A smoothing Newton method for a type of inverse semi-definite quadratic programming problem

Xiantao Xiao\textsuperscript{a,}\*, Liwei Zhang\textsuperscript{a}, Jianzhong Zhang\textsuperscript{b}

\textsuperscript{a}Department of Applied Mathematics, Dalian University of Technology, Dalian 116024, Liaoning, China
\textsuperscript{b}Department of Mathematics, City University of Hong Kong, Tat Chee Avenue, Hong Kong

Received 29 September 2007; received in revised form 24 December 2007

Abstract

We consider an inverse problem arising from the semi-definite quadratic programming (SDQP) problem. We represent this problem as a cone-constrained minimization problem and its dual (denoted ISDQD) is a semismoothly differentiable (SC\textsuperscript{1}) convex programming problem with fewer variables than the original one. The Karush–Kuhn–Tucker conditions of the dual problem (ISDQD) can be formulated as a system of semismooth equations which involves the projection onto the cone of positive semi-definite matrices. A smoothing Newton method is given for getting a Karush–Kuhn–Tucker point of ISDQD. The proposed method needs to compute the directional derivative of the smoothing projector at the corresponding point and to solve one linear system per iteration. The quadratic convergence of the smoothing Newton method is proved under a suitable condition. Numerical experiments are reported to show that the smoothing Newton method is very effective for solving this type of inverse quadratic programming problems.

© 2008 Elsevier B.V. All rights reserved.

MSC: 90C20; 90C22

Keywords: Semi-definite quadratic programming; Inverse optimization; Smoothing Newton method

1. Introduction

A typical optimization problem is a forward problem, in which there are usually parameters associated with decision variables in the objective function and constraints. When solving the typical optimization problem, the values of these parameters usually are available and we need to find an optimal solution to it. An inverse optimization problem is to find values of parameters which make the known solutions optimal and which differ from the given estimates as little as possible.

The interest in inverse optimization problems was initiated by the paper \cite{5} dealing with an inverse shortest path problem. In the past few years, a variety of inverse combinatorial optimization problems have been studied by researchers, see, for example, the survey paper \cite{8} and the references \cite{1,2,4,6,20}, etc. But for continuous optimization,
there are just a few papers on their inverse problems, except for linear programming [18,19] and for quadratic
programming [21].

In this paper, we consider a semi-definite quadratic programming problem of the form

\[
\text{SDQP}(G, c, A, B) \\
\begin{aligned}
\min & \quad f(x) := \frac{1}{2} x^T G x + c^T x \\
\text{s.t.} & \quad x \in \Omega_p := \{ x^T \in \mathbb{R}^n \mid Ax \preceq B \},
\end{aligned}
\] (1.1)

where \( G \in \mathbb{S}^n_+ \), \( \mathbb{S}^n \) denotes the space of \( n \times n \) symmetric matrices, \( \mathbb{S}^n_+ \) denotes the cone of \( n \times n \) positive semi-definite symmetric matrices. For any \( C, D \in \mathbb{S}^n \), denote \( \text{Tr}(C) \) the trace of \( C \), \( \langle C, D \rangle = \text{Tr}(C^T D) \), \( \| C \|_F = \sqrt{\langle C, C \rangle} \), \( C \succeq D \) if and only if \( C - D \in \mathbb{S}^n_+ \). \( A : \mathbb{R}^n \to \mathbb{S}^m \) is a linear operator and \( A^* : \mathbb{S}^m \to \mathbb{R}^n \) is the adjoint of \( A \), \( c \in \mathbb{R}^n \) and \( B \in \mathbb{S}^m \). We define \( A \) by

\[
Ax := \sum_{j=1}^{\ n} x_j A_j, \quad \forall x \in \mathbb{R}^n,
\]

then \( A^* \) is defined by

\[
A^*(X) := \begin{bmatrix}
\langle A_1, X \rangle \\
\langle A_2, X \rangle \\
\vdots \\
\langle A_n, X \rangle
\end{bmatrix}, \quad \forall X \in \mathbb{S}^m,
\]

where \( A_i \in \mathbb{S}^m \) for \( i = 1, \ldots, n \). For simplicity of notations, we introduce “SOL” as a mapping whose variables are problems, we denote \( \text{SOL}(P) \) to be the set of optimal solutions to a problem \( P \).

Given a feasible point \( x^0 \in \Omega_p \), which should be the optimal solution to Problem \( \text{SDQP}(G, c, A, B) \) and a pair \((G^0, c^0) \in \mathbb{S}^n \times \mathbb{R}^n \) which is an estimate to \((G, c)\). The inverse semi-definite quadratic programming (ISDQP) considered in this paper is to find a pair \((G, c) \in \mathbb{S}^n_+ \times \mathbb{R}^n \) to solve

\[
\text{ISDQP}(A, B) \\
\begin{aligned}
\min & \quad \frac{1}{2} \| (G, c) - (G^0, c^0) \|^2 \\
\text{s.t.} & \quad x^0 \in \text{SOL(\text{SDQP}(G, c, A, B))}, \\
& \quad (G, c) \in \mathbb{S}^n_+ \times \mathbb{R}^n,
\end{aligned}
\] (1.2)

where \( \| \cdot \| \) is defined by \( \| (G', c') \| := \sqrt{\text{Tr}(G'^T G') + c'^T c'} \) for \((G', c') \in \mathbb{S}^n \times \mathbb{R}^n \).

Problem (1.2) is a cone-constrained optimization problem with a quadratic objective function. The scale of this problem will be quite large when \( n \) is a large number as the number of its decision variables is \( n + n(n + 1)/2 \). Our main idea in this paper is that, instead of dealing with Problem (1.2) directly, we focus on solving its dual problem. The reason for doing this is that the dual is a \( \text{SC}^L \) convex programming problem with fewer \( (\leq n) \) decision variables than the original inverse quadratic problem, and its feasible set is a SDP cone. We consider the smoothing Newton method, developed by [17], for getting a Karush–Kuhn–Tucker point of the dual problem.

Throughout this paper the following notations will be used. We denote the symmetric square root of \( X \) by \( X^{1/2} \). Let \(|X| := (X^T)^{1/2} \) and \( \overline{X}_s(X) := (X + |X|)/2 \) for any \( X \in \mathbb{S}^n \). The Hadamard product of \( X \) and \( Y \) is denoted by \( X \circ Y \), namely \((X \circ Y)_{ij} := X_{ij} Y_{ij} \). Let \( I \) be the identity matrix of appropriate dimension.

This paper is organized as follows. In Section 2, we give some results from nonsmooth analysis which shall be used in our convergence analysis. Section 3 is devoted to deriving the dual of the inverse quadratic programming problem. In Section 4, we describe the smoothing Newton method for problem (3.9) and prove the global convergence and the quadratic convergence rate. Numerical results implemented by the smoothing Newton method are given in Section 5.
2. Preliminary

In this section, we recall some results on semismooth mappings and properties of some smoothing functions, which will be used in what follows. Let $X$ and $Y$ be two finite-dimensional real vector spaces. Let $\mathcal{O}$ be an open set in $X$ and $\Psi : \mathcal{O} \subseteq X \to Y$ be a locally Lipschitz continuous function on the open set $\mathcal{O}$. By Rademacher's theorem, $\Psi$ is almost everywhere Fréchet-differentiable in $\mathcal{O}$. We denote by $\mathcal{D}_\Psi$ the set of Fréchet-differentiable points of $\Psi$ in $\mathcal{O}$. Then, the Bouligand-subdifferential of $\Psi$ at $x \in \mathcal{O}$, denoted by $\partial_B \Psi (x)$, is

$$
\partial_B \Psi (x) := \left\{ \lim_{k \to \infty} \mathcal{J} (\Psi (x^k)) | x^k \in \mathcal{D}_\Psi, \ x^k \to x \right\},
$$

where $\mathcal{J} (\Psi (x^k))$ denotes the Jacobian of $\Psi$ at $x^k$. Clarke's generalized Jacobian of $\Psi$ at $x$ is the convex hull of $\partial_B \Psi (x)$, i.e.,

$$
\partial \Psi (x) = \text{conv} \{ \partial_B \Psi (x) \}.
$$

The following concept of semismoothness was first introduced in [10] for functionals and was extended in [11] to vector-valued functions.

