

A SQUARED SMOOTHING NEWTON METHOD FOR NONSMOOTH MATRIX EQUATIONS AND ITS APPLICATIONS IN SEMIDEFINITE OPTIMIZATION PROBLEMS*

JIE SUN[†], DEFENG SUN[‡], AND LIQUN QI[§]

Abstract. We study a smoothing Newton method for solving a nonsmooth matrix equation that includes semidefinite programming and the semidefinite complementarity problem as special cases. This method, if specialized for solving semidefinite programs, needs to solve only one linear system per iteration and achieves quadratic convergence under strict complementarity and nondegeneracy. We also establish quadratic convergence of this method applied to the semidefinite complementarity problem under the assumption that the Jacobian of the problem is positive definite on the affine hull of the critical cone at the solution. These results are based on the strong semismoothness and complete characterization of the B-subdifferential of a corresponding squared smoothing matrix function, which are of general theoretical interest.

Key words. matrix equations, Newton's method, nonsmooth optimization, semidefinite complementarity problem, semidefinite programming

AMS subject classifications. 65K05, 90C25, 90C33

DOI. 10.1137/S1052623400379620

1. Introduction.

1.1. Motivation. Let $\mathcal{S}(n_1, \dots, n_m)$ be the linear space of symmetric block-diagonal matrices with m blocks of sizes $n_k \times n_k$, $k = 1, \dots, m$, respectively, and let Ψ be a mapping from $\mathcal{S}(n_1, \dots, n_m)$ to $\mathcal{S}(n_1, \dots, n_m)$ itself. We consider the problem of finding a root of $\Psi(X) = 0$. This symmetric block-diagonal-matrix-valued equation problem (*matrix equation problem* for short) has many applications in optimization. For example, arising from Lyapunov stability analysis of systems under uncertainty [4, 23], we desire to know whether there exists an $n \times n$ symmetric matrix X such that the following system is feasible:

$$(1.1) \quad \begin{cases} \lambda X - (L_i X + X L_i) \succeq 0, & i = 1, \dots, k, \\ X - I \succeq 0, \end{cases}$$

where λ is a given constant, I, L_i , $i = 1, \dots, k$ are given $n \times n$ symmetric matrices, and for an arbitrary symmetric matrix Y we write $Y \succ 0$ and $Y \succeq 0$ if Y is positive definite and positive semidefinite, respectively. It is easy to convert (1.1) into a matrix equation problem. For $X \succeq 0$ we denote its symmetric square root by $X^{1/2}$. Let $|X| := (X^2)^{1/2}$ and $X_+ := (X + |X|)/2$ for any $X \in \mathcal{S}(n_1, \dots, n_m)$. Note that

*Received by the editors October 17, 2000; accepted for publication (in revised form) October 15, 2003; published electronically March 5, 2004.

<http://www.siam.org/journals/siopt/14-3/37962.html>

[†]School of Business and Singapore-MIT Alliance, National University of Singapore, Republic of Singapore (jsun@nus.edu.sg). The research of this author was partially supported by grant R314-000-028/042-112 of the National University of Singapore and a grant from Singapore-MIT Alliance.

[‡]Department of Mathematics, National University of Singapore, Republic of Singapore (matsundf@nus.edu.sg). The research of this author was partially supported by the Australian Research Council and grant R146-000-035-101 of the National University of Singapore.

[§]Department of Applied Mathematics, the Hong Kong Polytechnic University, Hong Kong, China (maqilq@polyu.edu.hk). The research of this author was supported by the Research Grant Council of Hong Kong.

$|X| - X = 0$ if and only if X is positive semidefinite. Let

$$\Psi(X) := \sum_{i=1}^k [|\lambda X - L_i X - X L_i| - \lambda X + L_i X + X L_i] + [|X - I| - X + I].$$

Then solving problem (1.1) is equivalent to solving the matrix equation $\Psi(X) = 0$. Note that this equation is not differentiable (in the sense of Fréchet), but is strongly semismooth [5, 32]. For the definition of semismooth matrix functions and some related topics see section 2 or references [5, 32] for more details.

Another application of matrix equations refers to semidefinite programming (SDP). As a modeling tool of optimization and a powerful relaxation form of some combinatorial optimization problems, SDP has received much attention in the research community in recent years. The website of semidefinite programming¹ contains a nice categorized list of papers in this area. Assuming strict feasibility of both primal and dual problems, a semidefinite program is equivalent to finding $X \succeq 0$, $S \succeq 0$, and $y \in \mathbb{R}^m$ such that

$$(1.2) \quad A_i \bullet X = b_i, \quad i = 1, \dots, m, \quad \sum_{i=1}^m y_i A_i + S = C, \quad X \bullet S = 0,$$

where \bullet denotes the matrix Frobenius inner product. It is shown by Tseng [35] that

$$(1.3) \quad X \succeq 0, \quad S \succeq 0, \quad X \bullet S = 0 \quad \iff \quad X - [X - S]_+ = 0.$$

Thus, system (1.2) can be rewritten as

$$(1.4) \quad A_i \bullet X = b_i, \quad i = 1, \dots, m, \quad \sum_{i=1}^m y_i A_i + S = C, \quad X - [X - S]_+ = 0,$$

which has the form of $\Psi(W) = 0$ with $W := \text{diag}(y_1, \dots, y_m, S, X)$ being a block-diagonal matrix.

A generalization of SDP—the semidefinite complementarity problem (SDCP)—can also be reformulated as a matrix equation. The SDCP is to find, for a given continuously differentiable mapping $F : \mathcal{S}(n_1, \dots, n_m) \rightarrow \mathcal{S}(n_1, \dots, n_m)$, an $X \in \mathcal{S}(n_1, \dots, n_m)$ such that

$$(1.5) \quad X \succeq 0, \quad F(X) \succeq 0, \quad X \bullet F(X) = 0.$$

By (1.3) this problem is equivalent to

$$(1.6) \quad X - [X - F(X)]_+ = 0.$$

A special case of the SDCP, where F is linear, was introduced by Kojima, Shindo, and Hara [19] and further studied in, e.g., [12, 13, 17, 18]. For the general (nonlinear) SDCP, Monteiro and Pang [21, 22] treated it as a constrained equation and introduced interior-point methods for solving the constrained equation. Tseng [35] introduced merit functions to reformulate the SDCP as an optimization problem. Chen and Tseng [6] studied noninterior continuation methods for solving the SDCP. Kanzow and Nagel [15] analyzed smoothing paths for the Karush–Kuhn–Tucker (KKT) system of

¹<http://www.zib.de/helmberg/semidef.html>

the SDP and proposed smoothing-type methods for solving the KKT system. Pang, Sun, and Sun [24] studied semismooth homeomorphisms and strong stability of the SDCP.

The interest in the nonlinear SDCP stems from the research on nonlinear semidefinite optimization problems. Shapiro [29] studied first- and second-order perturbation analysis of nonlinear semidefinite optimization problems. Jarre [14] gave an interior-point method for solving nonconvex semidefinite programs. Fares, Noll, and Apkarian [7] investigated a sequential SDP approach for a variety of problems in optimal control, which can be cast as minimizing a linear objective function subject to linear matrix inequality constraints and nonlinear matrix equality constraints. Leibfritz and Mostafa [20] proposed an interior-point constrained trust-region method for a special class of nonlinear SDP problems. Tseng [36] conducted a convergence analysis for an infeasible interior-point trust-region method for nonlinear semidefinite programs.

In this paper we study a smoothing Newton method for solving a nonsmooth matrix equation that includes the SDP and the SDCP as special cases. In particular, for the SDP, this method achieves quadratic convergence under strict complementarity and nondegeneracy. For the SDCP, quadratic convergence is proved under the condition that the Jacobian of the problem is positive definite on the affine hull of the critical cone at the solution. The strict complementarity condition is not assumed here. To establish these results, we investigate the strong semismoothness and the Bouligand-subdifferential (B-subdifferential) of the so-called squared smoothing matrix function, which are of their own theoretical interest.

The study on smoothing Newton methods can be traced back to a nonsmooth version of Newton’s method by Qi and Sun [27] for solving nonsmooth vector valued equations. It was later found that smoothing techniques could be applied to the nonsmooth Newton method to improve its computational performance. Many researchers have contributed to this area, see, for example, [11] and the references therein. The basic idea of the smoothing Newton method is to replace the nonsmooth equation $\Psi(X) = 0$ by a smoothing equation $G(\varepsilon, X) = 0$, where $G : \mathbb{R} \times \mathcal{S}(n_1, \dots, n_m) \rightarrow \mathcal{S}(n_1, \dots, n_m)$, such that

$$G(\varepsilon, Y) \rightarrow \Psi(X) \quad \text{as } (\varepsilon, Y) \rightarrow (0, X).$$

Here the function G is required to be continuously differentiable at (ε, X) unless $\varepsilon = 0$. The classical damped Newton method can then be used to solve $G(\varepsilon, X) = 0$ as $\varepsilon \downarrow 0$ to get a solution of $\Psi(X) = 0$. Computational results show that this type of method is quite efficient in solving vector complementarity problems [37].

For $\varepsilon \in \mathbb{R}$ and $X \in \mathcal{S}(n_1, \dots, n_m)$, the *squared smoothing function* $\Phi : \mathbb{R} \times \mathcal{S}(n_1, \dots, n_m) \rightarrow \mathcal{S}(n_1, \dots, n_m)$ is defined by

$$(1.7) \quad \Phi(\varepsilon, X) := (\varepsilon^2 I + X^2)^{1/2}, \quad (\varepsilon, X) \in \mathbb{R} \times \mathcal{S}(n_1, \dots, n_m).$$

Then, Φ is continuously differentiable at (ε, X) unless $\varepsilon = 0$, and for any $X \in \mathcal{S}(n_1, \dots, n_m)$,

$$[Y + \Phi(\varepsilon, Y)]/2 \rightarrow X_+ \quad \text{as } (\varepsilon, Y) \rightarrow (0, X).$$

Thus we can use Φ to construct smoothing functions for nonsmooth systems (1.4) and (1.6). We show that the smoothing function

$$(1.8) \quad G(\varepsilon, X) := X - [X - F(X) + \Phi(\varepsilon, X - F(X))] / 2$$

can be used to design a quadratically convergent algorithm for (1.4) and (1.6). We note that Chen and Tseng [6] have developed a nice smoothing Newton method for the SDCP and reported promising computational results. The difference between our paper and theirs is that we show the strong semismoothness of the smoothing function, which can be utilized to establish quadratic convergence, whereas paper [6] did not prove the strong semismoothness of the smoothing function. As a result, paper [6] needs the strict complementarity assumption and the convergence rate proved there is only superlinear, whereas we obtain quadratic rate of convergence without this assumption for the SDCP.

