A Majorized Penalty Approach for Calibrating Rank Constrained Correlation Matrix Problems

Yan Gao* and Defeng Sun†

This version: May 05, 2010

Abstract

In this paper, we aim at finding a nearest correlation matrix to a given symmetric matrix, measured by the componentwise weighted Frobenius norm, with a prescribed rank and bound constraints on its correlations. This is in general a non-convex and difficult problem due to the presence of the rank constraint. To deal with this difficulty, we first consider a penalized version of this problem and then apply the essential ideas of the majorization method to the penalized problem by solving iteratively a sequence of least squares correlation matrix problems without the rank constraint. The latter problems can be solved by a recently developed quadratically convergent smoothing Newton-BiCGStab method. Numerical examples demonstrate that our approach is very efficient for obtaining a nearest correlation matrix with both rank and bound constraints.

Key words: correlation matrix, penalty method, majorization, least squares, Newton's method

1 Introduction

In recent years, we have witnessed a lot of interests from the finance and insurance industries in finding a nearest correlation matrix[1] whose rank is not more than a given positive integer r. In response to these needs, the quantitative finance community has proposed a variety of nearest correlation matrix problems with rank conditions. Wu [61], Zhang and Wu [63], and Brigo and Mercurio [7] considered such a problem for pricing interest rate derivatives under the LIBOR and swap market models. The factor models of basket options, collateralized debt obligations (CDOs), portfolio risk models (VaR), and multivariate time series discussed by Lillo and Mantegna [37] rely on low rank nearest correlation matrices. A correlation matrix of low rank

*Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076, Republic of Singapore. Email: yangao@nus.edu.sg.

†Department of Mathematics and Risk Management Institute, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076, Republic of Singapore. The author's research was supported by the Academic Research Fund under Grant R-146-000-104-112. Email: matsundf@nus.edu.sg.

[1]A correlation matrix, a commonly used concept in statistics, is a real symmetric and positive semidefinite matrix whose diagonal entries are all ones.

is particularly useful in the Monte Carlo simulation for solving derivatives pricing problems as a model with low factors can significantly reduce the cost of drawing random numbers. Beyond quantitative finance, the rank constrained nearest correlation matrix problems also occur in many engineering fields, see for examples, [9, 13, 29, 55].

Let $\mathcal{S}^n$ and $\mathcal{S}^n_+$ be the space of $n \times n$ symmetric matrices and the cone of positive semidefinite matrices in $\mathcal{S}^n$, respectively. Denote the Frobenius norm induced by the standard trace inner product $\langle \cdot, \cdot \rangle$ in $\mathcal{S}^n$ by $\| \cdot \|$. Let C be a given matrix in $\mathcal{S}^n$ and $H \in \mathcal{S}^n$ a given weight matrix whose entries are nonnegative. Then the rank constrained nearest correlation matrix problem (rank-NCM) can be formulated as follows:

$$\begin{array}{ll} \min & \frac{1}{2}\|H \circ (X - C)\|^2 \\ \text{s.t.} & X_{ii} = 1, \quad i = 1, \ldots, n\,, \\ & X \in \mathcal{S}^n_+\,, \\ & \text{rank}(X) \le r\,, \end{array} \tag{1}$$

where "$\circ$" denotes the Hadamard product, i.e., $(A \circ B)_{ij} = A_{ij}B_{ij}$, $i, j = 1, \ldots, n$ and $r \in \{1, \ldots, n\}$ is a given integer. The weight matrix H is introduced by adding larger weights to correlations that are better estimated or are of higher confidence in their correctness. Zero weights are usually assigned to those correlations that are missing or not estimated. See [45] for more discussions.

The rank-NCM problem (1) has been investigated by many researchers. In [56], Simon gave a comprehensive literature review and summarized thirteen methods for solving the rank-NCM problem (1) and its many different variations. Here we will only briefly discuss several methods which are most relevant to our approach to be introduced in this paper.

We start with mentioning the method of "principal component analysis" (PCA). This method truncates the spectral decomposition of the symmetric matrix C to a positive semidefinite matrix by taking the first r largest eigenvalues of C. Its modified version (mPCA), perhaps firstly introduced by Flurry [21], is to take account of the unit diagonal constraints via a normalization procedure. The mPCA method is very popular in the financial industry due to its simplicity and has been widely implemented by many financial institutions for obtaining a correlation matrix with the required rank. The major drawback of the mPCA approach is that it only produces a non-optimal feasible solution to problem (1). Nevertheless, it can be used as a good initial feasible point for other methods of solving the rank-NCM problem. In terms of finding an optimal solution, Zhang and Wu [63] and Wu [61] took an important step by using a Lagrange dual method to solve the rank-NCM problem (1) with equal weights, i.e., $H = E$, where E is a symmetric matrix whose entries are all ones. Under the assumptions that the given matrix C is a valid correlation matrix and the rth and $(r+1)$th eigenvalues (arranged in the non-increasing order in terms of their absolute values) of $C + \text{diag}(\bar{y})$ have different absolute values, where $\bar{y}$ is an optimal solution to the Lagrange dual problem of (1) and $\text{diag}(\bar{y})$ is a diagonal matrix whose diagonal is $\bar{y}$, Zhang and Wu [63] provided a way to get a global solution of problem (1). This global optimality checking is very rare in non-convex optimization. The Lagrange dual method is effective when the required rank r is large. One interesting question is to know if this method can handle the rank-NCM problem with general matrices C and H. The next major progress is achieved by Pietersz and Groenen [45] who proposed an innovative

row by row alternating majorization method. This method can be applied to problem (1) with an arbitrary symmetric nonnegative weight matrix H and is particularly efficient when r is small as its computational cost at each iteration is of the order $O(r^2n^2)$. In [25], Grubisic and Pietersz introduced a geometric programming approach for solving problem (1). This approach is applicable to any weight matrix H too, but its numerical performance is not so efficient as the majorization method of Pietersz and Groenen as far as we know. Another well studied method for solving problem (1) is the trigonometric parametrization method of Rebonato [50, 51, 52, 53], Brigo [6], Brigo and Mercurio [8] and Rapisarda et al. [49]. In this method, they first decompose $X = RR^T$ with $R \in \mathbb{R}^{n\times r}$ and then parameterize each row vector of R by trigonometric functions through spherical coordinates. The resulting problem is unconstrained, but highly nonlinear and non-convex. It is not clear to us if the problem can be efficiently solved in practice. The trigonometric parametrization method has been considered earlier for the cases without the rank constraint [39, 53]. A class of alternating direction methods, which are easy to implement, are also well studied by many researchers for solving the rank-NCM problem. For example, Morini and Webber [41] suggested an iterative algorithm called eigenvalue zeroing by iteration (EZI). This algorithm generally does not converge to a stationary point of the rank-NCM problem and cannot be extended to the case with a general weight matrix H. Very recently, Li and Qi [38] proposed a sequential semismooth Newton method for solving problem (1) with $H = E$. They formulate the problem as a bi-affine semidefinite programming and then use an augmented Lagrange method to solve a sequence of least squares problems. This approach can be effective when the required rank r is relatively large.

So far we have seen that unless $r \le O(\sqrt{n})$ in which case the majorization method of Pietersz and Groenen [45] is an excellent choice, there still lacks an efficient method for solving the rank-NCM problem (1). The target of this paper is to fill up this gap by developing an efficient method of finding a nearest correlation matrix X with a prescribed rank and bound constraints on its components in the following sense

$$\begin{array}{ll} \min & \theta(X) := \dfrac{1}{2}\|H \circ (X - C)\|^2 \\ \text{s.t.} & X_{ii} = 1,\ i = 1, \ldots, n\,, \\ & X_{ij} = e_{ij},\ (i,j) \in \mathcal{B}_e\,, \\ & X_{ij} \ge l_{ij},\ (i,j) \in \mathcal{B}_l\,, \\ & X_{ij} \le u_{ij},\ (i,j) \in \mathcal{B}_u\,, \\ & X \in \mathcal{S}^n_+\,, \\ & \mathrm{rank}(X) \le r\,, \end{array} \tag{2}$$

where $\mathcal{B}_e$, $\mathcal{B}_l$, and $\mathcal{B}_u$ are three index subsets of $\{(i,j)\,|\,1 \le i < j \le n\}$ satisfying $\mathcal{B}_e \cap \mathcal{B}_l = \emptyset$, $\mathcal{B}_e \cap \mathcal{B}_u = \emptyset$, $-1 \le e_{ij}, l_{ij}, u_{ij} \le 1$ for any $(i,j) \in \mathcal{B}_e \cup \mathcal{B}_l \cup \mathcal{B}_u$, and $-1 \le l_{ij} < u_{ij} \le 1$ for any $(i,j) \in \mathcal{B}_l \cap \mathcal{B}_u$. Denote the cardinalities of $\mathcal{B}_e$, $\mathcal{B}_l$, and $\mathcal{B}_u$ by q_e, q_l, and q_u, respectively. Let $p := n + q_e$ and $m := p + q_l + q_u$. Note that problem (2) is a generalization of problem (1) and for problem (2) to have a feasible solution, the required rank r cannot be arbitrarily chosen as in problem (1) when m is large. In some real-world applications in the financial industry, such bound constraints in problem (2) are imposed in the context of calibraing an improper correlation matrix. For example, in several financial models, such as the Monte Carlo VaR models, stress testing is necessary in order to determine the value of a bank's portfolio

via the stressed correlation matrices. This is usually done by fixing some of the correlations while stressing the remaining ones to be within a certain confidence level to reflect the stressing scenarios. This naturally leads to the nearest correlation matrix problem with both equality and inequality constraints. See [1, 2, 34] for related finance problems with inequality constraints. From numerical algorithmic point of view, however, there is no much progress in extending approaches from problem (1) to deal with the more challenging problem (2). Only recently, Simon [56] extended the majorization method of Pietersz and Groenen [45] by incorporating some equality constraints of the kind $X_{ij} = 0$. But unlike the case for the simpler problem (1), this extension can easily fail even the number of such constraints is not large. The main reason is that the desired monotone decreasing property of the objective function is no longer valid whenever the off-diagonal bounds exist. In this paper we will propose a majorized penalty approach to solve problem (2). Our main idea is to first consider a penalized version of this problem and then to apply the essential ideas of the majorization method to the penalized problem by solving a sequence of diagonal weighted nearest correlation matrix problems, which take the following form

$$\begin{array}{ll} \min & \frac{1}{2}\|D^{1/2}(X-G)D^{1/2}\|^2 \\ \text{s.t.} & X_{ii} = 1,\ i = 1,\ldots,n\,, \\ & X_{ij} = e_{ij},\ (i,j) \in \mathcal{B}_e\,, \\ & X_{ij} \geq l_{ij},\ (i,j) \in \mathcal{B}_l\,, \\ & X_{ij} \leq u_{ij},\ (i,j) \in \mathcal{B}_u\,, \\ & X \in \mathcal{S}^n_+\,, \end{array} \tag{3}$$

where $D^{1/2}$ is a positive definite diagonal matrix and G is some symmetric matrix. Several methods can be readily applied to solve problem (3), such as the projected gradient method of Boyd and Xiao [5], the augmented Lagrangian dual method of Qi and Sun [47], the inexact smoothing Newton-BiCGStab method of Gao and Sun [23] and the SQP-Newton method of Chen, Gao and Liu [10]. Among them, the quadratically convergent smoothing Newton-BiCGStab method is particularly suitable for the efficiency of our majorized penalty approach.

The remaining part of this paper are organized as follows. In Section 2, we present some preliminaries on matrix valued functions and symmetric convex functions defined on matrix spaces. These results will facilitate our subsequent analysis. In Section 3, we introduce our majorized penalty approach for the rank constrained least squares correlation matrix problem (2). In Section 4, we discuss the Lagrangian dual reformulation and the related issues in global optimality checking, which does not need C to be a valid correlation matrix as in [63, Theorem 4.5]. We report some numerical experiments in Section 5 and make our final conclusions including future research directions in Section 6.

2 Preliminaries

For subsequent discussions, in this section we introduce some basic properties of matrix valued functions and real valued symmetric convex functions defined on matrix spaces.

We shall first write problem (2) in a compact form to facilitate the discussions below. Recall that the cardinalities of $\mathcal{B}_e$, $\mathcal{B}_l$, and $\mathcal{B}_u$ are denoted by q_e, q_l, and q_u, respectively. Let $q := q_l + q_u$.

For any $(i,j) \in \{1,\ldots,n\} \times \{1,\ldots,n\}$, define $\mathcal{E}^{ij} \in \mathbb{R}^{n\times n}$ by

$$(\mathcal{E}^{ij})_{lk} := \begin{cases} 1 & \text{if } (l,k) = (i,j)\,, \\ 0 & \text{otherwise}\,, \end{cases} \quad l,k = 1,\ldots,n\,.$$

Denote $A^{ij} := \frac{1}{2}(\mathcal{E}^{ij} + \mathcal{E}^{ji})$. Let $\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^m$ be defined by

$$\mathcal{A}X := \begin{bmatrix} \{\langle A^{ii}, X\rangle\}_{i=1}^n \\ \{\langle A^{ij}, X\rangle\}_{(i,j)\in\mathcal{B}_e} \\ \{\langle A^{ij}, X\rangle\}_{(i,j)\in\mathcal{B}_l} \\ -\{\langle A^{ij}, X\rangle\}_{(i,j)\in\mathcal{B}_u} \end{bmatrix}, \quad X \in \mathcal{S}^n\,, \tag{4}$$

and

$$b := \begin{pmatrix} b_0 \\ \{e_{ij}\}_{(i,j)\in\mathcal{B}_e} \\ \{l_{ij}\}_{(i,j)\in\mathcal{B}_l} \\ -\{u_{ij}\}_{(i,j)\in\mathcal{B}_u} \end{pmatrix},$$

where $b_0 \in \mathbb{R}^n$ is the vector of all ones. Now problem (2) can be equivalently reformulated as

$$\begin{array}{ll} \min & \theta(X) = \dfrac{1}{2}\|H \circ (X - C)\|^2 \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \in \mathcal{S}^n_+\,, \\ & \text{rank}(X) \le r\,, \end{array} \tag{5}$$

where $\mathcal{Q} := \{0\}^p \times \mathbb{R}^q_+$ is a polyhedral convex cone with $1 \le p \le m$ and $p + q = m$. For any symmetric matrix $X \in \mathcal{S}^n$, we write $X \succeq 0$ and $X \succ 0$ to represent that X is positive semidefinite and positive definite, respectively. Let Ω_r denote the feasible set of (5), i.e.,

$$\Omega_r := \{X \in \mathcal{S}^n \,|\, \mathcal{A}X \in b + \mathcal{Q},\ X \succeq 0,\ \text{rank}(X) \le r\}\,. \tag{6}$$

In this paper we always assume that $\Omega_r \neq \emptyset$. Let k be a positive integer. We use $\mathcal{O}_k$ to denote the set of all orthogonal matrices in $\mathbb{R}^{k\times k}$, i.e.,

$$\mathcal{O}_k = \{Q \in \mathbb{R}^{k\times k} \mid Q^TQ = I\}\,,$$

where I is the identity matrix with appropriate dimensions. Let $\tau \subseteq \{1,\ldots,k\}$ be an index set and M be a k by k matrix. We denote the cardinality of τ by $|\tau|$ and the matrix containing the columns in M indexed by τ as M_τ. For any $v \in \mathbb{R}^k$, we use $\text{diag}(v)$ to denote the k by k diagonal matrix whose ith diagonal entry is v_i, $i = 1,\ldots,n$, $\|v\|$ to denote the 2-norm of v, and $\|v\|_0$ to denote the cardinality of the set $\{i \,|\, v_i \neq 0,\ i = 1,\ldots,k\}$. We also use $|v|$ to denote the column vector in $\mathbb{R}^k$ such that its ith component is defined by $|v|_i = |v_i|$, $i = 1,\ldots,k$.