**Definition 2.1.** Let $\Psi : \mathcal{O} \subseteq X \to Y$ be a locally Lipschitz continuous function on the open set $\mathcal{O}$. We say that $\Psi$ is semismooth at a point $x \in \mathcal{O}$ if

(i) $\Psi$ is directionally differentiable at $x$; and

(ii) for any $\Delta x \in X$ and $V \in \partial \Psi (x + \Delta x)$ with $\Delta x \to 0$,

$$
\Psi (x + \Delta x) - \Psi (x) - V (\Delta x) = o(\| \Delta x \|).
$$

Furthermore, $\Psi$ is said to be strongly semismooth at $x \in \mathcal{O}$ if $\Psi$ is semismooth at $x$ and for any $\Delta x \in X$ and $V \in \partial \Psi (x + \Delta x)$ with $\Delta x \to 0$,

$$
\Psi (x + \Delta x) - \Psi (x) - V (\Delta x) = O(\| \Delta x \|^2).
$$

Let $K$ be a closed convex set in $Y$. For instance, the convex set $K$ will be chosen as the convex cone $S^+_k$ or $\mathbb{R}_+^p$ in the following sections. It is well known [22] that the metric projector $\Pi_K (\cdot)$ for each element of $Y$ is Lipschitz continuous with the Lipschitz constant 1. Then for any $y \in Y$, $\partial \Pi_K (y)$ is well defined. Below is a lemma showing the general properties of $\partial \Pi_K (\cdot)$.

**Lemma 2.1 ([9, Proposition 1]).** Let $K \subseteq Y$ be a closed convex set. Then, for any $y \in Y$ and $V \in \partial \Pi_K (y)$, it holds that

(i) $V$ is self-adjoint.

(ii) $\langle d, V d \rangle \geq 0$, $\forall d \in Y$.

(iii) $\langle V d, d - V d \rangle \geq 0$, $\forall d \in Y$.

As our method needs to use the projection onto $S^+_k$, in addition to the above lemma about the projection on a general closed convex set $K$, we should know more properties about $\partial_B \Pi_{S^+_k} (\cdot)$. Let $\bar{Z} \in S^p$ and $\bar{Z}_+ := \Pi_{S^+_k} (\bar{Z})$.

Suppose that $\bar{Z}$ has the following spectral decomposition

$$
\bar{Z} = \bar{P} \Lambda \bar{P}^T,
$$

where $\Lambda$ is the diagonal matrix of eigenvalues of $\bar{Z}$ and $\bar{P}$ is a corresponding orthogonal matrix of the orthonormal eigenvectors. Then

$$
\bar{Z}_+ = \bar{P} \Lambda_+ \bar{P}^T,
$$

where $\Lambda_+$ is the diagonal matrix whose diagonal entries are the nonnegative parts of the respective diagonal entries of $\Lambda$. Define three index sets of positive, zero, and negative eigenvalues of $\bar{Z}$, respectively, as

$$
\alpha := \{ i | \lambda_i > 0 \}, \quad \beta := \{ i | \lambda_i = 0 \}, \quad \gamma := \{ i | \lambda_i < 0 \}.
$$
Write
\[
A = \begin{bmatrix}
A_\alpha & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & A_\gamma
\end{bmatrix}
\quad \text{and} \quad
\overline{P} = [ \overline{P}_\alpha \quad \overline{P}_\beta \quad \overline{P}_\gamma ]
\]
with \( \overline{P}_\alpha \in \mathbb{R}^{p \times |\alpha|}, \overline{P}_\beta \in \mathbb{R}^{p \times |\beta|}, \) and \( \overline{P}_\gamma \in \mathbb{R}^{p \times |\gamma|} \). Let \( \Theta \) be any matrix in \( S^p \) with entries
\[
\begin{cases}
\Theta_{ij} = 0, & \text{if } (i, j) \in \beta \times \beta, \\
\Theta_{ij} = \max \{ \lambda_i, 0 \} + \max \{ \lambda_j, 0 \}, & \text{if } (i, j) \in \alpha \times \beta, \\
\Theta_{ij} = [\lambda_i] + [\lambda_j], & \text{otherwise}.
\end{cases}
\]
(2.1)

The projection operator \( \Pi_{S^p}(\cdot) \) is directionally differentiable everywhere in \( S^p \) [3] and is a strongly semismooth matrix-valued function [15]. For any \( H \in S^p \), we have
\[
\Pi'_{S^p}(\overline{Z}; H) = \overline{P} \left[ \begin{array}{ccc}
\overline{P}_\alpha H \overline{P}_\alpha & \overline{P}_\alpha H \overline{P}_\beta & \Theta_{\alpha \gamma} \circ \overline{P}_\alpha H \overline{P}_\gamma \\
\overline{P}_\beta H \overline{P}_\alpha & \Pi'_{S^p}(\overline{P}_\beta H \overline{P}_\beta) & 0 \\
\overline{P}_\gamma H \overline{P}_\alpha \circ \Theta_{\gamma \alpha} & 0 & 0
\end{array} \right] \overline{P}^T,
\]
where “\( \circ \)” denotes the Hadamard product [15], and \( \Theta_{\alpha \gamma} \) is the submatrix of \( \Theta \) formed by the elements of the first \( |\alpha| \) rows and the last \( |\gamma| \) columns, and \( \Theta_{\gamma \alpha} \) has a similar meaning.

The following lemma on \( \partial_B \Pi'_{S^p}(\overline{Z}) \) is from [16].

**Lemma 2.2.** Let \( \Theta \in S^p \) satisfy (2.1). Then \( W \in \partial_B \Pi_{S^p}(\overline{Z}) \) if and only if there exists \( W_0 \in \partial_B \Pi_{S^p}(0) \) such that
\[
W(H) = \overline{P} \left[ \begin{array}{ccc}
\overline{P}_\alpha H \overline{P}_\alpha & \overline{P}_\alpha H \overline{P}_\beta & \Theta_{\alpha \gamma} \circ \overline{P}_\alpha H \overline{P}_\gamma \\
\overline{P}_\beta H \overline{P}_\alpha & \Pi'_{S^p}(\overline{P}_\beta H \overline{P}_\beta) & 0 \\
\overline{P}_\gamma H \overline{P}_\alpha \circ \Theta_{\gamma \alpha} & 0 & 0
\end{array} \right] \overline{P}^T, \quad \forall H \in S^p.
\]
(2.2)

Let \( Q \) be the set of all orthogonal matrices of order \( |\beta| \times |\beta| \). Let
\[
\mathcal{P} := \{ P \in \mathbb{R}^{p \times p} \mid P = [P_\alpha \quad P_\beta \quad P_\gamma] = [\overline{P}_\alpha \quad (\overline{P}_\beta Q) \quad \overline{P}_\gamma], \; Q \in \mathcal{Q}. \}
\]
Note that all \( P \in \mathcal{P} \) have the same \( P_\alpha \) and \( P_\gamma \). From the definition of \( \partial_B \Pi_{S^p}(0) \) and [7, Lemma 4.7], we know that if \( W_0 \in \partial_B \Pi_{S^p}(0) \), then there exist matrices \( Q \in \mathcal{Q} \) and \( \Omega \in S^{|\beta|} \) with entries \( \Omega_{ij} \in [0, 1] \) such that
\[
W_0(D) = Q(\Omega \circ (Q^T D Q)) Q^T, \quad \forall D \in S^{|\beta|}.
\]
Thus, by using Lemma 2.2 we obtain the following useful lemma, which does not need further explanation.