1.2. Notation and organization of the paper. The notation used is fairly standard. Generally, we use calligraphic letters for sets, capital letters for matrices and matrix functions, lowercase letters for vectors, and Greek letters for scalars and index sets, respectively. A diagonal matrix is denoted by $\text{diag}(\lambda_1, \dots, \lambda_n)$, where $\lambda_1, \dots, \lambda_n$ are the diagonal entries. Similarly, a block-diagonal matrix is written as $\text{diag}(B_1, \dots, B_m)$ with B_1, \dots, B_m being the block matrices.

Let α and β be two sets of indices. We designate by $A_{\alpha\beta}$ the submatrix of A whose row indices belong to α and whose column indices belong to β . In particular, A_{ij} stands for the (i, j) th entry of A . For matrices $A, B \in \mathcal{S}(n_1, \dots, n_m)$, the Frobenius inner product is defined as

$$A \bullet B := \text{Trace}(A^T B) = \text{Trace}(AB).$$

Consequently, the Frobenius norm of $A \in \mathcal{S}(n_1, \dots, n_m)$ is

$$\|A\| := (A \bullet A)^{1/2}.$$

The Hadamard product of A and B is denoted by $A \circ B$, namely, $(A \circ B)_{ij} := A_{ij}B_{ij}$ for all i and j . The 2-norm of a vector x is denoted by $\|x\|$. Let I be the identity matrix of appropriate dimension.

This paper is organized as follows. In section 2 we review some results on nonsmooth matrix functions and prove the strong semismoothness of Φ defined in (1.7). Section 3 is devoted to characterizing the B-subdifferential of Φ , which will be used in the sequel. We describe the squared smoothing Newton method in section 4. Applications of the smoothing Newton method to the SDP and SDCP are discussed in sections 5 and 6, respectively. Some final remarks are given in section 7.

2. Strong semismoothness of $\Phi(\varepsilon, X)$. This section is devoted to proving the strong semismoothness of the squared smoothing function Φ defined by (1.7). As a preparation we introduce some basic definitions and results on a general matrix function $\Psi : \mathcal{S}(n_1, \dots, n_m) \rightarrow \mathcal{S}_1$, where \mathcal{S}_1 is also a symmetric block-diagonal matrix space, but could be of different shape and size from $\mathcal{S}(n_1, \dots, n_m)$.

Suppose that $\Psi : \mathcal{S}(n_1, \dots, n_m) \rightarrow \mathcal{S}_1$ is a locally Lipschitz matrix function. According to [32], Ψ is differentiable almost everywhere. Denote the set of points at which Ψ is differentiable by D_Ψ and for any $X \in D_\Psi$, let $J\Psi(X)$ denote the Jacobian of Ψ at X . Let $\partial_B\Psi(X)$ be the B-subdifferential of Ψ at X defined by

$$(2.1) \quad \partial_B\Psi(X) = \left\{ \lim_{\substack{X^k \rightarrow X \\ X^k \in D_\Psi}} J\Psi(X^k) \right\},$$

and let $\partial\Psi(X)$ denote the convex hull of $\partial_B\Psi(X)$.

DEFINITION 2.1. Suppose that $\Psi : \mathcal{S}(n_1, \dots, n_m) \rightarrow \mathcal{S}_1$ is a locally Lipschitz matrix function. Ψ is said to be semismooth at $X \in \mathcal{S}(n_1, \dots, n_m)$ if Ψ is directionally differentiable at X and for any $V \in \partial\Psi(X + H)$ and $H \in \mathcal{S}(n_1, \dots, n_m)$,

$$\Psi(X + H) - \Psi(X) - V(H) = o(\|H\|).$$

Ψ is said to be strongly semismooth at X if Ψ is semismooth at X and

$$(2.2) \quad \Psi(X + H) - \Psi(X) - V(H) = O(\|H\|^2).$$

Instead of showing the strong semismoothness by definition, we will use the following result [32, Theorem 3.6].

THEOREM 2.2. Suppose that $\Psi : \mathcal{S}(n_1, \dots, n_m) \rightarrow \mathcal{S}_1$ is locally Lipschitz and directionally differentiable in a neighborhood of X . Then Ψ is strongly semismooth at X if and only if for any $X + H \in D_\Psi$,

$$(2.3) \quad \Psi(X + H) - \Psi(X) - J\Psi(X + H)(H) = O(\|H\|^2).$$

In order to show that $\Phi(\varepsilon, X)$ satisfies (2.3), we will first identify the differentiable points of Φ . We shall show that Φ is differentiable at (ε, X) if and only if $\varepsilon^2 I + X^2$ is nonsingular. Here we view Φ as a function from $\mathcal{S}(1, n)$ to $\mathcal{S} \equiv \mathcal{S}(n)$. This result easily can be extended to the general block-diagonal case. Unless stated otherwise, \mathcal{S} is assumed to be of this simple structure here and below.

For any $X \in \mathcal{S}$, let L_X be the Lyapunov operator

$$L_X(Y) := XY + YX \quad \forall Y \in \mathcal{S}$$

with L_X^{-1} being its inverse (if it exists at all).

For $X \in \mathcal{S}$, there exist an orthogonal matrix P and a diagonal matrix $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ of eigenvalues of X such that

$$(2.4) \quad X = P\Lambda P^T.$$

Define three index sets associated with the eigenvalues of matrix X :

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

By permuting the rows and columns of X if necessary, we assume that Λ can be written as

$$\Lambda = \begin{bmatrix} \Lambda_\alpha & 0 & 0 \\ 0 & \Lambda_\gamma & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

where Λ_α and Λ_γ are diagonal matrices with diagonal elements $\lambda_i, i \in \alpha$ and $\lambda_i, i \in \gamma$, respectively. Let $\kappa := \alpha \cup \gamma$. Define two diagonal matrices of order $|\kappa|$:

$$D := \begin{bmatrix} \Lambda_\alpha & 0 \\ 0 & \Lambda_\gamma \end{bmatrix}$$

and $|D| = (D^2)^{1/2}$, i.e.,

$$|D| = \begin{bmatrix} \Lambda_\alpha & 0 \\ 0 & |\Lambda_\gamma| \end{bmatrix}.$$

LEMMA 2.3. For $(\varepsilon, X) \in \mathbb{R} \times \mathcal{S}$, the following statements hold.

- (a) If $\varepsilon^2 I + X^2$ is nonsingular, then Φ is continuously differentiable at (ε, X) and $J\Phi(\varepsilon, X)$ satisfies the following equation:

$$(2.5) \quad J\Phi(\varepsilon, X)(\tau, H) = L_{\Phi(\varepsilon, X)}^{-1}(L_X(H) + 2\varepsilon\tau I) \quad \forall (\tau, H) \in \mathbb{R} \times \mathcal{S}.$$

In particular, in this case,

$$(2.6) \quad \|J\Phi(\varepsilon, X)(\tau, H)\| \leq \sqrt{n}|\tau| + \|H\|.$$

- (b) Φ is globally Lipschitz continuous and for any $(\varepsilon, X), (\tau, Y) \in \mathbb{R} \times \mathcal{S}$,

$$(2.7) \quad \|\Phi(\varepsilon, X) - \Phi(\tau, Y)\| \leq \sqrt{n}|\varepsilon - \tau| + \|X - Y\|.$$

- (c) Φ is directionally differentiable at $(0, X)$ and for $(\tau, H) \in \mathbb{R} \times \mathcal{S}$,

$$\Phi'((0, X); (\tau, H)) = P \begin{bmatrix} L_{|D|}^{-1}[D\tilde{H}_{\kappa\kappa} + \tilde{H}_{\kappa\kappa}D] & |D|^{-1}D\tilde{H}_{\kappa\beta} \\ \tilde{H}_{\kappa\beta}^T D|D|^{-1} & (\tau^2 I + \tilde{H}_{\beta\beta}^2)^{1/2} \end{bmatrix} P^T,$$

where $\tilde{H} := P^T H P$.

- (d) Φ is differentiable at (ε, X) if and only if $\varepsilon^2 I + X^2$ is nonsingular.

Proof. (a) For any $C \succ 0$, we have, by applying [35, Lemma 6.2] or direct calculation, that $(C^2 + W)^{1/2} - C = L_C^{-1}(W) + o(\|W\|)$ for all $W \in \mathcal{S}$ sufficiently small. Then, for $\varepsilon^2 I + X^2$ nonsingular (and hence positive definite), we have that

$$\begin{aligned} \Phi(\varepsilon + \tau, X + H) - \Phi(\varepsilon, X) &= (C^2 + W)^{1/2} - C \\ &= L_C^{-1}(L_X(H) + 2\varepsilon\tau I) + O(\tau^2 + \|H\|^2) + o(\|W\|), \end{aligned}$$

where $(\tau, H) \in \mathbb{R} \times \mathcal{S}$, $C := \Phi(\varepsilon, X)$, and $W := L_X(H) + 2\varepsilon\tau I + \tau^2 I + H^2$. Thus, Φ is differentiable at (ε, X) and

$$J\Phi(\varepsilon, X)(\tau, H) = L_C^{-1}(L_X(H) + 2\varepsilon\tau I).$$

By noting the fact that for all $(\varepsilon + \tau, X + H)$ sufficiently close to (ε, X) , $\Phi(\varepsilon + \tau, X + H)$ is positive definite, from the definition of L_{Φ}^{-1} we know that L_{Φ}^{-1} is continuous at (ε, X) . Hence, Φ is continuously differentiable at (ε, X) .

