, \ldots, m$ **do**
    $\mathbf{y}_i^{(k)} := \sum_{j=1}^{i-1} A_{ij}\mathbf{x}_j^{(k+1)} + \sum_{j=i}^{m} A_{ij}\mathbf{x}_j^{(k)}$
    $\lambda_i^{(k)} := \|\mathbf{y}_i^{(k)}\|_2$
    $\mathbf{x}_i^{(k+1)} := \frac{\mathbf{y}_i^{(k)}}{\lambda_i^{(k)}}$
  end for
end for

In order to obtain the global maximizer for the MCP, efforts were made along two different lines in the literature. The first effort is to establish the global optimal conditions for the MCP. One pertaining result was made in [5] where it was shown that if $m = 2$ or if $m > 2$ with A being a *positive* matrix, a solution $\mathbf{x}^*$ to the MEP is a global solution for the MCP if and only if $A - \Lambda^*$ is negative semi-definite. In general, even for $m = 3$, example was given in [5] to illustrate the complicated situation (see also Example 1 in Sect. 3). Most recently, [19] further establishes a necessary global optimal condition for the general case. In particular, it was shown that if $\mathbf{x}^*$ is a global maximizer, the corresponding multivariate eigenvalue λ_i^*, for $i = 1, 2, \ldots, m$, is not less than the largest eigenvalue of A_{ii}, *i.e.,* $\lambda_i^* \geq \sigma_1(A_{ii})$, and hence, the matrix $D - \Lambda^*$ is negative semi-definite, where

$$D = \text{diag}\,\{A_{11}, \ldots, A_{mm}\} \in \mathbb{R}^{n\times n}, \tag{1.8}$$

is the block diagonal matrix of A, and $\sigma_1(A) \geq \ldots \geq \sigma_n(A)$ stands for the ordered eigenvalues of A. Our second contribution in this paper is also made along this line by extending the global optimal condition for the positive matrix to the nonnegative matrix. In particular, we show that when A is a nonnegative matrix, the negative semi-definiteness of $A - \Lambda^*$ is still a necessary and sufficient condition for a solution pair $(\mathbf{x}^*, \Lambda^*)$ of the MEP to be a global solution for the MCP. The importance of this global optimal condition is that, when

A is nonnegative irreducible, we are able to prove that both the Horst-Jacobi algorithm and the Gauss-Seidel algorithm converge globally to the global maximizer of the MCP. This is our third contribution of this paper.

The other effort towards the global solutions of the MCP was made in order to establish some estimation for the global solution. The estimation, $\lambda_i^* \geq \sigma_1(A_{ii})$, for $i = 1, 2, \ldots, m$, established in [19] is also made along this line, which leads to an effective starting point strategy for both the Horst-Jacobi algorithm and the Gauss-Seidel algorithm. As far as the estimate $\lambda_i^* \geq \sigma_1(A_{ii})$ itself is concerned, however, it is apparently a relation between the multivariate eigenvalues with only the block diagonal matrices A_{ii}. Therefore, some new estimates can also be expected by picking up the information in the off-diagonal block $A_{ij} \in \mathbb{R}^{n_i \times n_j}$, $i \neq j$. Our last contribution of this paper is then to establish some new estimates for λ_i^* by utilizing the information of A as much as possible.

The paper is organized as follows. In the next section, we will extend the global optimal condition of the MCP from the positive matrix [5] to the nonnegative matrix. In Sect. 3, we will discuss the uniqueness of the multivariate eigenvalues and show that Λ^* is unique for the global solutions of the MCP (1.4) when $m = 2$, or when A is a nonnegative matrix. The uniqueness of the global maximizer also discussed. In Sect. 4, we shall prove the global convergence of both the Horst-Jacobi algorithm and the Gauss-Seidel algorithm to the global maximizer whenever A is a nonnegative irreducible matrix. Section 5 is dedicated to the comparison between the classical eigenvalue problem and the MEP for the nonnegative case. Some new estimates of multivariate eigenvalues related to the global maxima of the MCP are established, and concluding remarks are given in the last section.

2 Global optimality for the MCP with nonnegative A

Up to now, two iterative algorithms (the Horst-Jacobi algorithm and the Gauss-Seidel algorithm) have been proved to converge monotonically to solutions of the MEP. However, none of these algorithms can guarantee that the solution found is a global maximizer of the MCP. Therefore, whenever a solution of the MEP is in hand, we should rely on some global optimal conditions to check the optimality for MCP so that the statistical prediction is reliable. The first effort of establishing the global optimal condition was made by Hanafi and Ten Berge [5] and we restate their result as follows.

Theorem 2.1 *Suppose* $(\mathbf{x}^*, \Lambda^*)$ *is a solution pair to the MEP (1.6). For* $m = 2$, *or* $m > 2$ *with a positive matrix* A, $\mathbf{x}^*$ *is a global maximizer of the MCP (1.4) if and only if* $A - \Lambda^*$ *is negative semi-definite.*

To generalize this global optimal condition to the nonnegative case, we provide first some preliminary results. Note that the feasible set $\mathcal{M}$ of (1.4) can be cast as the smooth manifold $\prod_{i=1}^m S^{n_i-1}$ embedded in $\prod_{i=1}^m \mathbb{R}^{n_i}$ under the product topology, where S^{n_i-1} stands for the unit sphere in $\mathbb{R}^{n_i}$. For any $\mathbf{x} \in \mathcal{M}$, the tangent space $\mathcal{T}_{\mathbf{x}}\mathcal{M}$ can be expressed simply by an orthogonal projection

$$P_{\mathbf{x}} := \operatorname{diag}\left\{I^{[n_1]} - \mathbf{x}_1\mathbf{x}_1^\top, \ldots, I^{[n_m]} - \mathbf{x}_m\mathbf{x}_m^\top\right\} = I^{[n]} - Q_{\mathbf{x}} \in \mathbb{R}^{n\times n}, \tag{2.1}$$

in the form

$$\mathcal{T}_{\mathbf{x}}\mathcal{M} = \left\{P_{\mathbf{x}}\mathbf{z} \mid \mathbf{z} \in \mathbb{R}^n\right\}. \tag{2.2}$$

As a local maximizer, $\mathbf{x}^* \in \mathcal{M}$ must satisfy the following second-order necessary condition [19]:

Lemma 2.2 *Suppose $A \in \mathbb{R}^{n\times n}$ is symmetric. Then a second-order necessary condition for $(\mathbf{x}^*, \Lambda^*)$ to be a local solution pair of (1.4) is that the inequality*

$$\mathbf{z}^\top P_{\mathbf{x}^*} \left(A - \Lambda^*\right) P_{\mathbf{x}^*}\mathbf{z} \leq 0 \tag{2.3}$$

holds for all $\mathbf{z} \in \mathbb{R}^n$, where $P_{\mathbf{x}^*}$ is defined in (2.1).