Let $X \in \mathcal{S}^n$ be arbitrarily chosen. Suppose that X has the spectral decomposition

$$X = P\Lambda(X)P^T, \tag{7}$$

where $\Lambda(X) := \mathrm{diag}(\lambda(X))$, $\lambda_1(X) \geq \cdots \geq \lambda_n(X)$ are the eigenvalues of X being arranged in the decreasing order and $P \in \mathcal{O}_n$ is a corresponding orthogonal matrix of orthonormal eigenvectors of X. Define

$$\alpha := \{i \mid \lambda_i(X) > \lambda_r(X)\}, \quad \beta := \{i \mid \lambda_i(X) = \lambda_r(X)\}, \text{ and } \gamma := \{i \mid \lambda_i(X) < \lambda_r(X)\}$$

and write $P = [P_\alpha \; P_\beta \; P_\gamma]$.

For any $z \in \mathbb{R}^n$, let $s_r(z)$ be the sum of the r largest components of z, i.e.,

$$s_r(z) := \sum_{i=1}^{r} z_i^{\downarrow} = \max_{v \in \mathbb{R}^n} \{\, v^T z \mid \sum_{i=1}^{n} v_i = r \,,\; 0 \leq v_i \leq 1,\; i = 1, \ldots, n \,\}.$$

Let $x = \lambda(X)$. Let $\partial_B s_r(x)$ be the B-subdifferential of s_r at x, i.e,

$$\partial_B s_r(x) := \{ \lim_{x^k \to x} s_r'(x^k) \,,\; s_r(\cdot) \text{ is differentiable at } x^k \}.$$

Then, one can easily check that the B-subdifferential of $s_r(\cdot)$ at x is given by

$$\begin{array}{rcl} \partial_B s_r(x) & = & \{v \in \mathbb{R}^{1\times n} \,|\, v_i = 1 \text{ for } i \in \alpha \,,\; v_i = 0 \text{ for } i \in \gamma \,, \\ & & \quad v_i \in \{0,1\} \text{ for } i \in \beta \text{ and } \textstyle\sum_{i\in\beta} v_i = r - |\alpha| \big\}. \end{array} \tag{8}$$

The subdifferential of $s_r(\cdot)$ at x, which is the convex hull of $\partial_B s_r(x)$ (see [12, Theorem 2.5.1]), takes the form of

$$\begin{array}{rcl} \partial s_r(x) & = & \{v \in \mathbb{R}^{1\times n} \,|\, v_i = 1 \text{ for } i \in \alpha \,,\; v_i = 0 \text{ for } i \in \gamma \,, \\ & & \quad 0 \leq v_i \leq 1 \text{ for } i \in \beta \text{ and } \textstyle\sum_{i\in\beta} v_i = r - |\alpha| \big\}. \end{array} \tag{9}$$

Define $S_r : \mathcal{S}^n \to \mathbb{R}$ by

$$S_r(Z) := s_r(\lambda(Z)) \,, \tag{10}$$

where for any $Z \in \mathcal{S}^n$, $\lambda(Z)$ is the column vector containing all the eigenvalues $\lambda_1(Z) \geq \cdots \geq \lambda_n(Z)$ of Z. That is, for any $Z \in \mathcal{S}^n$, $S_r(Z)$ is the sum of the r largest eigenvalues of Z. It is well known that $S_r(\cdot)$ is a convex function [20] and the subdifferential of $S_r(\cdot)$ at X is well defined. By [59, 44, 35] and the structure of $\partial s_r(x)$ given in (9), we can fully characterize $\partial S_r(X)$ as follows

$$\partial S_r(X) = \{\, [\, P_\alpha \;\; P_\beta Q \;\; P_\gamma \,] \mathrm{diag}(v) \, [\, P_\alpha \;\; P_\beta Q \;\; P_\gamma \,]^T \mid v \in \partial s_r(\lambda(X)) \,, Q \in \mathcal{O}_{|\beta|} \,\}. \tag{11}$$

Since $S_r(\cdot)$ is (continuously) differentiable at X if and only if $\lambda_r(X) > \lambda_{r+1}(X)$ (cf. [35]), we know that the B-subdifferential $\partial_B S_r(X)$ of $S_r(\cdot)$ at X is given by

$$\partial_B S_r(X) = \{\, [\, P_\alpha \;\; P_\beta Q \;\; P_\gamma \,] \mathrm{diag}(v) \, [\, P_\alpha \;\; P_\beta Q \;\; P_\gamma \,]^T \mid v \in \partial_B s_r(\lambda(X)) \,, Q \in \mathcal{O}_{|\beta|} \,\}. \tag{12}$$

By noting that

$$\mathbb{R}^{1\times n} \ni (\underbrace{1, \ldots, 1}_{r}, 0, \ldots, 0) \in \partial_B s_r(x) \,,$$

one has

$$(P_\alpha P_\alpha^T + P_{\beta_1} P_{\beta_1}^T) \in \partial_B S_r(X) \subseteq \partial S_r(X)\,, \tag{13}$$

where $\beta_1 := \{|\alpha| + 1, \ldots, r\}$.

Denote

$$\mathcal{S}^n(r) := \{Z \in \mathcal{S}^n \mid \mathrm{rank}(Z) \leq r\}\,. \tag{14}$$

We next discuss the projection operators over the closed convex cone $\mathcal{S}^n_+$ and the closed non-convex cone $\mathcal{S}^n(r)$, respectively. Let $X \in \mathcal{S}^n$ have the spectral decomposition as in (7). Then, since $\mathcal{S}^n_+$ is a closed convex cone in $\mathcal{S}^n$, it follows that the optimization problem

$$\begin{array}{ll} \min & \frac{1}{2}\|Z - X\|^2 \\ \text{s.t.} & Z \in \mathcal{S}^n_+ \end{array} \tag{15}$$

has a unique optimal solution, which is called the metric projection of X over $\mathcal{S}^n_+$ under the Frobenius norm and is denoted by $\Pi_{\mathcal{S}^n_+}(X)$. It has long been known that $\Pi_{\mathcal{S}^n_+}(X)$ can be computed analytically (e.g., [57])

$$\Pi_{\mathcal{S}^n_+}(X) = P\mathrm{diag}\big((\lambda_1(X))_+, \ldots, (\lambda_n(X))_+\big)P^T\,, \tag{16}$$

where for any $y \in \mathbb{R}$, $y_+ := \max(0, y)$. For more properties about the metric projector $\Pi_{\mathcal{S}^n_+}(\cdot)$, see [23] and references therein. When it comes to the metric projection over the set $\mathcal{S}^n(r)$, much more analysis is involved due to the non-convex nature of $\mathcal{S}^n(r)$. We first make some notations for the convenience of the subsequent analysis.

Let $Y \in \mathcal{S}^n$ be arbitrarily chosen. Suppose that Y has the spectral decomposition

$$Y = U\Sigma(Y)U^T, \tag{17}$$

where $U \in \mathcal{O}_n$ is a corresponding orthogonal matrix of orthonormal eigenvectors of Y and $\Sigma(Y) := \mathrm{diag}(\sigma(Y))$ where $\sigma(Y) = (\sigma_1(Y), \ldots, \sigma_n(Y))^T$ is the column vector containing all the eigenvalues of Y being arranged in the non-increasing order in terms of their absolute values, i.e.,

$$|\sigma_1(Y)| \geq \cdots \geq |\sigma_n(Y)|\,,$$

and whenever the equality holds, the larger one comes first, i.e.,

$$\text{if } |\sigma_i(Y)| = |\sigma_j(Y)| \text{ and } \sigma_i(Y) > \sigma_j(Y), \text{ then } i < j\,.$$

Define

$$\bar{\alpha} := \{i \mid |\sigma_i(Y)| > |\sigma_r(Y)|\,\},\quad \bar{\beta} := \{i \mid |\sigma_i(Y)| = |\sigma_r(Y)|\,\},\quad \bar{\gamma} := \{i \mid |\sigma_i(Y)| < |\sigma_r(Y)|\,\},$$

$$\text{and}\quad \bar{\beta}^+ := \{i \mid \sigma_i(Y) = |\sigma_r(Y)|\,\},\quad \bar{\beta}^- := \{i \mid \sigma_i(Y) = -|\sigma_r(Y)|\,\}\,.$$

Write $U = [U_{\bar{\alpha}}\ U_{\bar{\beta}}\ U_{\bar{\gamma}}]$. Denote

$$\begin{array}{rll} \Psi_r(Y) := & \min & \frac{1}{2}\|Z - Y\|^2 \\ & \text{s.t.} & Z \in \mathcal{S}^n(r)\,. \end{array} \tag{18}$$

Denote the set of optimal solutions to (18) by $\Pi_{\mathcal{S}^n(r)}(Y)$, which is called the metric projection of Y over $\mathcal{S}^n(r)$. Define $V \in \mathcal{O}_n$ by

$$V = U\mathrm{diag}(v),$$

where for each $i \in \{1, \ldots, n\}$, $v_i = \sigma_i(Y)/|\sigma_i(Y)|$ if $\sigma_i(Y) \neq 0$ and $v_i = 1$ if otherwise. Then, we have

$$Y = U\mathrm{diag}(|\sigma(Y)|)V^T.$$

Define $Z^* \in \mathcal{S}^n$ by

$$Z^* := \sum_{i=1}^{r} |\sigma(Y)|_i U_i V_i^T = \sum_{i=1}^{r} |\sigma_i(Y)| U_i (v_i U_i^T) = \sum_{i=1}^{r} \sigma_i(Y) U_i U_i^T \,. \tag{19}$$

It is well known that (see, e.g., [58, Theorem 5.9]) Z^* is an optimal solution to

$$\begin{array}{ll} \min & \frac{1}{2}\|Z - Y\|^2 \\ \text{s.t.} & Z \in \mathbb{R}^{n\times n}\,,\ \mathrm{rank}(Z) \leq r\,. \end{array}$$

Thus, by using the fact that $Z^* \in \mathcal{S}^n(r)$, we have

$$Z^* \in \Pi_{\mathcal{S}^n(r)}(Y) \quad \text{and} \quad \Psi_r(Y) = \frac{1}{2}\sum_{i=r+1}^{n} \sigma_i^2(Y)\,. \tag{20}$$

By employing Fan's inequality (e.g., see [3, (IV.62)]),

$$\|Z - Y\| \geq \|\lambda(Z) - \lambda(Y)\|, \quad Z \in \mathcal{S}^n\,, \tag{21}$$

where the equality holds if and only if Y and Z admit a simultaneous ordered spectral decomposition, we have for any $\widehat{Z} \in \Pi_{\mathcal{S}^n(r)}(Y)$ that

$$\sum_{i=r+1}^{n} \sigma_i^2(Y) = \|\widehat{Z} - Y\|^2 \geq \|\lambda(\widehat{Z}) - \lambda(Y)\|^2 \geq \sum_{i=r+1}^{n} \sigma_i^2(Y)\,. \tag{22}$$

Then we can easily prove the following lemma, whose proof is dropped for the sake of brevity.

Lemma 2.1 *Let $Y \in \mathcal{S}^n$ have the spectral decomposition as in (17). Then the solution set $\Pi_{\mathcal{S}^n(r)}(Y)$ to problem (18) can be characterized as follows*

$$\begin{array}{rcl} \Pi_{\mathcal{S}^n(r)}(Y) & = & \left\{ [\,U_{\bar{\alpha}}\ U_{\bar{\beta}}Q_{\bar{\beta}}\ U_{\bar{\gamma}}\,]\mathrm{diag}(v)\,[\,U_{\bar{\alpha}}\ U_{\bar{\beta}}Q_{\bar{\beta}}\ U_{\bar{\gamma}}\,]^T \Big| \right. \\ & & \left. v \in \mathcal{V}\,, Q_{\bar{\beta}} = \left[\begin{array}{cc} Q_{\bar{\beta}^+} & 0 \\ 0 & Q_{\bar{\beta}^-} \end{array}\right], Q_{\bar{\beta}^+} \in \mathcal{O}_{|\bar{\beta}^+|}, Q_{\bar{\beta}^-} \in \mathcal{O}_{|\bar{\beta}^-|} \right\}, \end{array} \tag{23}$$

where

$$\begin{array}{rcl} \mathcal{V} := & \big\{ v \in \mathbb{R}^n \,|\, & v_i = \sigma_i(Y) \text{ for } i \in \bar{\alpha} \cup \bar{\beta}_1,\ v_i = 0 \text{ for } i \in (\bar{\beta} \setminus \bar{\beta}_1) \cup \bar{\gamma}\,, \\ & & \text{where } \bar{\beta}_1 \subseteq \bar{\beta} \text{ and } |\bar{\beta}_1| = r - |\bar{\alpha}| \ \big\}\,. \end{array} \tag{24}$$

Since $\Psi_r(Y)$ takes the same value as in (20) for any element in $\Pi_{\mathcal{S}^n(r)}(Y)$, for notational convenience, with no ambiguity, we use $\frac{1}{2}\|\Pi_{\mathcal{S}^n(r)}(Y) - Y\|^2$ to represent $\Psi_r(Y)$.

