Calibrating Least Squares Covariance Matrix Problems with Equality and Inequality Constraints

Yan Gao* and Defeng Sun†

June 12, 2008

Abstract

In many applications in finance, insurance, and reinsurance, one seeks a solution of finding a covariance matrix satisfying a large number of given linear equality and inequality constraints in a way that it deviates the least from a given symmetric matrix. One difficulty in finding an efficient method for solving this problem is due to the presence of the inequality constraints. In this paper, we propose to overcome this difficulty by reformulating the problem as a system of semismooth equations with two level metric projection operators. We then design an inexact smoothing Newton method to solve the resulted semismooth system. At each iteration, we use the BiCGStab iterative solver to obtain an approximate solution to the generated smoothing Newton linear system. Our numerical experiments confirm the high efficiency of the proposed method.

Key words: covariance matrix, smoothing Newton method, quadratic convergence.

AMS subject classifications: 49M29, 90C25, 90C30

1 Introduction

In this paper, we are interested in the following least squares covariance matrix (LSCM) problem

$$\begin{array}{ll} \min & \frac{1}{2}\|X-C\|^2 \\ \text{s.t.} & \langle A_i, X\rangle = b_i, \quad i=1,\ldots,p\,, \\ & \langle A_i, X\rangle \geq b_i, \quad i=p+1,\ldots,m\,, \\ & X \in \mathcal{S}^n_+\,, \end{array} \tag{1}$$

*Department of Mathematics, National University of Singapore, 2 Science Drive 2, Republic of Singapore. Email: yangao@nus.edu.sg.

†Department of Mathematics and Risk Management Institute, National University of Singapore, 2 Science Drive 2, Republic of Singapore. The author's research was supported by the Academic Research Fund under Grant R-146-000-104-112 and the Risk Management Institute under Grants R-703-000-004-720 and R-703-000-004-646, National University of Singapore. Email: matsundf@nus.edu.sg.

where $\mathcal{S}^n$ and $\mathcal{S}^n_+$ are, respectively, the space of $n \times n$ symmetric matrices and the cone of positive semidefinite matrices in $\mathcal{S}^n$, $\|\cdot\|$ is the Frobenius norm induced by the standard trace inner product $\langle \cdot, \cdot \rangle$ in $\mathcal{S}^n$, C and A_i, $i = 1, \ldots, m$ are given matrices in $\mathcal{S}^n$, and $b \in \mathbb{R}^m$.

Mathematically, the LSCM problem (1) can be equivalently written as

$$\begin{array}{ll} \min & t \\ \text{s.t.} & \langle A_i, X \rangle = b_i, \quad i = 1, \ldots, p\,, \\ & \langle A_i, X \rangle \geq b_i, \quad i = p+1, \ldots, m\,, \\ & t + 1 \geq \sqrt{(t-1)^2 + 2\|X - C\|^2}\,, \\ & X \in \mathcal{S}^n_+\,. \end{array} \tag{2}$$

Problem (2) is a linear optimization problem with linear equality/inequality, the second order cone, and the positive semidefinite cone constraints. This suggests that one may then use well developed and publicly available softwares, based on interior point methods (IPMs), such as SeDuMi [44], SDPT3 [48], and a few others to solve (2), and so the LSCM problem (1), directly. This is indeed feasible on a Pentium IV PC (the computing machine that we will use in our numerical experiments) as long as n is small (say 80 at most) and m is not too large (say $5,000$). The reason is that at each iteration these solvers require to formulate and solve a linear system with a dense Schur complement matrix (for example, see [6]) of the size $(m+1+\bar{n}) \times (m+1+\bar{n})$, where $\bar{n} := \frac{1}{2}n(n+1)$.

Realizing the difficulties in using IPMs to solve the LSCM problem, in two recent papers, Malick [29] and Boyd and Xiao [9] proposed, respectively, to apply classical quasi-Newton methods (in particular, the BFGS method) and the projected gradient method to the Lagrangian dual of problem (1) as the objective function in the corresponding Lagrangian dual (dual in short) problem is continuously differentiable. Unlike the IPMs, these two dual based approaches are relatively inexpensive at each iteration as the dual problem is of dimension m only. The overall numerical performance of these two approaches vary from problem to problem. They may take dozens of iterations for some testing examples and several hundreds or thousands for some others.

Historically, the Lagrangian dual based approach is known to the optimization and approximation theory communities for a long time and has been discussed extensively during the last three decades. Rockafellar's monograph [42] is an excellent source to start with. The LSCM problem (1) is a special case of the best approximation problem [14]

$$\begin{array}{ll} \min & \frac{1}{2}\|x - c\|^2 \\ \text{s.t.} & \mathcal{A}x \in b + Q\,, \\ & x \in K\,, \end{array} \tag{3}$$

where $\mathcal{X}$ is a real Hilbert space equipped with a scalar product $\langle \cdot, \cdot \rangle$ and its induced norm $\|\cdot\|$, $\mathcal{A} : \mathcal{X} \to \mathbb{R}^m$ is a bounded linear operator, $Q = \{0\}^p \times \mathbb{R}^q_+$ is a polyhedral convex cone, $1 \leq p \leq m$, $q = m - p$, and K is a closed convex cone in $\mathcal{X}$. See Deutsch [14] for a comprehensive treatment on the best approximation problems in inner product spaces. The Lagrangian dual

(see, e.g., Borwein and Lewis [8]) of the best approximation problem (3) takes the form

$$\begin{array}{ll} \max & -\frac{1}{2}\|\Pi_K(c+\mathcal{A}^*y)\|^2 + \langle b, y\rangle + \frac{1}{2}\|c\|^2 \\ \text{s.t.} & y \in Q^+ , \end{array} \tag{4}$$

where $\mathcal{A}^* : \mathbb{R}^m \to \mathcal{X}$ is the adjoint of $\mathcal{A}$, $Q^+ = \mathbb{R}^p \times \mathbb{R}^q_+$ is the dual cone of Q, and for any $x \in \mathcal{X}$, $\Pi_K(x)$ is the metric projection of x onto K, i.e., $\Pi_K(x)$ is the unique optimal solution to

$$\begin{array}{ll} \min & \frac{1}{2}\|u - x\|^2 \\ \text{s.t.} & u \in K . \end{array}$$

Define $\theta : \mathbb{R}^m \to \mathbb{R}$ by

$$\theta(y) := \frac{1}{2}\|\Pi_K(c+\mathcal{A}^*y)\|^2 - \langle b, y\rangle - \frac{1}{2}\|c\|^2, \quad y \in \mathbb{R}^m . \tag{5}$$

Note that $-\theta(\cdot)$ is the objective function in problem (4). This definition of $\theta(\cdot)$ is for the convenience of subsequent discussions. Such a defined function $\theta(\cdot)$ is a convex function [42, 8]. From Zarantonello [50], we know that $\theta(\cdot)$ is continuously differentiable (but not twice continuously differentiable) with

$$\nabla\theta(y) = \mathcal{A}\Pi_K(c+\mathcal{A}^*y) - b, \quad y \in \mathbb{R}^m$$

and that $\nabla\theta(\cdot)$ is Lipschitz continuous. Thus, the dual problem (4) turns out to be a smooth convex optimization problem with a simple constraint:

$$\begin{array}{ll} \min & \theta(y) \\ \text{s.t.} & y \in Q^+ . \end{array} \tag{6}$$

Due to its wide applications in interpolation, the approximation theory community paid a lot of attentions in the 80's and 90's of the last century to the best approximation problem (3) when the inequality constraints are absent (i.e., $q = 0$). See, for examples, [2, 16, 26, 30, 31], to name only a few. For a summary, see [32]. In the case when $q = 0$, Micchelli and Utreras in their influential paper [31] suggested that a quasi-Newton method be used to solve the dual problem (6), which is an unconstrained convex optimization problem. In [15], Deutsch, Li, and Ward introduced a steepest descent method for solving the same problem. One particular well-studied best approximation problem is the convex best interpolation problem, where $K = \{x \in L_2[0,1] \,|\, x \geq 0 \text{ a.e. on } [0,1]\}$. For the dual of the convex best interpolation problem, Newton's method appears to be very efficient [26, 2, 16] despite the fact that $\theta(\cdot)$ may not be twice differentiable. By using the superlinear (quadratic) convergence theory of Clarke's generalized Jacobian based Newton methods for solving semismooth equations (an important subclass of nonsmooth equations) established by Kummer [27] and Qi and Sun [38], Dontchev, Qi, and Qi solved the myth on the performance of Newton's method for solving the dual of the convex best approximation problem in [17, 18].

Since the metric projection operator $\Pi_{\mathcal{S}^n_+}(\cdot)$ over the cone $\mathcal{S}^n_+$ has been proved to be strongly semismooth in [45], the effectiveness of Newton's method for the convex best interpolation problem inspired Qi and Sun [34] to study a quadratically convergent inexact semismooth Newton method to solve the following nearest correlation matrix problem under the W-weight (after an equivalent transformation):

$$\begin{array}{ll} \min & \dfrac{1}{2}\|X - C\|^2 \\ \text{s.t.} & (W^{-1/2}XW^{-1/2})_{ii} = 1,\ i = 1, \ldots, n\,, \\ & X \in \mathcal{S}^n_+\,, \end{array} \tag{7}$$

where the positive definite matrix $W \in \mathcal{S}^n$ is given as the W-weight to the problem and $W^{1/2}$ is the positive square root of W. The nearest correlation matrix problem (7) mainly comes from the finance and insurance/reinsurance industries. Higham first formulated this problem in [24] and considered to apply Dysktra's alternating projection algorithm [19] to solve it. Numerical experiments conducted in [34, 7] demonstrate clearly that Newton's method outperforms the alternating projection method and the BFGS method. In Section 5, we can see that Newton's method combined with a conjugate gradient (CG) iterative solver is much faster than the alternating projection method and the BFGS method when W is a randomly generated diagonal matrix.

The inexact semismooth Newton method introduced in [34] can certainly be used to solve, at least conceptually, the LSCM problem (1) in the absence of inequality constraints (i.e., $m = p$), but is not applicable to the cases with inequality constraints (i.e., $m > p$). The reason is that whenever there are inequality constraints in (1), its corresponding dual problem in (6) with $K = \mathcal{S}^n_+$ is no longer an unconstrained convex optimization problem. One the other hand, the projected gradient method used by Boyd and Xiao [9] can converge at best linearly while the rate of convergence of a BFGS type method suggested by Malick [29] is still an open question due to the fact that $\theta(\cdot)$ fails to be twice continuously differentiable (for a discussion on the convergence of BFGS type methods for problems of such a nature, see Chen [11]). Thus, a natural question arises: can one still expect a fast convergent numerical method for solving the least squares covariance matrix problem (1) with both equality and inequality constraints? This paper will give an affirmative answer to this question by introducing a quadratically convergent inexact smoothing Newton method.

Just as the case for semsimooth Newton methods, smoothing Newton methods for solving nonsmooth equations were mainly developed by the optimization community, in particular, the complementarity community (see [21] for an introduction on complementarity problems and on smoothing Newton methods), and have not been well received outside the optimization field. In this paper, we shall take the LSCM problem (1) as an example to introduce smoothing Newton methods to the numerical linear algebra community.

The paper is organized as follows. In Section 2, we present preliminaries on matrix valued functions. Section 3 presents a general introduction on an inexact smoothing Newton method for solving nonsmooth equations and its convergence analysis. In section 4, we apply the introduced inexact smoothing Newton method to the LSCM problem (1). We report our numerical results in Section 5 and make our final conclusions in Section 6.

2 Preliminaries on matrix-valued functions

For subsequent discussions, in this section we introduce some basic properties of matrix valued functions related to the LSCM problem (1) and its dual.

Let $\mathcal{F}$ denote the feasible set of of problem (1). Assume that $\mathcal{F} \neq \emptyset$. Then problem (1) has a unique optimal solution $\overline{X}$. Let $q = m - p$ and $Q = \{0\}^p \times \mathbb{R}^q_+$. Denote $\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^m$ by

$$\mathcal{A}(X) := \begin{bmatrix} \langle A_1, X\rangle \\ \vdots \\ \langle A_m, X\rangle \end{bmatrix}, \quad X \in \mathcal{S}^n\,.$$

For any symmetric $X \in \mathcal{S}^n$, we write $X \succeq 0$ and $X \succ$ to represent that X is positive semidefinite and positive definite, respectively. Then

$$\mathcal{F} = \{X \in \mathcal{S}^n \,|\, \mathcal{A}(X) \in b + Q,\ X \succeq 0\}$$

and the dual problem of (1) takes the form

$$\begin{array}{ll} \min & \theta(y) := \dfrac{1}{2}\|\Pi_{\mathcal{S}^n_+}(C + \mathcal{A}^* y)\|^2 - \langle b, y\rangle - \dfrac{1}{2}\|C\|^2 \\ \text{s.t.} & y \in Q^+ = \mathbb{R}^p \times \mathbb{R}^q_+\,. \end{array} \tag{8}$$

The objective function $\theta(\cdot)$ in (8) is a continuously differentiable convex function with

$$\nabla\theta(y) = \mathcal{A}\Pi_{\mathcal{S}^n_+}(C + \mathcal{A}^* y) - b, \quad y \in \mathbb{R}^m\,,$$

where the adjoint $\mathcal{A}^* : \mathbb{R}^m \to \mathcal{S}^n$ takes the form

$$\mathcal{A}^*(y) = \sum_{i=1}^m y_i A_i, \quad y \in \mathbb{R}^m\,. \tag{9}$$

One classical dual approach described by Rockafellar in [42, Page 4], when specialized to problem (1), is to first find an optimal solution $\bar{y}$, if it exists, to the dual problem (8), and then to obtain the unique optimal solution $\overline{X}$ to problem (1) via $\overline{X} = \Pi_{\mathcal{S}^n_+}(C + \mathcal{A}^*\bar{y})$. See Malick [29] and Boyd and Xiao [9] for the worked out details.

In order to apply an optimization method to solve (8), we need the following generalized Slater condition to hold for problem (1):

$$\begin{cases} \{A_i\}_{i=1}^p \text{ are linearly independent,} \\ \exists\ X^0 \in \mathcal{F} \text{ such that } \ \langle A_i, X^0\rangle > b_i,\ i = p+1, \ldots, m \text{ and } X^0 \succ 0\,. \end{cases} \tag{10}$$

The next proposition is a straightforward application of [42, Theorems 17 & 18].