Note that Lemma 2.2 is the representation of a fundamental theory ([12], Theorem 12.5) for the constrained optimization. To see this more clearly, we first define the Lagrange function related to the MCP (1.4)

$$L(\mathbf{x}, \lambda_1, \ldots, \lambda_m) = \mathbf{x}^\top A\mathbf{x} - \sum_{i=1}^{m} \lambda_i \left(\|\mathbf{x}_i\|_2^2 - 1\right)$$

where $\lambda_1, \ldots, \lambda_m$ are the Lagrange multipliers. As we have already pointed out that the first-order necessary condition for $(\mathbf{x}^*, \Lambda^*)$ to be a local solution pair of (1.4) is

$$\begin{cases} \nabla_{\mathbf{x}} L(\mathbf{x}^*, \lambda_1^*, \ldots, \lambda_m^*) = 0 \\ \|\mathbf{x}_i^*\|_2 = 1, \end{cases} \quad \text{or} \quad \begin{cases} A\mathbf{x}^* = \Lambda^*\mathbf{x}^* \\ \|\mathbf{x}_i^*\|_2 = 1, \quad \text{for } i = 1, 2, \ldots, m, \end{cases}$$

which is precisely the MEP (1.6). The second-order necessary condition ([12], Theorem 12.5) is then

$$\mathbf{y}^\top \nabla^2_{\mathbf{xx}} L(\mathbf{x}^*, \lambda_1^*, \ldots, \lambda_m^*)\mathbf{y} \leq 0 \qquad \forall \mathbf{y} \in \mathcal{T}_{\mathbf{x}^*}\mathcal{M}. \tag{2.4}$$

By a simple manipulation, it is not difficult to see that

$$\nabla^2_{\mathbf{xx}} L(\mathbf{x}^*, \lambda_1^*, \ldots, \lambda_m^*) = 2(A - \Lambda^*)$$

and consequently, by taking the advantage of the representation of $\mathcal{T}_{\mathbf{x}^*}\mathcal{M}$ of (2.2), the second-order necessary condition (2.4) is precisely (2.3). With the aid of Lemma 2.2, we have the following result.

Theorem 2.3 *Suppose $A \in \mathbb{R}^{n\times n}$ is symmetric and nonnegative, and suppose $(\mathbf{x}^*, \Lambda^*)$ is a solution pair to the MEP (1.6). Then $\mathbf{x}^*$ is a global maximizer of the MCP (1.4) if and only if $A - \Lambda^*$ is negative semi-definite.*

Proof It has been proved in [5] that the negative semi-definiteness of $A - \Lambda^*$ is a sufficient global optimal condition for the MCP (1.4). For the necessary part, we first denote $|\mathbf{x}^*|$ as the vector whose elements are the absolute values of $\mathbf{x}^*$'s. Observing that $|\mathbf{x}^*| \in \mathcal{M}$ and

$$r(|\mathbf{x}^*|) = |\mathbf{x}^*|^\top A|\mathbf{x}^*| \geq {\mathbf{x}^*}^\top A\mathbf{x}^* = r(\mathbf{x}^*),$$

we know that $|\mathbf{x}^*|$ is also a global maximizer of the MCP (1.4), and thereby, there is a diagonal matrix, say Ω, such that

$$A|\mathbf{x}^*| = \Omega|\mathbf{x}^*|, \quad \Omega = \operatorname{diag}\{\omega_1 I^{[n_1]}, \omega_2 I^{[n_2]}, \ldots, \omega_m I^{[n_m]}\} \in \mathbb{R}^{n\times n}. \tag{2.5}$$

Observe that $A\mathbf{x}^* = \Lambda^*\mathbf{x}^*$ yields

$$\sum_{j=1}^{n} a_{ij}\mathbf{x}^*(j) = \Lambda_{ii}^*\mathbf{x}^*(i), \quad i = 1, 2, \ldots, n,$$

where $\mathbf{x}^*(j) \in \mathbb{R}$ represents the j-th element of $\mathbf{x}^*$ to distinct the block vector $\mathbf{x}^*_{\mathbf{j}} \in \mathbb{R}^{n_j}$. Therefore,

$$\sum_{j=1}^{n} a_{ij}|\mathbf{x}^*(j)| \geq |\sum_{j=1}^{n} a_{ij}\mathbf{x}^*(j)| = \Lambda^*_{ii}|\mathbf{x}^*(i)|, \quad i = 1, 2, \ldots, n, \tag{2.6}$$

since $\lambda_i^* \geq \sigma_1(A_{ii}) > 0$ (see [19]). Consequently, it follows from (2.5) and (2.6) that

$$\Omega|\mathbf{x}^*| = A|\mathbf{x}^*| \geq_e \Lambda^*|\mathbf{x}^*|, \text{ or } (\Omega - \Lambda^*)|\mathbf{x}^*| \geq_e 0. \tag{2.7}$$

Here for two vectors $\mathbf{a}$ and $\mathbf{c} \in \mathbb{R}^n$, the symbol $\mathbf{a} \geq_e \mathbf{c}$ means $a_i \geq c_i$ for all $i = 1, 2, \ldots, n$. The inequality (2.7) together with $\|\mathbf{x}^*_i\|_2 = 1$ for $i = 1, 2, \ldots, m$, immediately implies

$$\omega_i \geq \lambda_i^*, \quad \forall i = 1, 2, \ldots, m. \tag{2.8}$$

On the other hand, it follows that $\sum_{i=1}^m \lambda_i^* = \sum_{i=1}^m \omega_i = r(\mathbf{x}^*)$, which together with (2.8) leads to $\Lambda^* = \Omega$.

Now, if $A - \Lambda^*$ is not negative semi-definite, then the largest eigenvalue $\sigma_1(A - \Lambda^*) > 0$, and there must be a corresponding nonnegative eigenvector $\mathbf{b} \geq_e 0$. Therefore,

$$\mathbf{b}^\top|\mathbf{x}| = 0, \quad \text{or} \quad \mathbf{b}_i^\top|\mathbf{x}_i| = 0, \quad \text{for } i = 1, 2, \ldots, m.$$

Moreover, Lemma 2.2 implies the following inequality:

$$\begin{aligned} 0 &\geq \mathbf{b}^\top P_{|\mathbf{x}^*|}(A - \Lambda^*)P_{|\mathbf{x}^*|}\mathbf{b} \\ &= \mathbf{b}^\top(A - \Lambda^*)\mathbf{b} - \mathbf{b}^\top Q_{|\mathbf{x}^*|}(A - \Lambda^*)\mathbf{b} \\ &\quad -\mathbf{b}^\top(A - \Lambda^*)Q_{|\mathbf{x}^*|}\mathbf{b} + \mathbf{b}^\top Q_{|\mathbf{x}^*|}(A - \Lambda^*)Q_{|\mathbf{x}^*|}\mathbf{b} \\ &= \mathbf{b}^\top(A - \Lambda^*)\mathbf{b} = \sigma_1(A - \Lambda^*) > 0, \end{aligned} \tag{2.9}$$

where

$$Q_{|\mathbf{x}^*|} := \text{diag}\left\{|\mathbf{x}^*{}_1||\mathbf{x}^*{}_1|^\top, \ldots, |\mathbf{x}^*{}_m||\mathbf{x}^*{}_m|^\top\right\}.$$

