# COMPUTING ABSOLUTE MAXIMUM CORRELATION

LEI-HONG ZHANG AND MOODY T. CHU

Abstract. Optimizing correlations between sets of variables is an important task in many areas of applications. There are plenty of algorithms for computing the maximum correlation. Most disappointedly, however, these methods typically cannot guarantee attaining the absolute maximum correlation which would have significant impact on practical applications. This paper makes two contributions. Firstly, some distinctive traits of the absolute maximum correlation are characterized. By exploiting these attributes, it is possible to propose an effective starting point strategy that significantly increases the likelihood of attaining the absolute maximum correlation. Numerical testing of the classical Horst algorithm with the starting point strategy seems to evince the potency of this approach. Secondly, the Horst algorithm is but one aggregated Jacobi-type power method. Following the innate iterative structure, a generalization to the Gauss-Seidel formulation is proposed as a natural improvement on the power method. Monotone convergence of the Gauss-Seidel algorithm is proved. When combined with the starting point strategy, the newer Gauss-Seidel approach leads to faster computation of the absolute maximum correlation.

## 1. Introduction

Canonical correlation analysis (CCA) is an important tool for assessing the relationship between sets of variables. Considerable research efforts have been devoted to the development of theory and techniques for CCA. The notion of CCA has been applied to areas such as cluster analysis, data classification, pattern recognition, principal component analysis, and bioinformatics. Some general treatments as well as practicalities on this subject can be found in treatises [5, 6, 8, 9, 12, 15, 18]. One very early development of CCA is the maximal correlation problem (MCP) proposed by Hotelling [10, 11] where the goal is to find the linear combination of one set of variables that correlates maximally with the linear combination of another set of variables. If this maximal correlation can be satisfactorily established, then we have the advantage of using one set of variables to predict the other. As a necessary condition to the solution of the MCP, the multivariate eigenvalue problem (MEP) arises, which we delineate below.

A concise brief about the statistical background of MCP and its relationship to the MEP can be found in [4, Section 2]. Here we only introduce the notation for the convenience of subsequent discussions. Given a set of positive integers

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

2000 *Mathematics Subject Classification.* Primary 62H20, 15A12; Secondary 65F10, 65K05.

*Key words and phrases.* Multivariate statistics, canonical correlation, multivariate eigenvalue problem, power method, Gauss-Seidal method, global maximizer.

This research was supported in part by the National Science Foundation under grants DMS-0505880 and DMS-0732299 and NIH Roadmap for Medical Research grant 1 P20 HG003900-01.

with $\sum_{i=1}^{m} n_i = n$, partition a matrix $A \in \mathbb{R}^{n\times n}$ and a vector $\mathbf{x} \in \mathbb{R}^n$ into blocks

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

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

with $A_{ij} \in \mathbb{R}^{n_i \times n_j}$ and $\mathbf{x}_i \in \mathbb{R}^{n_i}$, respectively. Under the context of multivariate statistical analysis, the matrix $A$ is usually assumed to be symmetric and positive definite. The assumption of positive definiteness generally is not needed in most results derived in this paper. We shall carefully mark out this assumption only when it is necessary. The maximum correlations in the MCP correspond to the extreme values of the equality constrained optimization problem [4, 9]:

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

There are many local maximizers, and the emphasis of this paper is on finding the *global* maximizer. Upon employing the Lagrange multiplier theory, it is easy to see that the first order optimality condition for (1.4) is the existence of real scalars, namely, the Lagrange multipliers, $\lambda_1, \ldots, \lambda_m$ and a vector $\mathbf{x} \in \mathbb{R}^n$ such that the system of equations

$$\begin{cases} A\mathbf{x} &= \Lambda\mathbf{x}, \\ \|\mathbf{x}_i\|_2 &= 1, \quad i = 1, 2, \cdots, m, \end{cases} \tag{1.5}$$

is satisfied, where

$$\Lambda := \text{diag}\{\lambda_1 I^{[n_1]}, \lambda_2 I^{[n_2]}, \cdots, \lambda_m I^{[n_m]}\} \tag{1.6}$$

with $I^{[n_i]} \in \mathbb{R}^{n_i \times n_i}$ denoting the $n_i \times n_i$ identity matrix [4, 18]. We shall exploit this condition again in Section 2. The system (1.5) is referred to hereby as the MEP with partition $\mathcal{P}$.

We mention in passing that the MEP, as a generalization of the classical eigenvalue problem, is of mathematical interest in its own right. The MEP arises also in entirely different non-statistical settings. For instance, the MEP finds applications in the perturbation analysis of linear dynamical systems subject to additive bounded noises [19]. A special bivariate eigenvalue problem arises from identification of finite impulse response systems [13]. See also [1] for a related linear complementary problem.

To our knowledge, a direct tackling of the MCP itself (and, hence, obtaining a solution in closed form) is possible only for the case when $m = 2$ and $A_{11} = I^{[n_1]}$ and $A_{22} = I^{[n_2]}$. An argument can be found in [8]. Higher dimensional problems can be solved perhaps only by means of iteration on the MEP. In that regard, just like the power method plays a fundamental role in the numerical procedures for solving the classical eigenvalue problem, an aggregated power method that iterates on blocks of $A$ has been proposed by Horst [8] as a general means for solving the MEP numerically. So that the paper is self-contained, we summarize the iterative scheme as Algorithm 1 below. While this algorithm has been used in practice for decades, its convergence theory was established three decates later in [4].

**Algorithm 1** The Horst-Jacobi algorithm for the MEP [8].

Given $\mathbf{x}^{(0)} \in \mathbb{R}^n$,
**for** $k = 0, 1, \cdots$, **do**
  **for** $i = 1, 2, \cdots, m$ **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**

The recurrence structure in Algorithm 1 is of Jacobi-type (and thus is named Horst-Jacobi). In other words, the updating of the blocks $\mathbf{x}_i^{(k+1)}$, $i = 1, \ldots, m$, is completely independent of each other. The blocks in each sweep of $k$ can be calculated in parallel. Following the conventional treatment in matrix iteration analysis, it is natural to consider adopting the Gauss-Seidel-type iteration by injecting the newly computed updates into the computation. The modification, made at the definition of $\mathbf{y}_k^{(k)}$ in Algorithm 2, raises the hope of improved convergence behavior which is yet to be justified in the literature.

**Algorithm 2** The Gauss-Seidel algorithm for the MEP [4].

Given $\mathbf{x}^{(0)} \in \mathbb{R}^n$,
**for** $k = 0, 1, \cdots$, **do**
  **for** $i = 1, 2, \cdots, m$ **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**

One of our contributions in this paper is to analyze the dynamical behavior of the Gauss-Seidel algorithm. In particular, we prove the monotone convergence of the sequence $\{r(\mathbf{x}^{(k)})\}$ to an optimal value for (1.4) when $\{\mathbf{x}^{(k)}\}$ is generated by Algorithm 2.

The most troubling issue is that, even if the sequence $\{r(\mathbf{x}^{(k)})\}$ from either algorithm converges monotonically, we *cannot* be sure that the limit point of $\{\mathbf{x}^{(k)}\}$ is a global maximizer. As the constraints in (1.4) form a compact set, there must be at least one global maximizer for $r(\mathbf{x})$. On the other hand, the constraints are non-convex, suggesting the existence of multiple local maximizers. Indeed, in the generic case where $A$ has $n$ distinct eigenvalues, it has been proved that the MEP has precisely $\prod_{i=1}^{m}(2n_i)$ solutions [4], counting multiplicities. Unless the starting point is judiciously selected, numerical experiments indicate a substantially high probability that the iterates converge to local maximizers. Without the global maximizer, the maximal correlation would not be established, making the statistical prediction less reliable.

Being able to characterize global maximizers, therefore, would be highly desirable. Surprisingly, research results in this regard are not so extensive in the literature. To our knowledge, the only other pertaining effort was made in [5] where it was shown that if $m = 2$, or if $m > 2$ with $A$ being a *positive* matrix, then $\mathbf{x}^*$ is a global solution to (1.4) if and only if $(\Lambda^*, \mathbf{x}^*)$ is a solution pair to the MEP and the matrix $A - \Lambda^*$ is negative semi-definite. For the general case, however, examples were given in [5] to demonstrate the complexity in analysis.

Our main contribution in the current paper is to offer a strategy to improve the likelihood of finding a global maximizer. In particular, we are able to establish some distribution bounds for the *multivariate eigenvalues* $\lambda_1, \cdots, \lambda_m$ in terms of the eigenvalues of $A$ and of the block diagonal matrix

$$D = \operatorname{diag}\{A_{11}, \ldots, A_{mm}\} \in \mathbb{R}^{n\times n} \tag{1.7}$$

of $A$. By exploiting these relationships, we obtain important information for setting up an effective starting point strategy. Upon equipping the Gauss-Seidel algorithm with the starting point strategy, we think we have a much more successful algorithm in producing a global solution for the MCP. Extensive numerical testing seems to suggest two advantages of our approach — the effectiveness of our strategy in boosting up the probability of finding a global maximizer of (1.4) and the superior convergence of the Gauss-Seidel algorithm over the Horst-Jacobi algorithm on large scale problems.

The paper is organized as follows. In Section 2, we establish the first-order and second-order optimality conditions for (1.4). We also attain a perturbation result for the MEP through the notion of the classical Rayleigh quotient. These basic facts allow us to characterize in Section 3 the Lagrange multipliers $\lambda_1, \ldots, \lambda_m$ in terms of the eigenvalues of both $A$ and $D$. From these estimates of $\lambda_1, \ldots, \lambda_m$, we propose some initial point strategies in Section 4. The convergence dynamics of the Gauss-Seidel algorithm is studied in Section 5. Finally, we report in Section 6 numerical testing results to evidence the effectiveness of our theory and algorithm.