**Lemma 2.3.** For any \( W \in \partial_B \Pi_{S^p}(\overline{Z}) \), there exist two matrices \( P \in \mathcal{P} \) and \( \Theta \in S^p \) satisfying (2.1) such that
\[
W(H) = P(\Theta \circ (P^T H P)) P^T, \quad \forall H \in S^p.
\]
(2.3)

Finally we discuss the properties of an important function defined as follows: for \( \varepsilon \in \mathbb{R} \) and \( X \in S^n \), the *square smoothing function* \( \Phi : \mathbb{R} \times S^n \to S^n \), see [17], is defined by
\[
\Phi(\varepsilon, X) := (\varepsilon^2 I + X^2)^{1/2}, \quad \forall (\varepsilon, X) \in \mathbb{R} \times S^n.
\]
(2.4)

Then, \( \Phi \) is continuously differentiable at \( (\varepsilon, X) \) unless \( \varepsilon = 0 \) and for any \( Y \in S^n \),
\[
\Phi(\varepsilon, X) \to |Y|, \quad \text{as } (\varepsilon, X) \to (0, Y).
\]

For any \( X \in S^n \), let \( L_X \) be the Lyapunov operator:
\[
L_X(Y) := XY + YX, \quad \forall Y \in S^n.
\]
with $L_X^{-1}$ being its inverse (if it exists at all), i.e., for any $Y \in S^n$, $L_X^{-1}(Y)$ is the unique $Z \in S^n$ satisfying $XZ + ZX = Y$. The following result is proved in [17, Lemma 2.3, Theorem 2.5 and Proposition 3.1].

**Lemma 2.4.** For $(\varepsilon, X) \in \mathbb{R} \times S^n$, assume that there exist an orthogonal matrix $P$ and a matrix $\Lambda = \text{diag}(\mu_1, \ldots, \mu_n)$ of eigenvalues of $X$ such that $X = P\Lambda P^T$, the following statements hold.

1. If $\varepsilon^2 I + X^2$ is nonsingular, then $\Phi$ is continuously differentiable at $(\varepsilon, X)$, where $J \Phi(\varepsilon, X)$ satisfies the following equations

   \[ J \Phi(\varepsilon, X)(\tau, H) = L_{\Phi(\varepsilon, X)}^{-1}(L_X(H) + 2\varepsilon \tau I), \quad \forall (\tau, H) \in \mathbb{R} \times S^n, \]

   and for $i, j = 1, \ldots, n$,

   \[ (P^T J \Phi(\varepsilon, X)(\tau, H) P)_{ij} = \begin{cases} 
   \frac{(\varepsilon^2 + \mu_i^2 + \mu_j^2)}/2 + (\varepsilon^2 + \mu_i^2 + \mu_j^2)}/2, & \text{if } i \neq j, \\
   \frac{\mu_i (P^T H P)_{ij} + \varepsilon \tau}{(\varepsilon^2 + \mu_i^2 + \mu_j^2)}/2, & \text{otherwise}.
   \end{cases} \]

2. $\Phi$ is strongly semismooth at $(0, X)$.

3. For $(0, H) \in \mathbb{R} \times S^n$ and $V \in \partial_B \Phi(0, X)$, it holds that

   \[ V(0, H) = P(\Omega \circ P^T H P) P^T, \]

   where $\Omega \in S^n$ has entries

   \[ \Omega_{ij} = \begin{cases} 
   t \in [-1, 1], & \text{if } \mu_i = \mu_j = 0, \\
   \frac{\mu_i + \mu_j}{|\mu_i| + |\mu_j|}, & \text{otherwise}.
   \end{cases} \]

3. The dual problem

From the conventional duality theory, if $G \in S^n_+$, then $x^0 \in \text{SOL}(SDQP(G, c, A, B))$ if and only if there is a matrix $\Omega \in S^n_+$ such that

\[ c + Gx^0 + A^*(\Omega) = 0, \quad \Omega \in S^n_+, \quad Ax^0 \leq B, \quad \langle \Omega, Ax^0 - B \rangle = 0. \]

Let $Z^0 := Ax^0 - B$, then problem (1.2) can be equivalently expressed as follows

\[ \min \frac{1}{2} \|(G, c) - (G^0, c^0)\|^2 \]

s.t. \[ c + Gx^0 + A^*(\Omega) = 0, \]

\[ \langle \Omega, Z^0 \rangle = 0, \]

\[ (G, c, \Omega) \in S^n_+ \times \mathbb{R}^n \times S^n_+. \]

Let $r := \text{rank } Z^0$. Assume that $Z^0$ has the following spectral decomposition

\[ Z^0 = [P_r, P_{\bar{r}}] \begin{bmatrix} \Lambda_r & 0 \\ 0 & \Lambda_{\bar{r}} \end{bmatrix} \begin{bmatrix} P_r^T \\ P_{\bar{r}}^T \end{bmatrix}, \]

where $P := [P_r, P_{\bar{r}}] \in \mathbb{R}^{m \times m}$ is an orthogonal matrix with $P_r \in \mathbb{R}^{n \times r}$ and $P_{\bar{r}} \in \mathbb{R}^{m \times (m-r)}$, $\Lambda_r = \text{diag} 1 \leq i \leq r(\lambda_i)$, where $\lambda_i < 0, i = 1, \ldots, r$ are $r$ nonzero eigenvalues of $Z^0$. Define $\hat{M} := P^T M P$ for $M \in S^n$ with

\[ \hat{M} = \begin{bmatrix} \hat{M}_{rr} & \hat{M}_{r\bar{r}} \\ \hat{M}_{\bar{r}r} & \hat{M}_{\bar{r}\bar{r}} \end{bmatrix}, \quad \hat{M}_{rr} = P_r^T M P_r, \quad \hat{M}_{r\bar{r}} = P_r^T M P_{\bar{r}}, \quad \hat{M}_{\bar{r}r} = P_{\bar{r}}^T M P_r. \]
Then problem (3.1) is equivalent to
\[
\begin{align*}
\min & \quad \frac{1}{2} \|(G, c) - (G^0, c^0)\|^2 \\
\text{s.t.} & \quad c + Gx^0 + \hat{A}^*(\hat{\Omega}) = 0, \\
& \quad \hat{\Omega}_{ij} = 0, \quad i = 1, \ldots, r, \\
& \quad (G, c, \hat{\Omega}) \in S^n_+ \times \mathbb{R}^n \times S^m_+,
\end{align*}
\] (3.2)
where \( \hat{A}^*(\hat{\Omega}) := ((\hat{A}_1, \hat{\Omega}), \ldots, (\hat{A}_n, \hat{\Omega}))^T \).