Define $\Xi_r : \mathcal{S}^n \to \mathbb{R}$ by

$$\Xi_r(Z) = -\frac{1}{2}\|\Pi_{\mathcal{S}^n(r)}(Z) - Z\|^2 + \frac{1}{2}\|Z\|^2\,,\ Z \in \mathcal{S}^n\,. \tag{25}$$

Then we have

$$\Xi_r(Y) = \frac{1}{2}\sum_{i=1}^{r}\sigma_i^2(Y) = \frac{1}{2}\|\Pi_{\mathcal{S}^n(r)}(Y)\|^2\,,$$

where $\|\Pi_{\mathcal{S}^n(r)}(Y)\|$ is interpreted as $\|\overline{Z}\|$ for any $\overline{Z} \in \Pi_{\mathcal{S}^n(r)}(Y)$, e.g., the matrix Z^* defined by (19). By noting that for any $Z \in \mathcal{S}^n$, $\Xi_r(Z)$ can be reformulated as

$$\begin{aligned}\Xi_r(Z) &= \max_{X\in\mathcal{S}^n(r)}\left\{\frac{1}{2}\|Z\|^2 - \frac{1}{2}\|X - Z\|^2\right\}\\ &= \max_{X\in\mathcal{S}^n(r)}\left\{\langle X, Z\rangle - \frac{1}{2}\|X\|^2\right\},\end{aligned} \tag{26}$$

we know that $\Xi_r(\cdot)$ is a convex function as it is the maximum of infinitely many affine functions.

Proposition 2.2 *Let $Y \in \mathcal{S}^n$ have the spectral decomposition as in (17). Then*

$$\partial\Xi_r(Y) = \operatorname{conv}\Pi_{\mathcal{S}^n(r)}(Y)\,, \tag{27}$$

where for any set $\mathcal{W}$, $\operatorname{conv}\mathcal{W}$ denotes the convex hull of $\mathcal{W}$.

Proof. See the Appendix A. □

The first equation in (27) is particularly useful in developing a technique for global optimality checking in Section 4.

Remark 2.3 *Proposition 2.2 implies that when $|\sigma_r(Y)| > |\sigma_{r+1}(Y)|$, $\Xi_r(\cdot)$ is continuously differentiable near Y and $\Xi_r'(Y) = \Pi_{\mathcal{S}^n(r)}(Y) = \{Z^*\}$, where Z^* is defined in (19).*

Remark 2.4 *Since, for a given n by n symmetric positive definite matrix W, the following W-weighted problem*

$$\begin{aligned}\min\ & \frac{1}{2}\|W^{1/2}(Z - Y)W^{1/2}\|^2\\ \text{s.t.}\ & Z \in \mathcal{S}^n(r)\,,\end{aligned} \tag{28}$$

admits the solution set as $W^{-\frac{1}{2}}\Pi_{\mathcal{S}^n(r)}(W^{\frac{1}{2}}YW^{\frac{1}{2}})W^{-\frac{1}{2}}$, there is no difficulty to work out the corresponding results presented in Lemma 2.1 and Proposition 2.2 for this more general case.

3 The majorized penalty approach

The purpose of this section is to introduce our majorized penalty approach for solving the rank constrained weighted least squares problem (5). The essential idea is to first consider a penalized version of problem (5) and then to apply a majorization method to the penalized problem.

For a given continuous function $f : \mathbb{R}^n \to \mathbb{R}$ and a closed set $\Omega \subset \mathbb{R}^n$, the principle of a majorization method for minimizing $f(x)$ over Ω is to start with an initial point $x^0 \in \Omega$ and for each $k \geq 0$, to minimize $\hat{f}^k(x)$ over Ω to get x^{k+1}, where $\hat{f}^k(\cdot)$ is a majorization function of f at x^k, i.e., $\hat{f}^k(\cdot)$ satisfies

$$\hat{f}^k(x^k) = f(x^k) \quad \text{and} \quad \hat{f}^k(x) \geq f(x) \quad \forall\, x \in \Omega\,.$$

The efficiency of the above majorization method hinges on two key issues: i) the majorization functions should be simpler than the original function f so that the resulting minimization problems are easier to solve, and ii) they should not deviate too much from f in order to get fast convergence. These two often conflicting issues need to be addressed on a case by case basis to achieve best possible overall performance.

The idea of using a majorization function in optimization appeared as early as in Ortega and Rheinboldt [43, Section 8.3] for the purpose of doing line searches to decide a step length along a descent direction. This technique was quickly replaced by more effective inexact line search models such as the back tracking. The very first majorization method was introduced by de Leeuw[14, 15] and de Leeuw and Heiser [19] to solve multidimensional scaling problems. Since then much progress has been made on using majorization methods to solve various optimization problems [17, 18, 26, 27, 31, 32], to name only a few.

3.1 The penalty function

In this subsection, we shall introduce a penalty technique to deal with the non-convex rank constraint in (5). Given the fact that for any $X \in \mathcal{S}^n_+$, $\mathrm{rank}(X) \leq r$ if and only if $\lambda_{r+1}(X) + \ldots + \lambda_n(X) = 0$, we can equivalently rewrite (5) as follows

$$\begin{array}{rll} \bar{\theta} := \min & \theta(X) = \dfrac{1}{2}\|H \circ (X - C)\|^2 & \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, & \\ & X \succeq 0\,, & \\ & \lambda_{r+1}(X) + \ldots + \lambda_n(X) = 0\,. & \end{array} \tag{29}$$

Now we consider the following penalized problem by taking a trade-off between the rank constraint and the weighted least squares distance:

$$\begin{array}{ll} \min & \theta(X) + c(\lambda_{r+1}(X) + \ldots + \lambda_n(X)) \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \succeq 0\,, \end{array} \tag{30}$$

where $c > 0$ is a given penalty parameter that decides the allocated weight to the rank constraint in the objective function. By noting that for any $X \in \mathcal{S}^n$,

$$\sum_{i=r+1}^{n} \lambda_i(X) = \sum_{i=1}^{n} \lambda_i(X) - \sum_{i=1}^{r} \lambda_i(X) = \langle I, X\rangle - \sum_{i=1}^{r} \lambda_i(X)\,, \tag{31}$$

we can equivalently write problem (30) as

$$\begin{array}{ll} \min & f_c(X) := \theta(X) - cp(X) \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \succeq 0\,, \end{array} \tag{32}$$

where for any $X \in \mathcal{S}^n$,

$$p(X) := \sum_{i=1}^{r} \lambda_i(X) - \langle I, X\rangle\,. \tag{33}$$

Note that the penalized problem (32) is not equivalent to the original problem (5). Then the question is how much we can say about the solutions to (5) by solving the penalized problem (32). We will address this question in the following two propositions.

Let $X_c^* \in \mathcal{S}^n$ be a global optimal solution to the penalized problem (32).

Proposition 3.1 *If the rank of X_c^* is not larger than r, then X_c^* is a global optimal solution to problem (5).*

Proof. Assume the rank of X_c^* is not larger than r. Then X_c^* is a feasible solution to (5) and $p(X_c^*) = 0$. Let $X_r \in \mathcal{S}^n$ be any feasible point to (5). Thus, by noting that $p(X_r) = 0$, we have

$$\theta(X_c^*) = \theta(X_c^*) - cp(X_c^*) \le \theta(X_r) - cp(X_r) = \theta(X_r)\,.$$

This shows that the conclusion of this proposition holds. □

Proposition 3.1 says in the ideal situation when the rank of X_c^* is not larger than r, X_c^* actually solves the original problem (5). Though this ideal situation is always observed in our numerical experiments for a properly chosen penalty parameter $c > 0$, there is no theoretical guarantee that this is the case. However, when the penalty parameter c is large enough, $|p(X_c^*)|$ can be proven to be very small. To see this, let X^* be an optimal solution to the following least squares convex optimization problem

$$\begin{array}{ll} \min & \theta(X) \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \succeq 0\,. \end{array} \tag{34}$$

Proposition 3.2 *Let $\varepsilon > 0$ be a given positive number and $X_r \in \mathcal{S}^n$ a feasible solution to problem (5). Assume that $c > 0$ is chosen such that $\big(\theta(X_r) - \theta(X^*)\big)/c \le \varepsilon$. Then we have*

$$|\,p(X_c^*)\,| \le \varepsilon \quad \text{and} \quad \theta(X_c^*) \le \bar{\theta} - c|p(X_c^*)| \le \bar{\theta}. \tag{35}$$

Proof. By noting that X_r is feasible to the penalized problem (32) and $p(X_r) = 0$, we have

$$\theta(X_r) = \theta(X_r) - cp(X_r) = f_c(X_r) \ge f_c(X_c^*) = \theta(X_c^*) - cp(X_c^*) \ge \theta(X^*) - cp(X_c^*)\,,$$

which implies

$$|\,p(X_c^*)\,| = -p(X_c^*) \le \big(\theta(X_r) - \theta(X^*)\big)/c \le \varepsilon\,.$$

Let $\overline{X}$ be a global optimal solution to problem (5). Then from

$$\theta(\overline{X}) - cp(\overline{X}) = f_c(\overline{X}) \geq f_c(X_c^*) = \theta(X_c^*) - cp(X_c^*)$$

and the fact that $p(\overline{X}) = 0$, we obtain that $\theta(X_c^*) \leq \theta(\overline{X}) - c|p(X_c^*)| = \bar{\theta} - c|p(X_c^*)|$. The proof is completed. □

Proposition 3.2 says that an ε-optimal solution to the original problem (5) in the sense of (35) is guaranteed by solving the penalized problem (32) as long as the penalty parameter c is above some ε-dependent number. This provides the rationale to replace the rank constraint in problem (5) by the penalty function $-cp(\cdot)$ in problem (32).

Remark 3.3 *In Proposition 3.2, we need to choose a feasible point X_r to problem (5). That is equivalently to say that we need to find a global solution to*

$$\begin{array}{ll} \min & \lambda_{r+1}(X) + \ldots + \lambda_n(X) = -p(X) \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \succeq 0\,. \end{array} \tag{36}$$

To solve problem (36), one may use the majorization method to be introduced in next subsection. This corresponds to the case that $H = 0$. However, this is not needed in many situations when a feasible point to problem (5) is readily available. For example, the mPCA of X^* is such a choice if there are no bound constraints on the off-diagonal entries.

3.2 The majorized penalty approach

In this subsection, we focus on the penalized problem (32). Note that in problem (32) the objective function $f_c(\cdot) = \theta(\cdot) - cp(\cdot)$ is the difference of a convex quadratic function $\theta(\cdot)$ and a nonsmooth convex function $cp(\cdot)$. In order to design a majorization method to solve problem (32), we need first to find majorization functions of $\theta(\cdot)$ and $-p(\cdot)$. Let Ω denote the feasible set to problem (32), i.e.,

$$\Omega := \{X \in \mathcal{S}^n \,|\, \mathcal{A}X \in b + \mathcal{Q}\,,\ X \succeq 0\}\,.$$

For any X and Y in Ω, let $\hat{\theta}(X, Y)$ be defined by

$$\begin{array}{rcl} \hat{\theta}(X,Y) & := & \theta(Y) + \langle \nabla\theta(Y)\,,\, X - Y\rangle + \frac{1}{2}\|\widehat{H}_Y \circ (X - Y)\|^2 \\ & = & \theta(Y) + \langle H \circ H \circ (Y - C)\,,\, X - Y\rangle + \frac{1}{2}\|\widehat{H}_Y \circ (X - Y)\|^2\,, \end{array} \tag{37}$$

where $\mathcal{S}^n \ni \widehat{H}_Y \geq 0$ is a componentwise nonnegative symmetric matrix satisfying

$$\|H \circ (Z - Y)\|^2 \leq \|\widehat{H}_Y \circ (Z - Y)\|^2 \quad \forall\, Z \in \Omega\,. \tag{38}$$

Define $\hat{p} : \mathcal{S}^n \times \mathcal{S}^n \to \mathbb{R}$ by

$$\hat{p}(X, Y) := p(Y) + \langle W_Y, X - Y\rangle\,, \tag{39}$$

where W_Y is any element in $\partial_B p(Y)$ and $(X, Y) \in \mathcal{S}^n \times \mathcal{S}^n$. Thus, by the convexity of $p(\cdot)$, we know that for any given $Y \in \Omega$, $\hat{\theta}(\cdot, Y)$ and $-\hat{p}(\cdot, Y)$ are the majorization functions of $\theta(\cdot)$ and $-p(\cdot)$ at Y, respectively. Consequently, for any $Y \in \Omega$, the function $f_c(\cdot)$ is majorized at Y by

$$\hat{f}_c(\cdot, Y) := \hat{\theta}(\cdot, Y) - c\hat{p}(\cdot, Y)\,. \tag{40}$$

For any $X \in \Omega$, let $\mathcal{N}_\Omega(X)$ denote the normal cone of Ω at the point X:

$$\mathcal{N}_\Omega(X) := \{Z \in \mathcal{S}^n \,|\, \langle Z, Y - X\rangle \le 0 \;\forall\, Y \in \Omega\}.$$

A point $X \in \Omega$ is said to be a *stationary* point of problem (32) if

$$(\nabla\theta(X) + N_\Omega(X)) \cap (c\partial p(X)) \neq \emptyset$$

and a *B-stationary* point of problem (32) if

$$(\nabla\theta(X) + N_\Omega(X)) \cap (c\partial_B p(X)) \neq \emptyset\,.$$

A B-stationary point of problem (32) is always a stationary point of the problem itself and the converse is not necessarily true.

Now we can summarize our majorized penalty approach for solving problem (5) as follows.

A Majorized Penalty Approach (MPA)

Step 0. Select a penalty parameter $c > 0$. Replace the rank constraint in problem (5) by the penalty function $-cp(\cdot)$ and start to solve problem (32).

Step 1. Choose $X^0 \in \Omega$. Set $k := 0$.

Step 2. Choose $\mathcal{S}^n \ni \widehat{H}^k := \widehat{H}_{X_k} \ge 0$ satisfying (38) and $W^k := W_{X_k} \in \partial_B p(X^k)$. Compute the majorization functions $\hat{\theta}^k(\cdot)$ and $-\hat{p}^k(\cdot)$ of $\theta(\cdot)$ and $-p(\cdot)$ at X^k , respectively, as in (37) and (39), i.e.,

$$\hat{\theta}^k(\cdot) := \hat{\theta}(\cdot, X^k) \quad \text{and} \quad -\hat{p}^k(\cdot) := -\hat{p}(\cdot, X^k)\,.$$

Then $f_c(\cdot)$ is majorized at X^k by

$$\hat{f}_c^k(\cdot) := \hat{f}_c(\cdot, X^k) = \hat{\theta}^k(\cdot) - c\hat{p}^k(\cdot)\,.$$

Solve

$$\begin{array}{ll} \min & \hat{f}_c^k(X) \\ \text{s.t.} & X \in \Omega \end{array} \tag{41}$$

to get X^{k+1}.

Step 3. If $X^{k+1} = X^k$, stop; otherwise, set $k := k + 1$ and goto Step 2.