## 2. Optimality Conditions

Since the MCP is formulated as an equality constrained problem (1.4), we first explore the optimality conditions. The discussion in this section applies to any symmetric matrix $A$ in general without the assumption of positive definiteness.

Denote the unit sphere in $\mathbb{R}^{n_i}$ by $S^{n_i-1}$. The feasible set

$$\mathcal{M} := \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}_i\|_2 = 1,\ \mathbf{x}_i \in \mathbb{R}^{n_i}\} \tag{2.1}$$

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. For any $\mathbf{x} \in \mathcal{M}$, the tangent space $\mathcal{T}_{\mathbf{x}}\mathcal{M}$ can be expressed as

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

where

$$P_{\mathbf{x}} := \operatorname{diag}\left\{I^{[n_1]} - \mathbf{x}_1\mathbf{x}_1^\top, \ldots, I^{[n_m]} - \mathbf{x}_m\mathbf{x}_m^\top\right\} \in \mathbb{R}^{n\times n}. \tag{2.3}$$

The projected gradient of $r(\mathbf{x})$ therefore is given by

$$\operatorname{Proj}_{\mathcal{T}_{\mathbf{x}}\mathcal{M}}\nabla r(\mathbf{x}) = P_{\mathbf{x}}(\nabla r(\mathbf{x})) = 2(A - \Lambda(\mathbf{x}))\mathbf{x}, \tag{2.4}$$

where

$$\Lambda(\mathbf{x}) := \operatorname{diag}\left\{\lambda_1(\mathbf{x})I^{[n_1]},\ldots,\lambda_m(\mathbf{x})I^{[n_m]}\right\}, \tag{2.5}$$

$$\lambda_i(\mathbf{x}) := \mathbf{x}_i^\top \mathscr{A}_i\mathbf{x}, \quad i=1,\ldots,m, \tag{2.6}$$

and

$$\mathscr{A}_i := [A_{i1}, A_{i2}, \cdots, A_{im}] \in \mathbb{R}^{n_i\times n}$$

denotes the $i$th row block of $A$. It follows that the set

$$\mathcal{S} := \{\mathbf{x}\in\mathcal{M} \mid A\mathbf{x} = \Lambda(\mathbf{x})\mathbf{x}\} \tag{2.7}$$

contains all critical points for the constrained optimization problem (1.4). This constitutes the first-order optimality condition, which is precisely the MEP.

The following theorem characterizes the second-order optimality condition.

**Theorem 2.1.** *Suppose $A\in\mathbb{R}^{n\times n}$ is symmetric and $\mathbf{x}^*\in\mathcal{S}$. Then a second-order necessary condition for $\mathbf{x}^*$ to be a local maximizer of (1.4) is that the inequality*

$$\mathbf{z}^\top P_{\mathbf{x}^*}\left(A-\Lambda(\mathbf{x}^*)\right)P_{\mathbf{x}^*}\mathbf{z}\le 0 \tag{2.8}$$

*holds for all $\mathbf{z}\in\mathbb{R}^n$, where $P_{\mathbf{x}^*}$ and $\Lambda(\mathbf{x}^*)$ are defined in (2.3) and (2.5), respectively. In other words, the matrix $A-\Lambda(\mathbf{x}^*)$ restricted to the tangent space $\mathcal{T}_{\mathbf{x}^*}\mathcal{M}$ is negative semi-definite. Moreover, if $A-\Lambda(\mathbf{x}^*)$ restricted to the tangent space $\mathcal{T}_{\mathbf{x}^*}\mathcal{M}$ is negative definite, then $\mathbf{x}^*$ is a local maximizer.*

*Proof.* To derive the second-order optimality condition, the projected Hessian technique developed in [3] becomes handy. The idea is to first extend the projected gradient smoothly to the entire space $\mathbb{R}^n$. For $\operatorname{Proj}_{\mathcal{T}_{\mathbf{x}}\mathcal{M}}\nabla r(\mathbf{x})$ in (2.4), this can easily be done because the function

$$g(\mathbf{x}) := 2\left(A-\Lambda(\mathbf{x})\right)\mathbf{x}$$

can be defined for every $\mathbf{x}\in\mathbb{R}^n$. The action of the Fréchet derivative of $g(\mathbf{x})$ at any $\mathbf{v}\in\mathbb{R}^n$ is given by

$$\begin{aligned} g'(\mathbf{x})\mathbf{v} &= 2\left(A\mathbf{v}-\Lambda(\mathbf{x})\mathbf{v}\right)\\ &\quad -2\operatorname{diag}\left\{\left(\mathbf{x}_1^\top\mathscr{A}_1\mathbf{v}+\mathbf{v}_1^\top\mathscr{A}_1\mathbf{x}\right)I^{[n_1]},\ldots,\left(\mathbf{x}_m^\top\mathscr{A}_m\mathbf{v}+\mathbf{v}_m^\top\mathscr{A}_m\mathbf{x}\right)I^{[n_m]}\right\}\mathbf{x}. \end{aligned}$$

The bilinear action of $g'(\mathbf{x})$ therefore can be expressed as

$$\mathbf{v}^\top g'(\mathbf{x})\mathbf{v} = 2\left[\mathbf{v}^\top A\mathbf{v}-\mathbf{v}^\top\Lambda(\mathbf{x})\mathbf{v}-\sum_{i=1}^m\left(\mathbf{x}_i^\top\mathscr{A}_i\mathbf{v}+\mathbf{v}_i^\top\mathscr{A}_i\mathbf{x}\right)\mathbf{v}_i^\top\mathbf{x}_i\right].$$

For the critical point $\mathbf{x}^*$ to be a local maximizer, it is necessary that $\mathbf{v}^\top g'(\mathbf{x}^*)\mathbf{v}\le 0$ for any $\mathbf{v}\in\mathcal{T}_{\mathbf{x}^*}\mathcal{M}$. The assertion (2.8) thus follows from the fact that if $\mathbf{v}\in\mathcal{T}_{\mathbf{x}^*}\mathcal{M}$ then $\mathbf{v}_i^\top\mathbf{x}_i^*=0$ for all $i=1,\ldots,m$.

For the last part of the theorem, define

$$C(\mathbf{x}) = \begin{pmatrix}\|\mathbf{x}_1\|_2^2-1\\ \vdots\\ \|\mathbf{x}_m\|_2^2-1\end{pmatrix}.$$

Note that for any $\mathbf{x}\in\mathbb{R}^n$ satisfying $C(\mathbf{x})=\mathbf{0}$, we have $\mathbf{x}\in\mathcal{M}$ and $\operatorname{rank}(C'(\mathbf{x}))=m$, implying that $\mathbf{0}$ is a regular value of $C(\mathbf{x})$. Therefore, $\mathbf{x}^*$ must be a local maximizer if $A-\Lambda(\mathbf{x}^*)$ restricted to the tangent space $\mathcal{T}_{\mathbf{x}^*}\mathcal{M}$ is negative definite [7]. □

It is interesting to note that the objective function $r(\mathbf{x})$ can be regarded as a generalization of the classical Rayleigh quotient. Consequently, it is expected that a small perturbation around any critical point should give rise to a quadratic perturbation in $r(\mathbf{x})$. More specifically, we have the following observation.

**Theorem 2.2.** *Suppose $A \in \mathbb{R}^{n\times n}$ is symmetric and $\mathbf{x}^* \in \mathcal{S}$. Then for any $\mathbf{x} \in \mathcal{M}$, it is true that*

$$|r(\mathbf{x}) - r(\mathbf{x}^*)| \leq \|A - \Lambda(\mathbf{x}^*)\|_2 \cdot \|\mathbf{x} - \mathbf{x}^*\|_2^2. \tag{2.9}$$

*Proof.* Since $A\mathbf{x}^* = \Lambda(\mathbf{x}^*)\mathbf{x}^*$ and $\mathbf{x}^* \in \mathcal{M}$, we have

$$r(\mathbf{x}^*) = {\mathbf{x}^*}^\top A\mathbf{x}^* = \sum_{i=1}^m \lambda_i(\mathbf{x}^*).$$

For any $\mathbf{x} \in \mathbb{R}^n$, define $\Delta\mathbf{x} := \mathbf{x} - \mathbf{x}^*$. Then

$$\begin{aligned} r(\mathbf{x}) &= \mathbf{x}^\top A\mathbf{x} = r(\mathbf{x}^*) + 2{\mathbf{x}^*}^\top A\Delta\mathbf{x} + \Delta\mathbf{x}^\top A\Delta\mathbf{x} \\ &= r(\mathbf{x}^*) + 2{\mathbf{x}^*}^\top \Lambda(\mathbf{x}^*)\Delta\mathbf{x} + \Delta\mathbf{x}^\top A\Delta\mathbf{x}. \end{aligned}$$

If $\mathbf{x} \in \mathcal{M}$, then it must be that

$$\mathbf{x}_i^{*\top}\Delta\mathbf{x}_i = -\frac{1}{2}\Delta\mathbf{x}_i^\top\Delta\mathbf{x}_i, \quad i = 1, \ldots, m.$$

It follows that

$$2{\mathbf{x}^*}^\top\Lambda(\mathbf{x}^*)\Delta\mathbf{x} = 2\sum_{i=1}^m \lambda_i(\mathbf{x}^*)\mathbf{x}_i^{*\top}\Delta\mathbf{x}_i = -\sum_{i=1}^m \lambda_i(\mathbf{x}^*)\Delta\mathbf{x}_i^\top\Delta\mathbf{x}_i = -\Delta\mathbf{x}^\top\Lambda(\mathbf{x}^*)\Delta\mathbf{x},$$

and thus

$$\begin{aligned} r(\mathbf{x}) &= r(\mathbf{x}^*) - \Delta\mathbf{x}^\top\Lambda(\mathbf{x}^*)\Delta\mathbf{x} + \Delta\mathbf{x}^\top A\Delta\mathbf{x} \\ &= r(\mathbf{x}^*) + \Delta\mathbf{x}^\top(A - \Lambda(\mathbf{x}^*))\Delta\mathbf{x}. \end{aligned} \tag{2.10}$$

The assertion (2.9) is proved. □

The optimality conditions characterized above serve as basic tools for further refining our estimation of the multivariate eigenvalues. The relationship (2.10) will be particularly useful for deciding whether an objective value $r(\mathbf{x})$ can be increased or not. Equipped with these facts, we continue to carry out our analysis in the next section.