The last second equality in (2.9) holds because

$$Q_{|\mathbf{x}^*|}\mathbf{b} = \text{diag}\left\{(\mathbf{b}_1^\top|\mathbf{x}^*{}_1|)|\mathbf{x}^*{}_1|, \ldots, (\mathbf{b}_m^\top|\mathbf{x}^*{}_m|)|\mathbf{x}^*{}_m|\right\} = 0.$$

Therefore, the contradiction in (2.9) implies that $\sigma_1(A - \Lambda^*) = 0$ and our conclusion follows. □

3 Uniqueness of the global solution

3.1 Uniqueness of the multivariate eigenvalues

For the MCP (1.4), it is clear that the global maximizer $\mathbf{x}^*$ is not unique. This point can be quickly verified from the simple example where $A = \text{diag}\{I^{[n_1]}, I^{[n_2]}\} = I^{[n_1+n_2]}$ for which any $\mathbf{x} \in \mathcal{M}$ becomes a global maximizer. However, we will show in this subsection that when $m = 2$ or when A is a nonnegative matrix, the corresponding diagonal matrix Λ^* is unique. Because of this uniqueness, we then can call Λ^* the *dominant multivariate eigenvalue* of the MEP for these two cases. We provide a sufficient condition for the uniqueness of Λ^* first.

Theorem 3.1 *If there is a global maximizer* $\mathbf{x}^*$ *of the MCP such that the related matrix* $A - \Lambda^*$ *is negative semi-definite, then any global maximizer of the MCP is related to the unique multivariate eigenvalues.*

Proof Suppose $(\overline{\mathbf{x}}, \overline{\Lambda})$ is another arbitrary solution pair to the MEP such that $r(\overline{\mathbf{x}}) = r(\mathbf{x}^*)$. Note that

$$\overline{\mathbf{x}}^\top (A - \Lambda^*)\overline{\mathbf{x}} = r(\overline{\mathbf{x}}) - r(\mathbf{x}^*) = 0,$$

which together with the negative semi-definiteness of $A - \Lambda^*$ implies that $\overline{\mathbf{x}}$ is also an eigenvector corresponding to the largest eigenvalue $\sigma_1(A - \Lambda^*) = 0$, i.e.,

$$(A - \Lambda^*)\overline{\mathbf{x}} = 0, \quad \text{or} \quad A\overline{\mathbf{x}} = \Lambda^*\overline{\mathbf{x}}.$$

Moreover, it is clear that $A\overline{\mathbf{x}} = \overline{\Lambda}\overline{\mathbf{x}}$. Therefore, we have

$$(\overline{\Lambda} - \Lambda^*)\overline{\mathbf{x}} = 0.$$

Since $\overline{\mathbf{x}}_i \neq 0$, for $i = 1, 2 \ldots, m$, it must follow that $\overline{\Lambda} = \Lambda^*$, which completes the proof. □

Based on the global optimality for the case $m = 2$ and the case when A is nonnegative, the uniqueness of Λ^* in the global maxima of the MCP turns out to be evident. Moreover, the following corollary is another direct consequence of Theorem 3.1.

Corollary 3.2 *If there are two global solution pairs* $(\overline{\mathbf{x}}, \overline{\Lambda})$ *and* $(\mathbf{x}^*, \Lambda^*)$ *of the MCP with* $\overline{\Lambda} \neq \Lambda^*$, *then neither* $A - \overline{\Lambda}$ *nor* $A - \Lambda^*$ *is negative semi-definite.*

It is interesting to note that in general, nevertheless, the uniqueness of Λ^* *cannot* be guaranteed. A simple example is presented for demonstration.

Example 1 Consider the example when $m = 3$, $\mathcal{P} = \{1, 1, 1\}$ and

$$A = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 4 & -1 \\ 1 & -1 & 5 \end{pmatrix} \in \mathbb{R}^{3\times 3}. \tag{3.1}$$

As a solution pair $(\mathbf{x}, \Lambda)$ to the MEP, λ_i for $i = 1, 2, 3$ must satisfy

$$\begin{cases} \lambda_1 = 3 + \frac{\mathbf{x}_2}{\mathbf{x}_1} + \frac{\mathbf{x}_3}{\mathbf{x}_1}, \\ \lambda_2 = 4 + \frac{\mathbf{x}_1}{\mathbf{x}_2} - \frac{\mathbf{x}_3}{\mathbf{x}_2}, \\ \lambda_3 = 5 + \frac{\mathbf{x}_1}{\mathbf{x}_3} - \frac{\mathbf{x}_2}{\mathbf{x}_3}. \end{cases}$$

It then becomes very clear that

$$\max_{\mathbf{x}\in\mathcal{M}} r(\mathbf{x}) = \max_{\mathbf{x}\in\mathcal{M}} \left(12 + 2\frac{\mathbf{x}_2}{\mathbf{x}_1} + 2\frac{\mathbf{x}_3}{\mathbf{x}_1} - 2\frac{\mathbf{x}_3}{\mathbf{x}_2}\right) = 14,$$

for which two global maxima are $\mathbf{x}^* = (1, 1, 1)^\top$ and $\widetilde{\mathbf{x}} = (1, -1, 1)^\top$, with the corresponding $\Lambda^* = \text{diag}\{5, 4, 5\}$ and $\widetilde{\Lambda} = \text{diag}\{3, 4, 7\}$, respectively. Obviously, $\Lambda^* \neq \widetilde{\Lambda}$ and both $A - \Lambda^*$ and $A - \widetilde{\Lambda}$ are not negative semi-definite. This is also a simple example to illustrate that the global optimality for the case $m = 2$ or $m > 2$ with a nonnegative matrix A is not valid for the general case any more. By randomly choosing 10^7 starting points with elements from the uniform $[-0.5, 0.5]$ distribution for the Horst-Jacobi algorithm as the replacement of the exhaustive search, the example given in [5] only numerically shows the optimality condition, Theorem 2.1, is violated, whereas Example 1 clearly illustrates this.

3.2 Uniqueness of the global maximizer

As has been pointed out in Subsect. 3.1, that the global maximizer of the MCP (1.4) is not unique in general, even for the nonnegative matrix. However, the case when A is nonnegative *irreducible* enjoys additional interesting properties and deserves extra investigation. For the discussion on this topic, we need some preliminaries of the nonnegative irreducible matrix.

Definition 3.3 $B \in \mathbb{R}^{n\times n}$ is said to be a reducible matrix when there exists a permutation matrix $P \in \mathbb{R}^{n\times n}$ such that

$$P^\top BP = \begin{pmatrix} B_1 & B_2 \\ 0 & B_3 \end{pmatrix}, \quad \text{where } B_1 \text{ and } B_3 \text{ are both square.}$$

Otherwise B is said to be an irreducible matrix.

Based on this definition, Lemma 3.4 is straightforward.

Lemma 3.4 *Suppose $B \in \mathbb{R}^{n\times n}$ is an irreducible matrix, then for any diagonal matrix $\Upsilon \in \mathbb{R}^{n\times n}$, $B + \Upsilon$ is also irreducible.*

The following theorem is very useful in analyzing the uniqueness of the global maximizer for the nonnegative irreducible case. $\rho(B)$ here represents the *spectral radius* of a matrix B.