Noticing that relations \( \hat{\Omega} \in S^n_+, \hat{\Omega}_{ii} = 0, i = 1, \ldots, r \) imply \( \hat{\Omega}_{ir} = 0 \) and \( \hat{\Omega}_{ri} = 0 \), we have that \( \hat{\Omega} \) has the following form
\[
\begin{bmatrix}
0 & 0 & 0 \\
0 & \hat{\Omega}_{ir} & \hat{\Omega}_{ri}
\end{bmatrix}.
\]
Thus problem (3.2) is equivalent to
\[
\text{ISDPQ}(\mathcal{A}, B)
\]
\[
\begin{align*}
\min & \quad \frac{1}{2} \|(G, c) - (G^0, c^0)\|^2 \\
\text{s.t.} & \quad c + Gx^0 + \hat{A}^*_y(\hat{\Omega}_{ir}) = 0, \\
& \quad (G, c, \hat{\Omega}_{ir}) \in \mathcal{K} := S^n_+ \times \mathbb{R}^n \times S^{m-r}_+,
\end{align*}
\] (3.3)
where \( \hat{A}^*_y(\hat{\Omega}_{ir}) := ((\hat{A}_1, \hat{\Omega}_{ir}), \ldots, (\hat{A}_n, \hat{\Omega}_{ir}))^T \). As the dimension of the above problem is \( n(n + 1)/2 + n + (m - r)(m - r + 1)/2 \), quite big when \( n \) is large, it would be helpful to consider its dual. Since problem (3.3) is a convex programming problem and the generalized Slater constraint qualification obviously holds for (3.3). So, by the classical duality theory for convex programming [13, Theorems 17 and 18], there is no duality gap between problem (3.3) and its dual. Let \( L : S^n_+ \times \mathbb{R}^n \times S^{m-r}_+ \times \mathbb{R}^n \to \mathbb{R} \) be the Lagrangian of problem (3.3), defined by
\[
L(G, c, \hat{\Omega}_{ir}, y) := \frac{1}{2} \|(G, c) - (G^0, c^0)\|^2 + y^T (c + Gx^0 + \hat{A}^*_y(\hat{\Omega}_{ir})).
\]
The (Lagrange) dual problem of (3.3) is
\[
\sup_{y \in \mathbb{R}^n} v(y) := \inf_{(G, c, \hat{\Omega}_{ir}) \in \mathcal{K}} L(G, c, \hat{\Omega}_{ir}, y).
\] (3.4)

Lemma 3.1. The function \( v(y) \) defined in (3.4) has the following expression
\[
v(y) = \begin{cases} 
-\frac{1}{2} \|y\|^2 + c^0 y - \frac{1}{2} \| \Pi_{S^n_+} (\hat{G}(y)) \|^2 + \frac{1}{2} \| G^0 \|^2, & \text{if } \hat{A}^*_y y \geq 0, \\
-\infty, & \text{otherwise},
\end{cases}
\] (3.5)
where \( \hat{G}(y) := G^0 - \frac{\sum_{i=1}^n y_i (\hat{A}_i)}{2} + x^0 y^T \) and \( \hat{A}^*_y y := \sum_{i=1}^n y_i (\hat{A}_i)_{ir} \).

Proof. From the definition of \( v \), we have
\[
v(y) = \inf_{(G, c, \hat{\Omega}_{ir}) \in \mathcal{K}} \left\{ \frac{1}{2} \| c - c^0 \|^2 + y^T c + \frac{1}{2} \| G - G^0 \|^2 + y^T Gx^0 + y^T \hat{A}^*_y(\hat{\Omega}_{ir}) \right\}
\]
\[
= \begin{cases} 
\inf_{(G, c) \in S^n_+ \times \mathbb{R}^n} \left\{ \frac{1}{2} \| c - c^0 \|^2 + y^T c + \frac{1}{2} \| G - G^0 \|^2 + y^T Gx^0 \right\}, & \text{if } \hat{A}^*_y y \geq 0, \\
-\infty, & \text{otherwise},
\end{cases}
\]
\[
= \begin{cases} 
\inf_{c \in \mathbb{R}^n} \left\{ \frac{1}{2} \| c - c^0 \|^2 + y^T c \right\} + \inf_{G \in S^n_+} \left\{ \frac{1}{2} \| G - G^0 \|^2 + y^T Gx^0 \right\}, & \text{if } \hat{A}^*_y y \geq 0, \\
-\infty, & \text{otherwise}.
\end{cases}
\] (3.6)
As the unconstrained quadratic programming problem
\[
\min_{c \in \mathbb{R}^n} \frac{1}{2} \|c - c^0\|^2 + y^T c
\]
takes its minimum at
\[
c^*(y) = c^0 - y,
\]
we have that
\[
\inf_{c \in \mathbb{R}^n} \left\{ \frac{1}{2} \|c - c^0\|^2 + y^T c \right\} = -\frac{1}{2} \|y\|^2 + c^0 y.
\]
From the following expression
\[
\inf_{G \in S^n} \left\{ \frac{1}{2} \|G - G^0\|_F^2 + y^T G x^0 \right\}
\]
\[
= \inf_{G \in S^n} \left\{ \frac{1}{2} \left\| G - \tilde{G}(y) \right\|_F^2 + 2 \left( G^0, \frac{y x^0 T + x^0 y^T}{2} \right) - \frac{\| y x^0 T + x^0 y^T \|_F^2}{2} \right\},
\]
we know that the minimum value is reached at
\[
G^*(y) = \Pi_{S^n_+}(\tilde{G}(y)),
\]
and thus
\[
\inf_{G \in S^n} \left\{ \frac{1}{2} \|G - G^0\|_F^2 + y^T G x^0 \right\} = \frac{1}{2} \left\| \tilde{G}(y) - \Pi_{S^n_+}(\tilde{G}(y)) \right\|_F^2 - \frac{\| \tilde{G}(y) \|_F^2 + \| G^0 \|_F^2}{2},
\]
\[
= -\frac{1}{2} \| \Pi_{S^n_+}(\tilde{G}(y)) \|_F^2 + \frac{1}{2} \| G^0 \|_F^2.
\]
Therefore, the function \(\nu(y)\) has the expression (3.5). \(\Box\)

For the simplicity of notations, we let \(H_i = P^T_i A_i P_i\), for \(i = 1, \ldots, n\). Moreover, we define a linear operator \(\mathcal{H}\) by
\[
\mathcal{H}y := \sum_{i=1}^n y_i H_i, \quad \forall y \in \mathbb{R}^n,
\]
then its adjoint \(\mathcal{H}^*\) is given as
\[
\mathcal{H}^*(X) := \begin{bmatrix} \langle H_1, X \rangle \\ \langle H_2, X \rangle \\ \vdots \\ \langle H_n, X \rangle \end{bmatrix}, \quad \forall X \in S^p,
\]
where \(p = m - r\). Then it follows from (3.5) that the dual problem of (1.2) can be written as
\[
\text{ISDQD}(A, B)
\]
\[
\begin{align*}
\min_{\nu_0(y)} & \quad \nu_0(y) \\
\text{s.t.} & \quad \mathcal{H}y \succeq 0, \ y \in \mathbb{R}^n,
\end{align*}
\]
where
\[
\nu_0(y) := \frac{1}{2} \|y\|^2 - c^0 y + \frac{1}{2} \| \Pi_{S^n_+}(\tilde{G}(y)) \|_F^2 - \frac{1}{2} \| G^0 \|_F^2,
\]
Define $B y := \frac{y \sigma_0 + x_0^T y}{2}$, then $\tilde{G}(y) = G^0 - B y$. Obviously, $B : \mathbb{R}^d \to S^n$ is an linear operator and its adjoint $B^* : S^n \to \mathbb{R}^d$ is given by $B^* G = G x^0$. The function $v_0$ is continuously differentiable with

$$
\nabla v_0(y) = y - c^0 + \mathcal{J}(\tilde{G}(y))^\ast \Pi_{S^n_+}(\tilde{G}(y))
= y - c^0 - B^* \Pi_{S^n_+}(\tilde{G}(y)).
$$

Since the mapping $\Pi_{S^n_+}(\cdot)$ is a strongly semismooth mapping, then $v_0(\cdot)$ is a $SC^1$ function, and we can derive an inclusion relation on the generalized Hessian of $v_0(\cdot)$ in the following lemma.