## 3. Estimation of multivariate eigenvalues at a global maximizer

It is convenient to arrange the spectrum of the diagonal block $A_{ii} \in \mathbb{R}^{n_i\times n_i}$ in the descending order

$$\sigma_1(A_{ii}) \geq \cdots \geq \sigma_{n_i}(A_{ii}).$$

Henceforth let $(\overline{\Lambda}, \overline{\mathbf{x}})$ with $\overline{\Lambda} := \mathrm{diag}\{\overline{\lambda}_1 I^{[n_1]}, \ldots, \overline{\lambda}_m I^{[n_m]}\}$ typify any solution to the MEP. We first provide a lower bound for $\overline{\lambda}_i$, when $\overline{\mathbf{x}}$ is a local maximizer for (1.4). The following result still does not need positive definiteness for $A$.

**Theorem 3.1.** *Suppose that $A \in \mathbb{R}^{n\times n}$ is symmetric and $(\overline{\Lambda}, \overline{\mathbf{x}})$ is a solution to the MEP. If $\overline{\mathbf{x}}$ is a local maximizer for (1.4), then for any $i = 1, \ldots, m$, it is true that*

$$\overline{\lambda}_i \geq \sigma_{n_i}(A_{ii}), \tag{3.1}$$

*whenever $n_i > 1$.*

*Proof.* If $n_i > 1$, then there exists some $\mathbf{z}_i \in \mathbb{R}^{n_i}$ such that $(I^{[n_i]} - \overline{\mathbf{x}}_i\overline{\mathbf{x}}_i^\top)\mathbf{z}_i \neq 0$. Define

$$\mathbf{z} = (0, 0, \cdots, \mathbf{z}_i^\top, 0, \cdots, 0)^\top \in \mathbb{R}^n.$$

Since $\overline{\mathbf{x}}$ is a local minimizer, it is necessary by Theorem 2.1 that

$$\mathbf{z}_i^\top (I^{[n_i]} - \overline{\mathbf{x}}_i\overline{\mathbf{x}}_i^\top)(A_{ii} - \overline{\lambda}_i I^{[n_i]})(I^{[n_i]} - \overline{\mathbf{x}}_i\overline{\mathbf{x}}_i^\top)\mathbf{z}_i \leq 0.$$

It follows that

$$\begin{aligned} \overline{\lambda}_i \|(I^{[n_i]} - \overline{\mathbf{x}}_i\overline{\mathbf{x}}_i^\top)\mathbf{z}_i\|_2^2 &\geq \mathbf{z}_i^\top (I^{[n_i]} - \overline{\mathbf{x}}_i\overline{\mathbf{x}}_i^\top)A_{ii}(I^{[n_i]} - \overline{\mathbf{x}}_i\overline{\mathbf{x}}_i^\top)\mathbf{z}_i \\ &\geq \sigma_{n_i}(A_{ii})\|(I^{[n_i]} - \overline{\mathbf{x}}_i\overline{\mathbf{x}}_i^\top)\mathbf{z}_i\|_2^2. \end{aligned}$$

This completes the proof. □

The case $n_i = 1$ for some $i$ deserves special attention. Since $\pm 1$ are the only two possible choices for $\mathbf{x}_i$, the feasible set $\mathcal{M}$ is made of disjoint connected subsets.

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

$$A = \begin{pmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 2 & 0 \\ -\frac{1}{2} & 0 & 3 \end{pmatrix} \in \mathbb{R}^{3\times 3}.$$

It is easy to check that $\overline{\mathbf{x}} = (1, 0, 1)^\top$ is a local maximizer for (1.4). Nevertheless, we see that $\overline{\lambda}_1(\overline{\mathbf{x}}) = \overline{\mathbf{x}}_1^\top A_1 \overline{\mathbf{x}} = \frac{1}{2} < \sigma_1(A_{11}) = 1$, violating the estimate (3.1) asserted in Theorem 3.1.

Suppose now $\mathbf{x}^*$ is a global maximizer for (1.4). Then Theorem 3.1 has an interesting and important extension which provides a much sharper estimate of the absolute optimal value. Still, the positive definiteness is not needed.

**Theorem 3.2.** *Suppose that $A \in \mathbb{R}^{n\times n}$ is symmetric and $(\Lambda^*, \mathbf{x}^*)$ is a solution to the MEP. If $\mathbf{x}^*$ is a global maximizer of (1.4), then*

$$\lambda_i^* \geq \sigma_1(A_{ii}), \tag{3.2}$$

*for every $i = 1, \ldots, m$.*

*Proof.* We prove by contradiction. Suppose that there exists some $1 \leq i \leq m$ such that $\lambda_i^* < \sigma_1(A_{ii})$. Since we also have (3.1), we can assume that the eigenvalues of the matrix $A_{ii} - \lambda_i^* I^{[n_i]}$ are distributed on two sides of zero, say,

$$\eta_1 \geq \cdots \geq \eta_\ell > 0 \geq \eta_{\ell+1} \geq \cdots \geq \eta_{n_i},$$

for some $\ell \geq 1$. Let the corresponding orthogonal eigenvectors of $A_{ii} - \lambda_i^* I^{[n_i]}$ be denoted by $\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_\ell, \mathbf{w}_{\ell+1} \cdots, \mathbf{w}_{n_i}$. We analyze $\mathbf{x}_i^*$ in the following two cases.

**Case I**: Suppose $\mathbf{w}_1^\top \mathbf{x}_i^* \neq 0$. (This includes the special case when $n_i = 1$.) Consider the new vector

$$\hat{\mathbf{x}} := \begin{pmatrix} \mathbf{x}_1^* \\ \vdots \\ \mathbf{x}_{i-1}^* \\ \mathbf{x}_i^* - 2(\mathbf{w}_1^\top \mathbf{x}_i^*)\mathbf{w}_1 \\ \mathbf{x}_{i+1}^* \\ \vdots \\ \mathbf{x}_m^* \end{pmatrix} = \mathbf{x}^* - \underbrace{\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 2(\mathbf{w}_1^\top \mathbf{x}_i^*)\mathbf{w}_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{\mathbf{q}}. \tag{3.3}$$

Observe first that

$$\|\hat{\mathbf{x}}_i\|_2^2 = \|\mathbf{x}_i^* - 2(\mathbf{w}_1^\top \mathbf{x}_i^*)\mathbf{w}_1\|_2^2 = 1.$$

Thus $\hat{\mathbf{x}} \in \mathcal{M}$. Observe further by (2.10) that

$$\begin{aligned} r(\hat{\mathbf{x}}) - r(\mathbf{x}^*) &= \mathbf{q}^\top (A - \Lambda^*)\mathbf{q} = 4(\mathbf{w}_1^\top \mathbf{x}_i^*)^2 \mathbf{w}_1^\top (A_{ii} - \lambda_i^* I^{[n_i]})\mathbf{w}_1 \\ &= 4(\mathbf{w}_1^\top \mathbf{x}_i^*)^2 \eta_1 > 0, \end{aligned}$$

implying that $r(\hat{\mathbf{x}}) > r(\mathbf{x}^*)$. This would contradict with the assumption that $\mathbf{x}^*$ is a global solution to (1.4).

**Case II**: Suppose $\mathbf{w}_1^\top \mathbf{x}_i^* = 0$. Upon substituting the particular vector

$$\overline{\mathbf{z}} := (0, 0, \cdots, \mathbf{w}_1^\top, 0, \cdots, 0)^\top \in \mathbb{R}^n,$$

into (2.8), we see that

$$\overline{\mathbf{z}}^\top P_{\mathbf{x}^*}(A - \Lambda^*)P_{\mathbf{x}^*}\overline{\mathbf{z}} = \mathbf{w}_1^\top (A_{ii} - \lambda_i^* I^{[n_i]})\mathbf{w}_1 = \eta_1 > 0.$$

Based on the second-order necessary condition in Theorem 2.1, we concluse that $\mathbf{x}^*$ could not be even a local maximizer. This is again a contradiction.

The two cases are mutually exclusive. The assertion (3.2) must be true. □

We stress the importance of Theorem 3.2 by stating that a necessary qualification of $\mathbf{x}^*$ being a global maximizer is that each $\lambda_i^*$ must be no less than the largest eigenvalue of the corresponding diagonal block $A_{ii}$. A demonstration by the example in [5] would be instructive.