Theorem 3.4 *Let $\{X^k\}$ be the sequence generated by the MPA. Then $\{f_c(X^k)\}$ is a monotonically decreasing sequence. If $X^{k+1} = X^k$ for some integer $k \geq 0$, then X^{k+1} is a B-stationary point of problem (32). Otherwise, the infinite sequence $\{f_c(X^k)\}$ satisfies*

$$\frac{1}{2}\|\widehat{H}^k \circ (X^{k+1} - X^k)\|^2 \leq f_c(X^k) - f_c(X^{k+1}), \quad k = 0, 1, \ldots \tag{42}$$

Moreover, any accumulation point of the bounded sequence $\{X^k\}$ is a B-stationary point of problem (32) provided that there exist two positive number κ_1, κ_2 such that $\kappa_2 \geq \kappa_1 > 0$ and for all $k \geq 0$

$$\kappa_1 \leq \min_{i,j=1,\ldots,n} \widehat{H}^k_{ij} \leq \max_{i,j=1,\ldots,n} \widehat{H}^k_{ij} \leq \kappa_2 . \tag{43}$$

Proof. The monotone decreasing property of $\{f_c(X^k)\}$ follows easily from the so-called sandwich inequality ([16]) for the general majorization method

$$f_c(X^{k+1}) \leq \hat{f}^k_c(X^{k+1}) \leq \hat{f}^k_c(X^k) = f_c(X^k), \quad k = 0, 1, \ldots \tag{44}$$

We first consider the case that $X^{k+1} = X^k$ for some integer $k \geq 0$. Since X^{k+1} is an optimal solution to problem (41), one has

$$0 \in \nabla\hat{\theta}^k(X^{k+1}) - c\nabla\hat{p}^k(X^{k+1}) + N_\Omega(X^{k+1}) .$$

From the facts that

$$\nabla\hat{\theta}^k(X^{k+1}) = H \circ H \circ (X^k - C) + \widehat{H}^k \circ \widehat{H}^k \circ (X^{k+1} - X^k) = \nabla\theta(X^k) = \nabla\theta(X^{k+1})$$

and

$$\nabla\hat{p}^k(X^{k+1}) = W^k \in \partial_B p(X^k) = \partial_B p(X^{k+1}) ,$$

we obtain

$$0 \in \nabla\theta(X^{k+1}) - c\partial_B p(X^{k+1}) + N_\Omega(X^{k+1}) ,$$

which implies that X^{k+1} is a B-stationary point of problem (32).

Next we assume that $X^{k+1} \neq X^k$ for all $k \geq 0$. Then an infinite bounded sequence $\{X^k\}$ is generated as Ω is a bounded set. For each $k \geq 0$, since X^{k+1} solves the convex optimization problem (41), there exists $D^{k+1} \in \mathcal{N}_\Omega(X^{k+1})$ such that

$$\nabla\hat{f}^k_c(X^{k+1}) + D^{k+1} = \nabla\theta(X^k) + M^k \circ (X^{k+1} - X^k) - cW^k + D^{k+1} = 0 , \tag{45}$$

where $M^k := \widehat{H}^k \circ \widehat{H}^k$. Thus, from (44) we have for each $k \geq 0$ that

$$\begin{aligned}
& f_c(X^{k+1}) - f_c(X^k) \\
\leq\ & \hat{f}^k_c(X^{k+1}) - f_c(X^k) \\
=\ & \langle X^{k+1} - X^k, \nabla\theta(X^k)\rangle + \tfrac{1}{2}\langle X^{k+1} - X^k,\, M^k \circ (X^{k+1} - X^k)\rangle - c\langle X^{k+1} - X^k, W^k\rangle \\
=\ & -\langle X^{k+1} - X^k,\, M^k \circ (X^{k+1} - X^k) + D^{k+1}\rangle + \tfrac{1}{2}\langle X^{k+1} - X^k,\, M^k \circ (X^{k+1} - X^k)\rangle \\
=\ & -\tfrac{1}{2}\langle X^{k+1} - X^k,\, M^k \circ (X^{k+1} - X^k)\rangle + \langle X^k - X^{k+1}, D^{k+1}\rangle ,
\end{aligned}$$

which, together with the fact that $\langle X^k - X^{k+1}, D^{k+1}\rangle \le 0$ since $X^k \in \Omega$ and $D^{k+1} \in \mathcal{N}_\Omega(X^{k+1})$, shows that

$$f_c(X^{k+1}) - f_c(X^k) \le -\frac{1}{2}\langle X^{k+1} - X^k,\, M^k \circ (X^{k+1} - X^k)\rangle\,.$$

This shows that (42) holds.

To prove the remaining part of this theorem, we assume that $\overline{X}$ is an accumulation point of $\{X^k\}$ and that (43) holds. Let $\{X^{k_j}\}$ be a subsequence of $\{X^k\}$ such that $\lim_{j\to+\infty} X^{k_j} = \overline{X}$. Then from (42) we have

$$\lim_{i\to+\infty} \frac{1}{2}\sum_{k=0}^{i} \|\widehat{H}^k \circ (X^{k+1} - X^k)\|^2 \le \liminf_{i\to+\infty} \big(f_c(X^0) - f_c(X^{i+1})\big) \le f_c(X^0) < +\infty\,,$$

which implies $\lim_{k\to+\infty} \|\widehat{H}^k \circ (X^{k+1} - X^k)\| = 0$. By using the condition (43), we obtain that

$$\lim_{j\to+\infty} X^{k_j+1} = \lim_{j\to+\infty} X^{k_j} = \overline{X} \quad \text{and} \quad \lim_{j\to+\infty} M^{k_j} \circ (X^{k_j+1} - X^{k_j}) = 0\,.$$

Since $\{X^{k_j}\}$ is bounded, from convex analysis [54, Chap 24, Thm 24.7] we know that $\{W^{k_j}\}$ is also bounded. By taking a subsequence if necessary, we assume that there exists $\overline{W} \in \partial_B p(\overline{X})$ such that $\lim_{j\to+\infty} W^{k_j} = \overline{W}$. Therefore, we obtain from (45) that

$$\overline{D} := \lim_{j\to+\infty} D^{k_j+1} = \lim_{j\to+\infty} \big(-\nabla\theta(X^{k_j}) - M^{k_j} \circ (X^{k_j+1} - X^{k_j}) + cW^{k_j}\big) = -\nabla\theta(\overline{X}) + c\overline{W}\,.$$

Now in order to show that $\overline{X}$ is a B-stationary point of problem (32), we only need to show that

$$\overline{D} \in \mathcal{N}_\Omega(\overline{X})\,.$$

Suppose that $\overline{D} \notin \mathcal{N}_\Omega(\overline{X})$, i.e., there exists $\widetilde{X} \in \Omega$ such that $\langle \overline{D}, \widetilde{X} - \overline{X}\rangle > 0$. Since for each $k_j \ge 0$, $D^{k_j+1} \in \mathcal{N}_\Omega(X^{k_j+1})$, we have

$$\langle D^{k_j+1}, \widetilde{X} - X^{k_j+1}\rangle \le 0,$$

which, from the convergence of the two subsequences $\{D^{k_j+1}\}$ and $\{X^{k_j+1}\}$, gives rise to

$$\langle \overline{D}, \widetilde{X} - \overline{X}\rangle \le 0\,.$$

This is a contradiction. So the proof is completed. □

Note that in the MPA, we need to solve a sequence of problems in the form of (41), i.e.,

$$\begin{array}{ll} \min & \hat{f}_c^k(X) = \dfrac{1}{2}\|\widehat{H}^k \circ (X - X^k)\|^2 + \langle X, H \circ H \circ (X^k - C) - cW^k\rangle + g_c^k \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \succeq 0\,, \end{array} \tag{46}$$

where

$$g_c^k := \theta(X^k) - \big\langle H \circ H \circ (X^k - C),\, X^k\big\rangle - cp(X^k) + c\big\langle W^k,\, X^k\big\rangle\,.$$

Problem (46) is a convex optimization problem, which can be solved by known algorithms, e.g., the augmented Lagrangian method discussed in [47]. However, for the sake of easy computations, in our implementation, we always choose a positive vector $d \in \mathbb{R}^n$ such that $H_{ij} \leq \widehat{H}^k_{ij} = \sqrt{d_i d_j}$ for all $i, j \in \{1, \ldots, n\}$. Let $D = \mathrm{diag}(d)$. Then the objective function $\hat{f}^k_c(\cdot)$ in (46) can be equivalently written as

$$\begin{array}{rcl} \hat{f}^k_c(X) & = & \frac{1}{2}\|D^{1/2}(X - X^k)D^{1/2}\|^2 + \langle X, H \circ H \circ (X^k - C) - cW^k \rangle + g^k_c \\ & = & \frac{1}{2}\|D^{1/2}\big(X - (X^k + C^k)\big)D^{1/2}\|^2 + f_c(X^k) - \frac{1}{2}\|D^{1/2}C^k D^{1/2}\|^2\,, \quad X \in \mathcal{S}^n\,, \end{array}$$

where $C^k := D^{-1}\big(cW^k - H \circ H \circ (X^k - C)\big)D^{-1}$. By dropping the constant terms in $\hat{f}^k_c(X)$, we can equivalently write problem (46) as the following well-studied diagonally weighted least squares problem

$$\begin{array}{ll} \min & \dfrac{1}{2}\|D^{1/2}\big(X - (X^k + C^k)\big)D^{1/2}\|^2 \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \succeq 0\,, \end{array} \tag{47}$$

which can be solved efficiently by the recently developed smoothing Newton-BiCGStab method [23]. For the choice of $d \in \mathbb{R}^n$, one can simply take

$$d_1 = \ldots = d_n = \max\big\{\delta,\, \max\{\, H_{ij} \,|\, i, j = 1, \ldots, n\}\big\}\,, \tag{48}$$

where $\delta > 0$ is a small positive number. However, a better way is to choose $d \in \mathbb{R}^n$ as follows

$$d_i = \max\big\{\delta,\, \max\{\, H_{ij} \mid j = 1, \ldots, n\,\}\big\}\,, \quad i = 1, \ldots, n\,. \tag{49}$$

Remark 3.5 *The choice of d in (48) is simpler and will lead to an unweighted least squares problem. The disadvantage of this choice is that the resulting MPA generally takes more iterations to converge than the one obtained from the choice of (49) due to the fact that the error $\|H - dd^T\|$ is larger for the choice of (48). If H takes the form of hh^T for some column vector $\mathbb{R}^n \ni h > 0$, we can just take $\widehat{H}^k \equiv H$ for all $k \geq 1$. In this case, the majorization function of $\theta(\cdot)$ is itself.*

4 The Lagrangian dual problem

In this section, we shall study the Lagrangian dual of (5) in order to check the optimality of the solution obtained by the majorization penalty method introduced in the previous section. Note that (5) can be equivalently reformulated as [2]

$$\begin{array}{ll} \min & \dfrac{1}{2}\|H \circ (X - C)\|^2 + \dfrac{1}{2}\|H \circ (Z - C)\|^2 \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & Z - X = 0\,, \\ & X \in \mathcal{S}^n_+\,, \\ & \mathrm{rank}(Z) \leq r\,. \end{array} \tag{50}$$

[2]The optimal value of (50) is twice of the optimal value of (5).

The Lagrange function of (50) is

$$L(X,Z,y,Y) = \frac{1}{2}\|H\circ(X-C)\|^2 + \frac{1}{2}\|H\circ(Z-C)\|^2 + \langle b-\mathcal{A}X,\, y\rangle + \langle Z-X,\, Y\rangle\,,$$

where $(X,Z,y,Y)\in\mathcal{S}^n_+\times\mathcal{S}^n(r)\times\mathbb{R}^m\times\mathcal{S}^n$. The Lagrange dual problem of (50) then takes the form of

$$\max_{y\in\mathcal{Q}^*,\,Y\in\mathcal{S}^n} V(y,Y)\,, \tag{51}$$

where $\mathcal{Q}^* = \mathbb{R}^p\times\mathbb{R}^q_+$ is the dual cone of $\mathcal{Q}$ and $V(y,Y)$ is defined by

$$\begin{aligned} V(y,Y) &:= \inf_{X\in\mathcal{S}^n_+,\,Z\in\mathcal{S}^n(r)} \big\{L(X,Z,y,Y)\big\} \\ &= \inf_{X\in\mathcal{S}^n_+,\,Z\in\mathcal{S}^n(r)} \left\{\frac{1}{2}\|H\circ(X-C)\|^2 + \frac{1}{2}\|H\circ(Z-C)\|^2 + \langle b-\mathcal{A}X,\, y\rangle + \langle Z-X,\, Y\rangle\right\}. \end{aligned} \tag{52}$$

Suppose that $(\bar{y},\overline{Y})\in\mathcal{Q}^*\times\mathcal{S}^n$ is an optimal solution to (51). Then for any feasible $(\widehat{X},\widehat{Z})$ to (50), one has

$$\begin{aligned} \|H\circ(\widehat{X}-C)\|^2 &\geq \frac{1}{2}\|H\circ(\widehat{X}-C)\|^2 + \frac{1}{2}\|H\circ(\widehat{Z}-C)\|^2 + \langle b-\mathcal{A}\widehat{X},\,\bar{y}\rangle + \langle\widehat{Z}-\widehat{X},\,\overline{Y}\rangle \\ &\geq V(\bar{y},\overline{Y})\,, \end{aligned} \tag{53}$$

which implies that the dual solution $(\bar{y},\overline{Y})$ provides a valid lower bound for checking the optimality of the primal solution. When H is the matrix with all the entries equal to 1, we can further simplify (52) and write $V(y,Y)$ explicitly as

$$\begin{aligned} V(y,Y) &= \inf_{X\in\mathcal{S}^n_+,\,Z\in\mathcal{S}^n(r)} \left\{\frac{1}{2}\|X-C\|^2 + \frac{1}{2}\|Z-C\|^2 + \langle b-\mathcal{A}X,\, y\rangle + \langle Z-X,\, Y\rangle\right\} \\ &= \inf_{X\in\mathcal{S}^n_+,\,Z\in\mathcal{S}^n(r)} \left\{\frac{1}{2}\|X-(C+\mathcal{A}^*y+Y)\|^2 + \frac{1}{2}\|Z-(C-Y)\|^2 + \langle b,y\rangle \right. \\ &\qquad\qquad \left. -\frac{1}{2}\|C+\mathcal{A}^*y+Y\|^2 - \frac{1}{2}\|C-Y\|^2 + \|C\|^2\right\} \\ &= -\frac{1}{2}\|\Pi_{\mathcal{S}^n_+}(C+\mathcal{A}^*y+Y)\|^2 - \frac{1}{2}\|\Pi_{\mathcal{S}^n(r)}(C-Y)\|^2 + \langle b,y\rangle + \|C\|^2 \end{aligned}$$

where $\mathcal{A}^*$ is the adjoint of $\mathcal{A}$. For any $(y,Y)\in\mathbb{R}^m\times\mathcal{S}^n$, let $\Phi(y,Y) := -V(y,Y) + \|C\|^2$. Now we can rewrite the dual problem as follows

$$\begin{array}{ll} \min & \Phi(y,Y) = \dfrac{1}{2}\|\Pi_{\mathcal{S}^n_+}(C+\mathcal{A}^*y+Y)\|^2 + \dfrac{1}{2}\|\Pi_{\mathcal{S}^n(r)}(C-Y)\|^2 - \langle b,y\rangle \\ \text{s.t.} & y\in\mathcal{Q}^* = \mathbb{R}^p\times\mathbb{R}^q_+\,, \\ & Y\in\mathcal{S}^n\,. \end{array} \tag{54}$$

Remark 4.1 *When H takes the form of $H = hh^T$ for some column vector $h>0$ in $\mathbb{R}^n$, we can also derive a similar explicit expression for $V(y,Y)$ as in (54). For the general weight matrix H, we cannot reformulate (52) explicitly. However, we can still apply the majorized penalty method introduced early in this paper to compute $V(y,Y)$.*

Next we discuss the existence of the optimal solution to (54). For this purpose, we need the following Slater condition:

$$\begin{cases} \{\mathcal{A}_i\}_{i=1}^p \text{ are linearly independent,} \\ \text{there exists } X^0 \succ 0 \text{ such that } \mathcal{A}_j X^0 = b_j \text{ for } j = 1, \ldots, p\,, \\ \text{and } \mathcal{A}_j X^0 > b_j \text{ for } j = p+1, \ldots, m\,. \end{cases} \tag{55}$$

Lemma 4.2 *Assume that the Slater condition (55) holds. Then* $\langle b, \bar{y}\rangle < 0$ *for any* $0 \neq \bar{y} \in \mathcal{Q}^*$ *satisfying* $\mathcal{A}^*\bar{y} \preceq 0$.