Proposition 2.1 *Under the generalized Slater condition (10), the following hold:*

(i) *There exists at least one $\bar{y} \in Q^+$ that solves the dual problem (8). The unique solution to problem (1) is given by*

$$\overline{X} = \Pi_{\mathcal{S}^n_+}(C + \mathcal{A}^*\bar{y}). \tag{11}$$

(ii) *For every real number τ, the constrained level set $\{y \in Q^+ \,|\, \theta(y) \leq \tau\}$ is closed, bounded, and convex.*

Proposition 2.1 says that one should be able to use any gradient based optimization method to find an optimal solution to the convex problem (8), and thus solves problem (1), as long as the generalized Slater condition (10) holds. Note that for any given $y \in \mathbb{R}^m$, both $\theta(y)$ and $\nabla\theta(y)$ can be computed explicitly as the metric projector $\Pi_{\mathcal{S}^n_+}(\cdot)$ has long been known by statisticians to admit an analytic formula [43].

Since $\theta(\cdot)$ is a convex function, $\bar{y} \in Q^+$ solves problem (8) if and only if it satisfies the following variational inequality

$$\langle y - \bar{y}, \nabla\theta(\bar{y})\rangle \geq 0 \quad \forall\, y \in Q^+ \,. \tag{12}$$

Define $F : \mathbb{R}^m \to \mathbb{R}^m$ by

$$F(y) := y - \Pi_{Q^+}(y - \nabla\theta(y)), \quad y \in \mathbb{R}^m \,. \tag{13}$$

Then one can easily check that $\bar{y} \in Q^+$ solves (12) if and only if $F(\bar{y}) = 0$ [20]. Thus, solving the dual problem (8) is equivalent to solving the following equation

$$F(y) = 0 \,, \quad y \in \mathbb{R}^m \,. \tag{14}$$

Since both Π_{Q^+} and $\Pi_{\mathcal{S}^n_+}$ are globally Lipschitz continuous, F is globally Lipschitz continuous. This means that though one cannot use classical Newton method to solve (14), one can still use Clarke's generalized Jacobian based Newton methods [27, 36, 38]. The difficulty is that these Clarke's Jacobian based Newton methods cannot be globalized because $F(\cdot)$ is not the gradient mapping of any real valued function. In this paper, we shall introduce an inexact smoothing Newton method to overcome this difficulty. For this purpose, we need smoothing functions for $F(\cdot)$.

Next, we shall first discuss smoothing functions for the metric projector $\Pi_{\mathcal{S}^n_+}(\cdot)$. Let $X \in \mathcal{S}^n$. Suppose that X has the spectral decomposition

$$X = P\mathrm{diag}(\lambda_1, \ldots, \lambda_n)P^T, \tag{15}$$

where $\lambda_1 \geq \cdots \geq \lambda_n$ are the eigenvalues of X and P is a corresponding orthogonal matrix of orthonormal eigenvectors of X. Then, from [43],

$$\Pi_{\mathcal{S}^n_+}(X) = P\mathrm{diag}(\max(0, \lambda_1), \ldots, \max(0, \lambda_n))P^T \,. \tag{16}$$

Define

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

Write $P = [P_\alpha \; P_\beta \; P_\gamma]$ with P_α, P_β, and P_γ containing the columns in P indexed by α, β, and γ, respectively. Let $\phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be defined by

$$\phi(\varepsilon, t) = \big[\, t + \sqrt{\varepsilon^2 + t^2}\,\big]/2, \quad (\varepsilon, t) \in \mathbb{R} \times \mathbb{R} \,. \tag{17}$$

For any $\varepsilon \in \mathbb{R}$, let

$$\Phi(\varepsilon, X) := P \begin{bmatrix} \phi(\varepsilon, \lambda_1) & & \\ & \ddots & \\ & & \phi(\varepsilon, \lambda_n) \end{bmatrix} P^T . \tag{18}$$

Then, by matrix analysis [4, 25], we have

$$\Phi(\varepsilon, X) = \big[\, X + \sqrt{\varepsilon^2 I + X^2}\,\big]/2,$$

where we use I to represent the identity matrix of appropriate dimension. Note that when $\varepsilon = 0$, $\Phi(0, X) = \Pi_{\mathcal{S}^n_+}(X)$. By a famous result of Löwner [28], we know that when $\varepsilon \neq 0$ or $\beta = \emptyset$,

$$\Phi'_X(\varepsilon, X)(H) = P[\Omega(\varepsilon, \lambda) \circ (P^T H P)]P^T \quad \forall\, H \in \mathcal{S}^n , \tag{19}$$

where "$\circ$" denotes the Hadamard product, $\lambda = (\lambda_1, \ldots, \lambda_n)^T$, and the symmetric matrix $\Omega(\varepsilon, \lambda)$ is given by

$$\big[\,\Omega(\varepsilon, \lambda)\,\big]_{ij} = \begin{cases} \dfrac{\phi(\varepsilon, \lambda_i) - \phi(\varepsilon, \lambda_j)}{\lambda_i - \lambda_j} \in [0, 1] & \text{if } \lambda_i \neq \lambda_j , \\ \phi'_{\lambda_i}(\varepsilon, \lambda_i) \in [0, 1] & \text{if } \lambda_i = \lambda_j , \end{cases} \quad i, j = 1, \ldots, n . \tag{20}$$

When $\varepsilon \neq 0$ or $\beta = \emptyset$, the partial derivative of $\Phi(\cdot, \cdot)$ with respect to ε can be computed by

$$\Phi'_\varepsilon(\varepsilon, X) = P \mathrm{diag}(\phi'_\varepsilon(\varepsilon, \lambda_1), \cdots, \phi'_\varepsilon(\varepsilon, \lambda_n)) P^T .$$

Thus, $\Phi(\cdot, \cdot)$ is continuously differentiable around $(\varepsilon, X) \in \mathbb{R} \times \mathcal{S}^n$ if $\varepsilon^2 I + X^2$ is nonsingular. Furthermore, $\Phi(\cdot, \cdot)$ is globally Lipschitz continuous and strongly semismooth at any $(0, X) \in \mathbb{R} \times \mathcal{S}^n$ [46]. In particular, for any $\varepsilon \downarrow 0$ and $\mathcal{S}^n \ni H \to 0$, it holds that

$$\Phi(\varepsilon, X + H) - \Phi(0, X) - \Phi'(\varepsilon, X + H)(\varepsilon, H) = O(\|(\varepsilon, H)\|^2) . \tag{21}$$

Recall that for a locally Lipschitz continuous function Γ from a finite dimensional real Hilbert space $\mathcal{X}$ to $\mathbb{R}^n$, the B-subdifferential of Γ at $x \in \mathcal{X}$ in the sense of Qi [36] is defined by

$$\partial_B \Gamma(x) := \{V \,|\, V = \lim_{k \to \infty} \Gamma'(x^k),\ x^k \to x,\ x_k \in \mathcal{D}_\Gamma\} ,$$

where $\mathcal{D}_\Gamma$ is the set of points where Γ is Fréchet differentiable. The generalized Jacobian $\partial \Gamma(x)$ of Γ at x in the sense of Clarke [12] is just the convex hull of $\partial_B \Gamma(x)$.

Define $\Phi_{|\beta|} : \mathbb{R} \times \mathcal{S}^{|\beta|} \to \mathcal{S}^{|\beta|}$ by

$$\Phi_{|\beta|}(\varepsilon, Z) := \big[\, Z + \sqrt{\varepsilon^2 I + Z^2}\,\big]/2, \quad (\varepsilon, Z) \in \mathbb{R} \times \mathcal{S}^{|\beta|}. \tag{22}$$

As the case for $\Phi(\cdot, \cdot)$, the mapping $\Phi_{|\beta|}(\cdot, \cdot)$ is also Lipschitz continuous. Then the B-subdifferentials $\partial_B \Phi(0, X)$ of Φ at $(0, X)$ and $\partial_B \Phi_{|\beta|}(0, Z)$ of $\Phi_{|\beta|}$ at $(0, Z) \in \mathbb{R} \times \mathcal{S}^{|\beta|}$ in the sense of Qi [36] are both well defined and are completely characterized in [10].

The following result is proven in [10, Proposition 5].

Proposition 2.2 *Suppose that $X \in \mathcal{S}^n$ has the spectral decomposition as in* (15)*. Then $V \in \partial_B \Phi(0, X)$ if and only if there exists $V_{|\beta|} \in \partial_B \Phi_{|\beta|}(0, 0)$ such that for all $(\varepsilon, H) \in \mathbb{R} \times \mathcal{S}^n$,*

$$V(\varepsilon, H) = P \begin{bmatrix} P_\alpha^T H P_\alpha & P_\alpha^T H P_\beta & U_{\alpha\gamma} \circ (P_\alpha^T H P_\gamma) \\ (P_\alpha^T H P_\beta)^T & V_{|\beta|}(\varepsilon, P_\beta^T H P_\beta) & 0 \\ (P_\alpha^T H P_\gamma)^T \circ U_{\alpha\gamma}^T & 0 & 0 \end{bmatrix} P^T , \tag{23}$$

where $U \in \mathcal{S}^n$ is defined by

$$U_{ij} := \frac{\max\{\lambda_i, 0\} + \max\{\lambda_j, 0\}}{|\,\lambda_i\,| + |\,\lambda_j\,|}, \quad i, j = 1, \ldots, n, \tag{24}$$

where $0/0$ is defined to be 1.

In order to define smoothing functions for $F(\cdot)$, we need to define smoothing functions for $\Pi_{Q^+}(\cdot)$. This, however, can be done in many different ways. For simplicity, we shall only use the function ϕ given by (17) to define a smoothing function for $\Pi_{Q^+}(\cdot)$. Let $\psi : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$ be defined by

$$\psi_i(\varepsilon, z) = \begin{cases} z_i & \text{if } i = 1, \ldots, p, \\ \phi(\varepsilon, z_i) & \text{if } i = p+1, \ldots, m, \end{cases} \qquad (\varepsilon, z) \in \mathbb{R} \times \mathbb{R}^m . \tag{25}$$

The function ψ is obviously continuously differentiable around any $(\varepsilon, z) \in \mathbb{R} \times \mathbb{R}^m$ as long as $\varepsilon \neq 0$ and is strongly semismooth everywhere.

Now, we are ready to define a smoothing function for $F(\cdot)$ itself. Let

$$\Upsilon(\varepsilon, y) := y - \psi\left(\varepsilon, y - (\mathcal{A}\Phi(\varepsilon, C + \mathcal{A}^* y) - b)\right), \quad (\varepsilon, y) \in \mathbb{R} \times \mathbb{R}^m . \tag{26}$$

By the definitions of Υ, ψ, and Φ, we know that for any $y \in \mathbb{R}^m$, $F(y) = \Upsilon(0, y)$. We summarize several useful properties of Υ in the next proposition.

Proposition 2.3 *Let $\Upsilon : \mathbb{R} \times \mathbb{R}^m$ be defined by (26). Let $y \in \mathbb{R}^m$. Then it holds that*

(i) *Υ is globally Lipschitz continuous on $\mathbb{R} \times \mathbb{R}^m$.*

(ii) *Υ is continuously differentiable around (ε, y) when $\varepsilon \neq 0$. For any fixed $\varepsilon \in \mathbb{R}$, $\Upsilon(\varepsilon, \cdot)$ is a P_0-function, i.e., for any $(y, h) \in \mathbb{R}^m \times \mathbb{R}^m$ with $y \neq h$,*

$$\max_{y_i \neq h_i} (y_i - h_i)(\Upsilon_i(\varepsilon, y) - \Upsilon_i(\varepsilon, h)) \geq 0 . \tag{27}$$

(iii) *Υ is strongly semismooth at $(0, y)$. In particular, for any $\varepsilon \downarrow 0$ and $\mathbb{R}^m \ni h \to 0$ we have*

$$\Upsilon(\varepsilon, y + h) - \Upsilon(0, y) - \Upsilon'(\varepsilon, y + h) \begin{pmatrix} \varepsilon \\ h \end{pmatrix} = O(\|(\varepsilon, h)\|^2) .$$

(iv) *For any* $h \in \mathbb{R}^m$,

$$\partial_B \Upsilon(0,y)(0,h) \subseteq h - \partial_B \psi(0, y - \nabla\theta(y))(0, h - \mathcal{A}\partial_B \Phi(0, C + \mathcal{A}^* y)(0, \mathcal{A}^* h))\,.$$

Proof. (i) Since both ψ and Φ are globally Lipschitz continuous, Υ is also globally Lipschitz continuous.

(ii) From the definitions of ψ and Φ we know that Υ is continuously differentiable around $(\varepsilon, y) \in \mathbb{R} \times \mathbb{R}^m$ when $\varepsilon \neq 0$.

Since, by part (i), Υ is continuous on $\mathbb{R} \times \mathbb{R}^m$, we only need to show that for any $0 \neq \varepsilon \in \mathbb{R}$, $\Upsilon(\varepsilon, \cdot)$ is a P_0-function.

Fix $\varepsilon \neq 0$. Define $g_\varepsilon : \mathbb{R}^m \to \mathbb{R}^m$ by

$$g_\varepsilon(y) = \mathcal{A}\Phi(\varepsilon, C + \mathcal{A}^* y) - b, \quad y \in \mathbb{R}^m\,.$$

Then g_ε is continuously differentiable on $\mathbb{R}^m$. From (19) and (20), we have

$$\langle h, (g_\varepsilon)'(y)h\rangle = \langle h, \mathcal{A}\Phi'_X(\varepsilon, X)(\mathcal{A}^* h)\rangle = \langle \mathcal{A}^* h, \Phi'_X(\varepsilon, X)(\mathcal{A}^* h)\rangle \geq 0 \quad \forall\, h \in \mathbb{R}^m\,,$$

which implies that g_ε is a monotone function on $\mathbb{R}^m$. Let $(y, h) \in \mathbb{R}^m \times \mathbb{R}^m$ with $y \neq h$. Then there exists $i \in \{1, \ldots, m\}$ with $y_i \neq h_i$ such that

$$(y_i - h_i)((g_\varepsilon)_i(y) - (g_\varepsilon)_i(h)) \geq 0\,.$$

Furthermore, by noting that for any $z \in \mathbb{R}^m$,

$$\phi'_{z_i}(\varepsilon, z_i) \in [0, 1], \quad i = 1, \ldots, m\,,$$

we obtain that

$$(y_i - h_i)(\Upsilon_i(\varepsilon, y) - \Upsilon_i(\varepsilon, h)) \geq 0\,.$$

This shows that (27) holds.

(iii) Since it can be checked directly that the composite of strongly semismooth functions is still strongly semismooth [22], Υ is strongly semismooth at $(0, y)$.