**Examle 2.** Consider the case where

$$A = \begin{pmatrix} 45 & -20 & 5 & 6 & 16 & 3 \\ -20 & 77 & -20 & -25 & -8 & -21 \\ 5 & -20 & 74 & 47 & 18 & -32 \\ 6 & -25 & 47 & 54 & 7 & -11 \\ 16 & -8 & 18 & 7 & 21 & -7 \\ 3 & -21 & -32 & -11 & -7 & 70 \end{pmatrix} \in \mathbb{R}^{6\times 6}, \tag{3.4}$$

$m = 3$, and $\mathcal{P} = \{2, 2, 2\}$. Both of the following vectors

$$\begin{aligned} \mathbf{x}^* &= (0.4921, -0.8705, 0.8004, 0.5995, 0.5648, -0.8228)^\top, \\ \overline{\mathbf{x}} &= (-0.4003, 0.9164, 0.8847, 0.4661, 0.1191, -0.9929)^\top, \end{aligned}$$

together with the corresponding multivariate eigenvalues

$$\begin{aligned}\Lambda^* &= \text{diag}\{109.2864\ I^{[2]}, 179.7093\ I^{[2]}, 89.9667\ I^{[2]}\},\\ \overline{\Lambda} &= \text{diag}\{75.4245\ I^{[2]}, 114.1218\ I^{[2]}, 125.1093\ I^{[2]}\}\end{aligned}$$

solve the MEP [5]. Nonetheless, we note that

$$\sigma_1(A_{11}) = 86.6125, \quad \sigma_1(A_{22}) = 112.0521, \quad \sigma_1(A_{33}) = 70.9804.$$

Since $\sigma_1(A_{11}) = 86.6125 > \overline{\lambda}_1 = 75.4245$, by Theorem 3.2, we know that $\overline{\mathbf{x}}$ is not a global maximizer for (1.4).

The following two corollaries are in consequence of Theorem 3.2, but nicely generalize the results in [5].

**Corollary 3.3.** *Under the same assumptions of Theorem 3.2, let $D$ be the block diagonal matrix of diagonal blocks of $A$ (see definition (1.7)). If $\mathbf{x}^*$ is a global maximizer of (1.4), then $D - \Lambda^*$ and $A_{ii} - \lambda_i^* I^{[n_i]}$ are negative semi-definite for all $i = 1, \ldots, m$.*

**Corollary 3.4.** *Under the same assumptions of Theorem 3.2, let*

$$\mathbf{v} := (\mathbf{v}_1^\top, \mathbf{v}_2^\top, \cdots, \mathbf{v}_m^\top)^\top \in \mathbb{R}^n$$

*be the vector consisting of unit eigenvectors $\mathbf{v}_i \in \mathbb{R}^{n_i}$, $i = 1, \ldots, m$, each of which is associated with the largest eigenvalue of $A_{ii}$. If $\mathbf{x}^*$ is a global maximizer of (1.4), then $r(\mathbf{x}^*) \geq r(\mathbf{v})$.*

Two more remarks about the applications of Theorem 3.2 are worth noting. First, the inequality (3.2) holds even when $m = 1$ or $A$ is block diagonal. Secondly, the proof in Theorem 3.2 suggests a constructive way to increase the objective value $r(\mathbf{x})$ if $\overline{\mathbf{x}}$ is not yet a global maximizer, which we shall further exploit in Section 4.

Finally, recall the results in [5] asserting that the matrix $A - \Lambda^*$ being negative semi-definite is a sufficient condition for $x^* \in \mathcal{S}$ being a global maximizer of (1.4) and that it is also a necessary condition if $m = 2$ or if $m > 2$ with $A$ being positive. In contrast, our Theorem 3.2 and Corollary 3.3 put forward a necessary condition by using submatrices which are much refined and "localized".

The next theorem relates the multivariate eigenvalues $\lambda_i^*$ of $A$ to the eigenvalues of $A$ in an ordered way, when $A - \Lambda^*$ is negative semi-definite. Note that the positive definiteness of $A$ is now assumed.

**Theorem 3.5.** *Suppose $A \in \mathbb{R}^{n\times n}$ is symmetric and positive semi-definite. If $(\Lambda^*, \mathbf{x}^*)$ is a solution to the MEP such that $A - \Lambda^*$ is negative semi-definite, then*

$$\sigma_j(\Lambda^*) \geq \sigma_j(A),\ j = 1, 2, \cdots, n, \tag{3.5}$$

*where $\sigma_j(\cdot)$ stands for the $j$-th largest eigenvalue of the respective matrices.*

*Proof.* By the classical Monotonicity Theorem for eigenvalue [14, Page 322], it is true that

$$\sigma_1(A - \Lambda^*) \geq \sigma_j(A) + \sigma_{n-j+1}(-\Lambda^*),\ j = 1, 2, \cdots, n. \tag{3.6}$$

The positive semi-definiteness of $A$ and Theorem 3.2 imply that $\lambda_i^* \geq 0$ for all $i = 1, 2, \cdots, m$ and, consequently, $\sigma_{n-j+1}(-\Lambda^*) = -\sigma_j(\Lambda^*)$. Together with the negative semi-definiteness of $A - \Lambda^*$, we see from (3.6) that

$$0 \geq \sigma_1(A - \Lambda^*) \geq \sigma_j(A) - \sigma_j(\Lambda^*),\ j = 1, 2, \cdots, n,$$

which yields the inequalities (3.5). □

## 4. Starting point strategy

By taking advantage of the bounds established in the preceding section for the multivariate eigenvalues, we propose in this section two starting point strategies to help finding a global maximizer for the MCP.

Consider first the situation that at a given point $\overline{\mathbf{x}} \in \mathcal{S}$, there exists a block $A_{ii}$, $1 \le i \le m$, such that the largest eigenvalue $\eta_1$ of $A_{ii} - \overline{\lambda}_i I^{[n_i]}$ is positive. Let $\mathbf{w}_1$ be the unit eigenvector of $A_{ii} - \overline{\lambda}_i I^{[n_i]}$ associated with the eigenvalue $\eta_1$. Assume the scenario that $\mathbf{w}_1^\top \overline{\mathbf{x}}_i \neq 0$. Then the choice $\widetilde{\mathbf{x}} = \overline{\mathbf{x}} - \mathbf{q}$ with $\mathbf{q}$ being defined as in (3.3) is in $\mathcal{M}$ and leads to $r(\widetilde{\mathbf{x}}) = r(\overline{\mathbf{x}}) + 4(\mathbf{w}_1^\top \overline{\mathbf{x}}_i)^2 \eta_1 > r(\overline{\mathbf{x}})$. We then can repeat this process until either $A_{ii} - \overline{\lambda}_i I^{[n_i]}$ becomes negative semi-definite for all $1 \le i \le m$ (Note that this is not sufficient to say that we have reached a global maximizer, since Theorem 3.2 is only a necessary condition), or $\mathbf{w}_1^\top \overline{\mathbf{x}}_i = 0$. In the latter scenario, we consider a new vector of the form

$$\widehat{\mathbf{x}} := \begin{pmatrix} \overline{\mathbf{x}}_1 \\ \vdots \\ \overline{\mathbf{x}}_{i-1} \\ \alpha\overline{\mathbf{x}}_i + \beta\mathbf{w}_1 \\ \overline{\mathbf{x}}_{i+1} \\ \vdots \\ \overline{\mathbf{x}}_m \end{pmatrix} = \overline{\mathbf{x}} - \underbrace{\begin{pmatrix} 0 \\ \vdots \\ 0 \\ (1-\alpha)\overline{\mathbf{x}}_i - \beta\mathbf{w}_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{\widehat{\mathbf{q}}},$$

where real scalars $\alpha$ and $\beta$ are to be determined. It is easy to see that $\widehat{\mathbf{x}} \in \mathcal{M}$ if $\alpha^2 + \beta^2 = 1$. It then follows by (2.10) that

$$\begin{aligned} r(\widehat{\mathbf{x}}) - r(\overline{\mathbf{x}}) &= \widehat{\mathbf{q}}^\top (A - \overline{\Lambda})\widehat{\mathbf{q}} \\ &= (1-\alpha)^2 \overline{\mathbf{x}}_i{}^\top (A_{ii} - \overline{\lambda}_i I^{[n_i]})\overline{\mathbf{x}}_i + \beta^2 \eta_1, \end{aligned}$$

since $\mathbf{w}_1^\top (A_{ii} - \overline{\lambda}_i I^{[n_i]})\overline{\mathbf{x}}_i = \eta_1 \mathbf{w}_1^\top \overline{\mathbf{x}}_i = 0$. For convenience, denote

$$t_i := \overline{\mathbf{x}}_i{}^\top (A_{ii} - \overline{\lambda}_i I^{[n_i]})\overline{\mathbf{x}}_i.$$

Our task is to select $\alpha, \beta$ such that $\alpha^2 + \beta^2 = 1$ and

$$\begin{aligned} r(\widehat{\mathbf{x}}) - r(\overline{\mathbf{x}}) &= (1-\alpha)^2 t_i + \beta^2 \eta_1 = (1-\alpha)^2 t_i + (1-\alpha^2)\eta_1 \\ &= (\alpha - 1)[\alpha(t_i - \eta_1) - (t_i + \eta_1)] > 0. \end{aligned}$$

By the facts that $t_i - \eta_1 \le 0$ and $\eta_1 > 0$, it is easy to check that this quadratic form attains its maximum, which is positive, at some $\alpha^*$. More precisely, we claim that the optimal $\alpha^*$ is given by

$$\alpha^* = \begin{cases} \frac{t_i}{t_i - \eta_1}, & \text{if } -1 \le \frac{t_i}{t_i - \eta_1} < 1, \\ -1, & \text{if } \frac{t_i}{t_i - \eta_1} < -1, \\ 0, & \text{if } t_i = \eta_1. \end{cases}$$

From here, we see that $r(\widehat{\mathbf{x}})$ is an improvement.

In either scenarios, we can always construct a point better than $\overline{\mathbf{x}}$ so long as the condition (3.2) is not met. After repeating this process either finitely many times or until the condition (3.2) is satisfied, we then call upon either the Horst-Jacobi algorithm which monotonically converges [4] or the Gauss-Seidel algorithm whose

monotone convergence will be proved in the next section to produce yet an even better solution to the MEP. Results from our numerical experiments support the effectiveness of this starting point strategy in both reducing the number of iterations and finding the global maximizer.