Let P and Λ be defined as in (2.4). To prove (2.6), we first note that

$$L_X(H) + 2\varepsilon\tau I = P(L_{\Lambda}(P^T H P) + 2\varepsilon\tau I)P^T,$$

and for any $Y \in \mathcal{S}$,

$$L_C^{-1}(Y) = PL_{\Phi(\varepsilon, \Lambda)}^{-1}(P^T Y P)P^T.$$

Thus, we have

$$P^T J\Phi(\varepsilon, X)(\tau, H)P = L_{\Phi(\varepsilon, \Lambda)}^{-1}(L_{\Lambda}(P^T H P) + 2\varepsilon\tau I).$$

Hence, by direct calculation, for $i, j = 1, \dots, n$,

$$(P^T J\Phi(\varepsilon, X)(\tau, H)P)_{ij} = \begin{cases} (P^T H P)_{ij}(\lambda_i + \lambda_j) \left(\sqrt{\varepsilon^2 + \lambda_i^2} + \sqrt{\varepsilon^2 + \lambda_j^2}\right)^{-1} & \text{if } i \neq j, \\ (\lambda_i(P^T H P)_{ii} + \varepsilon\tau) (\varepsilon^2 + \lambda_i^2)^{-1/2} & \text{otherwise,} \end{cases}$$

which implies that

$$\sum_{i,j=1}^n \left((P^T J\Phi(\varepsilon, X)(\tau, H)P)_{ij} \right)^2 \leq n\tau^2 + \sum_{i,j=1}^n \left((P^T HP)_{ij} \right)^2.$$

Hence,

$$\begin{aligned} \| J\Phi(\varepsilon, X)(\tau, H) \|^2 &= \| P^T J\Phi(\varepsilon, X)(\tau, H)P \|^2 \\ &\leq n\tau^2 + \| P^T HP \|^2 = n\tau^2 + \| H \|^2. \end{aligned}$$

This completes the proof of part (a).

(b) By part (a) of this lemma, for $\varepsilon \neq 0$ and $\tau \neq 0$ we have that

$$\begin{aligned} &\| \Phi(\varepsilon, X) - \Phi(\tau, Y) \| \\ &= \| \Phi(|\varepsilon|, X) - \Phi(|\tau|, Y) \| \\ &= \left\| \int_0^1 J\Phi(|\tau| + t(|\varepsilon| - |\tau|), Y + t(X - Y))(|\varepsilon| - |\tau|, X - Y) dt \right\| \\ &\leq \sqrt{n} (|\varepsilon| - |\tau|) + \| X - Y \| \\ &\leq \sqrt{n} |\varepsilon - \tau| + \| X - Y \|. \end{aligned}$$

By a limiting process the above inequality is also true for $\varepsilon\tau = 0$. Hence, (2.7) holds.

(c) Let P and Λ be defined as in (2.4). For any $\tau \in \mathbb{R}$, $H \in \mathcal{S}$, and $t \in [0, \infty)$, let

$$\Delta(t) := \Phi(t\tau, X + tH) - \Phi(0, X)$$

and

$$\tilde{\Delta}(t) := P^T \Delta(t) P.$$

Then,

$$\begin{aligned} \tilde{\Delta}(t) &= P^T \Phi(t\tau, X + tH)P - P^T \Phi(0, X)P \\ &= (t^2\tau^2 I + (P^T(X + tH)P)^2)^{1/2} - |P^T X P| \\ &= (t^2\tau^2 I + (P^T X P + tP^T H P)^2)^{1/2} - |P^T X P| \\ &= (t^2\tau^2 I + (\Lambda + t\tilde{H})^2)^{1/2} - |\Lambda|, \end{aligned}$$

where $\tilde{H} := P^T H P$. Thus,

$$\tilde{\Delta}(t) = \left(|\Lambda|^2 + \tilde{W} \right)^{1/2} - |\Lambda|,$$

where

$$\tilde{W} := t^2\tau^2 I + t\Lambda\tilde{H} + t\tilde{H}\Lambda + t^2\tilde{H}^2$$

and

$$|\Lambda| = \begin{bmatrix} |D| & 0 \\ 0 & 0 \end{bmatrix}.$$

After simple computations we have that

$$(2.8) \quad \widetilde{W} = t \begin{bmatrix} D\widetilde{H}_{\kappa\kappa} + \widetilde{H}_{\kappa\kappa}D & D\widetilde{H}_{\kappa\beta} \\ \widetilde{H}_{\kappa\beta}^T D & 0 \end{bmatrix} + \begin{bmatrix} O(t^2) & O(t^2) \\ O(t^2) & t^2\tau^2 I + t^2[\widetilde{H}_{\kappa\beta}^T \widetilde{H}_{\kappa\beta} + \widetilde{H}_{\beta\beta}^2] \end{bmatrix}.$$

By Lemma 6.2 in Tseng [35], we have that

$$(2.9) \quad \widetilde{\Delta}(t)_{\kappa\kappa} = L_{|D|}^{-1}(\widetilde{W}_{\kappa\kappa}) + o(\|\widetilde{W}\|),$$

$$(2.10) \quad \widetilde{\Delta}(t)_{\kappa\beta} = |D|^{-1}\widetilde{W}_{\kappa\beta} + o(\|\widetilde{W}\|),$$

and

$$(2.11) \quad \widetilde{W}_{\beta\beta} = \widetilde{\Delta}(t)_{\kappa\beta}^T \widetilde{\Delta}(t)_{\kappa\beta} + \widetilde{\Delta}(t)_{\beta\beta}^2.$$

Hence,

$$(2.12) \quad \widetilde{\Delta}(t)_{\kappa\beta} = t|D|^{-1}D\widetilde{H}_{\kappa\beta} + o(t),$$

which implies that

$$(2.13) \quad \widetilde{\Delta}(t)_{\kappa\beta}^T \widetilde{\Delta}(t)_{\kappa\beta} = t^2 \widetilde{H}_{\kappa\beta}^T (|D|^{-1}D)^2 \widetilde{H}_{\kappa\beta} + o(t^2) = t^2 \widetilde{H}_{\kappa\beta}^T \widetilde{H}_{\kappa\beta} + o(t^2).$$

According to (2.9) and (2.8),

$$(2.14) \quad \widetilde{\Delta}(t)_{\kappa\kappa} = tL_{|D|}^{-1}(D\widetilde{H}_{\kappa\kappa} + \widetilde{H}_{\kappa\kappa}D) + o(t).$$

Since

$$\widetilde{W}_{\beta\beta} = t^2\tau^2 I + t^2[\widetilde{H}_{\kappa\beta}^T \widetilde{H}_{\kappa\beta} + \widetilde{H}_{\beta\beta}^2],$$

from (2.11) and (2.13), we obtain that

$$(2.15) \quad \widetilde{\Delta}(t)_{\beta\beta}^2 = t^2\tau^2 I + t^2 \widetilde{H}_{\beta\beta}^2 + o(t^2).$$

Furthermore, since $\widetilde{\Delta}(\tau)_{\beta\beta}$ is positive semidefinite (see the definition of $\widetilde{\Delta}(t)$), we know from (2.15) that $\widetilde{\Delta}(t)_{\beta\beta}$ is well defined and

$$(2.16) \quad \widetilde{\Delta}(t)_{\beta\beta} = t \left(\tau^2 I + \widetilde{H}_{\beta\beta}^2 + o(1) \right)^{1/2}.$$

Hence, from (2.14), (2.12), (2.16), and the continuity of $(\cdot)^{1/2}$,

$$\lim_{t \downarrow 0} \frac{\widetilde{\Delta}(t)}{t} = \begin{bmatrix} L_{|D|}^{-1}[D\widetilde{H}_{\kappa\kappa} + \widetilde{H}_{\kappa\kappa}D] & |D|^{-1}D\widetilde{H}_{\kappa\beta} \\ \widetilde{H}_{\kappa\beta}^T D|D|^{-1} & (\tau^2 I + \widetilde{H}_{\beta\beta}^2)^{1/2} \end{bmatrix},$$

which completes the proof of part(c).

(d) Only the “only if” part needs a proof. Obviously $\varepsilon^2 I + X^2$ is nonsingular at $\varepsilon \neq 0$. If Φ is differentiable at $(0, X)$, then part (c) of this lemma shows that $\Phi'((0, X); (\tau, H))$ is a linear function of (τ, H) only if $\beta = \emptyset$; i.e., only if X is nonsingular. \square

Lemma 2.3 shows that the squared smoothing matrix function Φ is directionally differentiable everywhere and globally Lipschitz continuous. It also shows that it is differentiable at $(\varepsilon, X) \in \mathbb{R} \times \mathcal{S}$ if and only if $\varepsilon^2 I + X^2$ is nonsingular.

The next result is vital in order to prove the strong semismoothness of Φ . By noting the fact that I and X can be simultaneously diagonalized, we may extend the proof used in [32, Lemma 4.12] from $|X|$ to Φ . Here we follow the outline of a simpler proof given in [5, Proposition 4.10].

LEMMA 2.4. *Let $X \in \mathcal{S}$. Then, for any $\tau \in \mathbb{R}$ and $H \in \mathcal{S}$ such that $\tau^2 I + (X + H)^2$ is nonsingular, Φ is differentiable at $(\tau, X + H)$ and*

$$(2.17) \quad \Phi(\tau, X + H) - \Phi(0, X) - J\Phi(\tau, X + H)(\tau, H) = O(\|\Delta Z\|^2),$$

where $\Delta Z := (\tau, H)$.

Proof. Let \mathcal{D} denote the space of $n \times n$ real diagonal matrices with nonincreasing diagonal entries. For each $Y \in \mathcal{S}$, define

$$\mathcal{O}_Y := \{P \in \mathcal{O} : P^T Y P \in \mathcal{D}\},$$

where $\mathcal{O} := \{P \in \mathbb{R}^{n \times n} : P^T P = I\}$.