(iv) Since both ψ and Φ are directionally differentiable, for any $(\varepsilon, y') \in \mathbb{R} \times \mathbb{R}^m$ such that Υ is Fréchet differentiable at (ε, y'),

$$\Upsilon'(\varepsilon, y')(0, h) = h - \psi'\left((\varepsilon, z'); \big(0, h - \mathcal{A}\Phi'((\varepsilon, C + \mathcal{A}^* y'); (0, \mathcal{A}^* h))\big)\right)\,,$$

which, together with the semismoothness of ψ and Φ, implies

$$\Upsilon'(\varepsilon, y')(0, h) \in h - \partial_B \psi(\varepsilon, z')\big(0, h - \mathcal{A}\partial_B \Phi(\varepsilon, C + \mathcal{A}^* y')(0, \mathcal{A}^* h)\big)\,,$$

where $z' := y' - (\mathcal{A}\Phi(\varepsilon, C + \mathcal{A}^* y') - b)$. By taking $(\varepsilon, y') \to (0, y)$ in the above inclusion, we complete the proof. □

3 An inexact smoothing Newton method

The purpose of this section is to introduce an inexact smoothing Newton method for solving the general nonsmooth equation

$$F(y) = 0, \quad y \in \mathbb{R}^m ,$$

where $F : \mathbb{R}^m \to \mathbb{R}^m$ is a locally Lipschitz continuous function, which is not necessarily the mapping defined in the last section. This inexact smoothing Newton method is largely modified from the exact smoothing Newton method constructed in [37] for solving complementarity and variational inequality problems. The motivation to introduce an inexact version is completely from the computational point of view because the costs of the exact smoothing Newton method for solving problems such as the LSCM problem (1) are prohibitive. We shall talk about more on this in the numerical section.

Let $G : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$ be a locally Lipschitz continuous function satisfying

$$G(\varepsilon, y') \to F(y) \quad \text{as} \quad (\varepsilon, y') \to (0, y) .$$

Furthermore, G is required to be continuously differentiable around any (ε, y) unless $\varepsilon = 0$. The existence of such a function G can be easily proven via convolution. Define $E : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R} \times \mathbb{R}^m$ by

$$E(\varepsilon, y) := \begin{bmatrix} \varepsilon \\ G(\varepsilon, y) \end{bmatrix}, \quad (\varepsilon, y) \in \mathbb{R} \times \mathbb{R}^m .$$

Then solving the nonsmooth equation $F(y) = 0$ is equivalent to solving the following smoothing-nonsmooth equation

$$E(\varepsilon, y) = 0 .$$

Our inexact smoothing Newton method is specifically designed for solving the later.

Define the merit function $\varphi : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}_+$ by

$$\varphi(\varepsilon, y) := \|E(\varepsilon, y)\|^2 , \quad (\varepsilon, y) \in \mathbb{R} \times \mathbb{R}^m .$$

Choose $r \in (0, 1)$. Let

$$\zeta(\varepsilon, y) := r \min\{1, \varphi(\varepsilon, y)\} , \quad (\varepsilon, y) \in \mathbb{R} \times \mathbb{R}^m .$$

Then the inexact smoothing Newton method can be described as follows.

Algorithm 3.1 (An inexact smoothing Newton method)

Step 0. *Let $\hat{\varepsilon} \in (0, \infty)$ and $\eta \in (0, 1)$ be such that*

$$\delta := \sqrt{2} \max\{r\hat{\varepsilon}, \eta\} < 1 .$$

Select constants $\rho \in (0, 1)$, $\sigma \in (0, 1/2)$, $\tau \in (0, 1)$, and $\hat{\tau} \in [1, \infty)$. Let $\varepsilon^0 := \hat{\varepsilon}$ and $y^0 \in \mathbb{R}^m$ be an arbitrary point. $k := 0$.

Step 1. *If $E(\varepsilon^k, y^k) = 0$, then stop. Otherwise, compute*

$$\zeta_k := r \min\{1, \varphi(\varepsilon^k, y^k)\} \quad \text{and} \quad \eta_k := \min\{\tau, \hat{\tau}\|E(\varepsilon^k, y^k)\|\} .$$

Step 2. *Solve the following equation*

$$E(\varepsilon^k, y^k) + E'(\varepsilon^k, y^k) \begin{bmatrix} \Delta\varepsilon^k \\ \Delta y^k \end{bmatrix} = \begin{bmatrix} \zeta_k \hat{\varepsilon} \\ 0 \end{bmatrix} \tag{28}$$

approximately such that

$$\|R_k\| \leq \min\{\eta_k \|G(\varepsilon^k, y^k) + G'_\varepsilon(\varepsilon^k, y^k)\Delta\varepsilon^k\|,\, \eta\|E(\varepsilon^k, y^k)\|\}\,, \tag{29}$$

where

$$\Delta\varepsilon^k := -\varepsilon^k + \zeta_k\hat{\varepsilon}$$

and

$$R_k := G(\varepsilon^k, y^k) + G'(\varepsilon^k, y^k) \begin{bmatrix} \Delta\varepsilon^k \\ \Delta y^k \end{bmatrix}.$$

Step 3. *Let l_k be the smallest nonnegative integer l satisfying*

$$\varphi(\varepsilon^k + \rho^l\Delta\varepsilon^k, y^k + \rho^l\Delta y^k) \leq [1 - 2\sigma(1-\delta)\rho^l]\varphi(\varepsilon^k, y^k)\,. \tag{30}$$

Define:

$$(\varepsilon^{k+1}, y^{k+1}) := (\varepsilon^k + \rho^{l_k}\Delta\varepsilon^k, y^k + \rho^{l_k}\Delta y^k)\,.$$

Step 4. *Replace k by $k+1$ and go to* **Step 1**.

Lemma 3.2 *If for some $(\tilde{\varepsilon}, \tilde{y}) \in \mathbb{R}_{++} \times \mathbb{R}^m$, $E'(\tilde{\varepsilon}, \tilde{y})$ is nonsingular, then there exist an open neighborhood $\mathcal{O}$ of $(\tilde{\varepsilon}, \tilde{y})$ and a positive number $\bar{\alpha} \in (0, 1]$ such that for any $(\varepsilon, y) \in \mathcal{O}$ and $\alpha \in [0, \bar{\alpha}]$, $\varepsilon \in \mathbb{R}_{++}$, $E'(\varepsilon, y)$ is nonsingular, and*

$$\varphi(\varepsilon + \alpha\Delta\varepsilon, y + \alpha\Delta y) \leq [\,1 - 2\sigma(1-\delta)\alpha]\varphi(\varepsilon, y)\,, \tag{31}$$

where $(\Delta\varepsilon, \Delta y) \in \mathbb{R} \times \mathbb{R}^m$ satisfies

$$\Delta\varepsilon = -\varepsilon + \zeta(\varepsilon, y)\hat{\varepsilon}$$

and

$$\left\| G(\varepsilon, y) + G'(\varepsilon, y) \begin{bmatrix} \Delta\varepsilon \\ \Delta y \end{bmatrix} \right\| \leq \eta\|E(\varepsilon, y)\|\,.$$

Proof. Since $\tilde{\varepsilon} \in \mathbb{R}_{++}$ and $E'(\tilde{\varepsilon}, \tilde{y})$ is nonsingular, there exists an open neighborhood $\mathcal{O}$ of $(\tilde{\varepsilon}, \tilde{y})$ such that for any $(\varepsilon, y) \in \mathcal{O}$, $\varepsilon \in \mathbb{R}_{++}$ and $E'(\varepsilon, y)$ is nonsingular.

For any $(\varepsilon, y) \in \mathcal{O}$, denote

$$R(\varepsilon, y) := G(\varepsilon, y) + G'(\varepsilon, y) \begin{bmatrix} \Delta\varepsilon \\ \Delta y \end{bmatrix}.$$

Then $(\Delta\varepsilon, \Delta y)$ is the unique solution of the following equation

$$E(\varepsilon, y) + E'(\varepsilon, y) \begin{bmatrix} \Delta\varepsilon \\ \Delta y \end{bmatrix} = \begin{bmatrix} \zeta(\varepsilon, y)\hat{\varepsilon} \\ R(\varepsilon, y) \end{bmatrix}.$$

Thus,

$$\begin{aligned}
&\left\langle \nabla\varphi(\varepsilon,y), \left[\begin{array}{c}\Delta\varepsilon\\ \Delta y\end{array}\right]\right\rangle = \left\langle 2\nabla E(\varepsilon,y)E(\varepsilon,y), \left[\begin{array}{c}\Delta\varepsilon\\ \Delta y\end{array}\right]\right\rangle\\
= &\left\langle 2E(\varepsilon,y), \left[\begin{array}{c}\zeta(\varepsilon,y)\hat{\varepsilon}\\ R(\varepsilon,y)\end{array}\right] - E(\varepsilon,y)\right\rangle\\
= &-2\varphi(\varepsilon,y) + 2\varepsilon\zeta(\varepsilon,y)\hat{\varepsilon} + 2\langle R(\varepsilon,y), G(\varepsilon,y)\rangle\\
\leq &-2\varphi(\varepsilon,y) + 2\varepsilon(r\hat{\varepsilon})\min\{1,\varphi(\varepsilon,y)\} + 2\eta\varphi(\varepsilon,y)^{1/2}\|G(\varepsilon,y)\|\,,
\end{aligned}$$

which, implies that if $\varphi(\varepsilon,y) > 1$ we have

$$\begin{aligned}
&\left\langle \nabla\varphi(\varepsilon,y), \left[\begin{array}{c}\Delta\varepsilon\\ \Delta y\end{array}\right]\right\rangle\\
\leq &-2\varphi(\varepsilon,y) + 2\varepsilon(r\hat{\varepsilon}) + 2\eta\varphi(\varepsilon,y)^{1/2}\|G(\varepsilon,y)\|\\
\leq &-2\varphi(\varepsilon,y) + 2\max\{r\hat{\varepsilon},\eta\}\big(\varepsilon + \varphi(\varepsilon,y)^{1/2}\sqrt{\varphi(\varepsilon,y)-\varepsilon^2}\big)\\
\leq &-2\varphi(\varepsilon,y) + 2\sqrt{2}\max\{r\hat{\varepsilon},\eta\}\varphi(\varepsilon,y)\\
= &\,2\big(\sqrt{2}\max\{\gamma\hat{\varepsilon}\eta\} - 1\big)\varphi(\varepsilon,y) \qquad (32)
\end{aligned}$$

and if $\varphi(\varepsilon,y) < 1$ we have

$$\begin{aligned}
&\left\langle \nabla\varphi(\varepsilon,y), \left[\begin{array}{c}\Delta\varepsilon\\ \Delta y\end{array}\right]\right\rangle\\
\leq &-2\varphi(\varepsilon,y) + 2\varepsilon(r\hat{\varepsilon})\varphi(\varepsilon,y) + 2\eta\varphi(\varepsilon,y)^{1/2}\|G(\varepsilon,y)\|\\
\leq &-2\varphi(\varepsilon,y) + 2\max\{r\hat{\varepsilon},\eta\}\varphi(\varepsilon,y)^{1/2}\big(\varepsilon\varphi(\varepsilon,y)^{1/2} + \sqrt{\varphi(\varepsilon,y)-\varepsilon^2}\big)\\
\leq &-2\varphi(\varepsilon,y) + 2\sqrt{2}\max\{r\hat{\varepsilon},\eta\}\varphi(\varepsilon,y)\\
= &\,2\big(\sqrt{2}\max\{r\hat{\varepsilon},\eta\} - 1\big)\varphi(\varepsilon,y)\,. \qquad (33)
\end{aligned}$$

Therefore, by inequalities (32) and (33), we have

$$\left\langle \nabla\varphi(\varepsilon,y), \left[\begin{array}{c}\Delta\varepsilon\\ \Delta y\end{array}\right]\right\rangle \leq -2\,(1-\delta)\,\varphi(\varepsilon,y)\,. \qquad (34)$$

By using the fact that $\nabla\varphi(\cdot,\cdot)$ is uniformly continuous on $\mathcal{O}$, we obtain from the Taylor expansion that

$$\varphi(\varepsilon+\alpha\Delta\varepsilon, y+\alpha\Delta y) = \varphi(\varepsilon,y) + \alpha\left\langle \nabla\varphi(\varepsilon,y), \left[\begin{array}{c}\Delta\varepsilon\\ \Delta y\end{array}\right]\right\rangle + o(\alpha) \quad \forall\,(\varepsilon,y)\in\mathcal{O}\,,$$

which, together with (34), implies that there exists a positive number $\bar{\alpha}\in(0,1]$ such that for all $\alpha\in[0,\bar{\alpha}]$, (31) holds. □

Let

$$\mathcal{N} := \{(\varepsilon,y)\,|\,\varepsilon \geq \zeta(\varepsilon,y)\hat{\varepsilon}\}\,. \qquad (35)$$

Proposition 3.3 *For each fixed $k \geq 0$, if $\varepsilon^k \in \mathbb{R}_{++}$, $(\varepsilon^k, y^k) \in \mathcal{N}$, and $E'(\varepsilon^k, y^k)$ is nonsingular, then for any $\alpha \in [0,1]$ such that*

$$\varphi(\varepsilon^k + \alpha\Delta\varepsilon^k,\, y^k + \alpha\Delta y^k) \leq [\,1 - 2\sigma(1-\delta)\alpha]\varphi(\varepsilon^k, y^k) \tag{36}$$

it holds that $(\varepsilon^k + \alpha\Delta\varepsilon^k,\, y^k + \alpha\Delta y^k) \in \mathcal{N}$.

Proof. We prove this proposition by considering the following two cases.