**Lemma 3.2.** The function $v_0(y)$ is continuously differentiable and strongly convex. $\nabla v_0(y)$ is strongly semismooth, and the generalized Hessian of $v_0(y)$ satisfies

$$
\partial^2 v_0(y) \subset I + B^* \partial \Pi_{S^n_+}(\tilde{G}(y)) B.
$$

**Lemma 3.3.** There exists a unique solution to Problem ISDQD(A, B). Let $y^*$ be the unique solution to ISDQD(A, B), then

$$
(G^*, c^*) = (\Pi_{S^n_+}(\tilde{G}(y^*)), c^0 - y^*)
$$

solves the original problem (1.2).

**Proof.** As $v_0$ is strongly convex and the constraint set of ISDQD(A, B) is given by a linear SDP constraint, Problem ISDQD(A, B) has a unique solution. Noticing that there is no duality gap between (3.3) and its dual, we obtain that if $(G^*, c^*, \hat{\Omega}_T^*)$ uniquely solves the following problem

$$
\min \left\{ L(G, c, \hat{\Omega}_T^*, y^*) : (G, c, \hat{\Omega}_T^*) \in \mathcal{K} \right\},
$$

\{(G^*, c^*, \hat{\Omega}_T^*)\} solves ISDQP(A, B). From (3.7) and (3.8), we have that

$$
c^* = c^*(y^*) = c^0 - y^* \text{ and } G^* = G^*(y^*) = \Pi_{S^n_+}(\tilde{G}(y^*)�).
$$

The proof is completed. □

4. **Smoothing Newton method**

This section focuses on the convergence analysis of the smoothing Newton method for getting a Karush–Kuhn–Tucker point of problem (3.9). As $v_0$ is strongly convex and the constraint set of ISDQD(A, B) is a SDP cone, the Karush–Kuhn–Tucker conditions are necessary and sufficient conditions for the solution to Problem (3.9). The Lagrangian of the problem ISDQD(A, B) is

$$
L(y, \Xi) := v_0(y) + \langle \Xi, \mathcal{H}y \rangle
$$

and its gradient is

$$
\nabla_y L(y, \Xi) = y - c^0 - B^* \Pi_{S^n_+}(\tilde{G}(y)) - \mathcal{H}^\ast \Xi,
$$

where $\Xi \in S^n$ is the Lagrange multiplier. The Karush–Kuhn–Tucker optimality conditions for the problem ISDQD(A, B) are

$$
\nabla_y L(y, \Xi) = 0,
$$
$$
\Xi \succeq 0, \quad \mathcal{H}y \succeq 0, \quad \langle \Xi, \mathcal{H}y \rangle = 0,
$$

which can be equivalently reformulated as

$$
\nabla_y L(y, \Xi) = y - c^0 - B^*(\tilde{G}(y) + |\tilde{G}(y)|)/2 - \mathcal{H}^\ast \Xi = 0,
$$
$$
\mathcal{H}y - \Pi_{S^n_+}(\mathcal{H}y - \Xi) = 0.
$$
Define $F : \mathbb{R}^n \times \mathcal{S}^p \to \mathbb{R}^n \times \mathcal{S}^p$ as follows

$$F(y, \Xi) = \begin{bmatrix} y - c_0^0 - B^*(\tilde{G}(y) + |\tilde{G}(y)|)/2 - \mathcal{H}_y^* \Xi \\ \mathcal{H}_y - \Pi_{\mathcal{S}^p}(\mathcal{H}_y - \Xi) \end{bmatrix}.$$  

Then the Karush–Kuhn–Tucker conditions are equivalent to $F(\tilde{y}, \tilde{\Xi}) = 0 \in \mathbb{R}^n \times \mathcal{S}^p$. The smoothing Newton method is based on a smoothing approximation to $F$, it requires the nonsingularity of elements in $\partial F(\tilde{y}, \tilde{\Xi})$. For this purpose, the following constraint nondegeneracy condition is needed. Let $\tilde{r} := \mathcal{H}_y - \tilde{\Xi}$ have the following spectral decomposition

$$\tilde{r} = QAQ^T,$$

and define three index sets of positive, zero, and negative eigenvalues of $T$, respectively, as

$$\alpha := \{i | \lambda_i > 0\}, \quad \beta := \{i | \lambda_i = 0\}, \quad \gamma := \{i | \lambda_i < 0\}.$$  

Then write

$$A = \begin{bmatrix} A_{\alpha} & 0 & 0 \\ 0 & 0_{\beta} & 0 \\ 0 & 0 & A_{\gamma} \end{bmatrix}, \quad Q = [Q_{\alpha} \quad Q_{\beta} \quad Q_{\gamma}].$$

and let $Q_{\alpha} := [Q_{\beta} \quad Q_{\gamma}]$ and $t := |\beta| + |\gamma|$.

**Assumption 4.1.** Let the set of vectors

$$\left\{ \begin{bmatrix} q_i^T \ H_1 q_j \\ \vdots \\ q_i^T \ H_t q_j \end{bmatrix} \middle| 1 \leq i \leq j \leq t \right\}$$

be linear independent, where $q_i, i = 1, \ldots, t$ is the $i$th column of $Q_{\alpha}$.

**Proposition 4.1.** Let Assumption 4.1 be satisfied. Then every element of $\partial F(\tilde{y}, \tilde{\Xi})$ is nonsingular.

**Proof.** Let $W$ be an arbitrary element in $\partial F(\tilde{y}, \tilde{\Xi})$. Assume that there exist $(\Delta y, \Delta \Xi) \in \mathbb{R}^n \times \mathcal{S}^p$ such that

$$W(\Delta y, \Delta \Xi) = 0.$$  

Then, we can find a $U \in \partial \Pi_{\mathcal{S}^p}(\tilde{G}(\tilde{y}))$ and a $V \in \partial \Pi_{\mathcal{S}^p}(\mathcal{H}(\tilde{y}) - \tilde{\Xi})$ such that

$$W(\Delta y, \Delta \Xi) = \begin{bmatrix} \Delta y + B^*U B \Delta y - \mathcal{H}_y^* \Delta \Xi \\ \mathcal{H}_y \Delta y - V(\mathcal{H}_y \Delta y - \Delta \Xi) \end{bmatrix} = 0.$$  

(4.13)

Let $\Delta B := \mathcal{H}_y \Delta y$, then $\Delta B = V(\Delta B - \Delta \Xi)$. Define $\Delta \tilde{B} := Q^T \Delta B Q$, $\Delta \tilde{\Xi} := Q^T \Delta \Xi Q$. We have the equality

$$\Delta \tilde{B} = \begin{bmatrix} \Delta \tilde{B}_{\alpha \alpha} - \Delta \tilde{\Xi}_{\alpha \alpha} & \Delta \tilde{B}_{\alpha \beta} - \Delta \tilde{\Xi}_{\alpha \beta} & \Theta_{\alpha \gamma} \circ (\Delta \tilde{B}_{\alpha \gamma} - \Delta \tilde{\Xi}_{\alpha \gamma}) \\ \Delta \tilde{B}_{\beta \beta} - \Delta \tilde{\Xi}_{\beta \beta} & V_0(\Delta \tilde{B}_{\beta \beta} - \Delta \tilde{\Xi}_{\beta \beta}) \\ \Theta_{\gamma \gamma} \circ (\Delta \tilde{B}_{\gamma \gamma} - \Delta \tilde{\Xi}_{\gamma \gamma}) & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

where $V_0 \in \partial \Pi_{\mathcal{S}^p}(0)$, $\Theta_{\alpha \gamma} = (\Theta_{ij})_{i \in \alpha, j \in \gamma}$ with

$$\Theta_{ij} = \frac{\lambda_i}{|\lambda_i| + |\lambda_j|}, \quad (i, j) \in \alpha \times \gamma.$$  