**Example 3.** Returning to Example 2 discussed earlier, we find that $\overline{\mathbf{x}}_1^\top \mathbf{w}_1 \neq 0$. Applying the scheme described above once, we obtain an updated point

$$\widetilde{\mathbf{x}} = (0.4655, -0.8851, 0.8847, 0.4661, 0.1191, -0.9929)^\top \in \mathcal{M}$$

at which $r(\widetilde{\mathbf{x}}) = 359.3493 > r(\overline{\mathbf{x}}) = 314.6556$. Using $\widetilde{\mathbf{x}}$ as the starting point, the Horst-Jacobi algorithm converges to the global maximizer $\mathbf{x}^*$.

Because both Algorithms 1 and 2 generate an increasing sequence $\{r(\mathbf{x}^{(k)})\}$ along the iteration, the larger the initial value $\{r(\mathbf{x}^{(0)})\}$ is, the better chance the sequence $\{\mathbf{x}^{(k)}\}$ converges to a global maximizer. With this in mind, we approximate the original matrix $A$ by its block diagonal $D$. Observe that the vector $\mathbf{v}$ defined in Lemma 3.4 is a global maximizer of the resulting approximate problem. We may thus think heuristically that $\mathbf{v}$ might be a reasonable approximation to the global maximizer. That is, setting $\mathbf{x}^{(0)} = \mathbf{v} \in \mathcal{M}$ might be another slightly easier-to-use staring point strategy. For large scale problems, perhaps we can adopt only a few diagonal blocks $A_{ii}$, say, those with relatively smaller sizes, and assign $\mathbf{x}_i^{(0)} = \mathbf{v}_i$ only at those chosen blocks. Indeed, this choice of $\mathbf{v}$ as the starting point is particularly attractive if $A$ is positive as we shall see below.

**Theorem 4.1.** *Suppose that $A$ is a symmetric and positive matrix. Then there is a unique positive vector $\mathbf{v} = (\mathbf{v}_1^\top, \cdots, \mathbf{v}_m^\top)^\top \in \mathcal{M}$, where $\mathbf{v}_i \in \mathbb{R}^{n_i}$ is the unit eigenvector of $A_{ii}$ corresponding to the largest eigenvalue $\sigma_1(A_{ii})$, such that $\lambda_i(\mathbf{v}) \geq \sigma_1(A_{ii})$ for every $i = 1, 2, \cdots, m$. The inequality holds strictly when $m > 1$.*

*Proof.* By the Perron-Frobenius Theorem [14, Page 473], the largest eigenvalue $\sigma_1(A_{ii})$ of $A_{ii}$ is the spectral radius of $A_{ii}$ for each $i = 1, 2, \cdots, m$. Also, there is a unique unit eigenvector $\mathbf{v}_i$ with positive elements associated with $\sigma_1(A_{ii})$. It is obvious that

$$\lambda_i(\mathbf{v}) = \mathbf{v}_i^\top A_i \mathbf{v} = \mathbf{v}_i^\top A_{ii} \mathbf{v}_i + \sum_{j \neq i}^{m} \mathbf{v}_i^\top A_{ij} \mathbf{v}_j \geq \sigma_1(A_{ii}),$$

since $\mathbf{v}_i, \mathbf{v}_j$ and $A_{ij}$ are all of positive elements for $i, j = 1, 2, \cdots, m$. □

It is clear that both stated starting point strategies can be combined easily and in all our numerical experiments in Section 6, we have employed this combined starting point strategy, which guarantees that the computed point satisfies the necessary global optimal condition Theorem 3.5. Testing results seem to suggest that this strategy always improves the likelihood of convergence to a global maximizer.

## 5. Convergence analysis

This section is momentarily independent of the preceding three sections. Our goal here is to investigate the dynamical behavior of the Gauss-Seidel Algorithm and show its convergence to a solution of the MEP, provided that $A$ is a generic, symmetric, and positive definite matrix. We shall eventually implement this algorithm together with the starting point strategy proposed earlier as a more promising numerical means for solving the MCP. The following analysis follows along the same path as that carried out for Algorithm 1 in [4]. The key point in the proof is to show

that the objective values $\{r(\mathbf{x}^{(k)})\}$ at the iterates generated by the Gauss-Seidel Algorithm is an increasing sequence.

The Horst-Jacobi algorithm can neatly be expressed in the compact form

$$A\mathbf{x}^{(k)} = \Lambda^{(k)}\mathbf{x}^{(k+1)},$$

with

$$\Lambda^{(k)} := \text{diag}\{\lambda_1^{(k)} I^{[n_1]}, \lambda_2^{(k)} I^{[n_2]}, \cdots, \lambda_m^{(k)} I^{[n_m]}\}, \tag{5.1}$$

which then motivates the idea of casting the scheme as a variant of the classical power method [4]. Similarly, the Gauss-Seidel algorithm can be written in matrix form as

$$(D+U)\mathbf{x}^{(k)} = (\Lambda^{(k)} - U^\top)\mathbf{x}^{(k+1)}, \tag{5.2}$$

where $A$ is split as

$$A = D + U + U^\top, \tag{5.3}$$

with $U$ being the strictly block upper triangular matrix of $A$. This compact form significantly facilitates the convergence analysis. For convenience, denote

$$\begin{aligned}
\mathbf{s}^{(k)} &:= \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}, \\
r^{(k)} &:= r(\mathbf{x}^{(k)}), \\
c^{(k)} &:= \sum_{i=1}^{m} \lambda_i^{(k)} = \mathbf{x}^{(k+1)^\top}\Lambda^{(k)}\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)^\top}\Lambda^{(k)}\mathbf{x}^{(k)}.
\end{aligned}$$

To assess the progress made by the iterations, we rewrite $r^{(k)}$ as

$$\begin{aligned}
r^{(k)} &= \mathbf{x}^{(k)^\top}\left(D + U + U^\top\right)\mathbf{x}^{(k)} \\
&= \mathbf{x}^{(k)^\top}\left(\Lambda^{(k)} - U^\top\right)\mathbf{x}^{(k+1)} + \mathbf{x}^{(k)^\top}U\mathbf{x}^{(k)} \\
&= \mathbf{x}^{(k)^\top}\Lambda^{(k)}\mathbf{x}^{(k+1)} - \mathbf{s}^{(k)^\top}U\mathbf{x}^{(k)} \\
&= \mathbf{x}^{(k)^\top}\Lambda^{(k)}\mathbf{s}^{(k)} - \mathbf{s}^{(k)^\top}U\mathbf{x}^{(k)} + c^{(k)} \\
&= \mathbf{s}^{(k)^\top}\left(\Lambda^{(k)} - U\right)\mathbf{x}^{(k)} + c^{(k)}. \qquad (5.4)
\end{aligned}$$

In the above, the second equation follows from substitution of (5.2) while the third and the fourth equations make use the definitions of $\mathbf{s}^{(k)}$ and $c^{(k)}$. We also can express $r^{(k+1)}$ as

$$\begin{aligned}
r^{(k+1)} &= \mathbf{x}^{(k+1)^\top}\left(D + U + U^\top\right)\mathbf{x}^{(k+1)} \\
&= \mathbf{x}^{(k+1)^\top}(D+U)\mathbf{x}^{(k+1)} - \mathbf{x}^{(k+1)^\top}(D+U)\mathbf{x}^{(k)} \\
&\quad + \mathbf{x}^{(k+1)^\top}(D+U)\mathbf{x}^{(k)} + \mathbf{x}^{(k+1)^\top}U^\top\mathbf{x}^{(k+1)} \\
&= \mathbf{x}^{(k+1)^\top}(D+U)\mathbf{s}^{(k)} + \mathbf{x}^{(k+1)^\top}\left(\Lambda^{(k)} - U^\top\right)\mathbf{x}^{(k+1)} \\
&\quad + \mathbf{x}^{(k+1)^\top}U^\top\mathbf{x}^{(k+1)} \\
&= \mathbf{x}^{(k+1)^\top}(D+U)\mathbf{s}^{(k)} + c^{(k)}, \qquad (5.5)
\end{aligned}$$

where again the relationship (5.2) is used to obtain the middle term in the third equation. We thus can calculate the

$$\begin{aligned}
\Delta r^{(k)} &:= r^{(k+1)} - r^{(k)} \\
&= \mathbf{x}^{(k+1)^\top}(D+U)\,\mathbf{s}^{(k)} - \mathbf{s}^{(k)^\top}\left(\Lambda^{(k)} - U\right)\mathbf{x}^{(k)} \\
&= \mathbf{x}^{(k+1)^\top}(D+U)\,\mathbf{s}^{(k)} - \mathbf{x}^{(k)^\top}(D+U)\,\mathbf{s}^{(k)} \\
&\quad + \mathbf{x}^{(k)^\top}(D+U)\,\mathbf{s}^{(k)} - \mathbf{s}^{(k)^\top}\left(\Lambda^{(k)} - U\right)\mathbf{x}^{(k)} \\
&= \mathbf{s}^{(k)^\top}(D+U)\,\mathbf{s}^{(k)} + \mathbf{x}^{(k)^\top}\left(D + U - \Lambda^{(k)} + U^\top\right)\mathbf{s}^{(k)} \\
&= \mathbf{s}^{(k)^\top}(D+U)\,\mathbf{s}^{(k)} + \mathbf{x}^{(k)^\top}\left(D + U^\top\right)\mathbf{s}^{(k)} - \mathbf{x}^{(k)^\top}(\Lambda^{(k)} - U)\mathbf{s}^{(k)} \\
&= \mathbf{s}^{(k)^\top}(D+U)\,\mathbf{s}^{(k)} + \mathbf{x}^{(k+1)^\top}\left(\Lambda^{(k)} - U\right)\mathbf{s}^{(k)} - \mathbf{x}^{(k)^\top}(\Lambda^{(k)} - U)\mathbf{s}^{(k)} \\
&= \mathbf{s}^{(k)^\top}(D+U)\,\mathbf{s}^{(k)} + \mathbf{s}^{(k)^\top}\left(\Lambda^{(k)} - U\right)\mathbf{s}^{(k)} \\
&= \mathbf{s}^{(k)^\top}\left(D + \Lambda^{(k)}\right)\mathbf{s}^{(k)}. \qquad (5.6)
\end{aligned}$$

If each diagonal block $A_{ii}$, $i = 1, \ldots, m$, is symmetric and positive definite (note that this assumption is much weaker than $A$ being positive-definiteness), then so is $D + \Lambda^{(k)}$. It follows that $r^{(k+1)} \geq r^{(k)}$ by (5.6). Clearly, the sequence $\{r^{(k)}\}$ is bounded above. Thus, the monotone sequence $\{r^{(k)}\}$ must converge. We have proved the following theorem.