Let $\lambda_1 \geq \dots \geq \lambda_n$ denote the eigenvalues of X . By [6, Lemma 3] or [33, Proposition 4.4], there exist scalars $\eta > 0$ and $\rho > 0$ such that

$$\min_{P \in \mathcal{O}_X} \|P - Q\| \leq \eta \|Y - X\| \quad \text{whenever } Y \in \mathcal{S}, \|Y - X\| \leq \rho, Q \in \mathcal{O}_Y.$$

If $\tau = 0$, then the left-hand side of (2.17) reduces to $\Psi(X + H) - \Psi(X) - J\Psi(X + H)(H)$, where for each $Y \in \mathcal{S}$, $\Psi(Y) := |Y|$. Then, it follows from [32, Lemma 4.12] that (2.17) holds.

Suppose $\tau \neq 0$. Let $\mu_1 \geq \dots \geq \mu_n$ denote the eigenvalues of $X + H$, and choose any $Q \in \mathcal{O}_{X+H}$. Then, by (2.18), there exists $P \in \mathcal{O}_X$ satisfying

$$\|P - Q\| \leq \eta \|H\|.$$

For simplicity, let R denote the left-hand side of (2.17), i.e.,

$$R := \Phi(\tau, X + H) - \Phi(0, X) - J\Phi(\tau, X + H)(\tau, H).$$

Letting $C := \Phi(\tau, X + H) = (\tau^2 I + (X + H)^2)^{1/2}$ and noting that $Q \in \mathcal{O}_C$, we obtain from Lemma 2.3 and the formula for L_C^{-1} given in [35, Page 171] that

$$\begin{aligned} J\Phi(\tau, X + H)(\tau, H) &= L_C^{-1}[(X + H)H + H(X + H) + 2\tau^2 I] \\ &= Q[\Xi \circ (Q^T((X + H)H + H(X + H))Q + 2\tau^2 I)]Q^T, \end{aligned}$$

where the matrix $\Xi \in \mathcal{S}$ has entries

$$\Xi_{ij} = 1/(\theta_i + \theta_j)$$

and $\theta_i = \sqrt{\tau^2 + \mu_i^2}$ is the i th eigenvalue of C . Then, letting $\tilde{R} := Q^T R Q$ and $\tilde{H} := Q^T H Q$, we have that

$$(2.18) \quad \tilde{R} = \Sigma - S^T \Lambda S - \Xi \circ (U + 2\tau^2 I),$$

where $\Sigma := \text{diag}(\sqrt{\tau^2 + \mu_1^2}, \dots, \sqrt{\tau^2 + \mu_n^2})$, $\Lambda := \text{diag}(\lambda_1, \dots, \lambda_n)$, $S := P^T Q$, and $U_{ij} := (\mu_i + \mu_j)\tilde{H}_{ij}$ for all i, j .

Since $\text{diag}(\mu_1, \dots, \mu_n) = Q^T(X+H)Q = S^T \text{diag}(\lambda_1, \dots, \lambda_n)S + \tilde{H}$, we have that

$$(2.19) \quad \sum_{k=1}^n S_{ki}S_{kj}\lambda_k + \tilde{H}_{ij} = \begin{cases} \mu_i & \text{if } i = j \\ 0 & \text{else,} \end{cases} \quad i, j = 1, \dots, n.$$

Since $S = P^T Q = (P - Q)^T Q + I$ and $\|P - Q\| \leq \eta\|H\|$, it follows that

$$(2.20) \quad S_{ij} = O(\|H\|) \quad \forall i \neq j.$$

Since $P, Q \in \mathcal{O}$, we have $S \in \mathcal{O}$ so that $S^T S = I$. This implies

$$(2.21) \quad 1 = S_{ii}^2 + \sum_{k \neq i} S_{ki}^2 = S_{ii}^2 + O(\|H\|^2), \quad i = 1, \dots, n,$$

and

$$(2.22) \quad \begin{aligned} 0 &= S_{ii}S_{ij} + S_{ij}S_{jj} + \sum_{k \neq i, j} S_{ki}S_{kj} \\ &= S_{ii}S_{ij} + S_{ji}S_{jj} + O(\|H\|^2) \quad \forall i \neq j. \end{aligned}$$

We now show that $\tilde{R} = O(\|\Delta Z\|^2)$, which, by $\|R\| = \|\tilde{R}\|$, would prove (2.17). For any $i \in \{1, \dots, n\}$, we have from (2.18) and (2.19) that

$$(2.23) \quad \begin{aligned} \tilde{R}_{ii} &= \sqrt{\tau^2 + \mu_i^2} - \sum_{k=1}^n S_{ki}^2|\lambda_k| - \frac{1}{2\theta_i}(2\tau^2 + 2\mu_i\tilde{H}_{ii}) \\ &= \sqrt{\tau^2 + \mu_i^2} - \sum_{k=1}^n S_{ki}^2|\lambda_k| - \frac{\tau^2}{\theta_i} - \frac{\mu_i}{\theta_i} \left(\mu_i - \sum_{k=1}^n S_{ki}^2\lambda_k \right) \\ &= \sqrt{\tau^2 + \mu_i^2} - S_{ii}^2|\lambda_i| - \frac{\tau^2}{\theta_i} - \frac{\mu_i}{\theta_i}(\mu_i - S_{ii}^2\lambda_i) + O(\|H\|^2) \\ &= \sqrt{\tau^2 + \mu_i^2} - (1 + O(\|H\|^2))|\lambda_i| - \frac{\tau^2}{\theta_i} - \frac{\mu_i}{\theta_i}(\mu_i - (1 + O(\|H\|^2))\lambda_i) + O(\|H\|^2) \\ &= \sqrt{\tau^2 + \mu_i^2} - |\lambda_i| - \frac{\tau^2}{\theta_i} - \frac{\mu_i}{\theta_i}(\mu_i - \lambda_i) + O(\|H\|^2) \\ &= f(\tau, \mu_i) - f(0, \lambda_i) - Jf(\tau, \mu_i)(\tau, \mu_i - \lambda_i) + O(\|H\|^2), \end{aligned}$$

where the third and fifth equalities use (2.20), (2.21), and the fact that $|\mu_i/\theta_i| \leq 1$. The last equality follows by defining $f(\tau, \mu) := \sqrt{\tau^2 + \mu^2}$. Since f is known to be strongly semismooth and, by a result of Weyl [2, page 63],

$$(2.24) \quad |\mu_i - \lambda_i| \leq \|H\| \quad \forall i,$$

the right-hand side of (2.23) is $O(\|\Delta Z\|^2)$. For any $i, j \in \{1, \dots, n\}$ with $i \neq j$, we have from (2.18) and (2.19) that

$$\tilde{R}_{ij} = - \sum_{k=1}^n S_{ki}S_{kj}|\lambda_k| - \Xi_{ij}(\mu_i + \mu_j)\tilde{H}_{ij}$$

$$\begin{aligned}
 &= -\sum_{k=1}^n S_{ki}S_{kj}|\lambda_k| + \Xi_{ij}(\mu_i + \mu_j) \sum_{k=1}^n S_{ki}S_{kj}\lambda_k \\
 &= -(S_{ii}S_{ij}|\lambda_i| + S_{ji}S_{jj}|\lambda_j|) + \Xi_{ij}(\mu_i + \mu_j)(S_{ii}S_{ij}\lambda_i + S_{ji}S_{jj}\lambda_j) + O(\|H\|^2) \\
 &= -((S_{ii}S_{ij} + S_{ji}S_{jj})|\lambda_i|) + S_{ji}S_{jj}(|\lambda_j| - |\lambda_i|) \\
 &\quad + \Xi_{ij}(\mu_i + \mu_j)((S_{ii}S_{ij} + S_{ji}S_{jj})\lambda_i + S_{ji}S_{jj}(\lambda_j - \lambda_i)) + O(\|H\|^2) \\
 &= -S_{ji}S_{jj}(|\lambda_j| - |\lambda_i| - \Xi_{ij}(\mu_i + \mu_j)(\lambda_j - \lambda_i)) + O(\|H\|^2) \\
 (2.25) \quad &= -S_{ji}S_{jj} \left(|\lambda_j| - |\lambda_i| - \frac{\mu_j + \mu_i}{\theta_j + \theta_i}(\lambda_j - \lambda_i) \right) + O(\|H\|^2),
 \end{aligned}$$

where the third and fifth equalities use (2.20), (2.22), and $\Xi_{ij}|\mu_i + \mu_j| \leq 1$. We have that

$$\begin{aligned}
 &|\lambda_j| - |\lambda_i| - \frac{\mu_j + \mu_i}{\theta_j + \theta_i}(\lambda_j - \lambda_i) \\
 &= |\lambda_j| - |\lambda_i| - \frac{\mu_j + \mu_i}{\theta_j + \theta_i}(\mu_j - \mu_i) - \frac{\mu_j + \mu_i}{\theta_j + \theta_i}(\lambda_j - \mu_j + \mu_i - \lambda_i) \\
 &= |\lambda_j| - |\lambda_i| - \frac{(\tau^2 + \mu_j^2) - (\tau^2 + \mu_i^2)}{\sqrt{\tau^2 + \mu_j^2} + \sqrt{\tau^2 + \mu_i^2}} - \frac{\mu_j + \mu_i}{\theta_j + \theta_i}(\lambda_j - \mu_j + \mu_i - \lambda_i) \\
 (2.26) \quad &= |\lambda_j| - |\lambda_i| - \left(\sqrt{\tau^2 + \mu_j^2} - \sqrt{\tau^2 + \mu_i^2} \right) - \frac{\mu_j + \mu_i}{\theta_j + \theta_i}(\lambda_j - \mu_j + \mu_i - \lambda_i).
 \end{aligned}$$

Since $|\mu_j + \mu_i|/(\theta_j + \theta_i) \leq 1$ and $||\lambda_k| - \sqrt{\tau^2 + \mu_k^2}| = | \|(0, \lambda_k)\| - \|(\tau, \mu_k)\| | \leq |(0, \lambda_k) - (\tau, \mu_k)| \leq |\tau| + |\lambda_k - \mu_k|$ for $k \in \{i, j\}$, we see from (2.24) that the right-hand side of (2.26) is $O(|\tau| + \|H\|)$. This, together with (2.20), implies the right-hand side of (2.25) is $O(\|H\|(|\tau| + \|H\|))$. The proof is completed. \square

According to Theorem 2.2 and Lemmas 2.3 and 2.4, we obtain the following main result of this section.