(4.14)

Then obviously we have

$$\Delta \tilde{B}_{\gamma \gamma} = 0, \quad \Delta \tilde{B}_{\beta \beta} = 0, \quad \Delta \tilde{\Xi}_{\alpha \alpha} = 0, \quad \Delta \tilde{\Xi}_{\alpha \beta} = 0,$$

and

$$\Delta \tilde{B}_{\gamma \gamma} = V_0(\Delta \tilde{B}_{\gamma \gamma} - \Delta \tilde{\Xi}_{\gamma \gamma}), \quad \Delta \tilde{B}_{\gamma \gamma} = \Theta_{\gamma \gamma} \circ (\Delta \tilde{B}_{\gamma \gamma} - \Delta \tilde{\Xi}_{\gamma \gamma}).$$  

(4.15)
From (4.14) and (4.15) we have that

\[ \Delta \tilde{B}_{ij} = -\frac{\lambda_i}{|\lambda_j|} \Delta \tilde{e}_{ij}, \quad (i, j) \in \alpha \times \gamma. \]

Noting the first equality of (4.15), we have from (iii) of Lemma 2.1 that

\[ \langle \Delta \tilde{B}_{\beta \beta}, -\Delta \tilde{e}_{\beta \beta} \rangle = \langle V_0(\Delta \tilde{B}_{\beta \beta} - \Delta \tilde{e}_{\beta \beta}), (\Delta \tilde{B}_{\beta \beta} - \Delta \tilde{e}_{\beta \beta}) - V_0(\Delta \tilde{B}_{\beta \beta} - \Delta \tilde{e}_{\beta \beta}) \rangle \geq 0. \]

Therefore, we obtain that

\[ \langle \Delta B, -\Delta \Xi \rangle = \langle \Delta \tilde{B}, -\Delta \tilde{e} \rangle \]
\[ = 2 \operatorname{Tr}(\Delta \tilde{B}_{\alpha \gamma} \Delta \tilde{e}_{\alpha \gamma} + \langle \Delta \tilde{B}_{\beta \beta}, -\Delta \tilde{e}_{\beta \beta} \rangle) \]
\[ \geq 2 \operatorname{Tr}(\Delta \tilde{B}_{\alpha \gamma} \Delta \tilde{e}_{\alpha \gamma}) \]
\[ \geq 0. \quad (4.16) \]

The first equality of (4.13) implies

\[ \langle \Delta y, \Delta y \rangle + \langle \Delta y, B^* U B \Delta y \rangle - \langle \Delta y, H^* \Delta \Xi \rangle = 0 \]

or equivalently

\[ -\langle \Delta y, \Delta y \rangle - \langle B \Delta y, U B \Delta y \rangle = \langle H \Delta y, -\Delta \Xi \rangle. \]

Therefore, from (ii) of Lemma 2.1 and (4.16), we have \( \Delta y = 0 \), and (4.13) leads to \( V(\Delta \Xi) = 0, H^* \Delta \Xi = 0 \). From the definition of \( V \) and \( V(\Delta \Xi) = 0 \) we have

\[ Q^T \alpha \Delta \Xi Q_\alpha = 0, \quad Q^T \alpha \Delta \Xi Q_\beta = 0, \quad Q^T \alpha \Delta \Xi Q_\alpha = 0. \quad (4.17) \]

Let the \((i, j)\)-element of \( Q^T \alpha \Delta \Xi Q_\alpha \) be \( \zeta_{ij} \), where \( \alpha = \beta \cup \gamma = \{1, \ldots, t\} \). Namely,

\[ Q^T \alpha \Delta \Xi Q_\alpha = \sum_{1 \leq i < j \leq t} (e_i e_j^T + e_j e_i^T) \zeta_{ij} + \sum_{i=1}^p \zeta_{ii} e_i e_i^T, \]

where \( e_i \) is the \( i \)th unit of \( \mathbb{R}^p \). It follows from (4.17) that

\[ 0 = H^* \Delta \Xi \]
\[ = \left( \begin{array}{c}
\langle H_1, \Delta \Xi \rangle \\
\vdots \\
\langle H_t, \Delta \Xi \rangle
\end{array} \right) \\
= \left( \begin{array}{c}
\langle Q^T H_1 Q, Q^T \Delta \Xi Q \rangle \\
\vdots \\
\langle Q^T H_t Q, Q^T \Delta \Xi Q \rangle
\end{array} \right) \\
= \left( \begin{array}{c}
\langle Q^T \alpha H_1 Q_\alpha, Q^T \alpha \Delta \Xi Q_\alpha \rangle \\
\vdots \\
\langle Q^T \alpha H_t Q_\alpha, Q^T \alpha \Delta \Xi Q_\alpha \rangle
\end{array} \right) \\
= \left( \begin{array}{c}
\sum_{1 \leq i < j \leq t} 2\zeta_{ij} q_i^T H_1 q_j + \sum_{i=1}^p \zeta_{ii} q_i^T H_1 q_i^T \\
\vdots \\
\sum_{1 \leq i < j \leq t} 2\zeta_{ij} q_i^T H_t q_j + \sum_{i=1}^p \zeta_{ii} q_i^T H_t q_i^T
\end{array} \right). \]
Assumption 4.1

Let (4.17) that 

\[ \text{Step 1. Compute } \Theta \text{ (4.17)} \]

Algorithm 4.1 is well defined, we need the following important conclusion which characterizes the nonsingularity of \( J_E \), which implies \( Q^T \Delta \Xi Q = 0 \) and \( \Delta \Xi = 0 \). The nonsingularity of \( W \) is proved.

Then we consider a smoothing Newton method for solving \( F(y, \Xi) = 0 \). Let \( G : \mathbb{R} \times \mathbb{R}^n \times S^p \to \mathbb{R}^n \times S^p \) be a smoothing approximation mapping defined by

\[
G(\epsilon, y, \Xi) = \begin{bmatrix} y - c^0 - \mathcal{H}^* \Xi - \frac{1}{2} \mathcal{B}^* (\tilde{G}(y) + \sqrt{\tilde{G}(y)^2 + \epsilon^2 I}) \\
\frac{1}{2} (\mathcal{H}y + \Xi - \sqrt{(\mathcal{H}y - \Xi)^2 + \epsilon^2 I}) \end{bmatrix}.
\]

Obviously \( \lim_{\epsilon \to 0} G(\epsilon, y, \Xi) = F(y, \Xi) \). The smoothing Newton method is based on solving

\[
E(\epsilon, y, \Xi) := \begin{bmatrix} \epsilon \\
G(\epsilon, y, \Xi) \end{bmatrix} = 0
\]

and uses the merit function \( \phi(Z) := \|E(\epsilon, y, \Xi)\|^2 \) for the line search, where \( Z = (\epsilon, y, \Xi) \). Let \( \tilde{\epsilon} > 0 \) and \( \eta \in (0, 1) \) be such that \( \eta \tilde{\epsilon} < 1 \). Define an auxiliary point \( \tilde{Z} \) by \( \tilde{Z} := (\tilde{\epsilon}, 0, 0) \in \mathbb{R} \times \mathbb{R}^n \times S^p \) and \( \theta : \mathbb{R} \times \mathbb{R}^n \times S^p \to \mathbb{R}_+ \) by \( \theta(Z) := \eta \min \{1, \phi(Z)\} \). The smoothing Newton method, proposed by [12,17], can be described as follows:

**Algorithm 4.1.**

1. **Step 1.** Select constants \( \delta \in (0, 1) \) and \( \sigma \in (0, 1/2) \). Let \( \epsilon^0 := \tilde{\epsilon}, (y^0, \Xi^0) \in \mathbb{R}^n \times S^p \) be an arbitrary point. Then let the initial point \( Z^0 = (\epsilon^0, y^0, \Xi^0) \) and \( k = 0 \).