Case 1. $\varphi(\varepsilon^k, y^k) > 1$. In this case, $\zeta_k = r$. Then, because $(\varepsilon^k, y^k) \in \mathcal{N}$ and $\zeta(\varepsilon, y) \leq r$ for any $(\varepsilon, y) \in \mathbb{R} \times \mathbb{R}^m$, for all $\alpha \in [0,1]$ satisfying (36) we have

$$\begin{array}{rl} & \varepsilon^k + \alpha\Delta\varepsilon^k - \zeta(\varepsilon^k + \alpha\Delta\varepsilon^k,\, y^k + \alpha\Delta y^k)\hat{\varepsilon} \\ \geq & (1-\alpha)\varepsilon^k + \alpha\zeta_k\hat{\varepsilon} - r\hat{\varepsilon} \\ \geq & (1-\alpha)\zeta_k\hat{\varepsilon} + \alpha\zeta_k\hat{\varepsilon} - r\hat{\varepsilon} \\ = & (1-\alpha)r\hat{\varepsilon} + \alpha r\hat{\varepsilon} - r\hat{\varepsilon} \\ = & 0\,. \end{array} \tag{37}$$

Case 2. $\varphi(\varepsilon^k, y^k) \leq 1$. Then, for any $\alpha \in [0,1]$ satisfying (36), we have

$$\varphi(\varepsilon^k + \alpha\Delta\varepsilon^k,\, y^k + \alpha\Delta y^k) \leq [\,1 - 2\sigma(1-\delta)\alpha]\varphi(\varepsilon^k, y^k) \leq 1\,. \tag{38}$$

So, for any $\alpha \in [0,1]$ satisfying (36),

$$\zeta(\varepsilon^k + \alpha\Delta\varepsilon^k,\, y^k + \alpha\Delta y^k) = r\varphi(\varepsilon^k + \alpha\Delta\varepsilon^k,\, y^k + \alpha\Delta y^k).$$

Hence, again because $(\varepsilon^k, y^k) \in \mathcal{N}$, by using the first inequality in (38), for any $\alpha \in [0,1]$ satisfying (36) we have

$$\begin{array}{rl} & \varepsilon^k + \alpha\Delta\varepsilon^k - \zeta(\varepsilon^k + \alpha\Delta\varepsilon^k, y^k + \alpha\Delta y^k)\hat{\varepsilon} \\ = & (1-\alpha)\varepsilon^k + \alpha\zeta_k\hat{\varepsilon} - r\varphi(\varepsilon^k + \alpha\Delta\varepsilon^k, y^k + \alpha\Delta y^k)\hat{\varepsilon} \\ \geq & (1-\alpha)\zeta_k\hat{\varepsilon} + \alpha\zeta_k\hat{\varepsilon} - r[\,1 - 2\sigma(1-\delta)\alpha]\varphi(\varepsilon^k, y^k)\hat{\varepsilon} \\ = & \zeta_k\hat{\varepsilon} - r[1 - 2\sigma(1-\delta)\alpha]\varphi(\varepsilon^k, y^k)\hat{\varepsilon} \\ = & r\varphi(\varepsilon^k, y^k)\hat{\varepsilon} - r[1 - 2\sigma(1-\delta)\alpha]\varphi(\varepsilon^k, y^k)\hat{\varepsilon} \\ = & 2r\sigma(1-\delta)\alpha\varphi(\varepsilon^k, y^k)\hat{\varepsilon} \\ \geq & 0\,. \end{array} \tag{39}$$

Thus, by combining (37) and (39), we have proved that for all $\alpha \in [0,1]$ satisfying (36),

$$(\varepsilon^k + \alpha\Delta\varepsilon^k,\, y^k + \alpha\Delta y^k) \in \mathcal{N}.$$

This completes our proof.

In order to discuss the global convergence of Algorithm 3.1 we need the following assumption.

Assumption 3.4 *For any $(\varepsilon, y) \in \mathbb{R}_{++} \times \mathbb{R}^n$, $E'(\varepsilon, y)$ is nonsingular.*

Theorem 3.5 *Suppose that Assumptions 3.4 is satisfied. Then Algorithm 3.1 is well defined and generates an infinite sequence* $\{(\varepsilon^k, y^k)\} \in \mathcal{N}$ *with the property that any accumulation point* $(\bar{\varepsilon}, \bar{y})$ *of* $\{(\varepsilon^k, y^k)\}$ *is a solution of* $E(\varepsilon, y) = 0$.

Proof. It follows from Lemma 3.2, Proposition 3.3, and Assumption 3.4 that Algorithm 3.1 is well defined and generates an infinite sequence $\{(\varepsilon^k, y^k)\} \in \mathcal{N}$.

From the design of Algorithm 3.1, $\varphi(\varepsilon^{k+1}, y^{k+1}) < \varphi(\varepsilon^k, y^k)$ for all $k \geq 0$. Hence, the two sequences $\{\varphi(\varepsilon^k, y^k)\}$ and $\{\zeta(\varepsilon^k, y^k)\}$ are monotonically decreasing. Since both $\varphi(\varepsilon^k, y^k)$ and $\zeta(\varepsilon^k, y^k)$ are nonnegative for $k \geq 0$, there exist $\bar{\psi} \geq 0$ and $\bar{\zeta} \geq 0$ such that $\varphi(\varepsilon^k, y^k) \to \bar{\varphi}$ and $\zeta(\varepsilon^k, y^k) \to \bar{\zeta}$ as $k \to \infty$.

Let $(\bar{\varepsilon}, \bar{y})$ be any accumulation point (if it exists) of $\{(\varepsilon^k, y^k)\}$. By taking a subsequence if necessary, we may assume that $\{(\varepsilon^k, y^k)\}$ converges to $(\bar{\varepsilon}, \bar{y})$. Then $\bar{\varphi} = \varphi(\bar{\varepsilon}, \bar{y})$, $\bar{\zeta} = \zeta(\bar{\varepsilon}, \bar{y})$, and $(\bar{\varepsilon}, \bar{y}) \in \mathcal{N}$.

Suppose that $\bar{\varphi} > 0$. Then, from $\zeta(\bar{\varepsilon}, \bar{y}) = r \min\{1, \varphi(\bar{\varepsilon}, \bar{y})\}$ and $(\bar{\varepsilon}, \bar{y}) \in \mathcal{N}$, we see that $\bar{\varepsilon} \in \mathbb{R}_{++}$. Thus, from Assumption 3.4, $E'(\bar{\varepsilon}, \bar{y})$ exists and is invertible. Hence, from Lemma 3.2, there exist an open neighborhood $\mathcal{O}$ of $(\bar{\varepsilon}, \bar{y})$ and a positive number $\bar{\alpha} \in (0, 1]$ such that for any $(\varepsilon, y) \in \mathcal{O}$ and all $\alpha \in [0, \bar{\alpha}]$, $\varepsilon \in \mathbb{R}_{++}$, $E'(\varepsilon, y)$ is invertible, and (31) holds. Therefore, there exists a nonnegative integer l such that $\rho^l \in (0, \bar{\alpha}]$ and $\rho^{l_k} \geq \rho^l$ for all k sufficiently large. Thus

$$\varphi(\varepsilon^{k+1}, y^{k+1}) \leq [1 - 2\sigma(1-\delta)\rho^{l_k}]\varphi(\varepsilon^k, y^k) \leq [1 - 2\sigma(1-\delta)\rho^l]\varphi(\varepsilon^k, y^k)$$

for all sufficiently large k. This contradicts the fact that the sequence $\{\varphi(\varepsilon^k, y^k)\}$ converges to $\bar{\varphi} > 0$. This contradiction shows that $\varphi(\bar{\varepsilon}, \bar{y}) = \bar{\varphi} = 0$. i.e., $E(\bar{\varepsilon}, \bar{y}) = 0$. The proof is completed. □

Theorem 3.6 *Suppose that Assumptions 3.4 is satisfied and that* $(\bar{\varepsilon}, \bar{y})$ *is an accumulation point of the infinite sequence* $\{(\varepsilon^k, y^k)\}$ *generated by Algorithm 3.1. Suppose that* E *is strongly semismooth at* $(\bar{\varepsilon}, \bar{y})$ *and that all* $V \in \partial_B E(\bar{\varepsilon}, \bar{y})$ *are nonsingular. Then the whole sequence* $\{(\varepsilon^k, y^k)\}$ *converges to* $(\bar{\varepsilon}, \bar{y})$ *quadratically, i.e.,*

$$\left\|(\varepsilon^{k+1} - \bar{\varepsilon}, y^{k+1} - \bar{y})\right\| = O\big(\|(\varepsilon^k - \bar{\varepsilon}, y^k - \bar{y})\|^2\big)\,. \tag{40}$$

Proof. First, from Theorem 3.5, $(\bar{\varepsilon}, \bar{y})$ is a solution of $E(\varepsilon, y) = 0$. Then, since all $V \in \partial_B E(\bar{\varepsilon}, \bar{y})$ are nonsingular, from [36], for all (ε^k, y^k) sufficiently close to $(\bar{\varepsilon}, \bar{y})$,

$$\|E'(\varepsilon^k, y^k)^{-1}\| = O(1)$$

and

$$\begin{aligned}
& \left\| \begin{pmatrix} \varepsilon^k \\ y^k \end{pmatrix} + \begin{pmatrix} \Delta\varepsilon^k \\ \Delta y^k \end{pmatrix} - \begin{pmatrix} \bar{\varepsilon} \\ \bar{y} \end{pmatrix} \right\| \\
= & \left\| \begin{pmatrix} \varepsilon^k \\ y^k \end{pmatrix} + E'(\varepsilon^k, y^k)^{-1} \left[\begin{pmatrix} r\varphi(\varepsilon^k, y^k)\hat{\varepsilon} \\ R_k \end{pmatrix} - E(\varepsilon, y) \right] - \begin{pmatrix} \bar{\varepsilon} \\ \bar{y} \end{pmatrix} \right\| \\
= & \left\| -E'(\varepsilon^k, y^k)^{-1} \left[E(\varepsilon^k, y^k) - E'(\varepsilon^k, y^k) \begin{pmatrix} \varepsilon^k - \bar{\varepsilon} \\ y^k - \bar{y} \end{pmatrix} - \begin{pmatrix} r\varphi(\varepsilon^k, y^k)\hat{\varepsilon} \\ R_k \end{pmatrix} \right] \right\| \\
= & O\left(\left\| E(\varepsilon^k, y^k) - E(\bar{\varepsilon}, \bar{y}) - E'(\varepsilon^k, y^k) \begin{pmatrix} \varepsilon^k - \bar{\varepsilon} \\ y^k - \bar{y} \end{pmatrix} \right\| \right) + O(\varphi(\varepsilon^k, y^k)) + O(\|R_k\|)\,.
\end{aligned} \tag{41}$$

Since E is locally Lipschitz continuous near $(\bar{\varepsilon}, \bar{y})$, for all (ε^k, y^k) close to $(\bar{\varepsilon}, \bar{y})$ we have

$$\varphi(\varepsilon^k, y^k) = \|E(\varepsilon^k, y^k) - E(\bar{\varepsilon}, \bar{y})\|^2 = O\big(\|(\varepsilon^k - \bar{\varepsilon}, y^k - \bar{y})\|^2\big) \tag{42}$$

and

$$\begin{aligned} \|R_k\| &\leq \eta_k \|G(\varepsilon^k, y^k) + G'_\varepsilon(\varepsilon^k, y^k)\Delta\varepsilon^k\| \\ &\leq O(\|E(\varepsilon^k, y^k)\|)\big(\|G(\varepsilon^k, y^k)\| + O(|\Delta\varepsilon^k|)\big) \\ &\leq O(\|E(\varepsilon^k, y^k) - E(\bar{\varepsilon}, \bar{y})\|^2)\,. \end{aligned} \tag{43}$$

Therefore, by using the assumption that E is strongly semismooth at $(\bar{\varepsilon}, \bar{y})$ and the relations (41), (42), and (43), we have for all (ε^k, y^k) sufficiently close to $(\bar{\varepsilon}, \bar{y})$ that

$$\|(\varepsilon^k, y^k) + (\Delta\varepsilon^k, \Delta y^k) - (\bar{\varepsilon}, \bar{y})\| = O\big(\|(\varepsilon^k, y^k) - (\bar{\varepsilon}, \bar{y})\|^2\big)\,. \tag{44}$$

Finally, since E is strongly semismooth at $(\bar{\varepsilon}, \bar{y})$ and that all $V \in \partial_B E(\bar{\varepsilon}, \bar{y})$ are nonsingular, we have for all (ε^k, y^k) sufficiently close to $(\bar{\varepsilon}, \bar{y})$ that

$$\|(\varepsilon^k, y^k) - (\bar{\varepsilon}, \bar{y})\| \leq O(\|E(\varepsilon^k, y^k)\|)\,,$$

which, together with (44) and the Lipschitz continuity of E, implies that

$$\varphi(\varepsilon^k + \Delta\varepsilon^k, y^k + \Delta y^k) = O(\varphi^2(\varepsilon^k, y^k))\,.$$

This shows that for all (ε^k, y^k) sufficiently close to $(\bar{\varepsilon}, \bar{y})$,

$$(\varepsilon^{k+1}, y^{k+1}) = (\varepsilon^k, y^k) + (\Delta\varepsilon^k, \Delta y^k)\,.$$

Thus, by using (44) we know that (40) holds. □

4 The least squares covariance matrix problem

In this section, we apply the general inexact smoothing Newton method developed in the last section to the least squares covariance matrix problem (1).

Let $F : \mathbb{R}^m \to \mathbb{R}^m$ be defined by (13). Let $\kappa \in (0, \infty)$ be a constant. Define $G : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$ by

$$G(\varepsilon, y) := \Upsilon(\varepsilon, y) + \kappa|\varepsilon| y\,, \quad (\varepsilon, y) \in \mathbb{R} \times \mathbb{R}^m\,, \tag{45}$$

where $\Upsilon : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$ is defined by (26). The reason for defining G by (45) is that for any $(\varepsilon, y) \in \mathbb{R} \times \mathbb{R}^m$ with $\varepsilon \neq 0$, $G'_y(\varepsilon, y)$ is a P-matrix (i.e., all its principal minors are positive), thus nonsingular while by part (ii) of Proposition 2.3 $\Upsilon'_y(\varepsilon, y)$ is only a P_0-matrix (i.e., all its principal minors are nonnegative), which may be singular.

Let $E : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R} \times \mathbb{R}^m$ be defined by

$$E(\varepsilon, y) := \begin{bmatrix} \varepsilon \\ G(\varepsilon, y) \end{bmatrix} = \begin{bmatrix} \varepsilon \\ \Upsilon(\varepsilon, y) + \kappa|\varepsilon| y \end{bmatrix}, \quad (\varepsilon, y) \in \mathbb{R} \times \mathbb{R}^m\,. \tag{46}$$

Let $\mathcal{N}$ be defined by (35). Next, we discuss convergent properties of Algorithm 3.1 when it is applied to solve $E(\varepsilon, y) = 0$.

Theorem 4.1 *Algorithm 3.1 is well defined and generates an infinite sequence* $\{(\varepsilon^k, y^k)\} \in \mathcal{N}$ *with the properties that any accumulation point* $(\bar{\varepsilon}, \bar{y})$ *of* $\{(\varepsilon^k, y^k)\}$ *is a solution of* $E(\varepsilon, y) = 0$ *and* $\lim_{k\to\infty} \varphi(\varepsilon^k, y^k) = 0$. *Additionally, if the generalized Slater condition (10) holds, then* $\{(\varepsilon^k, y^k)\}$ *is bounded.*

Proof. From part (ii) of Proposition 2.3 and the definitions of G and E we know that for any $(\varepsilon, y) \in \mathbb{R}_{++} \times \mathbb{R}^m$, $G'_y(\varepsilon, y)$, and so $E'(\varepsilon, y)$, is a P-matrix. Then from Theorem 3.5 we know that Algorithm 3.1 is well defined and generates an infinite sequence $\{(\varepsilon^k, y^k)\} \in \mathcal{N}$ with the property that any accumulation point $(\bar{\varepsilon}, \bar{y})$ of $\{(\varepsilon^k, y^k)\}$ is a solution of $E(\varepsilon, y) = 0$.