**Theorem 5.1.** *Suppose $A$ is symmetric and its diagonal blocks $A_{ii}$ are positive definite for $i = 1, \ldots, m$. Then the sequence $\{r(\mathbf{x}^{(k)})\}$ with $\{\mathbf{x}^{(k)}\}$ generated by the Gauss-Seidel algorithm is a monotone increasing sequence and converges.*

Trivially, $\Lambda^{(k)}$ is bounded for all $k$. From (5.6), there exists a constant $\kappa > 0$ such that

$$\Delta r(\mathbf{x}^{(k)}) \geq \kappa \|\mathbf{s}^{(k)}\|_2^2.$$

It follows that the sequence $\{\mathbf{s}^{(k)}\}$ converges to zero. As useful byproducts, we can derive convergence of various other objects in the iteration.

**Corollary 5.2.** *Under the same assumptions of Theorem 5.1, the residual vectors $\{\delta\mathbf{x}^{(k)}\}$ defined by*

$$\delta\mathbf{x}^{(k)} := A\mathbf{x}^{(k)} - \Lambda^{(k)}\mathbf{x}^{(k)} \qquad (5.7)$$

*with $\{\mathbf{x}^{(k)}\}$ generated by the Gauss-Seidel algorithm converges to zero.*

*Proof.* Using (5.2), it is easy to see from (5.7) that

$$\begin{aligned}
\delta\mathbf{x}^{(k)} &= A\mathbf{x}^{(k)} - \Lambda^{(k)}\mathbf{x}^{(k)} \\
&= (D + U + U^\top - \Lambda^{(k)})\mathbf{x}^{(k)} \\
&= (\Lambda^{(k)} - U^\top)\mathbf{x}^{(k+1)} - (\Lambda^{(k)} - U^\top)\mathbf{x}^{(k)} \\
&= (\Lambda^{(k)} - U^\top)\mathbf{s}^{(k)}. \qquad (5.8)
\end{aligned}$$

The limiting behavior of $\{\delta\mathbf{x}^{(k)}\}$ is now clear, since $\{\mathbf{s}^{(k)}\}$ converges to zero. □

**Corollary 5.3.** *Suppose $A$ is symmetric and $A_{ii}$ is positive definite for $i = 1, \ldots, m$. Suppose further that $A$ has $n$ distinct eigenvalues. Then the sequence $\{\Lambda^{(k)}\}$ converges as $k$ goes to infinity.*

*Proof.* By compactness, we can only be sure that $\{(\Lambda^{(k)}, \mathbf{x}^{(k)})\}$ has a convergent subsequence $\{(\Lambda^{(k_j)}, \mathbf{x}^{(k_j)})\}$ whose limit point is a solution to the MEP. On the other hand, recall that under the generic assumption the MEP has only finitely many solutions [4, Theorem 3.8], that is, there are only finitely many such limit points.

Partition the matrices $U^\top$ and $D+U$ into rows of blocks and denote

$$U^\top = \begin{pmatrix} (U^\top)_1 \\ \vdots \\ (U^\top)_m \end{pmatrix}, \text{ and } D+U = \begin{pmatrix} (D+U)_1 \\ \vdots \\ (D+U)_m \end{pmatrix}. \tag{5.9}$$

By the iterative scheme of the Gauss-Seidel algorithm (5.2) and (5.9), we have

$$\lambda_i^{(k)} = \|(U^\top)_i \mathbf{x}^{(k+1)} + (D+U)_i \mathbf{x}^{(k)}\|_2,$$

and thus

$$\begin{aligned} |\lambda_i^{(k+1)} - \lambda_i^{(k)}| &= | \, \|(U^\top)_i \mathbf{x}^{(k+2)} + (D+U)_i \mathbf{x}^{(k+1)}\|_2 \\ &\quad - \|(U^\top)_i \mathbf{x}^{(k+1)} + (D+U)_i \mathbf{x}^{(k)}\|_2 \, | \\ &\le \|(U^\top)_i \mathbf{s}^{(k+1)} + (D+U)_i \mathbf{s}^{(k)}\|_2 \\ &\le \|(U^\top)_i\|_2 \|\mathbf{s}^{(k+1)}\|_2 + \|(D+U)_i\|_2 \|\mathbf{s}^{(k)}\|_2. \end{aligned}$$

By Theorem 5.1, we see that $|\lambda_i^{(k+1)} - \lambda_i^{(k)}| \to 0$ as $k$ goes to infinity. This fact together with the finite cardinality of limit points is sufficient to conclude the convergence of $\{\Lambda^{(k)}\}$ by using the proposition in [4, Lemma 4.3]. □

Finally, we establish the convergence of the iterates $\{\mathbf{x}^{(k)}\}$.

**Theorem 5.4.** *Under the assumptions of Corollary 5.3, the sequence $\{\mathbf{x}^{(k)}\}$ from the Gauss-Seidel algorithm converges monotonically to a solution to the MEP.*

*Proof.* The proof basically is the same as that in Theorem 5.3. Observe from Theorem 5.1 that componentwise the difference $\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$ converges to zero as $k$ goes to infinity. With the help of the proposition in [4, Lemma 4.3] again, this is sufficient enough to conclude that $\{\mathbf{x}^{(k)}\}$ converges to a solution of the MEP. □

We conclude this section by pointing out that the updating of $\mathbf{x}_i^{(k)}$ in the Gauss-Seidel Algorithm can be carried out in any order. Algorithm 2 is a forward version in the sense that the updates $\mathbf{x}^{(k)}$ are calculated from $\mathbf{x}_1^{(k)}$ to $\mathbf{x}_m^{(k)}$. A backward Gauss-Seidel algorithm, for example, updates $\mathbf{x}^{(k)}$ from $\mathbf{x}_m^{(k)}$ to $\mathbf{x}_1^{(k)}$. It then can be expressed in the compact form

$$(D+U^\top)\mathbf{x}^{(k)} = (\Lambda^{(k)} - U)\mathbf{x}^{(k+1)}. \tag{5.10}$$

Convergence results follow by using the same arguments as before. We speculate that such a rearrangement of order in computation might have some advantages when the blocks in the partition $\mathcal{P}$ are of diverse sizes, but we will not report our investigation in this regard in this paper.

## 6. Numerical experiment

We have experimented with our proposed methods extensively. In this section we report some testing results to demonstrate two main points — the effectiveness of our (combined) starting point strategy in Section 4, and the superior convergence behavior of the Gauss-Seidel algorithm over the Horst-Jacobi algorithm. All of our tests are carried by the MATLAB 7.1 on a PC with Intel(R) Core(TM)2 Duo CPU E8400 @3.00GHz.

6.1. **The effectiveness of the starting point strategy.** We first examine the effect of the starting point strategy. For this purpose, we need the testing problems whose global maxima are known *a prior*. Two small size problems are already available in the literature.

**Example 4.** The matrix $A$ is given by

$$A = \begin{pmatrix} 4.3229 & 2.3230 & -1.3711 & -0.0084 & -0.7414 \\ 2.3230 & 3.1181 & 1.0959 & 0.1285 & 0.0727 \\ -1.3711 & 1.0959 & 6.4920 & -1.9883 & -0.1878 \\ -0.0084 & 0.1285 & -1.9883 & 2.4591 & 1.8463 \\ -0.7414 & 0.0727 & -0.1878 & 1.8463 & 5.8875 \end{pmatrix} \in \mathbb{R}^{5\times 5},$$

with $m = 2$ and $\mathcal{P} = \{2, 3\}$. This example is from [4].

**Example 5.** The matrix $A$ is of size $6 \times 6$ is already described in (3.4) with $m = 3$ and $\mathcal{P} = \{2, 2, 2\}$. This example is from [5].

The sizes of the preceding two problems are small enough that their global maximizers are characterizable through an exhaustive search. Both problems have already served as examples in [4] and [5]. It is known that when the Horst-Jacobi algorithm is applied to these two problems, the iterates have high probabilities of converging to local maxima. In fact, by randomly choosing $10^4$ starting points for the Horst-Jacobi algorithm and the Gauss-Seidel algorithm, Table 1 summarizes their sample probabilities (labelled as "% to Global"), out of the random tests, of convergence to a global maximizer, and their corresponding average numbers of iterations (labelled as "Avg. Iter. #") needed to meet the stopping criterion $\|\delta\mathbf{x}^{(k)}\| \le 10^{-6}$. On the other hand, when our combined starting point strategy is employed, both problems are solved successfully, and for the Horst-Jacobi algorithm moreover, the numbers of iterations for Examples 4 and 5 are 58 and 18 respectively, while for the Gauss-Seidel algorithm, the numbers of iterations for Examples 4 and 5 are 56 and 17, respectively.