2. **Step 2.** If \( E(Z^k) = 0 \), then stop; otherwise, let \( \theta_k := \theta(Z^k) \).

3. **Step 3.** Compute \( \Delta Z^k := (\Delta \epsilon^k, \Delta y^k, \Delta \Xi^k) \in \mathbb{R} \times \mathbb{R}^n \times S^p \) by

\[
E(Z^k) + J E(Z^k)(\Delta Z^k) = \theta_k \tilde{Z}.
\]

4. **Step 4.** Let \( l_k \) be the smallest nonnegative integer \( l \) satisfying

\[
\phi(\tilde{Z}^k + \delta^l \Delta Z^k) \leq (1 - 2\sigma (1 - \eta \tilde{\epsilon})^l) \phi(Z^k).
\]

Define \( Z^{k+1} = Z^k + \delta^l \Delta Z^k \).

5. **Step 5.** \( k := k + 1 \), go to Step 2.

To show that Algorithm 4.1 is well defined, we need the following important conclusion which characterizes the nonsingularity of \( J E(\epsilon, y, \Xi) \).

**Proposition 4.2.** For \( \epsilon \neq 0 \) and any \( (y, \Xi) \in \mathbb{R}^n \times S^p \), the F-derivative \( J E(\epsilon, y, \Xi) \) is nonsingular.

**Proof.** The smoothing function \( \Phi \) is defined by (2.4). Assume that for any \( (\Delta \epsilon, \Delta y, \Delta \Xi) \in \mathbb{R} \times \mathbb{R}^n \times S^p \), we have

\[
J E(\epsilon, y, \Xi)(\Delta \epsilon, \Delta y, \Delta \Xi) = 0,
\]

namely

\[
\begin{bmatrix} \Delta \epsilon \\
\Delta y - \mathcal{H}^* \Delta \Xi + \frac{1}{2} \mathcal{B}^* B \Delta y - \frac{1}{2} \mathcal{B}^* J \Phi(\epsilon, \tilde{G}(y))(\Delta \epsilon, -B \Delta y) \\
\frac{1}{2} [\mathcal{H} \Delta y + \Delta \Xi - J \Phi(\epsilon, \mathcal{H} y - \Xi)(\Delta \epsilon, \mathcal{H} \Delta y - \Delta \Xi)] \end{bmatrix} = 0,
\]

where

\[
\tilde{\zeta}_{ij} = \begin{cases} 2\zeta_{ij}, & i \neq j, \\
\zeta_{ii}, & i = j. \end{cases}
\]

It follows from Assumption 4.1 that \( \zeta_{ij} = 0, 1 \leq i \leq j \leq t \), which implies \( Q^T \Delta \Xi Q = 0 \). Therefore, combining with (4.17), we have \( Q^T \Delta \Xi Q = 0 \) and \( \Delta \Xi = 0 \). The nonsingularity of \( W \) is proved.
Lemma 2.4

(4.20) can be rewritten as

\[ \Delta e = 0 \]

\[ \Delta y - \mathcal{H}^* \Delta \Xi = \frac{1}{2} B^* J \Phi(\epsilon, \tilde{G}(y))(0, -B \Delta y) - \frac{1}{2} B^* B \Delta y \]

\[ \mathcal{H} \Delta y + \Delta \Xi = L_{\Phi(\epsilon, \mathcal{H} y - \Xi)}^{-1} (L(\mathcal{H} y - \Xi) (\mathcal{H} \Delta y - \Delta \Xi)). \]

Assume that \((\mathcal{H} y - \Xi)\) has the spectral decomposition as follows

\[ \mathcal{H} y - \Xi = \mathcal{Q} \Lambda_1 \mathcal{Q}^T, \]

where \(\lambda_i, i = 1, \ldots, p\) are the eigenvalues of \((\mathcal{H} y - \Xi)\) and \(\Lambda_1 = \text{diag}(\lambda_1, \ldots, \lambda_p)\). Let \(\Delta T = \mathcal{H} \Delta y\),

\[ \tilde{\Delta T} = \mathcal{Q}^T \Delta T \mathcal{Q}, \]

\[ \Delta \tilde{\Xi} = \mathcal{Q}^T \Delta \Xi \]

and \(\Omega_{\epsilon} \in \mathbb{R}^{p \times p}\) be defined by

\[ (\Omega_{\epsilon})_{ij} = -\frac{\lambda_i + \lambda_j + \sqrt{\lambda_i^2 + \epsilon^2 + \lambda_j^2 + \epsilon^2}}{\lambda_i - \lambda_j + \sqrt{\lambda_i^2 + \epsilon^2 + \lambda_j^2 + \epsilon^2}}, \quad \text{for } i, j = 1, \ldots, p. \]

Obviously \((\Omega_{\epsilon})_{ij} < 0, i, j = 1, \ldots, p\) when \(\epsilon \neq 0\).

Suppose there exist an orthogonal matrix \(P\) and a matrix \(A_2 = \text{diag}(\mu_1, \ldots, \mu_n)\) of eigenvalues of \(\tilde{G}(y)\) such that \(\tilde{G}(y) = P A_2 P^T\), then from Lemma 2.4, we have for \(i, j = 1, \ldots, n\),

\[ (P^T J \Phi(\epsilon, \tilde{G}(y))(0, -B \Delta y) P)_{ij} = \begin{cases} 
-\frac{(\mu_i + \mu_j)}{(\epsilon^2 + \mu_i^2)^{1/2} + (\epsilon^2 + \mu_j^2)^{1/2}}, & \text{if } i \neq j, \\
-\frac{\mu_i}{(\epsilon^2 + \mu_i^2)^{1/2}}, & \text{if } i = j. 
\end{cases} \]

Define \(T \in \mathbb{S}^n\) by

\[ T_{ij} = \begin{cases} 
\frac{-\mu_i + \mu_j}{(\epsilon^2 + \mu_i^2)^{1/2} + (\epsilon^2 + \mu_j^2)^{1/2}}, & \text{if } i \neq j, \\
\frac{\mu_i}{(\epsilon^2 + \mu_i^2)^{1/2}}, & \text{if } i = j. 
\end{cases} \]

then \(P^T J \Phi(\epsilon, \tilde{G}(y))(0, -B \Delta y) P = T \circ (P^T B \Delta y P)\). As \(T_{ij} < 1, i, j = 1, \ldots, n\), we have that

\[ (P^T B \Delta y P, P^T J \Phi(\epsilon, \tilde{G}(y))(0, -B \Delta y) P) \leq \|B \Delta y\|_T^2. \]  

(4.21)

The third equation of (4.20) can be rewritten as

\[ \Phi(\epsilon, \mathcal{H} y - \Xi)(\mathcal{H} \Delta y + \Delta \Xi) + (\mathcal{H} \Delta y + \Delta \Xi) \Phi(\epsilon, \mathcal{H} y - \Xi) = (\mathcal{H} y - \Xi)(\mathcal{H} \Delta y - \Delta \Xi) + (\mathcal{H} \Delta y - \Delta \Xi)(\mathcal{H} y - \Xi) \]

or equivalently as

\[ (\lambda^2 + \epsilon^2 I)^{1/2}(\Delta \tilde{T} + \Delta \tilde{\Xi}) + (\Delta \tilde{T} + \Delta \tilde{\Xi})(\lambda^2 + \epsilon^2 I)^{1/2} = \lambda(\Delta \tilde{T} - \Delta \tilde{\Xi}) + (\Delta \tilde{T} - \Delta \tilde{\Xi}) \lambda, \]

which can be expressed as \(\Delta \tilde{T} = \Omega_{\epsilon} \circ \Delta \tilde{\Xi}\).