Proof. Let $0 \neq \bar{y} \in \mathcal{Q}^*$ be such that $\mathcal{A}^*\bar{y} \preceq 0$. Since the Slater condition (55) is assumed to hold, there exists an $X^0 \succ 0$ such that $\mathcal{A}_j X^0 = b_j$ for $1 \leq j \leq p$ and $\mathcal{A}_j X^0 > b_j$ for $p+1 \leq j \leq m$.

Next, we prove $\langle b, \bar{y}\rangle < 0$ by considering the following two cases.

1). There exists an index $j \in \{p+1, \ldots, m\}$ such that $\bar{y}_j > 0$. For this index j, we have

$$\big(\mathcal{A}_j X^0 - b_j\big)\bar{y}_j > 0.$$

It then follows that

$$\langle b, \bar{y}\rangle = \sum_{i=1}^m b_i\bar{y}_i \leq \sum_{\substack{i=1\\ i\neq j}}^m \mathcal{A}_i X^0 \bar{y}_i + b_j\bar{y}_j < \langle \mathcal{A}X^0\,,\, \bar{y}\rangle = \langle X^0,\, \mathcal{A}^*\bar{y}\rangle \leq 0\,.$$

2). There does not exist $j \in \{p+1, \ldots, m\}$ such that $\bar{y}_j > 0$. Since $\bar{y} \in \mathcal{Q}^*$, this implies $\bar{y}_j = 0$ for $j = p+1, \ldots, m$. Then, from the assumptions that $\{\mathcal{A}_i\}_{i=1}^p$ are linearly independent and $\bar{y} \neq 0$, we obtain $\mathcal{A}^*\bar{y} = \sum_{i=1}^p \mathcal{A}_j^*\bar{y} \neq 0$. Therefore, by using the fact that $X^0 \succ 0$, we obtain that

$$\langle b, \bar{y}\rangle = \langle \mathcal{A}X^0\,,\, \bar{y}\rangle = \langle X^0,\, \mathcal{A}^*\bar{y}\rangle < 0\,.$$

In both cases, we have shown that $\langle b, \bar{y}\rangle < 0$. □

Proposition 4.3 *Assume that the Slater condition (55) holds. Then, for any constant* $\nu \in \mathbb{R}$, *the level set* $L_\nu := \big\{(y, Y) \in \mathcal{Q}^* \times \mathcal{S}^n \mid \Phi(y, Y) \leq \nu\big\}$ *is bounded.*

Proof. We prove the conclusion of this proposition by contradiction. Suppose that on the contrary that there exists a constant $\nu \in \mathbb{R}$ such that L_ν is unbounded. Then there exists a sequence $\{(y^k, Y^k)\} \in \mathcal{Q}^* \times \mathcal{S}^n$ such that $\Phi(y^k, Y^k) \leq \nu$ for all $k \geq 1$ and $\lim\limits_{k\to+\infty}(\|y^k\| + \|Y^k\|) = +\infty$.

We consider the following two cases:

1). There exists a positive number $\delta > 0$ such that $\limsup\limits_{k\to+\infty} \dfrac{\|Y^k\|}{\|y^k\|} \geq \delta > 0$. For each $k \geq 1$, define $C^k := C/\|Y^k\| - Y^k/\|Y^k\|$. By taking a subsequence if necessary, we may assume that $\|Y^k\| \to \infty$ as $k \to \infty$ and there exists a matrix $0 \neq \overline{Y} \in \mathcal{S}^n$ such that

$$Y^k/\|Y^k\| \to \overline{Y} \quad \text{and} \quad C^k \to -\overline{Y} \text{ as } k \to \infty\,.$$

Then, for all k sufficiently large, there exists $\bar{\delta} > 0$ such that

$$\|\Pi_{\mathcal{S}^n(r)}(C^k)\|^2 \geq \bar{\delta}\,.$$

Thus,

$$\begin{aligned}
\liminf_{k\to+\infty} \Phi(y^k, Y^k) &\geq \liminf_{k\to+\infty} \Big(\frac{1}{2}\|\Pi_{\mathcal{S}^n(r)}(C - Y^k)\|^2 - \langle b, y^k\rangle\Big) \\
&= \liminf_{k\to+\infty} \Big(\frac{1}{2}\|Y^k\|^2\|\Pi_{\mathcal{S}^n(r)}(C^k)\|^2 - \langle b, y^k\rangle\|Y^k\|/\|Y^k\|\Big) \\
&\geq \liminf_{k\to+\infty} \|Y^k\|\Big(\frac{1}{2}\|Y^k\|\|\Pi_{\mathcal{S}^n(r)}(C^k)\|^2 - \|b\|\|y^k\|/\|Y^k\|\Big) \\
&\geq \frac{\bar{\delta}}{4}\liminf_{k\to+\infty} \|Y^k\|^2 = +\infty\,.
\end{aligned}$$

2). $\limsup\limits_{k\to+\infty} \dfrac{\|Y^k\|}{\|y^k\|} = 0$. Without loss of generality, we may assume that $y^k \neq 0$ for each $k \geq 1$ and $||y^k|| \to \infty$ as $k \to \infty$. For $k \geq 1$, let $B^k := (C + \mathcal{A}^* y^k + Y^k)/\|y^k\|$. We assume, by taking a subsequence if necessary, that there exists $\bar{y} \neq 0$ such that

$$\lim_{k\to+\infty} \frac{y^k}{\|y^k\|} = \bar{y}\,.$$

Next we consider the following two subcases:

2.1). $\mathcal{A}^*\bar{y} \not\preceq 0$, i.e., $\mathcal{A}^*\bar{y}$ has at least one positive eigenvalue. It then follows that there exists a positive number $\delta > 0$ such that

$$\liminf_{k\to+\infty} \|\Pi_{\mathcal{S}^n_+}(B^k)\|^2 = \liminf_{k\to+\infty} \|\Pi_{\mathcal{S}^n_+}(\mathcal{A}^*\bar{y})\|^2 \geq \delta > 0\,.$$

Hence, we have

$$\begin{aligned}
\liminf_{k\to+\infty} \Phi(y^k, Y^k) &\geq \liminf_{k\to+\infty} \Big(\frac{1}{2}\|\Pi_{\mathcal{S}^n_+}(C + \mathcal{A}^* y^k + Y^k)\|^2 - \langle b, y^k\rangle\Big) \\
&\geq \liminf_{k\to+\infty} \|y^k\|\Big(\frac{1}{2}\|y^k\|\|\Pi_{\mathcal{S}^n_+}(B^k)\|^2 - \|b\|\Big) = +\infty\,.
\end{aligned}$$

2.2). $\mathcal{A}^*\bar{y} \preceq 0$. Then $\langle b, \bar{y}\rangle < 0$ follows immediately from Lemma 4.2. Therefore,

$$\liminf_{k\to+\infty} \Phi(y^k, Y^k) \geq \liminf_{k\to+\infty} \|y^k\|\big(-\langle b, y^k/\|y^k\|\rangle\big) \geq -\langle b, \bar{y}\rangle \liminf_{k\to+\infty} \|y^k\|/2 = +\infty\,.$$

In summary, we have shown that $\Phi(y^k, Y^k) \to +\infty$ as $k \to \infty$, which is a contradiction to our assumption that $\Phi(y^k, Y^k) \leq \nu$ for all $k \geq 1$. This contradiction shows that the conclusion of this proposition holds. □

Proposition 4.3 says that if the Slater condition (55) holds, the dual problem (54) always has optimal solutions. Let $(\bar{y}, \overline{Y}) \in \mathcal{Q}^* \times \mathcal{S}^n$ be an optimal solution to (54). Then we have

$$0 \in \partial\Phi(\bar{y}, \overline{Y}) + \mathcal{N}_{\mathcal{Q}^*}(\bar{y}) \times \{0\}. \tag{56}$$

Theorem 4.4 *The optimal solution $(\bar{y}, \overline{Y}) \in \mathcal{Q}^* \times \mathcal{S}^n$ to the the dual problem (54) satisfies*

$$b - \mathcal{A}\Pi_{\mathcal{S}^n_+}\big(C + \mathcal{A}^*\bar{y} + \overline{Y}\big) \in \mathcal{N}_{\mathcal{Q}^*}(\bar{y}) \tag{57}$$

and

$$\Pi_{\mathcal{S}^n_+}\big(C + \mathcal{A}^*\bar{y} + \overline{Y}\big) \in \mathrm{conv}\left\{\Pi^B_{\mathcal{S}^n(r)}(C - \overline{Y})\right\}, \tag{58}$$

where $\Pi^B_{\mathcal{S}^n(r)}(\cdot)$ is defined as in Lemma 2.1. Furthermore, if there exists a matrix $\overline{X} \in \Pi^B_{\mathcal{S}^n(r)}(C - \overline{Y})$ such that $\overline{X} = \Pi_{\mathcal{S}^n_+}\big(C + \mathcal{A}^*\bar{y} + \overline{Y}\big)$, then $\overline{X}$ and $(\bar{y}, \overline{Y})$ globally solve the primal problem (50) with $H = E$ and the corresponding dual problem (54), respectively and there is no duality gap between the primal and dual problems.

Proof. Recall that for $(y, Y) \in \mathcal{Q}^* \times \mathcal{S}^n$,

$$\begin{array}{rcl} \Phi(y, Y) & = & \frac{1}{2}\|\Pi_{\mathcal{S}^n_+}(C + \mathcal{A}^*y + Y)\|^2 + \frac{1}{2}\|\Pi_{\mathcal{S}^n(r)}(C - Y)\|^2 - \langle b, y\rangle \\ & = & \frac{1}{2}\|\Pi_{\mathcal{S}^n_+}(C + \mathcal{A}^*y + Y)\|^2 + \Xi_r(C - Y) - \langle b, y\rangle\,. \end{array}$$

It is well known that $\|\Pi_{\mathcal{S}^n_+}(\cdot)\|^2$ is continuously differentiable [62] and for any $Z \in \mathcal{S}^n$,

$$\nabla(\|\Pi_{\mathcal{S}^n_+}(Z)\|^2) = 2\,\Pi_{\mathcal{S}^n_+}(Z)\,.$$

Then, by using Proposition 2.2, we know that the sub-differential of $\Phi(\cdot,\cdot)$ at the optimal solution point $(\bar{y}, \overline{Y})$ can be written as

$$\partial\Phi(\bar{y}, \overline{Y}) = \left\{ \begin{array}{c} \big(\,\omega^T,\ \Gamma\,\big)\,\big|\,\omega = \mathcal{A}\Pi_{\mathcal{S}^n_+}\big(C + \mathcal{A}^*\bar{y} + \overline{Y}\big) - b\,, \\ \Gamma \in \Pi_{\mathcal{S}^n_+}\big(C + \mathcal{A}^*\bar{y} + \overline{Y}\big) - \mathrm{conv}\left\{\,\Pi^B_{\mathcal{S}^n(r)}(C - \overline{Y})\,\right\} \end{array} \right\}. \tag{59}$$

The inclusions (57) and (58) now follow directly from (56). Note that (57) implies that

$$\mathcal{A}\Pi_{\mathcal{S}^n_+}\big(C + \mathcal{A}^*\bar{y} + \overline{Y}\big) \in b + \mathcal{Q} \text{ and } \big\langle b - \mathcal{A}\Pi_{\mathcal{S}^n_+}\big(C + \mathcal{A}^*\bar{y} + \overline{Y}\big),\, \bar{y}\big\rangle = 0\,.$$

Moreover, if there exists a matrix $\overline{X} \in \Pi^B_{\mathcal{S}^n(r)}(C - \overline{Y})$ such that $\overline{X} = \Pi_{\mathcal{S}^n_+}\big(C + \mathcal{A}^*\bar{y} + \overline{Y}\big)$, then we know that $\overline{X}$ is feasible to the primal problem (50) and

$$\begin{array}{rcl} & & V(\bar{y}, \overline{Y}) \\ & = & \frac{1}{2}\|\overline{X} - C\|^2 + \frac{1}{2}\|\overline{Z} - C\|^2 + \langle b - \mathcal{A}\overline{X},\, \bar{y}\rangle + \langle \overline{Z} - \overline{X},\, \overline{Y}\rangle \\ & = & \frac{1}{2}\|\overline{X} - C\|^2 + \frac{1}{2}\|\overline{Z} - C\|^2\,, \end{array}$$

which, together with the fact that $(\bar{y}, \overline{Y})$ is feasible to the dual problem (54), completes the proof of the remaining part of the theorem. □

Note that Theorem 4.4 shows that, $\Pi_{\mathcal{S}^n_+}\big(C + \mathcal{A}^*\bar{y} + \overline{Y}\big)$ is feasible to the primal problem without the rank constraint meanwhile it is the convex combination of several matrices in $\Pi^B_{\mathcal{S}^n(r)}(C - \overline{Y})$, which consists of matrices of rank less than or equal to r. If $\Pi_{\mathcal{S}^n_+}\big(C + \mathcal{A}^*\bar{y} + \overline{Y}\big)$ happens to be an extreme point of $\mathrm{conv}\,\Pi_{\mathcal{S}^n(r)}(C - \overline{Y})$, the convex hull of all solutions of the metric projection problem (18) with $Y = C - \overline{Y}$, then $\Pi_{\mathcal{S}^n_+}\big(C + \mathcal{A}^*\bar{y} + \overline{Y}\big)$ globally solves the primal problem.