Since $\varphi(\varepsilon^k, y^k)$ is a decreasing sequence, $\lim_{k\to\infty} \varphi(\varepsilon^k, y^k)$ exists. Let

$$\bar{\varphi} := \lim_{k\to\infty} \varphi(\varepsilon^k, y^k) \geq 0\,.$$

If $\bar{\varphi} > 0$, then there exists an $\varepsilon' > 0$ such that $\varepsilon^k \geq \varepsilon'$ for all $k \geq 0$. For any $\upsilon \geq 0$, let

$$L_\upsilon := \{y \in \mathbb{R}^m \,|\, \|\Upsilon(\nu, y) + \kappa\nu y\| \leq \upsilon,\ \nu \in [\varepsilon', \hat{\varepsilon}]\}\,.$$

Then it is not difficult to prove that for any $\upsilon \geq 0$, L_υ is bounded. In fact, suppose that for some $\upsilon \geq 0$, L_υ is unbounded. Then there exist two sequences $\{z^l\}$ and $\{\nu^l\}$ such that $\lim_{l\to\infty} ||z^l|| = \infty$ and for all $l \geq 1$, $\varepsilon' \leq \nu^l \leq \hat{\varepsilon}$ and $\|\Upsilon(\nu^l, z^l) + \kappa\nu^l z^l\| \leq \upsilon$. By taking subsequences if necessary, we may assume that $\lim_{l\to\infty} \nu^l = \bar{\nu} \in [\varepsilon', \hat{\varepsilon}]$ and

$$i \in I^\infty \cup I^{-\infty} \cup I^\upsilon \quad \forall\, i \in \{1, \ldots, m\}\,,$$

where

$$\begin{aligned}
I^\infty &:= \{i \,|\, \lim_{l\to\infty} z_i^l = \infty,\ i = 1, \ldots, m\}\,,\\
I^{-\infty} &:= \{i \,|\, \lim_{l\to\infty} z_i^l = -\infty,\ i = 1, \ldots, m\}, \quad \text{and}\\
I^\upsilon &:= \{i \,|\, \{z_i^l\} \text{ is uniformly bounded},\ i = 1, \ldots, m\}\,.
\end{aligned}$$

Then, we have

$$\Upsilon_i(\nu^l, z^l) \to -\infty \quad \forall\, i \in I^\infty, \tag{47}$$

and

$$\Upsilon_i(\nu^l, z^l) \to \infty \quad \forall\, i \in I^{-\infty}\,. \tag{48}$$

For each $l \geq 1$, define $h^l \in \mathbb{R}^m$ as follows

$$h_i^l = \begin{cases} 0 & \text{if } i \in I^\infty \cup I^{-\infty}\,,\\ z_i^l & \text{if } i \in I^\upsilon\,, \end{cases} \qquad i = 1, \ldots, m\,.$$

Since, by part (ii) of Proposition 2.3, for any $l \geq 1$, $\Upsilon(\nu^l, \cdot)$ is a P_0-function, by further taking subsequences if necessary, we know that there exists $i \in I^\infty \cup I^{-\infty}$ (note that $h_j^l = z_j^l$ for all $j \in I^\upsilon$ and $l \geq 1$) such that

$$(z_i^l - h_i^l)(\Upsilon_i(\nu^l, z^l) - \Upsilon_i(\nu^l, h^l)) \geq 0 \quad \forall\, l \geq 1\,,$$

which is impossible in view of (47), (48), and the fact that $\{\Upsilon(\nu^l, h^l)\}$ is bounded (note that Υ is globally Lipschitz continuous). This shows that for any $\upsilon \geq 0$, L_υ is bounded, i.e.,

$$\{y \in \mathbb{R}^m \,|\, \|G(\varepsilon, y)\| \leq \upsilon,\ \varepsilon \in [\varepsilon', \hat{\varepsilon}]\}$$

is bounded. This implies that $\{(\varepsilon^k, y^k)\}$ is bounded. Thus, $\{(\varepsilon^k, y^k)\}$ has at least one accumulation point, which is a solution of $E(\varepsilon, y) = 0$, contradicting $\bar{\varphi} > 0$. Therefore, $\bar{\varphi} = 0$.

Suppose that the generalized Slater condition (10) holds. Then from Proposition 2.1 we know that the solution set of the dual problem is nonempty and compact. Thus, $E(\varepsilon, y) = 0$ also has a nonempty and compact solution set. Since part (ii) of Proposition 2.3 implies that E is a P_0-function, the boundedness of $\{(\varepsilon^k, y^k)\}$ follows directly from [39, Theorem 2.5]. □

Assume that the generalized Slater condition (10) holds. Let $(\bar{\varepsilon}, \bar{y})$ be an accumulation point of the infinite sequence $\{(\varepsilon^k, y^k)\}$ generated by Algorithm 3.1. Then, by Theorem 4.1, we know that $\bar{\varepsilon} = 0$ and $F(\bar{y}) = 0$, i.e., $\bar{y} \in Q^+ = \mathbb{R}^p \times \mathbb{R}^q_+$ is an optimal solution to the dual problem (8). Let $\overline{X} := \Pi_{\mathcal{S}^n_+}(C + \mathcal{A}^*\bar{y})$. By Proposition 2.1 we know that $\overline{X} \in \mathcal{S}^n_+$ is the unique optimal solution to problem (1).

For quadratic convergence of Algorithm 3.1, we need the concept of constraint nondegeneracy initiated by Robinson [40] and extensively developed by Bonnans and Shapiro [5]. This concept is a generalization of the well-known linear independence constraint qualification (or LICQ) used in nonlinear programming. For a given closed $K \in \mathcal{X}$, a finite dimensional real Hilbert space, as in convex analysis [41] we use $T_K(x)$ to denote the tangent cone of K at $x \in K$. The largest linear space contained in $T_K(x)$ is denoted by $\mathrm{lin}\big(T_K(x)\big)$. Let $\mathcal{I}$ be the identity mapping from $\mathcal{S}^n$ to $\mathcal{S}^n$. Then the constraint nondegeneracy is said to hold at $\overline{X}$ if

$$\begin{pmatrix} \mathcal{A} \\ \mathcal{I} \end{pmatrix} \mathcal{S}^n + \begin{pmatrix} \mathrm{lin}\big(T_Q(\mathcal{A}(\overline{X}) - b)\big) \\ \mathrm{lin}\big(T_{\mathcal{S}^n_+}(\overline{X})\big) \end{pmatrix} = \begin{pmatrix} \mathbb{R}^m \\ \mathcal{S}^n \end{pmatrix}, \tag{49}$$

where $Q = \{0\}^p \times \mathbb{R}^q_+$. Note that the constraint nondegenerate condition (49) is called the primal nondegeneracy in [1]. The linearity space $\mathrm{lin}\big(T_Q(\mathcal{A}(\overline{X}) - b)\big)$ in (49) can be computed directly

$$\mathrm{lin}\big(T_Q(\mathcal{A}(\overline{X}) - b)\big) = \{h \in \mathbb{R}^m \,|\, h_i = 0,\ i = 1, \ldots, p,\ i \in \mathrm{Ind}(\overline{X})\}, \tag{50}$$

where $\mathrm{Ind}(\overline{X})$ denotes the index set of active constraints at $\overline{X}$:

$$\mathrm{Ind}(\overline{X}) := \{i \,|\, \langle A_i, \overline{X}\rangle = b_i,\ i = p+1, \ldots, m\}.$$

Let s denote the number of elements in $\mathrm{Ind}(\overline{X})$. Without loss of generality, we assume that

$$\mathrm{Ind}(\overline{X}) = \{p+1, \ldots, p+s\}.$$

Define $\widehat{\mathcal{A}} : \mathcal{S}^n \to \mathbb{R}^{p+s}$ by

$$\widehat{\mathcal{A}}(X) := \begin{bmatrix} \langle A_1, X\rangle \\ \vdots \\ \langle A_{p+s}, X\rangle \end{bmatrix}, \quad X \in \mathcal{S}^n. \tag{51}$$

The adjoint of $\widehat{\mathcal{A}}$ is denoted by $\widehat{\mathcal{A}}^*$. By using (50), we can see that (49) is reduced to

$$\begin{pmatrix} \widehat{\mathcal{A}} \\ \mathcal{I} \end{pmatrix} \mathcal{S}^n + \begin{pmatrix} \{0\}^{p+s} \\ \mathrm{lin}\big(T_{\mathcal{S}^n_+}(\overline{X})\big) \end{pmatrix} = \begin{pmatrix} \mathbb{R}^{p+s} \\ \mathcal{S}^n \end{pmatrix},$$

which is equivalent to

$$\widehat{\mathcal{A}}\big(\mathrm{lin}\, T_{\mathcal{S}^n_+}(\overline{X})\big) = \mathbb{R}^{p+s}. \tag{52}$$

The characterization of $\mathrm{lin}\,\big(T_{\mathcal{S}^n_+}(\overline{X})\big)$ is more involved than $\mathrm{lin}\big(T_Q(\mathcal{A}(\overline{X}) - b)\big)$. Let $X := C + \mathcal{A}^*\bar{y}$. Suppose that X has the spectral decomposition as in (15). Then

$$\overline{X} = \Pi_{\mathcal{S}^n_+}(X) = P\mathrm{diag}(\max(0,\lambda_1),\ldots,\max(0,\lambda_n))P^T.$$

The tangent cone $T_{\mathcal{S}^n_+}(\overline{X})$, which was first characterized by Arnold [3], takes the form

$$T_{\mathcal{S}^n_+}(\overline{X}) = \{B \in \mathcal{S}^n \,\big|\, [\,P_\beta\ P_\gamma\,]^T B\,[\,P_\beta\ P_\gamma\,] \succeq 0\}.$$

Consequently,

$$\mathrm{lin}\,\big(T_{\mathcal{S}^n_+}(\overline{X})\big) = \big\{B \in \mathcal{S}^n \,\big|\, P_\beta^T B P_\beta = 0,\ P_\beta^T B P_\gamma = 0,\ P_\gamma^T B P_\gamma = 0\big\}. \tag{53}$$

Thus, from (52), the constraint nondegenerate condition (49) holds if and only if for any $h \in \mathbb{R}^{p+s}$,

$$P_\alpha^T \widehat{\mathcal{A}}^* h = 0 \Longleftrightarrow h = 0. \tag{54}$$

Lemma 4.2 *Let $\Phi : \mathbb{R}\times\mathcal{S}^n \to \mathcal{S}^n$ be defined by (18). Assume that the constraint nondegeneracy (49) holds at $\overline{X}$. Then for any $V \in \partial_B\Phi(0,\overline{X})$ we have*

$$\big\langle h, \widehat{\mathcal{A}}V(0,\widehat{\mathcal{A}}^*h)\big\rangle > 0 \quad \forall\, 0 \neq h \in \mathbb{R}^{p+s}. \tag{55}$$

Proof. Let $V \in \partial_B\Phi(0,\overline{X})$. Suppose that there exists $0 \neq h \in \mathbb{R}^{p+s}$ such that (55) fails to hold, i.e.,

$$\big\langle h, \widehat{\mathcal{A}}V(0,\widehat{\mathcal{A}}^*h)\big\rangle \leq 0.$$

Denote $H := \widehat{\mathcal{A}}^*h$. Then, by Proposition 2.2, there exists $V_{|\beta|} \in \partial_B\Phi_{|\beta|}(0,0)$ such that

$$V(0,H) = P\begin{bmatrix} P_\alpha^T H P_\alpha & P_\alpha^T H P_\beta & U_{\alpha\gamma}\circ(P_\alpha^T H P_\gamma) \\ (P_\alpha^T H P_\beta)^T & V_{|\beta|}(0, P_\beta^T H P_\beta) & 0 \\ (P_\alpha^T H P_\gamma)^T \circ U_{\alpha\gamma}^T & 0 & 0 \end{bmatrix} P^T,$$

where $U \in \mathcal{S}^n$ is defined by (24). Since $\langle P_\beta^T H P_\beta, V_{|\beta|}(0, P_\beta^T H P_\beta)\rangle \geq 0$ and $\big\langle h, \widehat{\mathcal{A}}V(0,\widehat{\mathcal{A}}^*h)\big\rangle \leq 0$, we obtain from $\big\langle h, \widehat{\mathcal{A}}V(0,\widehat{\mathcal{A}}^*h)\big\rangle = \langle H, V(0,H)\rangle$ that

$$P_\alpha^T H P_\alpha = 0,\ P_\alpha^T H P_\beta = 0, \text{ and } P_\alpha^T H P_\gamma = 0,$$

i.e.,

$$P_\alpha^T H = P_\alpha^T \widehat{\mathcal{A}}^* h = 0.$$

On the other hand, since the constraint nondegeneracy (49) holds at $\overline{X}$, from (54) we know that $h = 0$. This contradiction shows that for any $V \in \partial_B\Phi(0,X)$, (55) holds. □

Proposition 4.3 *Let $\Upsilon : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$ be defined by (26). Assume that the constraint nondegeneracy (49) holds at $\overline{X}$. Then for any $W \in \partial_B \Upsilon(0, \bar{y})$ we have*

$$\max_i h_i(W(0,h))_i > 0 \quad \forall\, 0 \neq h \in \mathbb{R}^m\,. \tag{56}$$

Proof. Let $W \in \partial_B \Upsilon(0, \bar{y})$. Suppose that there exists $0 \neq h \in \mathbb{R}^m$ such that (56) does not hold, i.e.,

$$\max_i h_i(W(0,h))_i \leq 0\,. \tag{57}$$

Then from part (iv) of Proposition 2.3 we know that there exist $D \in \partial_B \psi(0, \bar{z})$ and $V \in \partial_B \Phi(0, \overline{X})$ such that

$$W(0,h) = h - D(0, h - \mathcal{A}V(0, \mathcal{A}^* h)) = h - D(0,h) + D(0, \mathcal{A}V(0, \mathcal{A}^* h))\,, \tag{58}$$

where $\bar{z} := \bar{y} - \nabla\theta(\bar{y}) = \bar{y} - (\mathcal{A}\Phi(0, \overline{X}) - b)$. By simple calculations, we can see that there exists a nonnegative vector $d \in \mathbb{R}^m$ satisfying

$$d_i = \begin{cases} 1 & \text{if } 1 \leq i \leq p\,, \\ \in [0,1] & \text{if } p+1 \leq i \leq p+s\,, \\ 0 & \text{if } p+s+1 \leq i \leq m \end{cases}$$

such that for any $y \in \mathbb{R}^m$,

$$(D(0,y))_i = d_i y_i,\ i = 1, \ldots, m\,.$$

Thus, we obtain from (58) and (57) that

$$\begin{cases} h_i(\mathcal{A}V(0, \mathcal{A}^* h))_i \leq 0 & \text{if } 1 \leq i \leq p\,, \\ h_i(\mathcal{A}V(0, \mathcal{A}^* h))_i \leq 0 \text{ or } h_i = 0 & \text{if } p+1 \leq i \leq p+s\,, \\ h_i = 0 & \text{if } p+s+1 \leq i \leq m\,, \end{cases}$$

which, implies

$$\langle h, \mathcal{A}V(0, \mathcal{A}^* h)\rangle = \langle \hat{h}, \widehat{\mathcal{A}} V(0, \widehat{\mathcal{A}}^* \hat{h})\rangle \leq 0\,,$$

where $0 \neq \hat{h} \in \mathbb{R}^{p+s}$ is defined by $\hat{h}_i = h_i$, $i = 1, \ldots, p+s$. This, however, contradicts (55) in Lemma 4.2. This contradiction shows that (56) holds. □

Theorem 4.4 *Let $(\bar{\varepsilon}, \bar{y})$ be an accumulation point of the infinite sequence $\{(\varepsilon^k, y^k)\}$ generated by Algorithm 3.1. Assume that the constraint nondegeneracy (49) holds at $\overline{X}$. Then the whole sequence $\{(\varepsilon^k, y^k)\}$ converges to $(\bar{\varepsilon}, \bar{y})$ quadratically, i.e.,*

$$\left\| (\varepsilon^{k+1} - \bar{\varepsilon}, y^{k+1} - \bar{y}) \right\| = O\big(\|(\varepsilon^k - \bar{\varepsilon}, y^k - \bar{y})\|^2 \big)\,. \tag{59}$$

Proof. In order to apply Theorem 3.6 to obtain the quadratic convergence of $\{(\varepsilon^k, y^k)\}$, we only need to check that E is strongly semismooth at $(\bar{\varepsilon}, \bar{y})$ and that all $V \in \partial_B E(\bar{\varepsilon}, \bar{y})$ are nonsingular.