Theorem 3.5 *(The Perron-Frobenius Theorem, [13]) Let $B \in \mathbb{R}^{n\times n}$ be a nonnegative irreducible matrix. Then the following statements are valid.*

(1) $\rho(B) > 0$.
(2) $\rho(B)$ *is an eigenvalue of* B.
(3) There exists a positive eigenvector of B corresponding to the eigenvalue $\rho(B)$.
(4) The eigenvalue $\rho(B)$ is simple.

Lemma 3.6 *Suppose $B \in \mathbb{R}^{n\times n}$ is symmetric and nonnegative irreducible. Then for any diagonal matrix $\Upsilon = diag\{\Upsilon_{11}, \Upsilon_{22}, \ldots, \Upsilon_{nn}\} \in \mathbb{R}^{n\times n}$, there is a unique unit positive eigenvector corresponding to the largest eigenvalue $\sigma_1(B + \Upsilon)$ of $B + \Upsilon$. Moreover, the largest eigenvalue $\sigma_1(B + \Upsilon)$ is a simple eigenvalue.*

Proof Let $\zeta := \max_i\{|\Upsilon_{ii}|\}$. The conclusion then follows immediately from Lemma 3.4 and Theorem 3.5 applied to the matrix $B + \Upsilon + \zeta I^{[n]}$ which is nonnegative and irreducible. □

With the aid of these results, we are able to establish the following conclusion:

Theorem 3.7 *Suppose A is nonnegative irreducible, then there are only two solution pairs, $(\mathbf{x}^*, \Lambda^*)$ and $(-\mathbf{x}^*, \Lambda^*)$ for the MEP, such that $\mathbf{x}^*$ and $-\mathbf{x}^*$ are the global maxima for the MCP (1.4). Moreover, one of $\mathbf{x}^*$ and $-\mathbf{x}^*$ must be a positive vector.*

Proof Suppose $(\mathbf{x}^*, \Lambda^*)$ is an arbitrary global solution for the MCP. According to Theorem 3.1, Λ^* is unique in all global maxima of the MCP. Note that $\mathbf{x}^*$ is an eigenvector of $A - \Lambda^*$ associated with $\sigma_1(A - \Lambda^*) = 0$ which, by Lemma 3.6 is a simple eigenvalue. This shows that $\mathbf{x}^*$ and $-\mathbf{x}^*$ are the only two global maxima and one of them is a positive vector. □

An interesting open question is whether Theorem 3.1 is also a necessary condition. If this is true, we then have another necessary and sufficient condition to describe the global solution of the MCP. Our current results established seem to support this conjecture because the uniqueness of Λ^* and the negative semi-definiteness of $A - \Lambda^*$ happen simultaneously in either the case $m \leq 2$ or the case $m > 2$ with a nonnegative matrix A. Further research is needed to clarify the relationship between the uniqueness of Λ^* and the negative semi-definiteness of $A - \Lambda^*$.

4 Convergence to the global maximizer

Suppose now a current solution pair $(\overline{\mathbf{x}}, \overline{\Lambda})$ for the MEP is in hand. For $m = 2$, [5] has already specified a numerical scheme to construct a better point $\widetilde{\mathbf{x}} \in \mathcal{M}$ than $\overline{\mathbf{x}}$ for the MCP, *i.e.,* $r(\widetilde{\mathbf{x}}) > r(\overline{\mathbf{x}})$, in case $A - \overline{\Lambda}$ has a positive eigenvalue. For the case when A is a nonnegative matrix, a similar result is certainly very demanding. This section is dedicated to this task. As an interesting and surprising result, we can show that both the Horst-Jacobi algorithm and the Gauss-Seidel algorithm are able to globally converge to a global maximizer of the MCP whenever A is nonnegative irreducible. This important convergence result turns out to be obvious with the following result.

Lemma 4.1 *Suppose A is a nonnegative irreducible matrix, and $(\overline{\mathbf{x}}, \overline{\Lambda})$ is a solution pair for the MEP where $\overline{\mathbf{x}} \geq_e 0$, then $\overline{\mathbf{x}}$ is a global maximizer for the MCP (1.4).*

Proof By contradiction, if $(\overline{\mathbf{x}}, \overline{\Lambda})$ were not a global solution pair, by Theorem 2.3, $A - \overline{\Lambda}$ would not be negative semi-definite and there would exist an eigenvector $\mathbf{b}$ such that $(A - \overline{\Lambda})\mathbf{b} = \sigma_1(A - \overline{\Lambda})\mathbf{b}$ and $\sigma_1(A - \overline{\Lambda}) > 0$. By Theorem 3.5 applied to $(A - \overline{\Lambda} + \zeta I^{[n]})$, for $\zeta > 0$ sufficiently large, one has $\mathbf{b} >_e 0$. Since $\overline{\mathbf{x}}$ and $\mathbf{b}$ are eigenvectors associated with different eigenvalues, one must have $\mathbf{b}^\top \overline{\mathbf{x}} = 0$ which contradicts $\overline{\mathbf{x}} \geq_e 0$, $\overline{\mathbf{x}} \neq 0$ and $\mathbf{b} >_e 0$. □

Based on Lemma 4.1 and Theorem 3.7, consequently, we can establish an alternative global optimal condition to Theorem 2.3, which is obviously much more convenient and easier to check.

Corollary 4.2 *Suppose $A \in \mathbb{R}^{n\times n}$ is symmetric and nonnegative irreducible, and suppose $\mathbf{x}^*$ is a solution to the MEP (1.6). Then $\mathbf{x}^*$ is a global maximizer of the MCP (1.4) if and only if all elements of $\mathbf{x}^*$ are of the same sign.*

Proof The necessary part is already ensured by Theorem 3.7. For the sufficient part, we only need to observe that $|\mathbf{x}^*|$ is still a solution to the MEP whenever all elements of $\mathbf{x}^*$ are of the same sign. Then Lemma 4.1 ensures the result. □

By observing the detailed iteration in the Horst-Jacobi algorithm (Algorithm 1), and the Gauss-Seidel algorithm (Algorithm 2), it is clear that if the starting point $\mathbf{x}^{(0)} \geq_e 0$, the successive iterations $\mathbf{x}^{(k)} \geq_e 0$ for $k = 1, 2, \ldots$. When A is generic, convergence of $\{\mathbf{x}^{(k)}\}$ to a solution of the MEP is guaranteed for both algorithms ([2] and [19]), and thereby, convergence to the global solution of the MCP when A is additionally assumed to be nonnegative irreducible then becomes clear. Without the generic assumption of A, converging to a global maximizer relies on the uniqueness of the positive global solution $\mathbf{x}^* >_e 0$ as we shall see in the next theorem.