Corollary 4.5 *If* $|\sigma_r(C-\overline{Y})| > |\sigma_{r+1}(C-\overline{Y})|$, *then* $\overline{X} = \Pi_{\mathcal{S}^n_+}(C+\mathcal{A}^*\bar{y}+\overline{Y}) = \Pi_{\mathcal{S}^n(r)}(C-\overline{Y})$ *globally solves problem (5).*

Proof. By noting, from Remark 2.3, that the set $\Pi^B_{\mathcal{S}^n(r)}(C-\overline{Y})$ is a singleton when $|\sigma_r(C-\overline{Y})| > |\sigma_{r+1}(C-\overline{Y})|$, we derive the conclusion of this corollary directly from Theorem 4.4. □

Remark 4.6 *Theorem 4.4 and Corollary 4.5 also hold for the following* W*-weighted problem*

$$\begin{array}{ll} \min & \dfrac{1}{2}\|W^{1/2}(X-C)W^{1/2}\|^2 \\ \text{s.t.} & \mathcal{A}X \in b+\mathcal{Q}\,, \\ & X \in \mathcal{S}^n_+\,, \\ & \mathrm{rank}(X) \le r\,, \end{array} \tag{60}$$

where W *is a symmetric positive definite matrix. If the rank constraint* $\mathrm{rank}(X) \le r$ *is dropped from (60), then its dual problem can be simply reformulated, without requiring the additional matrix variable* Z*, as follows*

$$\begin{array}{ll} \min & \dfrac{1}{2}\big\|\Pi_{\mathcal{S}^n_+}\big(W^{1/2}(C+W^{-1}\mathcal{A}^*yW^{-1})W^{1/2}\big)\big\|^2 - \langle b,y\rangle - \dfrac{1}{2}\|W^{1/2}CW^{1/2}\|^2 \\ \text{s.t.} & y \in \mathcal{Q}^* = \mathbb{R}^p \times \mathbb{R}^q_+\,. \end{array} \tag{61}$$

The dual formulation (61) will be repeated used in our numerical experiments.

Remark 4.7 *Theorem 4.4 extends the globalization checking results of Zhang and Wu [63, Theorem 4.5] in several aspects:*

(E1). The matrix C *is no longer required to be a valid correlation matrix.*

(E2). The problem may have more general constraints including the simple lower and upper bound constraints.

(E3). The assumption $|\sigma_r(C-\overline{Y})| > |\sigma_{r+1}(C-\overline{Y})|$ *is weakened to include the situation that* $\sigma_r(C-\overline{Y})$ *may have multiplicity larger than 1.*

5 Numerical experiments

In this section, we shall first address several practical issues in the implementation of applying the majorized penalty approach (MPA) to problem (5) and then report our numerical results.

1. *The choice of the initial point* $X^0 \in \Omega$. Compute d as in (49). Let $D = \mathrm{diag}(d)$. We then apply the majorization method alternatively (first fix Z and then X) to approximately solve

$$\begin{array}{ll} \min & \dfrac{1}{2}\|H\circ(X-C)\|^2 + \dfrac{1}{2}\|H\circ(Z-C)\|^2 + \dfrac{\rho}{2}\|D^{1/2}(X-Z)D^{1/2}\|^2 \\ \text{s.t.} & \mathcal{A}X \in b+\mathcal{Q}\,, \\ & X \in \mathcal{S}^n_+\,, \\ & \mathrm{rank}(Z) \le r \end{array} \tag{62}$$

to obtain a feasible solution, say $(\widetilde{X}, \widetilde{Z})$, where $\rho > 0$ is initially set as 100 and is increased by 10 times at each step. The maximum number of steps is set as 10. Then we set $X^0 := \widetilde{X} \in \Omega$.

2. *The choice of the penalty parameter c.* Let X^* be an optimal solution to the convex problem (34). We choose the initial penalty parameter c to be

$$c := \min\left\{1,\, 0.25\big(\theta(X^0) - \theta(X^*)\big)/\max\left\{1,\, p(X^0) - p(X^*)\right\}\right\}.$$

Thereafter, c is updated as follows: when $|p(X^k)|/\max\{1, r\} > 0.1$, c is increased by 4 times; otherwise, c is increased by 1.4 times. The penalty parameter c will be kept unchanged if $|p(X^k)| \le 10^{-8}$.

3. *The choice of the algorithm for solving the subproblems (41).* The success of our MPA heavily relies on our ability in solving a sequence of majorized subproblems of the form (41). For this purpose, we use the well tested smoothing Newton-BiCGStab method developed in [23].

4. *The stopping criterion.* We terminate our algorithm if

$$|p(X^k)| \le 10^{-8} \quad \text{and} \quad \frac{|\sqrt{f_c(X^k)} - \sqrt{f_c(X^{k-1})}|}{\max\big(100, \sqrt{f_c(X^{k-1})}\big)} \le 10^{-5}.$$

We did our numerical experiments in `MATLAB 7.8.0 (R2009a)` running on a PC Intel (R) Core (TM) 2 of 3.16 GHz CPU each and 2.96 GB of RAM. The testing examples to be reported are given below.

Example 5.1 *Let $n = 500$ and the weight matrix $H = E$. For $i, j = 1, \ldots, n$, $C_{ij} = 0.5 + 0.5e^{-0.05|i-j|}$. The index sets are $\mathcal{B}_e = \mathcal{B}_l = \mathcal{B}_u = \emptyset$. This matrix C is a valid correlation matrix and has been used by a number of authors [6, 38].*

Example 5.2 *Let $n = 500$ and the weight matrix $H = E$. The matrix C is extracted from the correlation matrix which is based on a $10,000$ gene micro-array data set obtained from 256 drugs treated rat livers; see Natsoulis et al. [42] for details. The index sets are $\mathcal{B}_e = \mathcal{B}_l = \mathcal{B}_u = \emptyset$.*

Example 5.3 *Let $n = 500$. The matrix C is the same as in Example 5.1, i.e., $C = 0.5 + 0.5e^{-0.05|i-j|}$ for $i, j = 1, \ldots, n$. The index sets are $\mathcal{B}_e = \mathcal{B}_l = \mathcal{B}_u = \emptyset$. The weight matrix H is generated in the same way as in [47] such that all its entries are uniformly distributed in $[0.1, 10]$ except for 2×100 entries in $[0.01, 100]$.*

Example 5.4 *Let $n = 500$. The matrix C is the same as in Example 5.2. The index sets are $\mathcal{B}_e = \mathcal{B}_l = \mathcal{B}_u = \emptyset$. The weight matrix H is generated in the same way as in Example 5.3.*

Example 5.5 *The matrix C is an estimated 943×943 correlation matrix based on $100,000$ ratings for 1682 movies by 943 users. Due to missing data, the generated matrix G is not positive semi-definite [22]. This rating data set can be downloaded from* `http://www.grouplens.org/node/73`. *The index sets are $\mathcal{B}_e = \mathcal{B}_l = \mathcal{B}_u = \emptyset$. The weight matrix H is provided by T. Fushiki at Institute of Statistical Mathematics, Japan.*

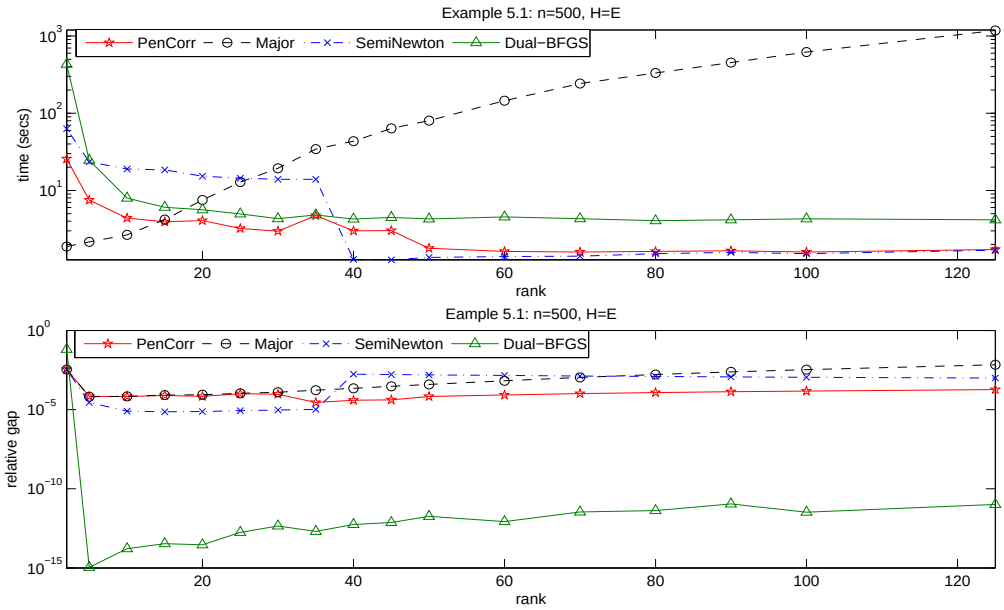

Figure 1: Example 5.1

Example 5.6 *The matrix C is obtained from the gene data sets with dimension $n = 1,000$ as in Example 5.2. The weight matrix H is the same as in Example 5.3. The index sets $\mathcal{B}_e$, $\mathcal{B}_l$, and $\mathcal{B}_u \subset \{(i,j) \,|\, 1 \le i < j \le n\}$ consist of the indices of $\min(\hat{n}_r, n-i)$ randomly generated elements at the ith row of X, $i = 1, \ldots, n$ with $\hat{n}_r = 5$ for $\mathcal{B}_e$ and $\hat{n}_r = 10$ for $\mathcal{B}_l$ and $\mathcal{B}_u$. We take $e_{ij} = 0$ for $(i,j) \in \mathcal{B}_e$, $l_{ij} = -0.1$ for $(i,j) \in \mathcal{B}_l$ and $u_{ij} = 0.1$ for $(i,j) \in \mathcal{B}_u$.*

Our numerical results are reported in Tables 1-5, where "*time*" and "*residue*" stand for the total computing time used (in seconds) and the residue $\sqrt{2\theta(X^k)}$ at the final iterate X^k of each algorithm, respectively. For the simplest rank-NCM problem (1) of equal weights (i.e., $H = E$), there are many algorithms to choose from. For the purpose of comparison, we only selected three most efficient ones from the literure: the dual approach of Zhang and Wu [63] and Wu [61] (C is required to be a valid correlation matrix), the majorization approach of Pietersz and Groenen [45], and the augmented Lagrangian approach of Li and Qi [38]. For the majorization approach and the augmented Lagrangian approach, we used the codes developed by the authors of [45] and [38]. They are referred to as `Major`[3] and `SemiNewton`, respectively, in Examples 5.1 and 5.2. For the dual approach of [63, 61], we used the BFGS implementation of Lewis and Overton [36] to solve the Lagrangian dual problem. This is denoted by `Dual-BFGS`. The `Dual-BFGS` solves the Lagrangian dual problem to get an approximate optimal dual solution y^k. This approximate optimal dual solution may not always be able to generate an optimal solution to the primal problem as the rth and $(r+1)$th eigenvalues (arranged in the non-increasing order in terms of their absolute values) of $C + \mathrm{diag}(y^k)$ may be of the same absolute values, but it does provide a valid lower bound for the optimal value of the primal problem. The final iterate of the

[3] `Majorw` is the corresponding code for solving the weighted cases.

Example 5.1	Major			SemiNewton			Dual-BFGS			PenCorr		
rank	*time*	*residue*	*relgap*	*time*	*residue*	*relgap*	*time*	*residue*	*relgap*	*time*	*residue*	*relgap*
2	1.9	1.564e2	3.4e-3	63.0	1.564e2	3.5e-3	432.0	1.660e2	6.5e-2	25.7	1.564e2	3.4e-3
5	2.2	7.883e1	6.5e-5	23.5	7.883e1	2.8e-5	24.6	7.883e1	1.1e-15	7.5	7.883e1	7.0e-5
10	2.7	3.869e1	6.9e-5	19.0	3.868e1	8.0e-6	8.0	3.868e1	1.7e-14	4.4	3.869e1	6.7e-5
15	4.2	2.325e1	8.3e-5	18.5	2.324e1	7.3e-6	6.0	2.324e1	3.4e-14	3.9	2.325e1	7.9e-5
20	7.5	1.571e1	8.8e-5	15.3	1.571e1	7.6e-6	5.6	1.571e1	2.9e-14	4.1	1.571e1	6.9e-5
25	12.8	1.145e1	1.1e-4	14.4	1.145e1	8.6e-6	5.0	1.145e1	1.8e-13	3.2	1.145e1	1.0e-4
30	19.4	8.797e0	1.3e-4	14.0	8.796e0	9.5e-6	4.3	8.795e0	4.4e-13	3.0	8.796e0	9.4e-5
35	34.4	7.020e0	1.7e-4	14.0	7.019e0	1.0e-5	4.8	7.019e0	2.0e-13	4.7	7.019e0	2.8e-5
40	43.4	5.766e0	2.2e-4	1.3	5.774e0	1.7e-3	4.3	5.764e0	5.6e-13	3.0	5.765e0	3.9e-5
45	63.6	4.843e0	3.0e-4	1.3	4.849e0	1.6e-3	4.5	4.841e0	7.4e-13	3.0	4.841e0	4.2e-5
50	80.1	4.141e0	4.0e-4	1.4	4.146e0	1.6e-3	4.3	4.139e0	1.8e-12	1.8	4.139e0	6.8e-5
60	145.0	3.156e0	6.7e-4	1.4	3.158e0	1.4e-3	4.5	3.153e0	8.4e-13	1.6	3.154e0	8.4e-5
70	243.0	2.507e0	1.1e-3	1.4	2.507e0	1.3e-3	4.3	2.504e0	3.4e-12	1.6	2.504e0	1.0e-4
80	333.0	2.053e0	1.6e-3	1.5	2.052e0	1.2e-3	4.1	2.050e0	4.2e-12	1.6	2.050e0	1.2e-4
90	452.0	1.722e0	2.4e-3	1.6	1.720e0	1.2e-3	4.2	1.718e0	1.1e-11	1.7	1.718e0	1.4e-4
100	620.0	1.471e0	3.3e-3	1.5	1.468e0	1.1e-3	4.3	1.467e0	3.3e-12	1.6	1.467e0	1.5e-4
125	1180.0	1.055e0	6.8e-3	1.7	1.049e0	9.9e-4	4.2	1.048e0	1.0e-11	1.7	1.048e0	1.8e-4