From the second equation of (4.20) and (4.21) we have

\[ (\Delta y, \Delta y - \mathcal{H}^* \Delta \Xi) = (\Delta y, B^* J \Phi(\epsilon, \tilde{G}(y))(0, -B \Delta y) - B^* B \Delta y) = (B \Delta y, J \Phi(\epsilon, \tilde{G}(y))(0, -B \Delta y) - B \Delta y) \]

\[ = (P^T B \Delta y P, P^T J \Phi(\epsilon, \tilde{G}(y))(0, -B \Delta y) P - B \Delta y) \leq 0 \]

and

\[ \Lambda \leq 0 \]

(4.22)
\begin{align*}
\langle \Delta y, \Delta y - \mathcal{H}^e \Delta \Xi \rangle &= \langle \Delta y, \Delta y \rangle - \langle \mathcal{H} \Delta y, \Delta \Xi \rangle \\
&= \langle \Delta y, \Delta y \rangle - \langle \Delta \tilde{T}, \Delta \tilde{\Xi} \rangle \\
&= \| \Delta y \|^2 - \langle \Omega_0 \circ \Delta \tilde{\Xi}, \Delta \tilde{\Xi} \rangle \\
&\geq 0. \quad (4.23)
\end{align*}

We have from (4.22) and (4.23) that $\Delta y = 0$ and $\Delta \Xi = 0$. Therefore $J(E, y, \Xi)$ is nonsingular. \qed

The following theorem gives the global convergence for Algorithm 4.1.

**Theorem 4.1.** As Algorithm 4.1 is well defined, the sequence $\{Z^k\}$ generated by Algorithm 4.1 converges to the solution of $E(Z) = 0$.

**Proof.** From (4.19), we know that

$$\infty > \sum_{k=0}^{\infty} (\phi(Z^k) - \phi(Z^{k+1})) \geq \sum_{k=0}^{\infty} (2\sigma (1 - \eta \bar{\varepsilon}) \delta^l) \phi(Z^k)$$

and $\phi(Z^k)$ is strict monotone decreasing, then we have $\{\phi(Z^k)\}$ converges to 0, which, together with the fact that $E(Z) = 0$ has a unique solution, implies $\{Z^k\}$ converges to the solution of $E(Z) = 0$. The proof is completed. \qed

We now state the quadratic convergence of Algorithm 4.1 in the following theorem. It is immediate from [17, Theorem 4.2].

**Theorem 4.2.** Let Assumption 4.1 be satisfied. Since the sequence $\{Z^k\}$ generated by Algorithm 4.1 converges to the solution of $E(Z) = 0$ and each element of $\delta F(Z^*)$ is nonsingular, then

$$\| Z^{k+1} - Z^* \| = O(\| Z^k - Z^* \|)$$

and

$$e^{k+1} = O((\bar{\varepsilon}^k)^2).$$

### 5. Numerical experiments

In this section, we report our numerical experiments of Algorithm 4.1 carried out in Matlab(R2007a) running on a PC Intel Pentium IV of 2.80 GHz CPU. In Step 3, as $J(E(Z))$ is nonsymmetric and its explicit form is complicated, we use CGS method (conjugate gradient square method) [14] to solve (4.18). We test the following class of problems:

**Problem 5.1.** Let $G^0$ and $e^0$ be a random $n \times n$ symmetric matrix and a random $n \times 1$ vector, respectively. $H_1, H_2, \ldots, H_n$ are $n$ random $p \times p$ symmetric matrices. For convenience, we set the elements of $x^0$ all 1.

In our numerical experiments, we choose the initial point $y^0$ with entries all zero and $\Xi^0$ as an identity matrix. The stopping criterion is Tol. := $\Phi(Z^k) < 10^{-5}$. We set other parameters in the algorithm as $\eta = 0.5, \sigma = 0.3, \delta = 0.5$.

### Table 1

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>10</td>
<td>1.4 s</td>
<td>5</td>
<td>22</td>
<td>$3.41 \times 10^2$</td>
<td>$1.12 \times 10^{-10}$</td>
</tr>
<tr>
<td>50</td>
<td>20</td>
<td>54.2 s</td>
<td>6</td>
<td>37</td>
<td>$3.20 \times 10^4$</td>
<td>$9.90 \times 10^{-8}$</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>2 m 24.5s</td>
<td>6</td>
<td>21</td>
<td>$3.35 \times 10^4$</td>
<td>$2.53 \times 10^{-12}$</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>1 m 21.0s</td>
<td>6</td>
<td>16</td>
<td>$3.28 \times 10^4$</td>
<td>$8.54 \times 10^{-6}$</td>
</tr>
<tr>
<td>50</td>
<td>500</td>
<td>12 m 48.8 s</td>
<td>2</td>
<td>3</td>
<td>$3.37 \times 10^4$</td>
<td>$8.00 \times 10^{-4}$</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>5 m 28.1s</td>
<td>32</td>
<td>172</td>
<td>$2.58 \times 10^5$</td>
<td>$5.04 \times 10^{-7}$</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>7 m 25.1s</td>
<td>7</td>
<td>29</td>
<td>$2.57 \times 10^5$</td>
<td>$5.25 \times 10^{-8}$</td>
</tr>
<tr>
<td>100</td>
<td>500</td>
<td>13 m 53.0 s</td>
<td>2</td>
<td>3</td>
<td>$2.56 \times 10^5$</td>
<td>$3.65 \times 10^{-10}$</td>
</tr>
<tr>
<td>100</td>
<td>1000</td>
<td>1 h 1 m 56.4 s</td>
<td>2</td>
<td>3</td>
<td>$2.54 \times 10^5$</td>
<td>$1.10 \times 10^{-14}$</td>
</tr>
</tbody>
</table>
Our numerical results are reported in Table 1, where Iter., Func., Res0, and Res* stand for, respectively, the number of iterations, the number of function evaluations, the residual $\Phi(\cdot)$ at the starting point and the residual $\Phi(\cdot)$ at the final iterate of implementation.

The numerical results reported in Table 1 indicate that our approach is highly effective. For example, when testing the problem with $n = 100$, $p = 1000$, we should solve a linear equation of almost $5 \times 10^5$ unknowns at each step. Considering both the scale of the problem and the cpu time spent, we think the result is satisfying.

Acknowledgements

The research of Xiantao Xiao and Liwei Zhang is supported by the National Natural Science Foundation of China under project No. 10771026 and by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry of China. The research of Jianzhong Zhang is supported by the University Grant Council of Hong Kong under the grant CERG CityU 9041091 and CUHK 103105.

References