Table 1. Performances of the Horst-Jacobi algorithm and the Guass-Seidel algorithm on Example 4 and Example 5.

| Example | The Horst-Jacobi algorithm | | The Guass-Seidel algorithm | |
|---|---|---|---|---|
| | Avg. Iter. # | % to Global | Avg. Iter. # | % to Global |
| Example 4 ($10^4$ random $x^{(0)}$'s) | 80.96 | 53.95 | 76.91 | 52.41 |
| Example 5 ($10^4$ random $x^{(0)}$'s) | 22.64 | 88.65 | 23.10 | 89.39 |

For problems of relatively larger sizes, though it would be harder to perform an thorough search among all local solutions to verify a global maximizer, to create

an MCP with the known global maximizer could still be realized in the following two different ways. The first way is based on the necessary and sufficient global optimality for the case when $m = 2$ or when $A$ is a positive matrix. However, for the later case, it has been proved recently that both algorithms are able to converge to the global solution whenever $\mathbf{x}^{(0)}$ is nonnegative [20], and therefore, we use the case $m = 2$ to demonstrate the effectiveness of the starting point strategy. By selecting the testing matrices from the Matrix Market[1], and by randomly choosing their corresponding partitions $\mathcal{P}$, we generate our first relatively large size testing problems set in Example 6. The other way to produce the testing problems is based on the solution of the following semi-definite programming (SDP):

$$\begin{aligned} &\min F(A) \\ &\left\{ \begin{array}{rcl} A\mathbf{x}^* & = & \Lambda^*\mathbf{x}^* \\ \Lambda^* - A & \succeq & 0 \\ A & \succeq & 0, \end{array} \right. \end{aligned} \tag{6.1}$$

with the preselected $\mathbf{x}^*$ and the corresponding $\Lambda^*$. It is clear according to the sufficient global optimality [5] that any feasible point $A$ of the previous problem (6.1) yields a global maximizer $\mathbf{x}^*$ of the MCP, and therefore, by randomly prescribing $\mathbf{x}^*, \Lambda^*$ and the objective function $F(A)$, we have the second set of testing problems described in Example 7.

**Example 6.** We choose the first 5 symmetric and positive definite matrices (see Table 2) in the set BCSSTRUC1 from the Harwell-Boeing collection as the original matrices $A$'s for the MCP. For each matrix in Table 2, we then randomly partition (1000 times) the size $n$ into two parts to form $\mathcal{P}$, each of which consequently corresponds to a particular MCP problem.

The behavior of both the Horst-Jacobi algorithm and the Gauss-Seidel algorithm starting from random $\mathbf{x}^{(0)}$'s, with and without the starting point strategy, is summarized in Table 3. We terminate the iterations whenever the residual reaches $\|\delta\mathbf{x}^{(k)}\|_2 \leq 10^{-6}$ or $k > 10000$. Under the column "% to Global" are the sample probabilities, out of the random tests, of convergence to a global maximizer. Under the column "Avg. Iter. #" are the average number of iterations needed to meet the stopping criteria. The rows marked by "Strategy active" mean that we have activated our combined starting point strategy.

TABLE 2. Summary of the testing problems in Example 6.

| Matrix name | Size | Condition number | 2-norm |
|---|---|---|---|
| BCSSTK01 | $48 \times 48$ | $1.6e+06$ | $3.0e+09$ |
| BCSSTK02 | $66 \times 66$ | $1.3e+04$ | $1.8e+04$ |
| BCSSTK03 | $112 \times 112$ | $9.5e+06$ | $2.0e+11$ |
| BCSSTK04 | $132 \times 132$ | $5.6e+06$ | $9.6e+06$ |
| BCSSTK05 | $153 \times 153$ | $3.5e+04$ | $6.2e+06$ |

[1] http://math.nist.gov/MatrixMarket/

TABLE 3. Performances of the Horst-Jacobi algorithm, the Guass-Seidel algorithm and the starting point strategy on Example 6.

| Example 6 | The Horst-Jacobi algorithm | | The Guass-Seidel algorithm | |
|---|---|---|---|---|
| Matrix name | Avg. Iter. # | % to Global | Avg. Iter. # | % to Global |
| BCSSTK01 | 1391.41 | 63.40 | 1369.72 | 63.60 |
| BCSSTK01 (*Strategy active*) | 1304.01 | 100.00 | 1280.31 | 100.00 |
| BCSSTK02 | 251.61 | 97.30 | 232.10 | 97.50 |
| BCSSTK02 (*Strategy active*) | 262.39 | 100.00 | 246.86 | 100.00 |
| BCSSTK03 | 9460.02 | 13.30 | 9557.82 | 12.70 |
| BCSSTK03 (*Strategy active*) | 6825.78 | 56.20 | 6486.28 | 55.30 |
| BCSSTK04 | 2951.21 | 92.00 | 2835.92 | 92.00 |
| BCSSTK04 (*Strategy active*) | 2450.22 | 94.30 | 2365.93 | 94.30 |
| BCSSTK05 | 424.72 | 82.90 | 406.58 | 83.80 |
| BCSSTK05 (*Strategy active*) | 555.26 | 100.00 | 542.64 | 100.00 |

**Example 7.** To form the testing problems in this case, we first generate the partition $\mathcal{P}_m = \{n_1, n_2, \cdots, n_m\}$ (for $m = 3, 5$ and 7) whose element $n_i$ is randomly chosen among $\{2, 3, \cdots, 10\}$. Then we generate the diagonal matrix $\Lambda^* = \text{diag}\{\lambda_1^* I^{[n_1]}, \cdots, \lambda_m^* I^{[n_m]}\}$ together with the corresponding random global solution $\mathbf{x}^*$ for the MCP, where $\lambda_i^*$ is a random number uniformly distributed in the interval $(100(i-1), 100i)$. It is clear that the choice of the objective function $F(A)$ has significant effect on the convergence behavior of algorithms (Note that $\Lambda^*$ itself is a feasible solution to (6.1)). For this reason, we have tried the following objective functions in our testing:

$$F_0(A) = [\,],\ F_1(A) = \text{trace}(AR),\ \text{and}\ F_2(A) = \|A - R\|_F.$$

By $F_0(A) = [\,]$, we mean that there is no objective function specified in (6.1) and the $n$-by-$n$ matrix $R$ in either $F_1(A)$ or $F_2(A)$ is randomly generated in each testing case whose elements are uniformly distributed in the interval $(0, \max(\lambda_1^*, \cdots, \lambda_m^*))$.

The free MATLAB-based toolbox `YALMIP`[2] is employed to solve the resulted problem (6.1). For the objective function $F_0(A)$, `YALMIP` then returns an arbitrary feasible point $A$ for (6.1). To effectuate the comparison, for each $m$, we generate 1000 testing problems by providing random $\mathbf{x}^*, \Lambda^*$ and the matrix $R$ for $F_1(A)$ and $F_2(A)$. Numerical results of both algorithms from randomly selected starting

[2]`YALMIP` is a convenient interface for multiple external optimization solvers and is a free MATLAB-based toolbox available at: http://control.ee.ethz.ch/~joloef/yalmip.php. It unifies and facilitates the different formats in semi-definite programming software. Here, we employ the `SDPT3` package [16, 17] as the solver for our problem (6.1). Refer to [16, 17] and the associated website for a list of commands and options.

points $\mathbf{x}^{(0)}$'s are documented in Table 4 with the same meaning of each term as in Table 3.

TABLE 4. Performances of the Horst-Jacobi algorithm, the Guass-Seidel algorithm and the starting point strategy on Example 7.

| Example 7 | | The Horst-Jacobi algorithm | | The Guass-Seidel algorithm | |
|---|---|---|---|---|---|
| | $m$ | Avg. Iter. # | % to Global | Avg. Iter. # | % to Global |
| $F_0(A)$ | $m = 3$ | 56.98 | 100.00 | 56.82 | 100.00 |
| | $m = 3$(*Strategy active*) | 66.10 | 100.00 | 66.60 | 100.00 |
| | $m = 5$ | 53.19 | 100.00 | 53.62 | 100.00 |
| | $m = 5$(*Strategy active*) | 54.01 | 100.00 | 56.12 | 100.00 |
| | $m = 7$ | 53.12 | 100.00 | 52.34 | 100.00 |
| | $m = 7$(*Strategy active*) | 63.14 | 100.00 | 62.57 | 100.00 |
| $F_1(A)$ | $m = 3$ | 627.55 | 80.10 | 589.32 | 78.30 |
| | $m = 3$(*Strategy active*) | 613.48 | 95.90 | 523.01 | 96.00 |
| | $m = 5$ | 714.36 | 76.80 | 668.56 | 75.30 |
| | $m = 5$(*Strategy active*) | 749.95 | 92.30 | 695.84 | 90.40 |
| | $m = 7$ | 1064.03 | 68.60 | 1015.60 | 68.50 |
| | $m = 7$(*Strategy active*) | 983.10 | 88.60 | 889.85 | 87.60 |
| $F_2(A)$ | $m = 3$ | 1182.30 | 76.80 | 1053.71 | 73.40 |
| | $m = 3$(*Strategy active*) | 1238.10 | 90.50 | 1128.10 | 90.80 |
| | $m = 5$ | 3401.61 | 60.80 | 3170.92 | 57.40 |
| | $m = 5$(*Strategy active*) | 3249.40 | 71.40 | 3016.42 | 71.80 |
| | $m = 7$ | 5908.93 | 42.50 | 5732.43 | 40.70 |
| | $m = 7$(*Strategy active*) | 5955.82 | 52.30 | 5620.13 | 51.00 |

Besides the preliminary evidence in the average number of iterations in these tested examples that the Gauss-Seidel algorithm generally converges faster than the Horst-Jacobi algorithm (more extensive numerical experiments will be carried out in the next subsection), the most interesting point is the evidence that when our starting point strategy is employed, not only the number of iterations could possibly be reduced in both algorithms, but also, and most importantly, the probability to global solution is also increased significantly. We think this feature should have consequential effect on applications.