Table 1: Numerical results for Example 5.1

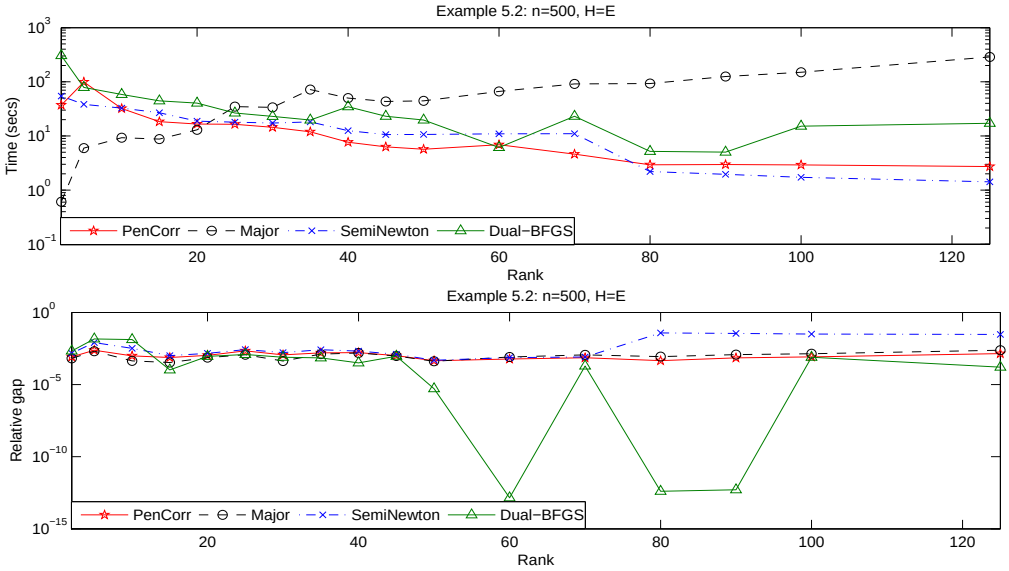

Figure 2: Example 5.2

Example 5.2	Major			SemiNewton			Dual-BFGS			PenCorr		
rank	*time*	*residue*	*relgap*	*time*	*residue*	*relgap*	*time*	*residue*	*relgap*	*time*	*residue*	*relgap*
2	0.6	2.858e2	6.5e-4	54.4	2.860e2	1.5e-3	304.5	2.862e2	2.1e-3	37.2	2.859e2	8.2e-4
5	6.0	1.350e2	2.0e-3	38.2	1.358e2	8.1e-3	78.8	1.367e2	1.5e-2	99.2	1.351e2	2.4e-3
10	9.3	6.716e1	4.4e-4	32.7	6.735e1	3.2e-3	58.3	6.802e1	1.3e-2	32.1	6.719e1	9.7e-4
15	8.8	4.097e1	3.4e-4	26.8	4.100e1	1.0e-3	44.6	4.096e1	1.0e-4	18.4	4.099e1	7.5e-4
20	13.0	2.842e1	7.3e-4	18.8	2.844e1	1.4e-3	40.4	2.842e1	8.9e-4	16.6	2.843e1	1.1e-3
25	34.9	2.149e1	1.2e-3	18.0	2.152e1	2.6e-3	26.6	2.149e1	1.2e-3	16.4	2.151e1	2.2e-3
30	33.7	1.693e1	4.3e-4	17.3	1.695e1	1.7e-3	23.0	1.694e1	7.8e-4	14.5	1.694e1	1.2e-3
35	71.8	1.379e1	1.3e-3	18.1	1.381e1	2.6e-3	19.7	1.378e1	7.1e-4	11.9	1.379e1	1.6e-3
40	50.0	1.151e1	1.5e-3	12.5	1.152e1	2.1e-3	34.7	1.145e1	3.2e-4	7.7	1.151e1	1.6e-3
45	43.3	9.733e0	9.6e-4	10.6	9.736e0	1.3e-3	23.1	9.733e0	9.2e-4	6.3	9.733e0	1.0e-3
50	44.5	8.318e0	4.1e-4	10.7	8.319e0	4.8e-4	19.7	8.315e0	5.1e-6	5.7	8.318e0	4.5e-4
60	66.5	6.214e0	8.1e-4	10.9	6.214e0	7.4e-4	6.1	6.209e0	1.4e-13	6.9	6.213e0	5.9e-4
70	91.2	4.733e0	1.1e-3	11.0	4.731e0	8.2e-4	23.1	4.728e0	1.9e-4	4.6	4.731e0	7.2e-4
80	93.0	3.663e0	8.7e-4	2.2	3.800e0	3.8e-2	5.2	3.660e0	4.0e-13	2.9	3.662e0	4.5e-4
90	125.0	2.865e0	1.2e-3	2.0	2.962e0	3.5e-2	5.0	2.862e0	5.1e-13	3.0	2.864e0	7.0e-4
100	150.0	2.255e0	1.4e-3	1.7	2.323e0	3.2e-2	15.1	2.254e0	7.8e-4	2.9	2.254e0	8.3e-4
125	288.6	1.269e0	2.4e-3	1.4	1.304e0	3.0e-2	17.1	1.266e0	1.6e-4	2.7	1.268e0	1.4e-3

Table 2: Numerical results for Example 5.2

	Example 5.3				Example 5.4			
	Majorw		PenCorr		Majorw		PenCorr	
rank	*time*	*residue*	*time*	*residue*	*time*	*residue*	*time*	*residue*
2	8.8	1.805e2	81.2	1.804e2	2.9	3.274e2	141.6	3.277e2
5	27.0	8.984e1	70.0	8.986e1	34.4	1.523e2	245.0	1.522e2
10	38.7	4.382e1	48.7	4.383e1	48.5	7.423e1	98.7	7.428e1
15	55.5	2.616e1	43.7	2.618e1	70.5	4.442e1	79.9	4.446e1
20	84.4	1.751e1	39.1	1.753e1	101.4	2.985e1	67.0	2.987e1
25	117.0	1.265e1	38.2	1.266e1	289.6	2.197e1	69.8	2.204e1
30	171.8	9.657e0	36.5	9.657e0	335.6	1.694e1	65.8	1.699e1
35	250.6	7.639e0	39.8	7.632e0	436.7	1.345e1	71.0	1.343e1
40	324.7	6.213e0	38.8	6.203e0	470.7	1.098e1	50.5	1.098e1
45	408.4	5.169e0	38.4	5.148e0	498.7	9.104e0	47.7	9.094e0
50	502.2	4.391e0	37.5	4.355e0	639.5	7.625e0	48.0	7.623e0
60	654.1	3.290e0	35.6	3.219e0	837.6	5.552e0	44.0	5.523e0
70	972.5	2.579e0	38.2	2.481e0	987.5	4.135e0	44.9	4.084e0
80	1274.9	2.090e0	42.6	1.959e0	1212.0	3.127e0	38.0	3.082e0
90	1526.9	1.740e0	44.0	1.588e0	1417.0	2.393e0	35.6	2.345e0
100	1713.7	1.478e0	40.9	1.310e0	1612.0	1.865e0	32.7	1.814e0
125	2438.1	1.052e0	44.6	8.591e-1	1873.0	1.030e0	27.7	9.748e-1

Table 3: Numerical results for Example 5.3 and 5.4

Example 5.5	Majorw		PenCorr	
rank	*time*	*residue*	*time*	*residue*
5	233.4	5.242e2	1534.9	5.273e2
10	706.5	3.485e2	1634.6	3.509e2
20	926.7	2.389e2	1430.2	2.398e2
50	2020.1	1.706e2	829.9	1.709e2
100	3174.3	1.609e2	537.5	1.611e2
150	3890.6	1.608e2	687.1	1.610e2
250	7622.5	1.608e2	694.2	1.610e2

Table 4: Numerical results for Example 5.5

Example 5.6	PenCorr	
rank	*time*	*residue*
20	11640.0	1.872e2
50	1570.0	1.011e2
100	899.0	8.068e1
250	318.3	7.574e1
500	326.3	7.574e1

Table 5: Numerical results for Example 5.6

`Dual-BFGS` is obtained by applying the modified PCA procedure to $C + \mathrm{diag}(y^k)$. Our own code is indicated by `PenCorr`. In Tables 1-2, "*relgap*" denotes the relative gap which is computed as

$$relgap := \frac{residue - \text{ lower bound}}{\max\{1, \text{ lower bound}\}},$$

where the lower bound is obtained by the `Dual-BFGS`. This "*relgap*" indicates the worst possible relative error from the global optimal value.

From Tables 1-2, we can see that even for the simplest rank-NCM problem (1) of equal weights (i.e., $H = E$), `PenCorr` is quite competitive in terms of computing time and solution quality except for small rank cases that `Major` is a clear winner. Examples 5.3, 5.4, and 5.5 belong to the rank-NCM problem (1) of general weights. For these three examples, we can see clearly from Tables 3-4 that `Majorw` performs better than `PenCorr` when the ranks are not large and loses its competitiveness quickly to `PenCorr` as the rank increases. When there are constraints on the off-diagonal parts as in Example 5.6, `PenCorr` seems to be the only viable approach.

6 Conclusions

In this paper, we proposed a majorized penalty approach for solving the rank constrained correlation matrix problem of the general form (5). Our approach is first to absorb the non-convex rank constraint into the objective function via a penalty technique by using the fact that for any $X \in \mathcal{S}^n_+$, $\mathrm{rank}(X) \le r$ if and only if $\lambda_{r+1}(X) + \ldots + \lambda_n(X) = 0$. Then we apply the majorization method to solve a sequence of recently well studied least squares problems in the form of (47). Numerical results indicate that our approach is able to handle both the rank and the bound constraints effectively, in particular in the situations when the rank is not very small. Though in order to make problem (5) feasible, one cannot ask the rank to be very small when there are a large number of bound constraints, it is still interesting to know if one can design a more efficient method to solve problem (5) with a small rank and a small number of bound constraints.

There are several directions that our approach presented here can be further researched. The first one is to replace the Frobenius norm by any other matrix norm, e.g., the matrix l_1 norm or the spectral norm, in the rank constrained nearest correlation matrix optimization problem (5). Another direction is to extend our approach to deal with even more complicated models on nearest correlation matrices of low rank in finance and risk management. For example, one may consider the following weighted version of the problem introduced by Werner and Schöttle in [60],

$$\begin{array}{ll} \min & \frac{1}{2}\|H \circ (X - C)\|^2 \\ \text{s.t.} & X_{ii} = 1, \quad i = 1, \ldots, n, \\ & X = B + D, \\ & D = \mathrm{diag}(d_1, \ldots, d_n), \\ & D \succeq 0, X \succeq 0, \\ & B \succeq 0, \ \mathrm{rank}(B) \le r\,. \end{array}$$

This type of problems comes from factor models in financial markets. See [60] for details. Finally, our majorized penalty approach can be extended to the following structured low rank matrix,

not necessary symmetric, approximation problem

$$\begin{array}{ll} \min & \frac{1}{2}\|H \circ (X - C)\|^2 \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & \text{rank}(X) \le r\,, \\ & X \in \mathbb{R}^{n_1 \times n_2}\,, \end{array} \tag{63}$$

where the matrices H and C are no longer required to be symmetric, the linear operator $\mathcal{A}$ is now defined from $\mathbb{R}^{n_1 \times n_2}$ to $\mathbb{R}^m$, $b \in \mathbb{R}^m$, and $\mathcal{Q}$ is a closed convex cone in $\mathbb{R}^m$. It is not clear at the moment if problem (63) has any specific application in finance or risk management. However, it has many applications in numerical linear algebra, engineering, and other fields. For a related survey article, see [11]. By using the fact that for any $X \in \mathbb{R}^{n_1 \times n_2}$ (without loss of generality, we assume $n_1 \le n_2$),

$$\text{rank}(X) \le r \Longleftrightarrow \sigma_{r+1}(X) + \ldots + \sigma_{n_1}(X) = 0 \Longleftrightarrow \sum_{i=1}^{n_1} \sigma_i(X) - \sum_{i=1}^{r} \sigma_i(X) = 0\,,$$

where $\sigma_i(X) \ge 0$ denotes the ith largest singular value of X, $i = 1, \ldots, n_1$, we derive the following analogue of problem (32)

$$\begin{array}{ll} \min & \frac{1}{2}\|H \circ (X - C)\|^2 + c\|X\|_* - c(\sigma_1(X) + \ldots + \sigma_r(X)) \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \in \mathbb{R}^{n_1 \times n_2}\,, \end{array}$$

where $c > 0$ is the penalty parameter and $\|X\|_* := \sigma_1(X) + \ldots + \sigma_{n_1}(X)$ is the nuclear norm of X. With slight modifications, one may use the majorized penalty method introduced in Section 3 to solve the above problem.

Acknowledgments.

We would like to thank all the participants of the Workshop on Risk Measures and Robust Optimization held at National University of Singapore from November 12 to 23, 2009 for their helpful comments and suggestions on this work. In particular, we thank Professor Lixin Wu at Hong Kong University of Science and Technology for discussions on the models of low rank matrices in financial industry, Dr. Hanqing Jin at Oxford University for the applications of this work in quantitative finance, and Drs. Ying Chen, Fermin Aldabe, and Chengjing Wang at the National University of Singapore on the covariance matrix data collection. Additionally, we also thank Professor Tadayoshi Fushiki at Institute of Statistical Mathematics for providing us the testing weight matrix H on the MovieLens problem, Dr Hou-Duo Qi at University of Southampton for making his code for solving the rank-NCM problem with equal weights available to us, Dr Min Li at Southeast University for her comments and suggestions on an earlier version of this paper.

References

[1] A. d'Aspremont, *Interest rate model calibration using semidefinite programming*, Applied Mathematical Finance 10 (2003), pp. 183-213.

[2] A. d'Aspremont, *Risk-Management method for the Libor market model using semidefinite programming*, Journal of Computational Finance 8 (2005), pp. 77-99.

[3] R. Bhatia, *Matrix Analysis*, Springer, New York, 1997.

[4] I. Borg and P. Groenen, *Modern Multidimensional Scaling*, Springer, New York, 1997.

[5] S. Boyd and L. Xiao, *Least-squares covariance matrix adjustment*, SIAM Journal on Matrix Analysis and Applications 27 (2005), pp. 532-546.

[6] D. Brigo, *A note on correlation and rank reduction*, working paper, 2002. Downloadable from `http://www.damianobrigo.it`.

[7] D. Brigo and F. Mercurio, *Interest rate models: theory and practice*, Springer-Verlag, Berlin, 2001.

[8] D. Brigo and F. Mercurio, *Calibrating LIBOR*, Risk Magazine 15 (2002), pp. 117–122.