Theorem 4.3 *Suppose A is a symmetric and nonnegative irreducible matrix. Then for any starting point $\mathbf{x}^{(0)} \geq_e 0$, both the sequences generated by the Horst-Jacobi algorithm and the Gauss-Seidel algorithm converge to a global maximizer $\mathbf{x}^* >_e 0$ of the MCP (1.4).*

Proof Suppose $\{(\mathbf{x}^{(k)}, \Lambda^{(k)})\}$ is generated from either the Horst-Jacobi algorithm or the Gauss-Seidel algorithm. From the convergence proofs of both these algorithms (see [2] and [19]), it follows that any convergent subsequence, say $\{(\mathbf{x}^{(k_j)}, \Lambda^{(k_j)})\}$, of $\{(\mathbf{x}^{(k)}, \Lambda^{(k)})\}$ converges to a corresponding solution $(\overline{\mathbf{x}}, \overline{\Lambda})$, of the MEP. Note that $\mathbf{x}^{(k_j)} \geq_e 0$ implies $\overline{\mathbf{x}} \geq_e 0$. Corollary 4.2 and Theorem 3.7 therefore, ensure $\overline{\mathbf{x}} = \mathbf{x}^*$ and $\overline{\Lambda} = \Lambda^*$. This shows that $\{(\mathbf{x}^{(k)}, \Lambda^{(k)})\}$ converges to $(\mathbf{x}^*, \Lambda^*)$. □

It is worth mentioning that if A is only nonnegative, the strong convergence in Theorem 4.3 can not be expected. This can be easily verified by the following simple example.

Example 2 Consider the example when $m = 2$, $\mathcal{P} = \{2, 1\}$ and

$$A = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \in \mathbb{R}^{3\times 3}. \tag{4.1}$$

Note that both

$$\left(\bar{\mathbf{x}}, \overline{\Lambda}\right) = ((0,1,1)^\top, \mathrm{diag}\{2,2,2\}), \quad \text{and} \quad \left(\widehat{\mathbf{x}}, \widehat{\Lambda}\right) = \left(\left(\frac{\sqrt{3}}{2}, \frac{1}{2}, 1\right)^\top, \mathrm{diag}\left\{3, 3, \frac{3}{2}\right\}\right),$$

are the solution pairs for the MEP (1.6), and hence the Horst-Jacobi algorithm or the Gauss-Seidel algorithm stops at these solutions if they are the initial values.

We now apply our established results to a real world problem given by Rohwer ([16], §8.8). In this example, 37 kindergarten students are selected in a low-socioeconomic-status area to investigate how well two paired associate learning tasks using action words versus nonaction words are related to three student achievement tests. Therefore, there are two sets of variables involved, and the sample correlation matrix is

$$R = \begin{pmatrix} 1.0000 & 0.7951 & 0.2617 & 0.6720 & 0.3390 \\ 0.7951 & 1.0000 & 0.3341 & 0.5876 & 0.3404 \\ 0.2617 & 0.3341 & 1.0000 & 0.3703 & 0.2114 \\ 0.6720 & 0.5876 & 0.3703 & 1.0000 & 0.3548 \\ 0.3390 & 0.3404 & 0.2114 & 0.3548 & 1.0000 \end{pmatrix}.$$

Following the formulation in [2] or [6], the MCP can be stated as follows.

Example 3 The matrix A is nonnegative irreducible and is given by

$$A = \begin{pmatrix} 1 & 0 & 0.2617 & 0.6191 & 0.1068 \\ 0 & 1 & 0.2078 & 0.0118 & 0.0746 \\ 0.2617 & 0.2078 & 1 & 0 & 0 \\ 0.6191 & 0.0118 & 0 & 1 & 0 \\ 0.1068 & 0.0746 & 0 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{5\times 5}, \tag{4.2}$$

and $m = 2$ and $\mathcal{P} = \{2, 3\}$.

By applying either the Horst-Jacobi algorithm (Algorithm 1) or the Gauss-Seidel algorithm (Algorithm 2) with a random nonnegative $\mathbf{x}^{(0)}$, a positive solution for the MEP

$$\mathbf{x}^* = (0.9869, 0.1615, 0.4236, 0.8897, 0.1705)^\top$$

is found. According to Theorem 4.3, we know that $\mathbf{x}^*$ is a global maximizer for the MCP. Indeed, we note that the matrix $A - \Lambda^*$ is negative semi-definite, and by using $\mathbf{x}^*$, we obtain the same linear combinations of variables as that provided in [16].

5 The MEP versus the classical eigenvalue problem for nonnegative matrices

In the previous sections, we have already seen the crucial role of the nonnegativity of A in analyzing the global solution of MCP. In this section, we shall further summarize some interesting analogies between the MEP and the classical eigenvalue problem for the nonnegative case.

It is well known that Perron (1907) discovered that the spectral radius $\rho(A)$ is a simple eigenvalue for a positive matrix A with a unique unit positive eigenvector, the so-called *Perron vector* $\mathbf{p} >_e 0$, and Frobenius (1912) contributed substantial extensions of Perron's results to cover the case of nonnegative matrices. Moreover, an alternative characterization of the spectral radius $\rho(A)$ and the Perron vector is the so-called Collatz-Wielandt Formula, which possesses the following form in the context of nonnegative matrices [11].

Theorem 5.1 *If A is nonnegative, then*

$$\rho(A) = \max_{\mathbf{x}\in\mathcal{N}} \min_{1\le j\le n, \mathbf{x}(j)\neq 0} \frac{[A\mathbf{x}]_j}{\mathbf{x}(j)}, \quad \text{where} \quad \mathcal{N} = \{\mathbf{x}|\mathbf{x} \ge_e 0 \quad \text{and } \mathbf{x} \neq 0\}, \tag{5.1}$$

where $[A\mathbf{x}]_j$ and $\mathbf{x}(j)$ denote the jth elements of $A\mathbf{x}$ and $\mathbf{x}$, respectively.

A straightforward generalization is like (5.2), which takes the feature that the objective function value $r(\mathbf{x})$ at a solution of the MEP is the sum of the corresponding multivariate eigenvalues; *i.e.,*

$$\max_{\mathbf{x}\in\mathcal{N}} \sum_{i=1}^{m} \min_{1\le j\le n_i, \mathbf{x}_i(j)\neq 0} \frac{[A_i\mathbf{x}]_j}{\mathbf{x}_i(j)}, \tag{5.2}$$

where $A_i = [A_{i1}, A_{i2}, \ldots, A_{im}] \in \mathbb{R}^{n_i\times n},\ i = 1, 2, \ldots, m.$

However, we need to point out a subtle flaw in (5.2) in that its maximum would possibly be infinite. An example is enough to demonstrate this flaw.

Example 4 Consider the example when $m = 2$, $\mathcal{P} = \{1, 1\}$ and

$$A = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 2 \end{pmatrix} \in \mathbb{R}^{2\times 2}. \tag{5.3}$$

It is easy to verify that in this case

$$\max_{\mathbf{x}\in\mathcal{N}} \sum_{i=1}^{m} \min_{1\le j\le n_i, \mathbf{x}_i(j)\neq 0} \frac{[A_i\mathbf{x}]_j}{\mathbf{x}_i(j)} = \max_{\mathbf{x}\in\mathcal{N}} (3 + \frac{\mathbf{x}(1)}{2\mathbf{x}(2)} + \frac{\mathbf{x}(2)}{2\mathbf{x}(1)}) \to +\infty.$$

This flaw, nonetheless, can be quickly fixed by using

$$\widehat{\mathcal{N}} = \{\mathbf{x}|\mathbf{x} \ge_e 0, \quad \text{and } \|\mathbf{x}_i\|_2 = 1,\ i = 1, 2, \ldots, m.\} = \mathcal{N} \cap \mathcal{M} \tag{5.4}$$

instead of $\mathcal{N}$, and the corresponding Collatz-Wielandt Formula for the MEP can be described as follows.