6.2. **The convergence speeds.** This subsection focuses on the comparison of the convergence speeds between the Horst-Jacobi algorithm and the Gauss-Seidel algorithm. As already shown in Subsection 6.1 that they roughly have the same probability in reaching the global solution, we therefore only need to compare their speeds in meeting the same stopping criterion $\|\delta\mathbf{x}^{(k)}\|_2 \leq 10^{-6}$. For this purpose, we think it is sufficient to simply feed both algorithms with randomly generated matrices together with the starting points and make a comparison between their average numbers of iterations as well as the CPU times. By fixing the matrix size

as $1000 \times 1000$ and the partitions

$$\mathcal{P}_m = \{\underbrace{\frac{1000}{m}, \frac{1000}{m}, \cdots, \frac{1000}{m}}_{m}\}, \text{ for } m = 2, 5, 10, 20, 100,$$

we recorded their corresponding average iteration numbers ("Avg. Iter. #" ) and the CPU times ("Avg. CPU(s)") over $10^4$ random tests in Table 5.

TABLE 5. The summary of iterations of the Horst-Jacobi and the Guass-Seidel algorithm, over $10^4$ random tests.

| | The Horst-Jacobi algorithm | | The Guass-Seidel algorithm | |
|---|---|---|---|---|
| $m$ | Avg. Iter. # | Avg. CPU(s) | Avg. Iter. # | Avg. CPU(s) |
| 2 | 1660.20 | 45.20 | 1394.41 | 28.91 |
| 5 | 2051.72 | 41.57 | 1551.81 | 27.93 |
| 10 | 1928.22 | 47.62 | 1389.50 | 34.19 |
| 20 | 2050.51 | 52.77 | 1394.60 | 39.84 |
| 100 | 1879.71 | 46.69 | 1249.30 | 39.17 |

A more clear demonstration to see the faster convergence of the Guass-Seidel algorithm is shown in Figure 1 where $10^4$ random $500 \times 500$ $A$'s are tested. The relationship between the average numbers of iterations (left) as well as the average CPU times (right) in terms of the partition numbers $m$ are plotted. As is shown in this figure, the Guass-Seidel algorithm uses much less iteration number and the CPU time than the Horst-Jacobi algorithm as $m$ increases.

## 7. Conclusions

Although it arises naturally from the Lagrangian principle for the maximal correlation problem which finds applications in statistical data analysis, the multivariate eigenvalue problem is itself an interesting generalization of the classical eigenvalue problem. This paper is not about solving the multivariate eigenvalue problem per se. Rather, we are interested in find some particular eigenpairs that maximize the correlations.

While the conventional power method applied to the classical eigenvalue problem is known to converge generically to the dominant eigenvector, the aggregated power method such as the Horst-Jacobi algorithm does not have that property, which has consequential impact on statistical applications. By analyzing the first-order and second-order optimal conditions of the MCP, we are able to characterize the multivariate eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_m$ for the MEP in terms of the eigenvalues of the original matrix $A$ and its block diagonal $D$. In particular, we are able to provide a lower bound for each of the multivariate eigenvalues at a global maximizer, which motivates an effective starting point strategy. Numerical tests seem to support that our approaches significantly increase the probability of finding a global maximizer.

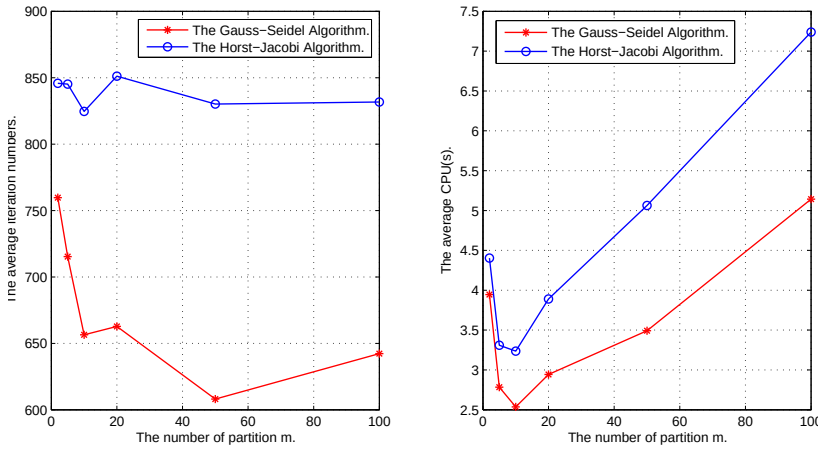


FIGURE 1. Experimental comparison of convergence speeds of the Gauss-Seidel algorithm and the Horst-Jacobi algorithm.

As an improved alternative, we also study analytically and test numerically the convergence behavior of the Gauss-Seidel algorithm. We find that in general the Gauss-Seidel algorithm does use less number of iterations than the conventional Horst-Jacobi algorithm.

Possible future research topics include the convergence analysis of the SOR algorithm proposed in [4] and of other formulations such as those rate acceleration tactics already developed for matrix iterative analysis. From the eigenvalue computation point of view, the notion of a multivariate shift, Rayleigh quotient type iteration scheme sketched in Algorithm 3 also seems appealing. Obviously, when

**Algorithm 3** A Rayleigh quotient type iteration for the MEP.

Given $\mathbf{x}^{(0)} \in \mathbb{R}^n$,
**for** $k = 0, 1, \cdots$, **do**
  Solve $\big(A - \Lambda(\mathbf{x}^{(k)})\big)\,\mathbf{y}^{(k)} := \mathbf{x}^{(k)}$
  **for** $i = 1, 2, \cdots, m$ **do**
    $\mathbf{x}_i^{(k+1)} := \dfrac{\mathbf{y}_i^{(k)}}{\|\mathbf{y}_i^{(k)}\|_2}$
  **end for**
**end for**

$m = 1$, Algorithm 3 reduces to the classical Rayleigh quotient iteration which is known to enjoy cubic convergence. Taking the concern for maximizing the correlations out of consideration, it will be of great interest to investigate whether some of the classical eigenvalue computation techniques can be generalized to solve this multivariate eigenvalue problem.

## REFERENCES

1. A. Amri and A. Seeger, *Spectral analysis of coupled linear complementarity problems,* Linear Alg. Appl., to appear, 2009.

2. P. J. Browne and B. D. Sleeman, *A numerical technique for multiparameter eigenvalue problems,* IMA J. Numer. Anal., **2**(1982), pp. 451-457.
3. M. T. Chu and K. R. Driessel, *The projected gradient method for least squares matrix approximations with spectral constraints,* SIAM J. Numer. Anal., **27**(1990), pp. 1050-1060.
4. M. T. Chu and J. L. Watterson, *On a multivariate eigenvalue problem: I. Algebraic theory and power method,* SIAM J. Sci. Comput., vol. **14**(1993), pp. 1089-1106.
5. M. Hanafi and J. M. F. Ten Berge, *Global optimality of the successive Maxbet algorithm,* Psychometrika, **68**(2003), pp. 97-103.
6. D. R. Hardoon, S. Szedmak and J. Shawe-Taylor, Canonical correlation analysis; An overview with application to learning methods, CSD-TR-03-02, University of London, 2003.
7. U. Helmke and J. B. Moore, *Optimization and dynamical systems*, Springer-Verlag, London, UK, 1994.
8. P. Horst, *Relations among m sets of measures,* Psychometrika, **26**(1961), pp. 129-149.
9. P. Horst, *Factor analysis of data matrices,* Holt Rinehart and Winston, New York, 1965.
10. H. Hotelling, *The most predictable criterion,* J. Educ. Pyschol., **26**(1935), pp. 139-142.
11. H. Hotelling, *Relations between two sets of variates,* Biometrika, **28**(1936), pp. 321-377.
12. J. R. Kettenring, *Canonical analysis of several sets of variables,* Biometrika, **58**(1971), pp. 433-451.
13. Z.Y. Liu, J. Qian, and S.F. Xu, *On the double eigenvalue problem,* Preprint, 2009, available online.
14. C. R. Rao and M. B. Rao, *Matrix algebra and its applications to statistics and econometrics,* World Scientific, 1998.
15. J. M. F. Ten Berge, *Generalized approaches to the MAXBET problem and the MAXDIFF problem, with applications to canonical correlations,* Psychometrika, **53** (1988), pp. 487-494.
16. K. C. Toh, M. J. Todd, and R. H. Tutuncu, *SDPT3 — a Matlab software package for semidefinite programming*, Optimization Methods and Software, **11** (1999), pp. 545-581.
17. R. H. Tutuncu, K. C. Toh, and M. J. Todd, *Solving semidefinite-quadratic-linear programs using SDPT3,* Mathematical Programming Ser. B, **95** (2003), pp. 189-217.
18. J. R. Van de Geer, *Linear relations among k sets of variables,* Psychometrika, **49**(1984), pp. 79-94.
19. T. L. Van Noorden and J. Barkmeijer, *The multivariate eigenvalue problem: a new application, theory and a subspace accelerated power method,* Universiteit Utrecht, preprint, 2008.
20. L.-H. Zhang, L.-Z. Liao and L.-M. Sun, *Towards the global solution of the maximal correlation problem*, in preprint.

Department of Applied Mathematics, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai 200433, P. R. China.
*E-mail address*: longzlh@gmail.com

Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205, USA.
*E-mail address*: chu@math.ncsu.edu