The strong semismoothness of E at $(\bar{\varepsilon}, \bar{y})$ follows directly from part (iii) of Proposition 2.3 and the fact that the modulus function $|\cdot|$ is strongly semismooth everywhere on $\mathbb{R}$. The nonsingularity of all matrices in $\partial_B E(\bar{\varepsilon}, \bar{y})$ can be proved as follows.

Let $V \in \partial_B E(\bar{\varepsilon}, \bar{y})$ be arbitrarily chosen. From Proposition 4.3 and the definition of E, we know that for any $0 \neq d \in \mathbb{R}^{m+1}$,

$$\max_i d_i (Vd)_i > 0\,,$$

which, by [13, Theorem 3.3.4], implies that V is a P-matrix, and so nonsingular. Then the proof is completed. □

Theorem 4.4 says that Algorithm 3.1 can achieve quadratic convergence under the assumption that the constraint nondegenerate condition (49) holds at $\overline{X}$. Next, we shall discuss about this assumption by considering the following special least squares covariance matrix problem

$$\begin{array}{ll} \min & \dfrac{1}{2}\|X - C\|^2 \\ \text{s.t.} & X_{ij} = e_{ij},\ (i,j) \in \mathcal{B}_e\,, \\ & X_{ij} \geq l_{ij},\ (i,j) \in \mathcal{B}_l\,, \\ & X_{ij} \leq u_{ij},\ (i,j) \in \mathcal{B}_u\,, \\ & X \in \mathcal{S}^n_+\,, \end{array} \tag{60}$$

where $\mathcal{B}_e$, $\mathcal{B}_l$, and $\mathcal{B}_u$ are three index subsets of $\{(i,j) \,|\, 1 \leq i \leq j \leq n\}$ satisfying $\mathcal{B}_e \cap \mathcal{B}_l = \emptyset$, $\mathcal{B}_e \cap \mathcal{B}_u = \emptyset$, and $l_{ij} < u_{ij}$ 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 p, q_l, and q_u, respectively. Let $m := p + q_l + q_u$. For any $(i,j) \in \{1,\ldots,n\} \times \{1,\ldots,n\}$, define $\mathcal{E}^{ij} \in \mathbb{R}^{n\times n}$ by

$$(\mathcal{E}^{ij})_{st} := \begin{cases} 1 & \text{if } (s,t) = (i,j)\,, \\ 0 & \text{otherwise}\,, \end{cases} \qquad s,t = 1,\ldots,n\,.$$

Thus, problem (60) can be written as a special case of (1) with

$$\mathcal{A}(X) := \begin{bmatrix} \{\langle A^{ij}, X\rangle\}_{(i,j)\in\mathcal{B}_e} \\ \{\langle A^{ij}, X\rangle\}_{(i,j)\in\mathcal{B}_l} \\ -\{\langle A^{ij}, X\rangle\}_{(i,j)\in\mathcal{B}_u} \end{bmatrix}, \quad X \in \mathcal{S}^n \tag{61}$$

and

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

where $A^{ij} := \frac{1}{2}(\mathcal{E}^{ij} + \mathcal{E}^{ji})$. Then, its dual problem takes the same form as (8) with $q := q_l + q_u$. The index set $\mathrm{Ind}(\overline{X})$ of active constraints at $\overline{X}$ now becomes

$$\mathrm{Ind}(\overline{X}) = \widehat{\mathcal{B}}_l \cup \widehat{\mathcal{B}}_u\,,$$

where

$$\widehat{\mathcal{B}}_l := \{\,(i,j) \in \mathcal{B}_l \,|\, \langle\, A^{ij}, \overline{X}\,\rangle = l_{ij}\} \quad \text{and} \quad \widehat{\mathcal{B}}_u := \{\,(i,j) \in \mathcal{B}_u \,|\, \langle\, A^{ij}, \overline{X}\,\rangle = u_{ij}\}\,.$$

Let s be the cardinality of $\mathrm{Ind}(\overline{X})$. Then the mapping $\widehat{\mathcal{A}} : \mathcal{S}^n \to \mathbb{R}^{p+s}$ defined by (51) takes the form

$$\widehat{\mathcal{A}}(X) := \begin{bmatrix} \{\langle A^{ij}, X\rangle\}_{(i,j)\in\mathcal{B}_e} \\ \{\langle A^{ij}, X\rangle\}_{(i,j)\in\widehat{\mathcal{B}}_l} \\ -\{\langle A^{ij}, X\rangle\}_{(i,j)\in\widehat{\mathcal{B}}_u} \end{bmatrix}.$$

Recall that the constraint nondegenerate condition (49) holds at $\overline{X}$ if and only if for any $h \in \mathbb{R}^{p+s}$, (54) holds. A particular case for (54) to hold is when $\mathcal{B}_e = \{(i,i) \,|\, i = 1, \ldots, n\}$, $\mathcal{B}_l \cup \mathcal{B}_u = \emptyset$, and $b > 0$ [34, 35]. Furthermore, if $\mathcal{B}_e$ has a band structure, (54) also holds as long as the corresponding band of the given matrix C is positive definite [35]. In general, the equivalent constraint nondegenerate condition (54) may fail to hold for problem (60). In [33], Qi establishes an interesting connection between the constraint nondegeneracy and the positive semidefinite matrix completions on chordal graphs.

5 Numerical results

In this section, we report our numerical experiments conducted for testing the efficiency of Algorithm 3.1. The main task of Algorithm 3.1 for solving the least squares covariance matrix problem (1), at the kth iterate, is to solve the linear system (28) with $E(\cdot)$ being defined by (46). In numerical implementation, we first obtain $\Delta\varepsilon^k = -\varepsilon^k + \zeta_k\hat{\varepsilon}$ by the first equation of (28), and then apply the BiCGStab iterative solver of Van der Vorst [49] to the resulted linear system

$$G'_y(\varepsilon^k, y^k)\Delta y^k = -G(\varepsilon^k, y^k) - G'_\varepsilon(\varepsilon^k, y^k)\Delta\varepsilon^k \tag{62}$$

to obtain Δy^k such that it satisfies (29). The cost of using the BiCGStab iterative solver to compute Δy^k will be analyzed next.

For the sake of convenience, in subsequent analysis we suppress the superscript k. By noting that $G(\varepsilon, y)$ and $\Upsilon(\varepsilon, y)$ are defined by (45) and (26), respectively, we obtain that

$$\begin{aligned} G'_y(\varepsilon, y)\Delta y &= \Upsilon'_y(\varepsilon, y)\Delta y + \kappa\varepsilon\Delta y \\ &= \Delta y - \psi'_z(\varepsilon, z)(\Delta y - \mathcal{A}\Phi'_X(\varepsilon, X)(\mathcal{A}^*\Delta y)) + \kappa\varepsilon\Delta y \\ &= \Delta y - \psi'_z(\varepsilon, z)\Delta y + \psi'_z(\varepsilon, z)(\mathcal{A}\Phi'_X(\varepsilon, X)(\mathcal{A}^*\Delta y)) + \kappa\varepsilon\Delta y\,, \end{aligned} \tag{63}$$

where $z := y - (\mathcal{A}\Phi(\varepsilon, X) - b)$, $X := C + \mathcal{A}^*y$, and $\mathcal{A}^*$ is given by (9). Let X have the spectral decomposition (15). Then, by (19),

$$\Phi'_X(\varepsilon, X)(\mathcal{A}^*\Delta y) = P[\Omega(\varepsilon, \lambda) \circ (P^T\mathcal{A}^*\Delta y P)]P^T, \tag{64}$$

where $\Omega(\varepsilon, \lambda)$ is given by (20).

From (63) and (64), we know that in order to compute the coefficient matrix $G'_y(\varepsilon, y)$ one needs $O(m^2n^3)$ flops. This implies that it is impractical to use direct methods to solve the linear system (62) even when n and m are not large, say $n = 100$ and $m = 1,000$. Given the fact that the coefficient matrix $G'_y(\varepsilon, y)$ is nonsymmetric when problem (1) has inequality constraints, i.e., $q \neq 0$, it is natural to choose the BiCGStab as our iterative solver for solving the linear system (62).

In order to speed up the convergence of the BiCGStab iterative solver, in the following we shall introduce a diagonal preconditioner. Define the vector $d \in \mathbb{R}^m$ as the diagonal part of the coefficient matrix $G'_y(\varepsilon, y)$, i.e., $d := \mathrm{diag}(G'_y(\varepsilon, y))$. Then from (63) and (64) we know for each $l \in \{1, \ldots, m\}$ that

$$d_l = [G'_y(\varepsilon, y)]_{ll} = [G'_y(\varepsilon, y)I_l]_l = 1 - \phi'_{z_l}(\varepsilon, z_l) + \phi'_{z_l}(\varepsilon, z_l)w_l + \kappa\varepsilon, \tag{65}$$

where $I_l \in \mathbb{R}^m$ denotes the lth column of the identity matrix I and w_l is defined by

$$\begin{aligned} w_l &:= [\,\mathcal{A}\Phi'_X(\varepsilon, X)(\mathcal{A}^* I_l)\,]_l \\ &= \langle\, I_l\,,\, \mathcal{A}\Phi'_X(\varepsilon, X)(\mathcal{A}^* I_l)\,\rangle \\ &= \langle\, \mathcal{A}^* I_l\,,\, \Phi'_X(\varepsilon, X)(\mathcal{A}^* I_l)\,\rangle \\ &= \langle\, \mathcal{A}^* I_l\,,\, P[\Omega(\varepsilon,\lambda) \circ (P^T \mathcal{A}^* I_l P)]P^T\rangle \\ &= \langle\, P^T A_l P\,,\, \Omega(\varepsilon,\lambda) \circ (P^T A_l P)\rangle. \end{aligned} \tag{66}$$

Let $D \in \mathbb{R}^{m\times m}$ be defined by $D := \mathrm{diag}(d)$. On the one hand, the diagonal matrix D is an obvious diagonal preconditioner for the linear system (62). On the other hand, from (65) and (66), we know that for a large m it is expensive to compute the diagonal matrix D because its computation needs $O(mn^3)$ flops. Of course, the cost for computing D can be reduced if most of the matrices A_l's are sparse. For instance, in problem (60), for each $l \in \{1, \ldots, m\}$, the matrix A_l takes the form

$$A_l = \frac{1}{2}(\mathcal{E}^{i_l j_l} + \mathcal{E}^{j_l i_l}),$$

where

$$(i_l, j_l) \in \begin{cases} \mathcal{B}_e & \text{if } l = 1, \ldots, p, \\ \mathcal{B}_l & \text{if } l = p+1, \ldots, p+q_l, \\ \mathcal{B}_u & \text{if } l = p+q_l+1, \ldots, m. \end{cases}$$

By taking account of such a special structure of A_l, w_l can be further simplified as follows

$$\begin{aligned} w_l &= \langle\, P^T A_l P\,,\, \Omega(\varepsilon,\lambda) \circ (P^T A_l P)\rangle \\ &= \frac{1}{2}[\, a_{i_l}^2 \Omega(\varepsilon,\lambda)(a_{j_l}^2)^T + (a_{i_l} \circ a_{j_l})\Omega(\varepsilon,\lambda)(a_{i_l} \circ a_{j_l})^T], \quad l = 1, \ldots, m, \end{aligned} \tag{67}$$

where "$\circ$" denotes the Hadamard product of two vectors, a_i is the ith row of P, and $a_i^2 := a_i \circ a_i$, $i = 1, \ldots, n$. Thus, in this case the diagonal matrix D can be computed directly from (65) and (67) with a reduced cost of $O(mn^2)$ flops. However, when m is much larger than n, say $m = O(n^2)$, this cost is still too expensive. In our numerical implementation, we use an estimated diagonal matrix of D as our diagonal preconditioner. This estimated diagonal matrix can be obtained in $O(n^3)$ flops as follows.

From (67) and the fact that

$$0 \le (a_{i_l} \circ a_{j_l})\Omega(\varepsilon,\lambda)(a_{i_l} \circ a_{j_l})^T \le a_{i_l}^2 \Omega(\varepsilon,\lambda)(a_{j_l}^2)^T, \tag{68}$$

we have the following inequality

$$\frac{1}{2} a_{i_l}^2 \Omega(\varepsilon,\lambda)(a_{j_l}^2)^T \le w_l \le a_{i_l}^2 \Omega(\varepsilon,\lambda)(a_{j_l}^2)^T\,, \quad l = 1, \ldots, m\,. \tag{69}$$

Let $\tilde{w} \in \mathbb{R}^m$ and $\tilde{d} \in \mathbb{R}^m$ be defined by

$$\tilde{w}_l := a_{i_l}^2 \Omega(\varepsilon,\lambda)(a_{j_l}^2)^T\,, \quad l = 1, \ldots, m, \tag{70}$$

and

$$\tilde{d}_l := 1 - \phi'_{z_l}(\varepsilon, z_l) + \phi'_{z_l}(\varepsilon, z_l)\tilde{w}_l + \kappa\varepsilon\,, \quad l = 1, \ldots, m\,, \tag{71}$$

respectively. Then we have the following relationship between d and $\tilde{d}$.