THEOREM 2.5. *The squared smoothing matrix function Φ is strongly semismooth at $(0, X) \in \mathbb{R} \times \mathcal{S}$.*

The theorem above provides a basis for quadratic convergence of the squared smoothing Newton method for the SDCP, which is to be discussed in section 5.

3. Properties of the B-subdifferential of Φ . In this section, we shall discuss some properties of the B-subdifferential of the squared smoothing function Φ at $(0, X) \in \mathbb{R} \times \mathcal{S}$. These properties play a key role in the proof of nonsingularity of the Jacobians arising from the SDP and the SDCP. Assume that X has the eigen-decomposition as in (2.4), i.e.,

$$X = P\Lambda P^T,$$

where P is an orthogonal matrix and Λ is the diagonal matrix of eigenvalues of X and has the form

$$\Lambda = \begin{bmatrix} \Lambda_\alpha & 0 & 0 \\ 0 & \Lambda_\gamma & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Partition the orthogonal matrix P according to

$$P = [W_\alpha \ W_\gamma \ W_\beta],$$

with $W_\alpha \in \mathbb{R}^{n \times |\alpha|}$, $W_\gamma \in \mathbb{R}^{n \times |\gamma|}$, and $W_\beta \in \mathbb{R}^{n \times |\beta|}$.

Recall that the critical cone of $\mathcal{S}_+ := \{X \succeq 0 : X \in \mathcal{S}\}$ at $X \in \mathcal{S}$ is defined as

$$\mathcal{C}(X; \mathcal{S}_+) := \mathcal{T}(X_+; \mathcal{S}_+) \cap (X_+ - X)^\perp,$$

where $\mathcal{T}(X_+; \mathcal{S}_+)$ is the tangent cone of \mathcal{S}_+ at X_+ and $(X_+ - X)^\perp$ is the subset of matrices in \mathcal{S} that are orthogonal to $(X_+ - X)$ under the matrix Frobenius inner product. The critical cone can be completely described [3, 9] by

$$(3.1) \quad \mathcal{C}(X; \mathcal{S}_+) = \{Y \in \mathcal{S} : W_\gamma^T Y W_\gamma = 0, W_\gamma^T Y W_\beta = 0, W_\beta^T Y W_\beta \succeq 0\}.$$

Consequently, the affine hull of $\mathcal{C}(X; \mathcal{S}_+)$, which we denote by $\mathcal{L}(X; \mathcal{S}_+)$, is the linear subspace

$$\{Y \in \mathcal{S} : W_\gamma^T Y W_\gamma = 0, W_\gamma^T Y W_\beta = 0\}.$$

PROPOSITION 3.1. *For any $(0, H) \in \mathbb{R} \times \mathcal{S}$ and $V \in \partial_B \Phi(0, X)$, it holds that*

$$(3.2) \quad V(0, H) = P(\Omega \circ P^T H P)P^T,$$

$$(3.3) \quad H + V(0, H) \in \mathcal{L}(X; \mathcal{S}_+),$$

and

$$(3.4) \quad [H - V(0, H)] \bullet [H + V(0, H)] \geq 0,$$

where the matrix $\Omega \in \mathcal{S}$ has entries

$$\Omega_{ij} = \begin{cases} t \in [-1, 1] & \text{if } (i, j) \in \beta \times \beta, \\ \frac{\lambda_i + \lambda_j}{|\lambda_i| + |\lambda_j|} & \text{otherwise.} \end{cases}$$

Proof. Let $V \in \partial_B \Phi(0, X)$. By Lemma 2.3 and the definition of the elements in $\partial_B \Phi(0, X)$, it follows that there exists a sequence $\{(\varepsilon^k, X^k)\}$ converging to $(0, X)$ with $(\varepsilon^k)^2 I + (X^k)^2$ being nonsingular such that

$$V(0, H) = \lim_{k \rightarrow \infty} J\Phi(\varepsilon^k, X^k)(0, H) = \lim_{k \rightarrow \infty} L_{C^k}^{-1}(L_{X^k}(H)),$$

where $C^k := \Phi(\varepsilon^k, X^k)$. Let $X^k = P^k \Lambda^k (P^k)^T$ be the orthogonal decomposition of X^k , where Λ^k is the diagonal matrix of eigenvalues of X^k and P^k is a corresponding orthogonal matrix. Without loss of generality, by taking subsequences if necessary, we may assume that $\{P^k\}$ is a convergent sequence with limit $P = \lim_{k \rightarrow \infty} P^k$ and $\Lambda = \lim_{k \rightarrow \infty} \Lambda^k$ (clearly $X = P \Lambda P^T$). Then,

$$\lim_{k \rightarrow \infty} \Lambda_\beta^k = 0.$$

For any $H \in \mathcal{S}$ with $\tilde{H}^k := (P^k)^T H P^k$, we have that

$$L_{C^k}(J\Phi(\varepsilon^k, X^k)(0, H)) = L_{X^k}(H);$$

i.e.,

$$((\varepsilon^k)^2 I + (\Lambda^k)^2)^{1/2} \tilde{U}^k + \tilde{U}^k ((\varepsilon^k)^2 I + (\Lambda^k)^2)^{1/2} = \Lambda^k \tilde{H}^k + \tilde{H}^k \Lambda^k,$$

where $\tilde{U}^k := (P^k)^T [J\Phi(\varepsilon^k, X^k)(0, H)] P^k$. By denoting $\tilde{C}^k := ((\varepsilon^k)^2 I + (\Lambda^k)^2)^{1/2}$, we have that

$$\begin{aligned} & \begin{bmatrix} \tilde{C}_{\alpha\alpha}^k \tilde{U}_{\alpha\alpha}^k + \tilde{U}_{\alpha\alpha}^k \tilde{C}_{\alpha\alpha}^k & \tilde{C}_{\alpha\alpha}^k \tilde{U}_{\alpha\gamma}^k + \tilde{U}_{\alpha\gamma}^k \tilde{C}_{\alpha\alpha}^k & \tilde{C}_{\alpha\alpha}^k \tilde{U}_{\alpha\beta}^k + \tilde{U}_{\alpha\beta}^k \tilde{C}_{\alpha\alpha}^k \\ \tilde{C}_{\gamma\gamma}^k \tilde{U}_{\gamma\alpha}^k + \tilde{U}_{\gamma\alpha}^k \tilde{C}_{\gamma\gamma}^k & \tilde{C}_{\gamma\gamma}^k \tilde{U}_{\gamma\gamma}^k + \tilde{U}_{\gamma\gamma}^k \tilde{C}_{\gamma\gamma}^k & \tilde{C}_{\gamma\gamma}^k \tilde{U}_{\gamma\beta}^k + \tilde{U}_{\gamma\beta}^k \tilde{C}_{\gamma\gamma}^k \\ \tilde{C}_{\beta\beta}^k \tilde{U}_{\beta\alpha}^k + \tilde{U}_{\beta\alpha}^k \tilde{C}_{\beta\beta}^k & \tilde{C}_{\beta\beta}^k \tilde{U}_{\beta\gamma}^k + \tilde{U}_{\beta\gamma}^k \tilde{C}_{\beta\beta}^k & \tilde{C}_{\beta\beta}^k \tilde{U}_{\beta\beta}^k + \tilde{U}_{\beta\beta}^k \tilde{C}_{\beta\beta}^k \end{bmatrix} \\ &= \begin{bmatrix} \Lambda_\alpha^k \tilde{H}_{\alpha\alpha}^k + \tilde{H}_{\alpha\alpha}^k \Lambda_\alpha^k & \Lambda_\alpha^k \tilde{H}_{\alpha\gamma}^k + \tilde{H}_{\alpha\gamma}^k \Lambda_\alpha^k & \Lambda_\alpha^k \tilde{H}_{\alpha\beta}^k + \tilde{H}_{\alpha\beta}^k \Lambda_\alpha^k \\ \Lambda_\gamma^k \tilde{H}_{\gamma\alpha}^k + \tilde{H}_{\gamma\alpha}^k \Lambda_\gamma^k & \Lambda_\gamma^k \tilde{H}_{\gamma\gamma}^k + \tilde{H}_{\gamma\gamma}^k \Lambda_\gamma^k & \Lambda_\gamma^k \tilde{H}_{\gamma\beta}^k + \tilde{H}_{\gamma\beta}^k \Lambda_\gamma^k \\ \Lambda_\beta^k \tilde{H}_{\beta\alpha}^k + \tilde{H}_{\beta\alpha}^k \Lambda_\beta^k & \Lambda_\beta^k \tilde{H}_{\beta\gamma}^k + \tilde{H}_{\beta\gamma}^k \Lambda_\beta^k & \Lambda_\beta^k \tilde{H}_{\beta\beta}^k + \tilde{H}_{\beta\beta}^k \Lambda_\beta^k \end{bmatrix}. \end{aligned}$$

For each k , define the matrix $\Omega^k \in \mathcal{S}$ with entries

$$\Omega_{ij}^k = \left(\sqrt{(\varepsilon^k)^2 + (\lambda_i^k)^2} + \sqrt{(\varepsilon^k)^2 + (\lambda_j^k)^2} \right)^{-1} (\lambda_i^k + \lambda_j^k), \quad i, j = 1, \dots, n.$$