Theorem 5.2 *If A is nonnegative, then the global maximum r^* of the MCP (1.4) can be expressed as*

$$r^* = \sum_{i=1}^{m} \lambda_i^* = \max_{\mathbf{x}\in\widehat{\mathcal{N}}} \sum_{i=1}^{m} \min_{1\le j\le n_i, \mathbf{x}_i(j)\neq 0} \frac{[A_i\mathbf{x}]_j}{\mathbf{x}_i(j)}, \tag{5.5}$$

where $\widehat{\mathcal{N}}$ is given by (5.4).

Table 1 Correspondences between the classical eigenvalue problem and the MEP for nonnegative irreducible matrices

The classical eigenvalue problem	The MEP
$\rho(A)$	Λ^*
The Perron vector, $\mathbf{p} >_e 0$, is the unique unit positive eigenvector	$\mathbf{x}^* >_e 0$ is the unique positive solution of the MEP
$A\mathbf{p} = \rho(A)\mathbf{p}$	$A\mathbf{x}^* = \Lambda^*\mathbf{x}^*$
$A - \rho(A)I^{[n]}$ is negative semi-definite	$A - \Lambda^*$ is negative semi-definite
$\sigma_1(A - \rho(A)I^{[n]}) = 0$ is a simple eigenvalue	$\sigma_1(A - \Lambda^*) = 0$ is a simple eigenvalue
$\rho(A) = \max_{\|\mathbf{x}\|_2=1} \mathbf{x}^\top A\mathbf{x}$	$\sum_{i=1}^m \lambda_i^* = \max_{\mathbf{x}\in\mathcal{M}} \mathbf{x}^\top A\mathbf{x}$
The Collatz-Wielandt Formula (5.1)	The Collatz-Wielandt Formula (5.5)

Proof For any $\mathbf{x} \in \widehat{\mathcal{N}}$, we denote

$$r_i(\mathbf{x}) = \min_{1\le j\le n_i, \mathbf{x}_i(j)\neq 0} \frac{[A_i\mathbf{x}]_j}{\mathbf{x}_i(j)}.$$

Then it follows that

$$r_i(\mathbf{x})\mathbf{x}_i \le_e A_i\mathbf{x},$$

and hence

$$r_i(\mathbf{x}) \le \mathbf{x}_i^\top A_i\mathbf{x}, \quad \text{and} \quad \sum_{i=1}^m r_i(\mathbf{x}) \le \sum_{i=1}^m \mathbf{x}_i^\top A_i\mathbf{x} = \mathbf{x}^\top A\mathbf{x} \le r^*.$$

On the other hand, Theorem 3.7 ensures that $\mathbf{x}^* \in \widehat{\mathcal{N}}$ achieves the maximum, i.e., $\sum_{i=1}^m r_i(\mathbf{x}^*) = r^*$, and our claim (5.5) follows directly. □

To sum up, we list the detailed correspondences between the classical eigenvalue problem and the MEP for nonnegative irreducible matrices in Table 1.

6 Estimation of multivariate eigenvalues

Let $(\mathbf{x}^*, \Lambda^*)$ be a typical global solution pair to the MCP where $\Lambda^* = \mathrm{diag}\{\lambda_1^* I^{[n_1]}, \ldots, \lambda_m^* I^{[n_m]}\}$. We focus on the estimation of λ_i^*, $i = 1, 2, \ldots, m$, in this section.

Recall that in [19], a lower bound

$$\lambda_i^* \ge \sigma_1(A_{ii}), \quad i = 1, 2, \ldots, m, \tag{6.1}$$

has been established. This estimate is proved by contradiction; that is, if (6.1) is not true, then we can either show that the second-order necessary condition Lemma 2.2 is violated or construct, by using the dominant eigenvector of A_{ii} corresponding to $\sigma_1(A_{ii})$, a new feasible point with larger objective function value. While (6.1) has already shown its importance in setting up an effective starting point strategy for both the Horst-Jacobi algorithm and the Gauss-Seidel algorithm, the estimate itself mainly uses the information in the diagonal block matrices A_{ii}'s, ignoring the contribution of the off-diagonal block matrices A_{ij}'s. To establish some new estimates in terms of the information in the off-diagonal block matrices of A, we first establish a general result in Lemma 6.1 which is related with the partition of the positive definite matrix A.

Lemma 6.1 *Suppose $A \in \mathbb{R}^{n\times n}$ is symmetric positive semi-definite and has a partition (1.2), then*

$$\|A_{ij}\|_2 \le \sqrt{\|A_{ii}\|_2 \cdot \|A_{jj}\|_2}, \quad (i \ne j), \tag{6.2}$$

$$\|A_{ij}\|_2 \le \frac{1}{2}(\|A_{ii}\|_2 + \|A_{jj}\|_2), \tag{6.3}$$

$$\max_{i,j} \|A_{ij}\|_2 = \max_i \|A_{ii}\|_2. \tag{6.4}$$

Proof For (6.2) with $i \ne j$, let us consider the vector

$$\bar{\mathbf{v}} = (0, \ldots, 0, t\mathbf{v}_i^\top, 0, \ldots, 0, \mathbf{v}_j^\top, 0, \ldots, 0)^\top \in \mathbb{R}^n, \quad t \in \mathbb{R},$$

where $\mathbf{v}_i \in \mathbb{R}^{n_i}$ and $\mathbf{v}_j \in \mathbb{R}^{n_j}$ are the left and right singular vectors of A_{ij} corresponding to its largest singular value $\sigma_1(A_{ij}) = \|A_{ij}\|_2$; *i.e.*, $\mathbf{v}_i^\top A_{ij}\mathbf{v}_j = \|A_{ij}\|_2$. It then follows that

$$\begin{aligned} 0 \le \bar{\mathbf{v}}^\top A \bar{\mathbf{v}} &= t^2\mathbf{v}_i^\top A_{ii}\mathbf{v}_i + \mathbf{v}_j^\top A_{jj}\mathbf{v}_j + 2t\mathbf{v}_i^\top A_{ij}\mathbf{v}_j \\ &\le t^2\|A_{ii}\|_2 + \|A_{jj}\|_2 + 2t\|A_{ij}\|_2, \quad \forall t \in \mathbb{R}. \end{aligned}$$

Therefore, one must have

$$4\|A_{ij}\|_2^2 - 4\|A_{ii}\|_2 \cdot \|A_{jj}\|_2 \le 0,$$

which leads to (6.2), and thus (6.3) can immediately be obtained *via* the arithmetic-geometric mean inequality. (6.4) is also straightforward from (6.2). □

This lemma is a direct generalization of Theorem 4.2.6 [4] and the following corollary provides one of its straightforward applications for estimating λ_i^*.