Proposition 5.1 *Let $d \in \mathbb{R}^m$ and $\tilde{d} \in \mathbb{R}^m$ be defined by (65) and (71), respectively. Then it holds that*

$$\frac{1}{2} \leq \frac{d_l}{\tilde{d}_l} \leq 1\,, \quad l = 1, \ldots, m\,. \tag{72}$$

Proof. From (69) and (70), we obtain that

$$\frac{1}{2} \leq \frac{w_l}{\tilde{w}_l} \leq 1\,, \quad l = 1, \ldots, m\,,$$

which, together with (65), (71), and the fact that for each $l \in \{1, \ldots, m\}$, $\phi'_{z_l}(\varepsilon, z_l) \in [0, 1]$ shows that (72) holds. □

Let $\widetilde{D} := \mathrm{diag}(\tilde{d})$. Proposition 5.1 says that $\widetilde{D}$ is a good estimation of the diagonal part of $G'_y(\varepsilon, y)$. Furthermore, computing the diagonal matrix $\widetilde{D}$ only requires $O(n^3)$ flops, which is independent of m. Hence, in our numerical experiments we use the diagonal matrix $\widetilde{D}$ instead of D as our diagonal preconditioner.

Remark 5.2 *For the least squares covariance matrix problem (1), if all A_i's are rank 1 matrices, we can still apply the above technique to obtain a diagonal preconditioner with a cost of $O(n^3)$.*

In our numerical experiments we compare our inexact smoothing Newton method, which is referred as `Smoothing` in our numerical results, with the following methods for solving the least squares covariance matrix problem with simple constraints (60):

(i) The projected gradient method (`PGM`) of Boyd and Xiao [9]. At the kth step, the `PGM` computes

$$y^{k+1} := \Pi_{Q^+}(y^k - \rho \nabla\theta(y^k))\,,$$

where $\rho \in (0, 2)$. Note that in this case the Lipschitz modulus of $\nabla\theta(y)$ is one. We take $\rho = 1.8$ as it performs better than $\rho = 1$. The maximum number of iterations is set as $5,000$ and the stopping criterion is

$$Res_k := \|\nabla\theta(y^k)\| \leq 10^{-5}.$$

Note that when there are no inequality constraints, i.e., $m = p$, the `PGM` reduces to the gradient method (`GM`) of Higham [24].

(ii) The `BFGS-SQP` method of Chen [11] and Malick [29]. In [29], Malick suggested using a `BFGS` type method to solve the dual problem (8). Here we adopt a `BFGS-SQP` approach introduced by Chen [11] for solving a more general LC^1 problem. For the dual problem (8), at the kth step, the `BFGS-SQP` method solves the following quadratic program

$$\begin{array}{ll} \min & \langle\, \nabla\theta(y^k),\, \Delta y^k \rangle + \dfrac{1}{2} \langle\, \Delta y^k, (B_k + \epsilon_k I)\Delta y^k \rangle \\ \text{s.t.} & y^k + \Delta y^k \in Q^+\,. \end{array} \tag{73}$$

We use Algorithm 3.1 to compute the subproblem (73). The maximum number of iterations is set as $2,000$ and the stopping criterion is the same as in the `PGM`. For cases with equality constraints only, instead of using our own code, we run the Matlab package – *SDLS*, which is an implementation of the `BFGS` method for the dual problem and is written by Henrion and Malick [23].

(iii) The inexact semismooth Newton method of Qi and Sun [34], which is referred as `Semismooth` in our numerical results. This algorithm is particularly designed for the nearest correlation matrix problem. In order to improve its performance and for comparison purpose, we introduced a diagonal preconditioner by using the same technique as in our inexact smoothing Newton method. The stopping criterion is

$$Res_k := \|\nabla\theta(y^k)\| \leq 10^{-6}.$$

(iv) The inexact interior point method of Toh, Tütüncü, and Todd [47] for solving the W-weighted nearest correlation matrix problem (7). This inexact interior point method, referred as `IP-NCM` here, uses the preconditioned PSQMR as the iterative solver and can solve equality constrained cases when applied to problem (60). The stopping criterion is $Res_k \leq 10^{-7}$. Note that the "*Res*" defined in the `IP-NCM` method is the relative residue rather than the absolute residue of the resulted Karush-Kuhn-Tucker system of problem (60).

The stopping criterion chosen for Algorithm 3.1 is

$$Res_k := \|E(\varepsilon^k,\, y^k)\| \leq 10^{-6}\,.$$

The maximum number of BiCGStab steps at each iteration is set as 200. The other parameters used in Algorithm 3.1 are set as $r = 0.2$, $\hat{\varepsilon} = 0.01$, $\eta = 0.5$, $\rho = 0.5$, $\sigma = 0.5 \times 10^{-6}$, $\tau = 0.01$, $\hat{\tau} = 0.5$, and $\kappa = 0.01$.

We implemented all algorithms in `MATLAB 7.1` running on a PC Intel Pentium IV of 2.40 GHz CPU and 512 MB of RAM. The testing examples are given below.

Example 5.1 *The matrix C is the 387×387 1-day correlation matrix (as of June 15, 2006) from the lagged datasets of RiskMetrics* (`www.riskmetrics.com/stddownload_edu.html`). *For the test purpose, we perturb C to*

$$C := (1-\alpha)C + \alpha R,$$

where $\alpha \in (0,1)$ and R is a randomly generated symmetric matrix with entries in $[-1,1]$. The `MATLAB` *code for generating the random matrix R is:* `R = 2.0*rand(387,387)-ones(387,387); R = triu(R)+triu(R,1)'; for i=1:387; R(i,i) = 1; end`. *Here we take $\alpha = 0.1$ and the index sets*

$$\mathcal{B}_e := \{(i,i)\,|\,i = 1,\ldots,387\},\ \ \mathcal{B}_l \cup \mathcal{B}_u = \emptyset\,,\ \textit{and}\ e_{ii} = 1\ \textit{for}\ (i,i) \in \mathcal{B}_e.$$

Note that this example corresponds exactly to the nearest correlation matrix problem.

Example 5.2 *All the data are the same as in Example 5.1 except that $e_{ii} \in [0,1]$ for $(i,i) \in \mathcal{B}_e$ are randomly generated. This example corresponds to the W-weighted nearest correlation problem (7) when the weight matrix W is a randomly generated diagonal matrix. It can also come from the local correlation stress testing [35].*

Example 5.3 *All the data are the same as in Example 5.1 except that R is a randomly generated symmetric matrix with entries in* $[0,1]$: `R = rand(387,387); R = triu(R)+triu(R,1)'; for i=1:387; R(i,i) = 1; end`.

Example 5.4 *The matrix C is a randomly generated $n \times n$ symmetric matrix with entries in* $[-1,1]$: `C = 2.0*rand(n,n)-ones(n,n); C = triu(C)+triu(C,1)'; for i=1:n; C(i,i) = 1; end`. *The index sets are the same as in Example 5.1, i.e.,*

$$\mathcal{B}_e = \{(i,i) \,|\, i = 1, \ldots, n\} \text{ and } \mathcal{B}_l \cup \mathcal{B}_u = \emptyset \,.$$

We test the following two cases: a) $e_{ii} = 1$, $(i,i) \in \mathcal{B}_e$ and b) $e_{ii} \in [0,1]$, $(i,i) \in \mathcal{B}_e$ are randomly generated. For each of the two cases, we take $n = 500,\ 1,000$, and $2,000$, respectively.

Example 5.5 *Let $n = 387$. The matrix C is the same as in Example 5.1, i.e., C is the perturbed 387×387 correlation matrix. The index sets $\mathcal{B}_l \cup \mathcal{B}_u = \emptyset$ and $\mathcal{B}_e$ is defined as $\mathcal{B}_e := \mathcal{B}_{e_1} \cup \mathcal{B}_{e_2}$ with*

$$\mathcal{B}_{e_1} := \{(i,i) \,|\, i = 1, \ldots, n\} \text{ and } \emptyset \neq \mathcal{B}_{e_2} \subset \{(i,j) \,|\, 1 \leq i < j \leq n\} \,,$$

where $\mathcal{B}_{e_2}$ consists of the indices of $\min(n_r, n-i)$ randomly generated elements at the ith row of X, $i = 1, \ldots, n$ with $n_r = 1, 5$, and 10. For all $(i,j) \in \mathcal{B}_{e_2}$, $e_{ij} = 0$. We consider the following two cases: a) $e_{ii} = 1$, $(i,i) \in \mathcal{B}_{e_1}$ and b) $e_{ii} \in [0,1]$, $(i,i) \in \mathcal{B}_{e_1}$ are randomly generated.

Example 5.6 *All the data are the same as in Example 5.5 except that the matrix C is a randomly generated $n \times n$ symmetric matrix with entries in $[-1,1]$. We test for $n = 500,\ 1,000$, and $2,000$, respectively.*

Example 5.7 *Let $n = 387$. The data C, $\mathcal{B}_e$, and e_{ii} are the same as in Example 5.1, i.e., C is the perturbed 387×387 correlation matrix and $e_{ii} = 1$ for $(i,i) \in \mathcal{B}_e = \{(i,i) \,|\, i = 1, \ldots, n\}$. The index sets $\mathcal{B}_l$, $\mathcal{B}_u \subset \{(i,j) \,|\, 1 \leq i < j \leq n\}$ 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 $\hat{n}_r$ taking the following values: a) $\hat{n}_r = 1$ $(q_l = q_u = 386)$; b) $\hat{n}_r = 2$ $(q_l = q_u = 771)$; c) $\hat{n}_r = 5$ $(q_l = q_u = 1,920)$; d) $\hat{n}_r = 10$ $(q_l = q_u = 3,815)$; e) $\hat{n}_r = 20$ $(q_l = q_u = 7,530)$. We take $l_{ij} = -0.1$ for $(i,j) \in \mathcal{B}_l$ and $u_{ij} = 0.1$ for $(i,j) \in \mathcal{B}_u$.*

Example 5.8 *All the data are the same as in Example 5.7 except that*

$$e_{ii} = \alpha + (1-\alpha)\omega, \quad (i,i) \in \mathcal{B}_e \,,$$

where $\alpha = 0.1$ and ω is a randomly generated number in $[0,1]$.

Example 5.9 *The matrix C is a randomly generated $n \times n$ symmetric matrix with entries in $[-1,1]$. The index sets $\mathcal{B}_e$, $\mathcal{B}_l$, and $\mathcal{B}_u$ are the same as in Example 5.7 with $\hat{n}_r = 1,\ 5$, and 10. We consider the following two cases: a) $e_{ii} = 1$, $(i,i) \in \mathcal{B}_e$ and b) $e_{ii} = \alpha + (1-\alpha)\omega$, $(i,i) \in \mathcal{B}_e$, where $\alpha = 0.1$ and ω is a randomly generated number in $[0,1]$. We test for $n = 500,\ 1,000$, and $2,000$, respectively.*

	Example 5.1			**Example 5.2**			**Example 5.3**		
Method	*Iter*	*cputime*	*Res*	*Iter*	*cputime*	*Res*	*Iter*	*cputime*	*Res*
`GM`	37	0:44	1.0e-5	5000*	1:23:00	3.3e-2	5000*	1:17:41	3.9e-2
`BFGS`	16	0:33	7.4e-6	1203	34:35	9.8e-6	1217	25:31	8.7e-6
`Semismooth`	5	0:09	1.6e-7	12	0:21	4.0e-8	13	0:21	3.5e-9
`IP-NCM`	11	1:18	6.8e-9	18	3:07	2.6e-8	19	2:55	4.1e-8
`Smoothing`	5	0:10	3.2e-7	12	0:23	1.0e-7	13	0:21	1.2e-8

Table 1: Numerical results for Examples 5.1, 5.2, and 5.3

Example 5.4		**n=500**			**n=1000**			**n=2000**		
Method	*Case*	*Iter*	*cputime*	*Res*	*Iter*	*cputime*	*Res*	*Iter*	*cputime*	*Res*
`GM`	a)	57	1:53	8.5e-6	79	19:28	8.6e-6	109	2:55:10	9.5e-6
	b)	5000*	2:31:32	5.6e-3	2891	>10 `hrs`	2.4e-2	406	>10 `hrs`	2.8e-1
`BFGS`	a)	19	0:48	7.4e-6	21	06:02	9.9e-6	24	48:48	6.8e-6
	b)	683	23:22	1.0e-5	1400	4:51:46	1.0e-5	407	>10 `hrs`	1.3e-2
`Semismooth`	a)	5	0:15	8.0e-7	6	2:04	7.5e-9	6	14:00	6.8e-8
	b)	13	0:34	6.7e-9	13	4:06	2.2e-8	14	33:56	3.3e-9
`IP-NCM`	a)	10	1:22	2.4e-8	10	8:48	1.2e-8	`out of memory`		
	b)	12	2:02	2.2e-8	12	11:30	4.1e-8	−−		
`Smoothing`	a)	6	0:18	2.0e-9	6	2:10	1.4e-8	6	14:46	1.5e-7
	b)	13	0:35	1.5e-8	13	4:21	5.6e-8	14	34:26	9.8e-9

Table 2: Numerical results for Example 5.4

Our numerical results are reported in Tables 1-6. "*Iter*" and "*Res*" stand for the number of total iterations and the residue at the final iterate of an algorithm, respectively. " $*$ " means that an algorithm reaches the set maximum number of iterations before the accuracy is achieved while "`out of memory`" means that our computer runs out of memory. The maximum cputime set for each algorithm is 10 hours. So "$>$ 10 `hrs`" indicates that an algorithm is terminated after 10 hours.

Example 5.5		**Case a)**			**Case b)**		
Method	n_r	*Iter*	*cputime*	*Res*	*Iter*	*cputime*	*Res*
`GM`	1	857	11:30	1.0e-5	5000*	1:19:27	2.0e-2
	5	1038	14:41	9.9e-6	5000*	1:25:03	1.3e-2
	10	1034	15:47	1.0e-5	5000*	1:09:47	1.2e-2
`BFGS`	1	87	1:26	9.3e-6	1164	37:44	9.3e-6
	5	87	3:12	9.7e-6	866	1:02:02	9.9e-6
	10	82	20:32	9.8e-6	870	1:18:47	9.8e-6
`Semismooth`	1	7	0:20	5.6e-7	14	1:01	3.2e-8
	5	8	0:27	1.7e-8	14	1:06	1.2e-8
	10	8	0:29	1.8e-8	14	0:58	1.7e-8
`IP-NCM`	1	11	1:15	7.5e-8	16	2:49	5.7e-8
	5	11	1:51	3.7e-8	14	3:10	4.3e-8
	10		`out of memory`			`out of memory`	
`Smoothing`	1	8	0:24	9.5e-9	14	1:04	8.0e-8
	5	8	0:30	3.7e-8	14	1:09	3.6e-8
	10	8	0:30	3.4e-8	14	0:56	5.3e-8

Table 3: Numerical results for Example 5.5

From the numerical results reported in Tables 1-6, we can see that for cases with equality constraints only (Tables 1-4), Newton-type methods (`Semimsooth`, `IP-NCM`, and `Smoothing`) can achieve higher accuracy than the gradient based methods (`GM` and `BGFS`). The `Smoothing` method is comparable to the well-tested `Semismooth` method in terms of cputime. Both the `Smoothing` and `Semismooth` methods are faster (2-8 times) than the `IP-NCM` method for achieving similar accuracies. For some cases, the `IP-NCM` method ran out of memory. The performance of the `BFGS` method varies from case to case. It is more competitive than the `IP-NCM` method for the tested unweighted nearest correlation matrix problems, but is much slower than the `IP-NCM` method for the weighted cases. The `GM` method is easy to implement, but it may converge extremely slow or not converge at all due to truncated numerical errors. For cases with both equality and inequality constraints (Tables 5-6), we only have three methods – `PGM`, `BFGS-SQP`, and `Smoothing` to test. The performance of the `BFGS-SQP` method is very poor partially because it has to solve a quadratic programming problem at each iteration and needs much more memory than the `PGM` and `Smoothing` methods. In all tested examples with both equality and inequality constraints, the `Smoothing` method is much more efficient.