Since $\{\Omega^k\}$ is bounded, by taking a subsequence if necessary, we assume that $\{\Omega^k\}$ is a convergent sequence and that

$$\lim_{k \rightarrow \infty} \Omega^k = \Omega.$$

Hence, it follows that

$$\lim_{k \rightarrow \infty} \tilde{U}^k = \lim_{k \rightarrow \infty} \Omega^k \circ \tilde{H}^k = \Omega \circ P^T H P,$$

which proves (3.2). Let $\tilde{H} := P^T H P$. Then, we obtain that

$$P^T V(0, H) P = \begin{bmatrix} \tilde{H}_{\alpha\alpha} & \Omega_{\alpha\gamma} \circ \tilde{H}_{\alpha\gamma} & \tilde{H}_{\alpha\beta} \\ \tilde{H}_{\alpha\gamma}^T \circ \Omega_{\alpha\gamma}^T & -\tilde{H}_{\gamma\gamma} & -\tilde{H}_{\gamma\beta} \\ \tilde{H}_{\alpha\beta}^T & -\tilde{H}_{\gamma\beta}^T & \Omega_{\beta\beta} \circ \tilde{H}_{\beta\beta} \end{bmatrix}.$$

Let $\mathbf{E} \in \mathcal{S}$ be the matrix whose entries are all ones. Thus,

$$(3.5) = \begin{bmatrix} 2\tilde{H}_{\alpha\alpha} & (\Omega_{\alpha\gamma} + \mathbf{E}_{\alpha\gamma}) \circ \tilde{H}_{\alpha\gamma} & 2\tilde{H}_{\alpha\beta} \\ \tilde{H}_{\alpha\gamma}^T \circ (\Omega_{\alpha\gamma} + \mathbf{E}_{\alpha\gamma})^T & 0 & 0 \\ 2\tilde{H}_{\alpha\beta}^T & 0 & (\Omega_{\beta\beta} + \mathbf{E}_{\beta\beta}) \circ \tilde{H}_{\beta\beta} \end{bmatrix}$$

and

$$(3.6) \quad P^T[H - V(0, H)]P = \begin{bmatrix} 0 & (\mathbf{E}_{\alpha\gamma} - \Omega_{\alpha\gamma}) \circ \tilde{H}_{\alpha\gamma} & 0 \\ \tilde{H}_{\alpha\gamma}^T \circ (\mathbf{E}_{\alpha\gamma} - \Omega_{\alpha\gamma})^T & 2\tilde{H}_{\gamma\gamma} & 2\tilde{H}_{\gamma\beta} \\ 0 & 2\tilde{H}_{\gamma\beta}^T & (\mathbf{E}_{\beta\beta} - \Omega_{\beta\beta}) \circ \tilde{H}_{\beta\beta} \end{bmatrix}.$$

Hence, from (3.5), we get that

$$W_\gamma^T[H + V(0, H)]W_\gamma = 0 \quad \text{and} \quad W_\beta^T[H + V(0, H)]W_\beta = 0,$$

which proves (3.3).

By noting the fact that $\Omega_{ij} \in [-1, 1]$ for all $i, j = 1, \dots, n$, from (3.5) and (3.6) we obtain that

$$\begin{aligned} & [H - V(0, H)] \bullet [H + V(0, H)] \\ &= (P^T[H - V(0, H)]P) \bullet (P^T[H + V(0, H)]P) \\ &= \sum_{i \in \alpha, j \in \gamma} 2(1 - \Omega_{ij})(1 + \Omega_{ij})\tilde{H}_{ij}^2 + \sum_{i \in \beta, j \in \beta} (1 - \Omega_{ij})(1 + \Omega_{ij})\tilde{H}_{ij}^2 \\ &\geq 0, \end{aligned}$$

which proves (3.4). This completes the proof. \square

4. The squared smoothing Newton method. Let $\Psi : \mathcal{S}(n_1, \dots, n_m) \rightarrow \mathcal{S}(n_1, \dots, n_m)$ be locally Lipschitz continuous. Let $G : \mathbb{R} \times \mathcal{S}(n_1, \dots, n_m) \rightarrow \mathcal{S}(n_1, \dots, n_m)$ be an approximate function of Ψ such that G is continuously differentiable at $(\varepsilon, X) \in \mathbb{R} \times \mathcal{S}(n_1, \dots, n_m)$ unless $\varepsilon = 0$ and

$$\lim_{(\varepsilon, Y) \rightarrow (0, X)} G(\varepsilon, Y) = \Psi(X).$$

The existence of such a G was proved in [31] for vector-valued functions. It can be easily extended to matrix-valued functions by making use of the isometry between \mathbb{R}^n and $\mathcal{S}(n_1, \dots, n_m)$. For the SDP and the SDCP, there are many choices for G . In particular, a computationally efficient form for the SDCP is

$$(4.1) \quad G(\varepsilon, X) := X - [X - F(X) + \Phi(\varepsilon, X - F(X))]/2.$$

The *squared smoothing Newton method*, in particular, solves the auxiliary equation

$$(4.2) \quad E(\varepsilon, X) := \begin{bmatrix} \varepsilon \\ G(\varepsilon, X) \end{bmatrix} = 0$$

and uses the merit function $\phi(Z) := \varepsilon^2 + \|G(Z)\|^2$ for the line search, where $Z := (\varepsilon, X)$.

Let $\bar{\varepsilon} \in \mathbb{R}_{++}$ and $\eta \in (0, 1)$ be such that $\eta\bar{\varepsilon} < 1$. Define an auxiliary point \bar{Z} by

$$\bar{Z} := (\bar{\varepsilon}, 0) \in \mathbb{R} \times \mathcal{S}(n_1, \dots, n_m)$$

and $\theta : \mathbb{R} \times \mathcal{S}(n_1, \dots, n_m) \mapsto \mathbb{R}_+$ by

$$\theta(Z) := \eta \min\{1, \phi(Z)\}.$$

Let

$$\mathcal{N} := \{ Z = (\varepsilon, X) \in \mathbb{R} \times \mathcal{S}(n_1, \dots, n_m) : \varepsilon \geq \theta(Z)\bar{\varepsilon} \}.$$

ALGORITHM 4.1.

Step 0. Select constants $\delta \in (0, 1)$ and $\sigma \in (0, 1/2)$. Let $\varepsilon^0 := \bar{\varepsilon}$, $X^0 \in \mathcal{S}(n_1, \dots, n_m)$ be an arbitrary point and $k := 0$.

Step 1. If $E(Z^k) = 0$, then stop. Otherwise, let $\theta_k := \theta(Z^k)$.

Step 2. Compute $\Delta Z^k := (\Delta \varepsilon^k, \Delta X^k) \in \mathbb{R} \times \mathcal{S}(n_1, \dots, n_m)$ by

$$(4.3) \quad E(Z^k) + JE(Z^k)(\Delta Z^k) = \theta_k \bar{Z}.$$

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

$$(4.4) \quad \phi(Z^k + \delta^l \Delta Z^k) \leq [1 - 2\sigma(1 - \eta\bar{\varepsilon})\delta^l] \phi(Z^k).$$

Define $Z^{k+1} := Z^k + \delta^{l_k} \Delta Z^k$.

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

THEOREM 4.2. Assume that

- (i) for every $k \geq 0$, if $\varepsilon^k \in \mathbb{R}_{++}$ and $Z^k \in \mathcal{N}$, then $JE(Z^k)$ is nonsingular; and
- (ii) for any accumulation point $Z^* = (\varepsilon^*, X^*)$ of $\{Z^k\}$, if $\varepsilon^* > 0$ and $Z^* \in \mathcal{N}$, then $JE(Z^*)$ is nonsingular.

Then an infinite sequence $\{Z^k\} \subset \mathcal{N}$ is generated by Algorithm 4.1 and each accumulation point Z^* of $\{Z^k\}$ is a solution of $E(Z) = 0$. Moreover, if E is strongly semismooth at Z^* and if all $V \in \partial_B E(Z^*)$ are nonsingular, then the whole sequence $\{Z^k\}$ converges to Z^* ,

$$(4.5) \quad \|Z^{k+1} - Z^*\| = O(\|Z^k - Z^*\|^2),$$

and

$$(4.6) \quad \varepsilon^{k+1} = O((\varepsilon^k)^2).$$

The vector version of the above convergence result is proved in [26], where the smoothing parameter is a vector rather than a scalar. However, the proof was independent of the dimension of the parameter vector. Therefore, with a slight revision if necessary, its matrix version can be established similarly. For brevity we omit the proof.

The key conditions for quadratic convergence of Algorithm 4.1 are: (a) the strong semismoothness of the smoothing function E and (b) the nonsingularity of all $V \in \partial_B E(Z^*)$. (In [26], $\partial E(Z^*)$, rather than $\partial_B E(Z^*)$, was used. However, it is easy to check whether the convergence properties are still valid if we replace $\partial E(Z^*)$ by $\partial_B E(Z^*)$ in the analysis.) In the subsequent sections we will provide sufficient conditions for (b) to hold in the cases of SDP and SDCP where (a) is naturally implied by the strong semismoothness of Φ .

5. Application to the SDP. In this section we shall show how to use Algorithm 4.1 to solve (1.4), which constitutes the optimality conditions of the SDP. For this purpose, we assume that $\{A_i\}_{i=1}^m$ are linearly independent, i.e., any $\alpha \in \mathbb{R}^m$ satisfying $\sum_{i=1}^m \alpha_i A_i = 0$ implies $\alpha_i = 0$, $i = 1, \dots, m$.