[9] J.P. Burge, D.G. Luenberger and D.L. Wenger, *Estimation of structured covariance matrices*, Proceedings of the IEEE 70 (1982), pp. 963–974.

[10] Y.D. Chen, Y. Gao, and Y.-J. Liu, *An inexact SQP Newton method for convex SC^1 minimization problems*, to appear in Journal of Optimization Theory and Applications.

[11] M. T. Chu, R. E. Funderlic, and R. J. Plemmons, *Structured low rank approximation*, Linear Algebra and its Applications 366 (2003), pp. 157–172.

[12] F.H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, 1983.

[13] G. Cybenko, *Moment problems and low rank Toeplitz approximatons*, Circuits, Systems, and Signal Processing 1 (1983), pp. 245–366.

[14] J. de Leeuw, *Applications of convex analysis to multidimensional scaling*. In J. R. Barra, F. Brodeau, G. Romier, and B. van Cutsem (Eds.), Recent developments in statistics, Amsterdam, The Netherlands, 1977, pp. 133-145.

[15] J. de Leeuw, *Convergence of the majorization method for multidimensional scaling*, Journal of classification 5 (1988), pp. 163–180.

[16] J. de Leeuw, *Fitting distances by least squares*, technical report, University of California, Los Angeles, 1993.

[17] J. de Leeuw, *Block relaxation algorithms in statistics*. In H. H. Bock, W. Lenski and M. M. Richter (Eds.), Information Systems and Data Analysis, Springer-Verlag., Berlin, 1994, pp. 308–325.

[18] J. de Leeuw, *A decomposition method for weighted least squares low-rank approximation of symmetric matrices*, Department of Statistics, UCLA, April 2006. Available at `http://repositories.cdlib.org/uclastat/papers/2006041602`.

[19] J. de Leeuw and W.J. Heiser, *Convergence of correction matrix algorithms for multidimensional scaling*. In J. C. Lingoes, I. Borg and E. E. C. I. Roskam (Eds.), Geometric Representations of Relational Data, Mathesis Press, 1977, pp. 735–752.

[20] K. Fan, *On a theorem of Weyl concerning eigenvalues of linear transformations*, Proceedings of the National Academy of Science of U.S.A. 35 (1949), pp. 652–655.

[21] B. Flury, *Common Principal Components and Related Multivariate Models*, John Wiley & Sons, New York, 1988.

[22] T. Fushiki, *Estimation of positive semidefinite correlation matrices by using convex quadratic semidefinite programming*, Neural Computation 21 (2009), pp. 2028-2048.

[23] Y. Gao and D. F. Sun, *Calibrating least squares semidefinite programming with equality and inequality constraints*, SIAM Journal on Matrix Analysis and Applications 31 (2009), pp. 1432–1457.

[24] I. Grubišić, *Interest Rate Theory: The BGM Model*, master thesis, Leiden University, August 2002. Available at `http://www.math.uu.nl/people/grubisic`.

[25] I. Grubišić and R. Pietersz, *Efficient rank reduction of correlation matrices*, Linear Algebra and Its Applications 422 (2007), pp. 629–653.

[26] W.J. Heiser, *A generalized majorization method for least squares multidimensional scaling of pseudodistance that may be negative*, Psychometrika 56 (1991), pp. 7–27.

[27] W.J. Heiser, *Convergent computation by iterative majorization: theory and applications in multidimensional data analysis*, In W. J. Krzanowski (Ed.), Recent Advances in Descriptive Multivariate Analysis, Oxford University Press, Oxford, 1995, pp. 157–189.

[28] N.J. Higham, *Computing the nearest correlation matrix – a problem from finance*, IMA Journal of Numerical Analysis 22 (2002), pp. 329–343.

[29] W. Hoge, *A subspace identification extension to the phase correlation method*, IEEE Transactions on Medical Imaging 22 (2003), pp. 277–280.

[30] A.N. Kercheval, *On Rebonato and Jäckel's parametrization method for finding nearest correlation matrices*, International Journal of Pure and Applied Mathematics 45 (2008), pp. 383–390.

[31] H.A.L. Kiers, *Majorization as a tool for optimizing a class of matrix functions*, Psychometrika 55 (1990), pp. 417–428.

[32] H.A.L. Kiers, *Setting up alternating least squares and iterative majorization algorithm for solving various matrix optimization problems*, Computational Statistics & Data Analysis 41 (2002), pp. 157–170.

[33] D.L. Knol and J.M.F. ten Berge, *Least-squares approximation of an improper matrix by a proper one*, Psychometrika 54 (1989), pp. 53–61.

[34] A.B. Kurtulan, *Correlations in economic capital models for pension fund pooling*, Master Thesis, Tilburg University, December 2009.

[35] A.S. Lewis, *Derivatives of spectral functions*, Mathematics of Operations Research 21 (1996), pp. 576–588.

[36] A.S. Lewis and M.L. Overton, *Nonsmooth optimization via BFGS*, 2008. The MATLAB software is downloadable at `http://cs.nyu.edu/overton/software/index.html`.

[37] F. Lillo and R.N. Mantegna, *Spectral density of the correlation matrix of factor models: A random matrix theory approach*, Physical Review E 72 (2005), pp. 016219-1–016219-10.

[38] Q.N. Li and H.D. Qi, *A sequential semismooth Newton method for the nearest low-rank correlation matrix problem*, technical report, University of Southampton, September 2009.

[39] P.M. Lurie and M.S. Goldberg, *An approximate method for sampling correlated variables from partially-specified distributions*, Management Science (44) 1998, pp. 203–218.

[40] S.K. Mishra, *Optimal solution of the nearest correlation matrix problem by minimization of the maximum norm*, Munich Personal RePEc Archive, August 2004. Available at `http://mpra.ub.uni-muenchen.de/1783`.

[41] M. Morini and N. Webber, *An EZI method to reduce the rank of a correlation matrix in financial modelling*, Applied Mathematical Finance 13 (2009), pp. 309–331.

[42] G. Natsoulis, C. I Pearson, J. Gollub, B. P. Eynon, J. Ferng, R. Nair, R. Idury, M. D Lee, M. R Fielden, R. J Brennan, A. H Roter and K. Jarnagin, *The liver pharmacological and xenobiotic gene response repertoire*, Molecular Systems Biology 4 (2008), pp. 1–12.

[43] J.M. Otega and W.C. Rheinboldt, *Iterative solutions of nonlinear equations in several variables*, Academic Press, New York, 1970.

[44] M. Overton and R.S. Womersley, *Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices*, Mathematical Programming 62 (1993), pp. 321–357.

[45] R. Pietersz and P. Groenen, *Rank reduction of correlation matrices by majorization*, Quantitative Finance 4 (2004), pp. 649–662.

[46] H.-D. Qi and D.F. Sun, *A quadratically convergent Newton method for computing the nearest correlation matrix*, SIAM Journal on Matrix Analysis and Applications 28 (2006), pp. 360–385.

[47] H.-D. Qi and D.F. Sun, *An augmented Lagrangian dual approach for the H-weighted nearest correlation matrix problem*, preprint, to appear in: IMA Journal of Numerical Analysis.

[48] H.-D. Qi and D.F. Sun, *Correlation stress testing for value-at-risk: an unconstrained convex optimization approach*, Computational Optimization and Applications 45 (2010), pp. 427–462.

[49] F. Rapisarda, D. Brigo and F. Mercurio, *Parametrizing correlations: a geometric interpretation*, IMA Journal of Management Mathematics (18) 2007, pp. 55–73.

[50] R. Rebonato, *Calibrating the BGM model*, Risk Magazine (1999), pp. 74–79.

[51] R. Rebonato, *On the simultaneous calibration of multifactor lognormal interest rate models to black volatilities and to the correlation matrix*, Journal of Computational Finance 2 (1999), pp. 5–27.

[52] R. Rebonato, *Morden pricing of interest-rate derivatives*, Princeton University Press, New Jersey, 2002.

[53] R. Rebonato and P. Jackel, *The most general methodology to create a valid correlation matrix for risk management and option pricing purposes*, The Journal of Risk 2 (1999), pp. 17–27.

[54] R. T. Rockafellar, *Convex Analyis*, Princeton University Press, Princeton, 1970.

[55] D. Simon, *Reduced order kalman filtering without model reduction*, Control and Intelligent Systems 35 (2007), pp. 169–174.

[56] D. Simon, *A Majorization Algorithm for Constrained Correlation Matrix Approximation*, technical report, Cleveland State University, January 2009.

[57] N. C. Schwertman and D. M. Allen, *Smoothing an indefinite variance-covariance matrix*, Journal of Statistical Computation and Simulation 9 (1979), pp. 183–194.

[58] L. N. Trefethen and D. Bau, III, *Numerical Linear Algebra*, SIAM, Philadephia, 1997.

[59] G.A. Watson, *On matrix approximation problems with Ky Fan k norms*, Numerical Algorithms 5 (1993), pp. 263–272.

[60] R. Werner and K. Schöttle, *Calibration of correlation matrices - SDP or not SDP*, technical report, Munich University of Technology, 2007.

[61] L.X. Wu, *Fast at-the-money calibration of the LIBOR market model using Lagrange multipliers*, Journal of Computational Finance 6 (2003), pp. 39–77.

[62] E.H. Zarantonello, *Projections on convex sets in Hilbert space and spectral theory I and II.* In E. H. Zarantonello (Ed.), Contributions to Nonlinear Functional Analysis, Academic Press, New York, 1971, pp. 237–424.

[63] Z.Y. Zhang and L.X. Wu, *Optimal low-rank approximation to a correlation matrix*, Linear Algebra and Its Applications 364 (2003), pp. 161–187.

Appendix A. Proof of Proposition 2.2.

For any $z \in \mathbb{R}^n$, define

$$\xi_r(z) = \max_{x \in \mathcal{F}_r} \left\{ \frac{1}{2}\|z\|^2 - \frac{1}{2}\|x - z\|^2 \right\} = \max_{x \in \mathcal{F}_r} \left\{ \langle x, z \rangle - \frac{1}{2}\|x\|^2 \right\}, \tag{64}$$

where $\mathcal{F}_r := \{x \in \mathbb{R}^n \,|\, \|x\|_0 \le r\}$. Then $\xi_r(\cdot)$ is a convex function and its sub-differential is well defined. Let $y = \sigma(Y)$. Thus

$$\xi_r(y) = \max_{x \in \mathcal{F}_r} \left\{ \frac{1}{2}\|y\|^2 - \frac{1}{2}\|x - y\|^2 \right\} = \frac{1}{2}\|\sigma(Y)\|^2 - \min_{x \in \mathcal{F}_r} \frac{1}{2}\|x - y\|^2 . \tag{65}$$

Denote the solution set of (65) by $\mathcal{F}_r^*$. Define $\alpha := \{i \,|\, x_i \neq 0\}$ and $\beta := \{i \,|\, x_i = 0\}$ for any given $x \in \mathcal{F}_r$. It then follows that

$$\sum_{i=1}^{n} (x_i - y_i)^2 = \sum_{i \in \alpha} (x_i - y_i)^2 + \sum_{i \in \beta} (x_i - y_i)^2 \ge \sum_{i \in \beta} y_i^2 \ge \sum_{i=r+1}^{n} y_i^2 , \tag{66}$$

where the last inequality is from the facts that $|\beta| \ge n - r$ and the non-increasing order of y in terms of the absolute value. Therefore, we know that

$$\xi_r(y) = \frac{1}{2}\sum_{i=1}^{r} y_i^2 \quad \text{and} \quad \mathcal{F}_r^* = \mathcal{V}, \tag{67}$$

where $\mathcal{V}$ is defined in (24). From convex analysis [54], we can easily derive that

$$\partial \xi_r(y) = \operatorname{conv} \mathcal{V}$$

and that $\xi_r(\cdot)$ is differentiable at y if and only if $|\sigma(Y)|_r > |\sigma(Y)|_{r+1}$. In the latter case,

$$\partial \xi_r(y) = \{\nabla \xi_r(y)^T\} = \big\{ v \in \mathbb{R}^{1 \times n} \,|\, v_i = \sigma_i(Y) \text{ for } 1 \le i \le r \text{ and } v_i = 0 \text{ for } r+1 \le i \le n \big\}.$$

Since the convex function $\xi_r(\cdot)$ is symmetric, i.e., $\xi_r(z) = \xi_r(Sz)$ for $z \in \mathbb{R}^n$ and any permutation matrix S, for $Z \in \mathcal{S}^n$ we can rewrite $\Xi_r(Z)$ as

$$\Xi_r(Z) = \xi_r(\lambda(Z)) = \xi_r(\sigma(Z)) ,$$

where $\lambda_1(Z) \geq \ldots \geq \lambda_n(Z)$ are the eigenvalues of Z being arranged in the non-increasing order. By [35, Theorem 1.4], we know that $\Xi_r(\cdot)$ is differentiable at $Y \in \mathcal{S}^n$ if and only if $\xi_r(\cdot)$ is differentiable at $\lambda(Y)$ and

$$\partial \Xi_r(Y) = \{P\mathrm{diag}(v)P^T \,|\, v \in \partial \xi_r(\lambda(Y))\,,\; P \in \mathcal{O}_n,\; P\mathrm{diag}(\lambda(Y))P^T = Y\,\}.$$

Thus $\Xi_r(\cdot)$ is differentiable at Y if and only if $|\sigma(Y)|_r > |\sigma(Y)|_{r+1}$. In the latter case,

$$\partial \Xi_r(Y) = \{\Xi_r'(Y)\} = \big\{U\mathrm{diag}(v)U^T \,|\, v_i = \sigma_i(Y) \text{ for } 1 \leq i \leq r \text{ and } \; v_i = 0 \text{ for } r+1 \leq i \leq n\}.$$

Let the B-subdifferential of $\Xi_r(\cdot)$ at Y be defined by

$$\partial_B \Xi_r(Y) = \big\{ \lim_{Y^k \to Y} \Xi_r'(Y^k)\,,\; \Xi_r(\cdot) \text{ is differentiable at } Y^k \big\}.$$

Then we can easily check that

$$\partial_B \Xi_r(Y) = \Pi_{\mathcal{S}^n(r)}(Y)\,, \tag{68}$$

where we used the fact that the two matrices $\sum_{i\in\bar{\alpha}} \sigma_i(Y)U_iU_i^T$ and $\sum_{i\in\bar{\gamma}} \sigma_i(Y)U_iU_i^T$ are independent of the choices of $U \in \mathcal{O}_n$ satisfying (17). Thus, by Theorem 2.5.1 in [12], one has

$$\partial \Xi_r(Y) = \mathrm{conv}\, \partial_B \Xi_r(Y) = \mathrm{conv}\, \Pi_{\mathcal{S}^n(r)}(Y)\,.$$

The proof is completed. □