Corollary 6.2 *Suppose $\mathbf{x}^*$ is a global maximizer of the MCP, then the corresponding multivariate eigenvalues satisfy*

$$\max_{i,j}\{\lambda_i^*, \lambda_j^*\} \ge \|A_{ij}\|_2, \quad i, j = 1, 2, \ldots, m. \tag{6.5}$$

Proof For any $i, j = 1, 2, \ldots, m$, (6.1) and (6.2) yields

$$\lambda_i^* + \lambda_j^* \ge \|A_{ii}\|_2 + \|A_{jj}\|_2 \ge 2\|A_{ij}\|_2,$$

and consequently the estimation (6.5). □

Furthermore, Lemma 6.1 is also very helpful in establishing the following estimation.

Theorem 6.3 *Suppose A is symmetric positive definite and $(\mathbf{x}^*, \Lambda^*)$ is a solution pair to the MEP in which $\mathbf{x}^*$ is a global maximizer for the MCP. If $A - \Lambda^*$ is negative semi-definite, then*

$$r(\mathbf{x}^*) \ge \frac{1}{m-1}\sum_{i>j}[\sigma_{n_i}(A_{ii}) + \sigma_{n_j}(A_{jj}) + 2\|A_{ij}\|_2], \tag{6.6}$$

$$\|A_{ij}\|_2^2 \le (\lambda_i^* - \sigma_{n_i}(A_{ii}))(\lambda_j^* - \sigma_{n_j}(A_{jj})), \quad i \ne j, \tag{6.7}$$

$$\lambda_i^* - \sigma_1(A_{ii}) \le \left[\sum_{j\ne i}(\lambda_j^* - \sigma_{n_j}(A_{jj}))^{\frac{1}{2}}\right]^2, \quad i = 1, 2, \ldots, m. \tag{6.8}$$

Proof For (6.6), we note that $\Lambda^* - A$ is positive semi-definite and

$$\|\lambda_i^* I^{[n_i]} - A_{ii}\|_2 = \lambda_i^* - \sigma_{n_i}(A_{ii}), \quad i = 1, 2, \ldots, m, \tag{6.9}$$

and it follows from Lemma 6.1 that

$$2\|A_{ij}\|_2 \le \lambda_i^* - \sigma_{n_i}(A_{ii}) + \lambda_j^* - \sigma_{n_j}(A_{jj}), \quad i, j = 1, 2, \ldots, m;$$

or

$$\lambda_i^* + \lambda_j^* \ge 2\|A_{ij}\|_2 + \sigma_{n_i}(A_{ii}) + \sigma_{n_j}(A_{jj}), \quad i, j = 1, 2, \ldots, m.$$

Therefore,

$$(m-1)r(\mathbf{x}^*) = \sum_{i>j}(\lambda_i^* + \lambda_j^*) \ge \sum_{i>j}[2\|A_{ij}\|_2 + \sigma_{n_i}(A_{ii}) + \sigma_{n_j}(A_{jj})],$$

which is (6.6).

The inequality (6.7) is straightforward based on Lemma 6.1 and (6.9).

For (6.8), it is trivially true if $\lambda_i^* I^{[n_i]} - A_{ii}$ is singular. Therefore, we consider only the nonsingular case. For any $j \ne i$, the following submatrix of $(\Lambda^* - A)$

$$\begin{pmatrix} \lambda_i^* I^{[n_i]} - A_{ii} & -A_{ij} \\ -A_{ji} & \lambda_j^* I^{[n_j]} - A_{jj} \end{pmatrix} \in \mathbb{R}^{(n_i+n_j)\times(n_i+n_j)},$$

is positive semi-definite. Applying Proposition 1.3.2 in [1], we can express $-A_{ij}$ as

$$-A_{ij} = (\lambda_i^* I^{[n_i]} - A_{ii})^{\frac{1}{2}} K_{ij} (\lambda_j^* I^{[n_j]} - A_{jj})^{\frac{1}{2}}, \tag{6.10}$$

where $K_{ij} \in \mathbb{R}^{n_i \times n_j}$ satisfies $\|K_{ij}\|_2 \le 1$. Thus from (6.10) and $A\mathbf{x}^* = \Lambda^* \mathbf{x}^*$, one has

$$\lambda_i^* \mathbf{x}_i^* = \sum_{j \ne i} A_{ij}\mathbf{x}_j^* + A_{ii}\mathbf{x}_i^* = -\sum_{j\ne i}(\lambda_i^* I^{[n_i]} - A_{ii})^{\frac{1}{2}} K_{ij}(\lambda_j^* I^{[n_j]} - A_{jj})^{\frac{1}{2}}\mathbf{x}_j^* + A_{ii}\mathbf{x}_i^*,$$

and hence

$$(\lambda_i^* I^{[n_i]} - A_{ii})\mathbf{x}_i^* = -\sum_{j\ne i}(\lambda_i^* I^{[n_i]} - A_{ii})^{\frac{1}{2}} K_{ij}(\lambda_j^* I^{[n_j]} - A_{jj})^{\frac{1}{2}}\mathbf{x}_j^*.$$

The nonsingularity of $\lambda_i^* I^{[n_i]} - A_{ii}$ further gives rise to

$$\mathbf{x}_i^* = -(\lambda_i^* I^{[n_i]} - A_{ii})^{-\frac{1}{2}} \sum_{j\ne i} K_{ij}(\lambda_j^* I^{[n_j]} - A_{jj})^{\frac{1}{2}}\mathbf{x}_j^*,$$

and

$$\begin{aligned} 1 &= \|(\lambda_i^* I^{[n_i]} - A_{ii})^{-\frac{1}{2}} \sum_{j\ne i} K_{ij}(\lambda_j^* I^{[n_j]} - A_{jj})^{\frac{1}{2}}\mathbf{x}_j^*\|_2 \\ &\le (\lambda_i^* - \sigma_1(A_{ii}))^{-\frac{1}{2}} \sum_{j\ne i}(\lambda_j^* - \sigma_{n_j}(A_{jj}))^{\frac{1}{2}}, \end{aligned}$$

which is exactly (6.8). □

It is clear, based on the global optimal condition, that Theorem 6.3 is applicable for both the case $m = 2$ and the case $m > 2$ with a nonnegative matrix A. In particular, (6.8) interestingly provides an upper bound for the estimation (6.1). In the rest of this section, we will focus on the case $m = 2$.

Firstly, if $m = 2$, (6.8) reads as

$$\lambda_1^* - \sigma_1(A_{11}) \le \lambda_2^* - \sigma_{n_2}(A_{22}), \quad (6.11)$$
$$\lambda_2^* - \sigma_1(A_{22}) \le \lambda_1^* - \sigma_{n_1}(A_{11}). \quad (6.12)$$

Thus, if $\sigma_1(A_{11}) \le \sigma_{n_2}(A_{22})$ (or $\sigma_1(A_{22}) \le \sigma_{n_1}(A_{11})$), we have $\lambda_1^* \le \lambda_2^*$ (or $\lambda_2^* \le \lambda_1^*$). Moreover, a refinement of (6.11) and (6.12) can be made as follows.