Define $\mathcal{A} : \mathcal{S} \rightarrow \mathbb{R}^m$ as

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

Then solving (1.4) is equivalent to finding a solution to

$$(5.1) \quad \Psi(X, y, S) := \begin{bmatrix} \mathcal{A}(X) - b \\ \sum_{i=1}^m y_i A_i + S - C \\ X - [X - S]_+ \end{bmatrix} = 0, \quad (X, y, S) \in \mathcal{S} \times \mathbb{R}^m \times \mathcal{S}.$$

Define $G : \mathbb{R} \times \mathcal{S} \times \mathbb{R}^m \times \mathcal{S} \rightarrow \mathbb{R}^m \times \mathcal{S} \times \mathcal{S}$ as

$$(5.2) \quad G(\varepsilon, X, y, S) := \begin{bmatrix} \mathcal{A}(X) - b \\ \sum_{i=1}^m y_i A_i + S - C \\ X - [X - S + \Phi(\varepsilon, X - S)] / 2 \end{bmatrix}.$$

Then G is continuously differentiable at (ε, X, y, S) with $\varepsilon \neq 0$. Let

$$(5.3) \quad E(\varepsilon, X, y, S) := \begin{bmatrix} \varepsilon \\ G(\varepsilon, X, y, S) \end{bmatrix}.$$

Hence, finding a solution of $\Psi(X, y, S) = 0$ is equivalent to finding a solution of $E(\varepsilon, X, y, S) = 0$.

Similar smoothing functions for the SDP were first used in [6] and very recently in [15]. Based on these smoothing functions, smoothing Newton methods were also designed in [6, 15]. The major differences between our method and those in [6, 15] in the context of SDP are (i) our algorithm needs to solve only one linear system per iteration while the methods in [6, 15] need to solve two; (ii) quadratic convergence has been established for our algorithm while only superlinear convergence has been established for methods in [6, 15]; and (iii) numerical results are reported in [6, 15] while our paper is focused on theoretical analysis.

The next result shows that $JE(\varepsilon, X, Y, S)$ is nonsingular at $(\varepsilon, X, y, S) \in \mathbb{R} \times \mathcal{S} \times \mathbb{R}^m \times \mathcal{S}$ with $\varepsilon \neq 0$. Similar proofs can be found in [6, 15, 34].

PROPOSITION 5.1. *For any $(\varepsilon, X, y, S) \in \mathbb{R} \times \mathcal{S} \times \mathbb{R}^m \times \mathcal{S}$ with $\varepsilon \neq 0$, $JE(\varepsilon, X, Y, S)$ is nonsingular.*

Proof. By Lemma 2.3, we know that $JE(\varepsilon, X, Y, S)$ exists. Suppose that there exists $(\tau, H, z, T) \in \mathbb{R} \times \mathcal{S} \times \mathbb{R}^m \times \mathcal{S}$ such that

$$JE(\varepsilon, X, Y, S)(\tau, H, z, T) = 0;$$

i.e.,

$$(5.4) \quad \begin{bmatrix} \tau \\ \mathcal{A}(H) \\ \sum_{i=1}^m z_i A_i + T \\ H - [H - T + J\Phi(\varepsilon, X - S)(\tau, H - T)] / 2 \end{bmatrix} = 0,$$

which implies that

$$\tau = 0 \quad \text{and} \quad 2H - [H - T + J\Phi(\varepsilon, X - S)(0, H - T)] = 0.$$

Hence, by Lemma 2.3,

$$2H - \left[H - T + L_{\Phi(\varepsilon, X-S)}^{-1} L_{(X-S)}(H - T) \right] = 0,$$

which implies that

$$L_{\Phi(\varepsilon, X-S)}(H + T) = L_{(X-S)}(H - T);$$

i.e.,

$$\begin{aligned} & (\varepsilon^2 I + (X - S)^2)^{1/2} (H + T) + (H + T) (\varepsilon^2 I + (X - S)^2)^{1/2} \\ &= (X - S)(H - T) + (H - T)(X - S). \end{aligned}$$

Since $X - S \in \mathcal{S}$, there exist an orthogonal matrix P and a diagonal matrix Λ of eigenvalues of $X - S$ such that

$$X - S = P\Lambda P^T.$$

By denoting $\tilde{H} := P^T H P$ and $\tilde{T} := P^T T P$, we have that

$$(\varepsilon^2 I + \Lambda^2)^{1/2} (\tilde{H} + \tilde{T}) + (\tilde{H} + \tilde{T})(\varepsilon^2 I + \Lambda^2)^{1/2} = \Lambda(\tilde{H} - \tilde{T}) + (\tilde{H} - \tilde{T})\Lambda.$$

Hence,

$$\tilde{H} + \tilde{T} = \Omega \circ (\tilde{H} - \tilde{T}),$$

where the matrix $\Omega \in \mathcal{S}$ has entries

$$\Omega_{ij} = \left(\sqrt{\varepsilon^2 + \lambda_i^2} + \sqrt{\varepsilon^2 + \lambda_j^2} \right)^{-1} (\lambda_i + \lambda_j), \quad i, j = 1, \dots, n.$$

Thus,

$$\tilde{H} = \tilde{\Omega} \circ \tilde{T},$$

where the matrix $\tilde{\Omega} \in \mathcal{S}$ has entries

$$\tilde{\Omega}_{ij} = \left(\lambda_i + \lambda_j - \sqrt{\varepsilon^2 + \lambda_i^2} - \sqrt{\varepsilon^2 + \lambda_j^2} \right)^{-1} \left(\lambda_i + \lambda_j + \sqrt{\varepsilon^2 + \lambda_i^2} + \sqrt{\varepsilon^2 + \lambda_j^2} \right),$$

where $i, j = 1, \dots, n$. From (5.4), we know that

$$A_i \bullet H = 0, \quad i = 1, \dots, m, \quad \text{and} \quad \sum_{i=1}^m z_i A_i + T = 0,$$

which implies that

$$T \bullet H = \sum_{i=1}^m z_i A_i \bullet H + T \bullet H = \left(\sum_{i=1}^m z_i A_i + T \right) \bullet H = 0.$$

Hence,

$$0 = T \bullet H = \tilde{T} \bullet \tilde{H} = \tilde{T} \bullet (\tilde{\Omega} \circ \tilde{T}),$$

which, together with the fact that $\tilde{\Omega}_{ij} < 0$ for all i and j , implies that $\tilde{T} = 0$. Thus,

$$\tilde{H} = \tilde{\Omega} \bullet \tilde{T} = 0 \quad \text{and} \quad T = H = 0.$$

From the linear independence of $\{A_i\}_{i=1}^m$ and that fact $\sum_{i=1}^m z_i A_i + T = 0$, we can conclude that $z = 0$. This shows that $JE(\varepsilon, X, y, S)$ is nonsingular. \square

Proposition 5.1 shows that Algorithm 4.1 is well defined when it is applied to the SDP. We state it formally in the following theorem. Its proof is a direct application of Theorem 4.2 and Proposition 5.1.

THEOREM 5.2. *If Algorithm 4.1 is applied to the SDP, then an infinite sequence $\{Z^k\}$ is generated and each accumulation point Z^* of $\{Z^k\}$ is a solution of $E(Z) = 0$.*

For local convergence analysis of Algorithm 4.1 for the SDP, we need the nonsingularity of $\partial_B E(Z^*)$ at a solution Z^* of $E(Z) = 0$. Next, we discuss a sufficient condition to guarantee the nonsingularity of $\partial_B E(Z^*)$ at a *strict complementary and nondegenerate* solution $Z^* = (0, X^*, y^*, S^*)$ of $E(Z) = 0$; i.e., Z^* satisfies the following two conditions: (a) $X^* + S^* \succ 0$ and (b) for any $(H, z, T) \in \mathcal{S} \times \mathbb{R}^m \times \mathcal{S}$ satisfying

$$\mathcal{A}(H) = 0, \quad \sum_{i=1}^m z_i A_i + T = 0, \quad \text{and} \quad X^* T + H S^* = 0,$$

it holds that $H = T = 0$. Condition (a) is called the strict complementarity, under which E is continuously differentiable at Z^* . Condition (b) was first introduced by Kojima, Shida, and Shindoh [16] for local analysis of interior-point methods. Conditions (a) and (b) are also used in noninterior-point methods for solving the SDP [6, 15]. See [1] for a discussion on strict complementarity and nondegeneracy conditions in the SDP.

PROPOSITION 5.3. *Let $Z^* = (0, X^*, y^*, S^*) \in \mathbb{R} \times \mathcal{S} \times \mathbb{R}^m \times \mathcal{S}$ be a strict complementary and nondegenerate solution of $E(Z) = 0$. Then $JE(Z^*)$ is nonsingular.*

Proof. Since (X^*, y^*, S^*) is a solution to the SDP, we have that

$$X^* \succeq 0, \quad S^* \succeq 0, \quad X^* S^* = S^* X^* = 0,$$

which implies that there exists an orthogonal matrix P such that

$$X^* = P \Delta P^T \quad \text{and} \quad S^* = P \Sigma P^T,$$

where $\Delta = \text{diag}(\delta_1, \dots, \delta_n)$ and $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$ are two positive semidefinite diagonal matrices and $\delta_i \sigma_i = 0$, $i = 1, \dots, n$, where $\delta_1, \dots, \delta_n$ and $\sigma_1, \dots, \sigma_n$ are eigenvalues of X^* and S^* , respectively. By using the fact that $X^* + S^* \succ 0$, we also have that

$$\delta_i + \sigma_i > 0, \quad i = 1, \dots, n.$$

Denote $\Lambda := \Delta - \Sigma$. Then, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ is nonsingular and

$$X^* - S^* = P \Lambda P^T,$$

where $\lambda_i = \delta_i - \sigma_i$, $i = 1, \dots, n$.