6 Conclusions

In this paper, we proposed a quadratically convergent inexact smoothing Newton method – Algorithm 3.1 for solving the least squares covariance matrix problem (1). Our approach is to reformulate the problem as a smoothing-nonsmooth system of equations. Under mild conditions, we showed that the reformulated system enjoys a desirable non-singularity property, which is vital for applying the BiCGStab iterative solver to the resulted smoothing Newton linear system.

Example 5.6			**n=500**			**n=1000**			**n=2000**		
Method	*Case*	n_r	*Iter*	*cputime*	*Res*	*Iter*	*cputime*	*Res*	*Iter*	*cputime*	*Res*
GM	a)	1	122	3:35	9.4e-6	164	35:20	9.5e-6	222	5:43:47	9.6e-6
		5	177	5:26	1.0e-6	222	48:38	9.9e-6	282	7:21:23	1.0e-5
		10	216	6:58	9.9e-6	268	58:36	9.7e-6	336	8:41:07	1.0e-5
	b)	1	5000*	2:11:30	4.8e-3	2953	>10 hrs	2.0e-2	393	>10 hrs	3.1e-1
		5	5000*	2:57:42	7.1e-3	2785	>10 hrs	2.1e-2	249	>10 hrs	4.8e-1
		10	5000*	2:27:22	7.3e-3	2887	>10 hrs	2.3e-2	402	>10 hrs	3.0e-1
BFGS	a)	1	31	1:06	8.6e-6	35	8:42	8.1e-6	40	2:00:42	6.7e-6
		5	41	2:53	7.1e-6	out of memory			out of memory		
		10	out of memory			−−			−−		
	b)	1	551	18:33	9.9e-6	1363	5:07:11	9.7e-6	200	>10 hrs	5.8e-2
		5	1137	1:37:19	9.6e-6	out of memory			out of memory		
		10	out of memory			−−			−−		
Semismooth	a)	1	6	0:19	6.6e-8	6	2:05	2.3e-7	6	14:22	8.6e-7
		5	6	0:22	6.4e-8	6	2:26	3.7e-7	7	19:37	6.5e-9
		10	6	0:28	2.6e-7	6	2:33	8.4e-7	7	20:28	3.6e-8
	b)	1	13	0:40	9.2e-9	18	5:55	4.5e-7	14	33:59	1.2e-8
		5	13	1:03	6.5e-7	13	5:30	5.5e-8	14	1:03:28	6.1e-9
		10	13	1:02	7.3e-9	15	7:02	2.3e-7	16	50:09	9.4e-8
IP-NCM	a)	1	10	1:26	3.6e-8	10	8:54	4.4e-8	out of memory		
		5	10	2:24	4.8e-8	out of memory			−−		
		10	out of memory			−−			−−		
	b)	1	12	2:07	2.7e-8	13	13:59	3.9e-8	out of memory		
		5	12	4:03	9.0e-8	out of memory			−−		
		10	out of memory			−−			−−		
Smoothing	a)	1	6	0:20	1.4e-7	6	2:08	4.5e-7	7	17:14	6.7e-9
		5	6	0:23	1.4e-7	7	2:51	8.5e-9	7	19:52	3.4e-8
		10	6	0:25	5.6e-7	7	3:11	1.3e-8	7	21:18	1.2e-7
	b)	1	13	0:40	2.0e-8	18	6:01	1.0e-6	14	34:26	3.8e-8
		5	14	1:11	8.2e-9	13	5:35	1.3e-7	14	1:03:17	1.9e-8
		10	13	1:04	2.6e-8	15	7:21	7.6e-7	16	52:50	2.8e-7

Table 4: Numerical results for Example 5.6

		Example 5.7			Example 5.8		
Method	*Case*	*Iter*	*cputime*	*Res*	*Iter*	*cputime*	*Res*
PGM	a)	687	10:31	9.9e-6	5000*	1:08:37	3.3e-4
	b)	868	13:36	9.9e-6	5000*	1:05:38	4.2e-5
	c)	813	11:54	1.0e-5	5000*	1:12:19	4.5e-5
	d)	841	13:11	1.0e-5	5000*	1:18:23	7.9e-5
	e)	848	13:20	1.0e-5	5000*	1:24:16	7.6e-5
BFGS-SQP	a)	122	9:27	9.4e-6	399	1:11:31	9.9e-6
	b)	139	56:49	1.0e-5	415	5:47:26	9.7e-6
	c)		out of memory			out of memory	
	d)		−−			−−	
	d)		−−			−−	
Smoothing	a)	8	0:39	2.0e-8	12	1:01	1.0e-7
	b)	8	0:59	1.9e-8	12	0:54	1.5e-8
	c)	8	0:34	6.1e-8	18	1:32	1.4e-8
	d)	8	0:37	8.9e-8	18	2:38	7.9e-8
	e)	9	1:06	2.5e-7	25	3:26	7.7e-9

Table 5: Numerical results for Example 5.7 and Example 5.8

Example 5.9			**n=500**			**n=1000**			**n=2000**		
Method	*Case*	$\lvert\hat{n}_r$	*Iter*	*cputime*	*Res*	*Iter*	*cputime*	*Res*	*Iter*	*cputime*	*Res*
PGM	a)	1	105	3:26	9.6e-6	139	33:38	9.3e-6	189	5:33:36	9.5e-6
		5	131	4:27	9.9e-6	166	44:20	9.6e-6	213	6:23:57	9.5e-6
		10	150	5:16	9.7e-6	180	47:57	9.9e-6	229	7:02:37	9.7e-6
	b)	1	809	27:53	9.9e-6	1166	4:07:01	1.0e-5	382	>10 hrs	3.8e-2
		5	936	32:47	9.9e-6	1167	4:13:12	9.9e-6	372	>10 hrs	4.0e-2
		10	890	31:27	9.9e-6	1114	2:03:46	9.9e-6	371	>10 hrs	1.5e-2
BFGS-SQP	a)	1	44	5:46	9.5e-6	48	59:15	8.9e-6	out of memory		
		5	out of memory			out of memory			−−		
		10	−−			−−			−−		
	b)	1	123	29:57	9.6e-6	143	5:29:49	9.8e-6	out of memory		
		5	out of memory			out of memory			−−		
		10	−−			−−			−−		
Smoothing	a)	1	7	0:27	9.6e-9	8	3:26	3.3e-9	8	25:29	2.8e-8
		5	7	0:31	9.6e-9	8	3:51	8.8e-8	9	29:33	1.5e-8
		10	8	0:39	6.7e-9	9	4:46	6.8e-9	9	30:37	7.3e-7
	b)	1	8	0:34	2.7e-7	9	3:28	6.0e-8	10	28:03	1.6e-8
		5	9	0:40	2.1e-7	10	3:59	4.0e-8	10	29:49	4.3e-8
		10	10	0:49	9.4e-9	11	5:08	5.5e-8	13	46:39	1.5e-8

Table 6: Numerical results for Example 5.9

Our conducted numerical results[1] clearly demonstrated that Algorithm 3.1 is very efficient for solving the least squares covariance matrix problem with simple constraints (60), for which we introduced a simple diagonal preconditioner with a low cost.

References

[1] F. Alizadeh, J.-P. A. Haeberly, and M.L. Overton, *Complementarity and nondegenracy in semidefinite programming*, Mathematical Programming 77 (1997), pp. 111-128.

[2] L.-E. Andersson and T. Elfving, *An algorithm for constrained interpolation*, SIAM Journal on Scientific and Statistical Computing 8 (1987), pp. 1012–1025.

[3] V.I. Arnold, *On matrices depending on parameters*, Russian Mathematical Surveys 26 (1971), pp. 29–43.

[4] R. Bhatia, *Matrix Analysis*, Springer-Verlag (New York, 1997).

[5] J.F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.

[6] B. Borchers and J.G. Young, *Implementation of a primal-dual method for SDP on a shared memory parallel architecture*, Computational Optimization and Applications 37 (2007), pp. 355-369.

[7] R. Borsdorf, *A Newton algorithm for the nearest correlation matrix*, Master Thesis, School of Mathematics, University of Manchester, 2007.

[8] J. Borwein and A. S. Lewis, *Partially finite convex programming, part II: Explicit lattice models*, Mathematical Programming 57 (1992), pp. 49–83.

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

[10] Z.X. Chan and D.F. Sun, *Constraint nondegeneracy, strong regularity, and nonsingulairty in semidefinite programming*, SIAM Journal on Optimization 19 (2008), pp. 370–396.

[11] X.J. Chen, *Convergence of the BFGS method for LC^1 convex constrained optimization*, SIAM Journal on Control and Optimization 34 (1996), pp. 2051–2063.

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

[13] R.W. Cottle, J.-S. Pang, and R.E. Stone, *The Linear Complementarity Problem*, Aacdemic Press, Boston, 1992.

[14] F. Deutsch, *Best Approximation in Inner Product Spaces*, CMS Books in Mathematics 7, Springer, New York, 2001.

[1] The `MATLAB` code is available from `http://www.math.nus.edu.sg/~matsundf/`

[15] F. Deutsch, W. Li, and J.D. Ward, *A dual approach to constrained interpolation from a convex subset of Hilbert space*, Journal of Approximation Theory 90 (1997), pp. 385–414.

[16] A. L. Dontchev and B. D. Kalchev, *Duality and well-posedness in convex interpolation*, Numereical Functional Analysis and Optimization 10 (1989), pp. 673-689.

[17] A. L. Dontchev, H.-D. Qi, and L. Qi, *Convergence of Newton's method for convex best interpolation*, Numerische Mathematik 87 (2001), pp. 435–456.

[18] A. L. Dontchev, H.-D. Qi, and L. Qi, *Quadratic convergence of Newton's method for convex interpolation and smoothing*, Constructive Approximation 19 (2003), pp. 123–143.

[19] R.L. Dykstra, *An algorithm for restricted least squares regression*, Journal of the American Statistical Association 78 (1983), pp. 837–842.

[20] B.C. Eaves, *On the basic theorem for complemenarity*, Mathematical Programming 1 (1971), pp. 68–75.

[21] F. Facchinei and J.S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems: Volumes I & II*, Springer, New York, 2003.

[22] A. Fischer, *Solution of monotone complementarity problems with locally Lipschitzian functions*, Mathematical Programming 76 (1997), pp. 513–532.

[23] D. Henrion and J. Malick, *SDLS: a Matlab package for solving conic least-squares problems*, `http://www.optimization-online.org/DB_HTML/2007/09/1783.html`.

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

[25] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[26] L. D. Irvine, S. P. Marin and P. W. Smith, *Constrained interpolation and smoothing*, Constructive Approximation 2 (1986), pp. 129–151.

[27] B. Kummer, *Newton's method for nondifferentiable functions*, Advances in Mathematical Optimization, 114–125, Math. Res., 45, Akademie-Verlag, Berlin, 1988.

[28] K. Löwner, *Über monotone matrixfunktionen*, Mathematische Zeitschrift 38 (1934), pp. 177–216.

[29] J. Malick, *A dual approach to semidefinite least-squares problems*, SIAM Journal on Matrix Analysis and Applications 26 (2004), pp. 272–284.

[30] C. A. Micchelli, P. W. Smith, J. Swetits and J. D. Ward, *Constrained L_p approximation*, Constructive Approximation 1 (1985), pp. 93–102.

[31] C. A. Micchelli and F. I. Utreras, *Smoothing and interpolation in a convex subset of a Hilbert space,* SIAM Journal on Scientific and Statistical Computing 9 (1988), pp. 728–747.

[32] H.-D. Qi, *Some theoretical aspects on Newton's method for constrained best interpolation*, Continuous Optimization: Current Trends and Modern Applications, pp. 23–49, Springer, New York, 2005.

[33] H.-D. Qi, *Positive semidefinite matrix completions on chordal graphs and the constraint nondegeneracy in semidefinite programming*, Technical Report, School of Mathematics, University of Southampton, UK, 2008.

[34] 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.

[35] H.-D. Qi and D.F. Sun, *Correlation stress testing for value-at-risk: an unconstrained convex optimization approach*, Technical Report, Department of Mathematics, National University of Singapore, 2007.

[36] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Mathematics of Operations Research 18 (1993), pp. 227–244.

[37] L. Qi, D.F. Sun, and G. Zhou, *A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities*, Mathematical Programming 87 (2000), pp. 1-35.

[38] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Mathematical Programming 58 (1993), pp. 353–367.

[39] G. Ravindran and M.S. Gowda, *Regularization of P_0-functions in box variational inequality problems*, SIAM Journal on Optimization 11 (2000/01), pp. 748–760.

[40] S.M. Robinson, *Local structure of feasible sets in nonlinear programming, Part II: Nondegeneracy*, Mathematical Programming Study 22 (1984), pp. 217–230.

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

[42] R.T. Rockafellar, *Conjugate Duality and Optimization*, SIAM, Philadelphia, 1974.

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

[44] J.F. Sturm, *Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones*, Optimization Methods and Software 11 & 12 (1999), pp. 625–653.

[45] D.F. Sun and J. Sun, *Semismooth matrix valued functions*, Mathematics of Operations Research 27 (2002), pp. 150–169.

[46] J. Sun, D.F. Sun, and L. Qi, *A squared smoothing Newton method for nonsmooth matrix equations and its applications in semidefinite optimization problems*, SIAM Journal on Optimization 14 (2003), pp. 783–806.

[47] K.C. Toh, R.H. Tütüncü, and M.J. Todd, *Inexact primal-dual path-following algorithms for a special class of convex quadratic SDP and related problems*, Pacific Journal on Optimization 3 (2007), pp. 135–164.

[48] R.H. Tütüncü, K.C. Toh, and M.J. Todd, *Solving semidefinite-quadratic-linear programs using SDPT3*, Mathematical Programming 95 (2003), pp. 189–217.

[49] H.A. van der Vorst, *BI-CGSTAB: a fast and smoothly converging variant of BI-CG for the solution of nonsymmetric linear systems*, SIAM Journal on Scientific and Statistical Computing 13 (1992), pp. 631-644.

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