Theorem 6.4 *Suppose* $\mathbf{x}^*$ *is a global maximizer for the MCP with* $m = 2$*. Then the corresponding multivariate eigenvalues* λ_1^* *and* λ_2^* *satisfy*

$$(\lambda_1^* - \sigma_1(A_{11}))(\lambda_2^* - \sigma_1(A_{22})) \le \|A_{12}\|_2^2 \le (\lambda_1^* - \sigma_{n_1}(A_{11}))(\lambda_2^* - \sigma_{n_2}(A_{22})). \quad (6.13)$$

Proof The second inequality is already clear from (6.7). For the first inequality, if either $\lambda_1^* I^{[n_1]} - A_{11}$ or $\lambda_2^* I^{[n_2]} - A_{22}$ is singular, it is trivially true. We then assume they are both nonsingular.

Note that

$$\Lambda^* - A = \begin{pmatrix} (\lambda_1^* I^{[n_1]} - A_{11})^{\frac{1}{2}} & \\ & (\lambda_2^* I^{[n_2]} - A_{22})^{\frac{1}{2}} \end{pmatrix} E \times \begin{pmatrix} (\lambda_1^* I^{[n_1]} - A_{11})^{\frac{1}{2}} & \\ & (\lambda_2^* I^{[n_2]} - A_{22})^{\frac{1}{2}} \end{pmatrix},$$

where

$$E := \begin{pmatrix} I^{[n_1]} & C \\ C^\top & I^{[n_2]} \end{pmatrix} \in \mathbb{R}^{n \times n}, \quad \text{and}$$
$$C := -(\lambda_1^* I^{[n_1]} - A_{11})^{-\frac{1}{2}} A_{12} (\lambda_2^* I^{[n_2]} - A_{22})^{-\frac{1}{2}} \in \mathbb{R}^{n_1 \times n_2}.$$

The positive semi-definiteness of $\Lambda^* - A$ implies the positive semi-definiteness of E whose eigenvalues are exactly

$$1 - \sigma_1(C) \le \ldots \le 1 - \sigma_q(C) \le 1 = \ldots = 1 \le 1 + \sigma_q(C) \le \ldots \le 1 + \sigma_1(C)$$

with $|n_1 - n_2|$ ones in the middle, where $q := \min\{n_1, n_2\}$ and $\sigma_i(C)$ here represents the i-th largest singular value of C. Thus it must follow that $1 = \sigma_1(C)$, and hence,

$$1 = \| - (\lambda_1^* I^{[n_1]} - A_{11})^{-\frac{1}{2}} A_{12} (\lambda_2^* I^{[n_2]} - A_{22})^{-\frac{1}{2}} \|_2$$
$$\le \|A_{12}\|_2 (\lambda_1^* - \sigma_1(A_{11}))^{-\frac{1}{2}} (\lambda_2^* - \sigma_1(A_{22}))^{-\frac{1}{2}},$$

which is exactly the first inequality in (6.13). □

7 Concluding remarks

In this paper, towards the global solutions of the MCP, which is extremely desired from a practical point of view, we put our main efforts into the global optimality, the uniqueness and the estimation of multivariate eigenvalues, $\lambda_1^*, \lambda_2^*, \ldots, \lambda_m^*$. The theoretical achievements of the global optimal conditions and the uniqueness of the global solutions in the nonnegative case not only offer us a much more clear picture of the MCP and the related MEP, but also helps us to establish the global convergence of both the Horst-Jacobi algorithm and the Gauss-Seidel algorithm to a global maximizer. By comparing the classical eigenvalue problem with the MEP, we have observed many nice similarities for nonnegative matrices. The estimation established in Sect. 6, lastly, picks up the information in the off-diagonal matrices of A, and provides us more information on the global solution of the MCP.

However, some problems still remain open. The uniqueness of Λ^* and the global optimality are interesting issues that are worthy to be generalized. Moreover, efficient numerical methods for the MCP with a general matrix A should still be pursued in the future. The success of the Horst-Jacobi algorithm and the Gauss-Seidel algorithm for a nonnegative irreducible matrix A relies on the nice properties of the global maximizer of the MCP and the core engine in choosing the starting points. This success again shows the power of some special starting point strategy in finding the global maximizer of the MCP, and encourages us to investigate further for other cases.

Acknowledgments The first author would like to thank Professor Moody T. Chu for introducing the maximal correlation and multivariate eigenvalue problems during his visit to the North Carolina State University. The authors also would like to thank the Associate Editor and one anonymous referee for their helpful suggestions and comments on the earlier version of this paper.

References

1. Bhatia, R.: Positive Definite Matrices. Princeton University Press, Princeton (2007)
2. Chu, M.T., Watterson, J.L.: On a multivariate eigenvalue problem: I. Algebraic theory and power method. SIAM J. Sci. Comput. **14**, 1089–1106 (1993)
3. Golub, G.H., Liao, L.-Z.: Continuous methods for extreme and interior eigenvalue problems. Linear Algebra Appl. **415**, 31–51 (2006)
4. Golub, G.H., Van Loan, C.F.: Matrix Computations. 3rd edn. Johns Hopkins University Press, Baltimore, MD (1996)
5. Hanafi, M., Ten Berge, J.M.F.: Global optimality of the successive Maxbet algorithm. Psychometrika **68**, 97–103 (2003)
6. Horst, P.: Relations among m sets of measures. Psychometrika **26**, 129–149 (1961)
7. Horst, P.: Factor Analysis of Data Matrices. Holt Rinehart and Winston, New York (1965)
8. Hotelling, H.: The most predictable criterion. J. Educ. Pyschol. **26**, 139–142 (1935)
9. Hotelling, H.: Relations between two sets of variates. Biometrika **28**, 321–377 (1936)
10. Kettenring, J.R.: Canonical analysis of several sets of variables. Biometrika **58**, 433–451 (1971)
11. Meyer, C.D.: Matrix Analysis and Applied Linear Algebra. SIAM, Philadelphia (2000)
12. Nocedal, J., Wright, S.J.: Numerical Optimization, Springer Series in Operations Research. 2nd edn. Springer, New York (2006)
13. Rao, C.R., Rao, M.B.: Matrix Algebra and its Applications to Statistics and Econometrics. World Scientific, River Edge (1998)
14. Sigg, C., Fischer, B., Ommer, B., Roth, V., Buhmann, J.: Nonnegative CCA for audiovisual source separation. In: IEEE International Workshop on Machine Learning for Signal Processing'07, IEEE, 2007
15. Ten Berge, J.M.F.: Generalized approaches to the MAXBET problem and the MAXDIFF problem, with applications to canonical correlations. Psychometrika **53**, 487–494 (1988)
16. Timm, N.H.: Applied Multivariate Analysis. Springer, New York (2002)
17. Van de Geer, J.R.: Linear relations among k sets of variables. Psychometrika **49**, 79–94 (1984)
18. Van Noorden, T.L., Barkmeijer, J.: The Multivariate Eigenvalue Problem: A New Application, Theory and a Subspace Accelerated Power Method. Universiteit Utrecht, preprint (2008)
19. Zhang, L.-H., Chu, M.T.: Computing absolute maximum correlation, in preprint. Available: http://www4.ncsu.edu/~mtchu/Research/Papers/Readme.html