Suppose that there exists $(\tau, H, z, T) \in \mathbb{R} \times \mathcal{S} \times \mathbb{R}^m \times \mathcal{S}$ such that

$$JE(0, X^*, y^*, S^*)(\tau, H, z, T) = 0.$$

We have that $\tau = 0$ and

$$(5.5) \quad \begin{bmatrix} \mathcal{A}(H) \\ \sum_{i=1}^m z_i A_i + T \\ H + T - J\Phi(0, X^* - S^*)(0, H - T) \end{bmatrix} = 0.$$

In particular, from the third equality of (5.5), we obtain that

$$P^T(H + T)P - P^T J\Phi(0, X^* - S^*)(0, H - T)P = 0,$$

which, together with Proposition 3.1, implies that

$$\tilde{H} + \tilde{T} = P^T J\Phi(0, X^* - S^*)(0, H - T)P = \Omega \circ (\tilde{H} - \tilde{T}),$$

where $\tilde{H} := P^T H P$, $\tilde{T} = P^T T P$, and $\Omega \in \mathcal{S}$ has entries

$$\Omega_{ij} = \frac{\lambda_i + \lambda_j}{|\lambda_i| + |\lambda_j|}, \quad i, j = 1, \dots, n.$$

Hence,

$$(5.6) \quad (E - \Omega) \circ \tilde{H} + \tilde{T} \circ (E + \Omega) = 0,$$

where $E \in \mathcal{S}$ denotes the matrix whose entries are all ones. Denote two index sets

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

By noting the fact that $\lambda_i = \delta_i$ if $\lambda_i > 0$ and $\lambda_i = -\sigma_i$ if $\lambda_i < 0$ and $\alpha \cup \gamma = \{1, \dots, n\}$, from (5.6) we have that

$$\tilde{T}_{ij} = 0 \quad \forall (i, j) \in \alpha \times \alpha;$$

$$\tilde{H}_{ij}\sigma_j + \tilde{T}_{ij}\delta_i = 0 \quad \forall (i, j) \in \alpha \times \gamma$$

and

$$\tilde{H}_{ij} = 0 \quad \forall (i, j) \in \gamma \times \gamma.$$

Thus,

$$\Delta\tilde{T} + \tilde{H}\Sigma = 0;$$

i.e.,

$$X^*T + HS^* = 0,$$

which, together with the first and second equalities of (5.5) and the nondegeneracy assumption at Z^* , shows that

$$H = T = 0.$$

The linear independence of $\{A_i\}_{i=1}^m$ and the fact that $T = 0$ imply $z = 0$. Hence, $JE(Z^*)$ is nonsingular. \square

We can now state quadratic convergence of Algorithm 4.1 for solving the SDP, which does not require a proof.

THEOREM 5.4. *If an accumulation point Z^* of $\{Z^k\}$ generated by Algorithm 4.1 for solving the SDP is a strict complementary and nondegenerate solution of $E(Z) = 0$, then the whole sequence $\{Z^k\}$ converges to Z^* with*

$$(5.7) \quad \|Z^{k+1} - Z^*\| = O(\|Z^k - Z^*\|^2)$$

and

$$(5.8) \quad \varepsilon^{k+1} = O((\varepsilon^k)^2).$$

In the above theorem for the SDP, we need the nondegeneracy to prove quadratic convergence of Algorithm 4.1. In the next section, we shall show that, for the SDCP, this assumption can be replaced by the positive definiteness of the Jacobian of the problem on a certain subspace.

6. Application to the SDCP. In this section, we shall deduce quadratic convergence of the squared smoothing Newton method in solving the SDCP. We first prove a result on the generalized Jacobian for a composite function.

PROPOSITION 6.1. *Let $\mathcal{S}, \mathcal{S}_1$, and \mathcal{S}_2 be symmetric block-diagonal matrix spaces. Let $F : \mathcal{S} \rightarrow \mathcal{S}_1$ be continuously differentiable on an open neighborhood \mathcal{N} of \bar{X} and $\Psi : \mathcal{S}_1 \rightarrow \mathcal{S}_2$ be locally Lipschitz continuous and semismooth on an open neighborhood of $F(\bar{X})$. Then, for any $H \in \mathcal{S}$, it holds that*

$$(6.1) \quad \partial_B \Upsilon(\bar{X})(H) \subseteq \partial_B \Psi(F(\bar{X})) JF(\bar{X})(H),$$

where for any $X \in \mathcal{N}$, $\Upsilon(X) := \Psi(F(X))$.

Proof. Since Υ is locally Lipschitz continuous, by Rademacher’s theorem (see [28, page 403]), Υ is differentiable almost everywhere in \mathcal{N} . For any $V \in \partial_B \Upsilon(\bar{X})$, there exists a sequence of differentiable points $\{X^k\} \subset \mathcal{N}$ of Υ converging to \bar{X} such that

$$V = \lim_{k \rightarrow \infty} J\Upsilon(X^k).$$

Since Ψ is directionally differentiable on an open neighborhood of $F(\bar{X})$, for any $H \in \mathcal{S}$,

$$J\Upsilon(X^k)(H) = \Psi'(F(X^k); JF(X^k)(H)).$$

Since Ψ is semismooth at $F(X^k)$, there exists a $W \in \partial_B \Psi(F(X^k))$ such that [25]

$$\Psi'(F(X^k); JF(X^k)(H)) = W JF(X^k)(H).$$

Thus,

$$J\Upsilon(X^k)(H) \in \partial_B \Psi(F(X^k)) JF(X^k)(H),$$

which, together with the upper semicontinuity of ∂_B (see [25]), implies that

$$\lim_{k \rightarrow \infty} J\Upsilon(X^k)(H) \in \partial_B \Psi(F(\bar{X})) JF(\bar{X})(H).$$

This proves (6.1). \square

In the following analysis, we assume that $F : \mathcal{S} \rightarrow \mathcal{S}$ is continuously differentiable and $E : \mathbb{R} \times \mathcal{S} \rightarrow \mathbb{R} \times \mathcal{S}$ is defined as

$$(6.2) \quad E(\varepsilon, X) = \begin{bmatrix} \varepsilon \\ G(\varepsilon, X) \end{bmatrix}, \quad (\varepsilon, X) \in \mathbb{R} \times \mathcal{S},$$

where $G : \mathbb{R} \times \mathcal{S} \rightarrow \mathcal{S}$ is defined by (4.1); i.e.,

$$G(\varepsilon, X) = X - [X - F(X) + \Phi(\varepsilon, X - F(X))] / 2$$

and for any $Y \in \mathcal{S}$,

$$\Phi(\varepsilon, Y) = (\varepsilon^2 I + Y^2)^{1/2}.$$

Then solving the SDCP is equivalent to solving the following equation:

$$(6.3) \quad E(\varepsilon, X) = 0.$$

The next result is on the nonsingularity of the B-subdifferential of E at $(0, X) \in \mathbb{R} \times \mathcal{S}$.

PROPOSITION 6.2. *Suppose that for a given $X \in \mathcal{S}$, the Jacobian $JF(X)$ of F at X is positive definite on the linear subspace $\mathcal{L}(X - F(X); S_+)$, the affine hull of $\mathcal{C}(X - F(X); S_+)$. Then all $U \in \partial_B E(0, X)$ are nonsingular.*

Proof. Let U be an element of $\partial_B E(0, X)$. Assume that $(\tau, H) \in \mathbb{R} \times \mathcal{S}$ is such that $U(\tau, H) = 0$. Then, from the definition of the B-subdifferential of E , we know that $\tau = 0$ and there exists a $W \in \partial_B G(0, X)$ such that $W(0, H) = 0$. By Proposition 6.1, there exists a $V \in \partial_B \Phi(0, X - F(X))$ such that

$$W(0, H) = H - [H - JF(X)(H) + V(0, H - JF(X)(H))] / 2,$$

which, together with the fact that $W(0, H) = 0$, implies that

$$2H - [H - JF(X)(H)] - V(0, H - JF(X)(H)) = 0.$$

Let $\bar{H} := H - JF(X)(H)$. We have that

$$(6.4) \quad 2H = \bar{H} + V(0, \bar{H})$$

and that

$$2[\bar{H} + JF(X)((\bar{H} + V(0, \bar{H}))/2)] - \bar{H} - V(0, \bar{H}) = 0;$$

i.e.,

$$\bar{H} - V(0, \bar{H}) + JF(X)(\bar{H} + V(0, \bar{H})) = 0,$$

which implies that

$$(6.5) \quad \begin{aligned} & [\bar{H} + V(0, \bar{H})] \bullet [\bar{H} - V(0, \bar{H})] \\ & + [\bar{H} + V(0, \bar{H})] \bullet [JF(X)(\bar{H} + V(0, \bar{H}))] = 0. \end{aligned}$$

By Proposition 3.1, (6.5), and the assumption that $JF(X)$ is positive definite on $\mathcal{L}(X - F(X); S_+)$, we conclude that

$$\bar{H} + V(0, \bar{H}) = 0,$$

which, together with (6.4), implies that $H = 0$. This shows that for any $(\tau, H) \in \mathbb{R} \times \mathcal{S}$

satisfying $U(\tau, H) = 0$, one has $(\tau, H) = 0$. Hence, U is nonsingular. The proof is completed. \square

Finally, we can state quadratic convergence of the squared smoothing Newton method for solving the SDCP.

THEOREM 6.3. *Suppose that $F : \mathcal{S} \rightarrow \mathcal{S}$ is continuously differentiable on \mathcal{S} . Suppose that for each $X \in \mathcal{S}$, $JF(X)$ is positive semidefinite. Then an infinite sequence $\{Z^k\}$ is generated by Algorithm 4.1 for solving (6.3) and each accumulation point Z^* of $\{Z^k\}$ is a solution of $E(Z) = 0$. Moreover, if $JF(\cdot)$ is Lipschitz continuous around X^* and $JF(X^*)$ is positive definite on the linear subspace $\mathcal{L}(X^* - F(X^*); S_+)$, the affine hull of $\mathcal{C}(X^* - F(X^*); S_+)$, then the whole sequence $\{Z^k\}$ converges to Z^* ,*

$$(6.6) \quad \|Z^{k+1} - Z^*\| = O(\|Z^k - Z^*\|^2),$$

and

$$(6.7) \quad \varepsilon^{k+1} = O((\varepsilon^k)^2).$$

Proof. For any $\varepsilon \neq 0$ and $X \in \mathcal{S}$, by Lemma 2.3, E is continuously differentiable at (ε, X) . It is easy to check that $JE(\varepsilon, X)$ is nonsingular if and only if $JG(\varepsilon, X)(0, H) = 0$ implies $H = 0$. It has been shown by Chen and Tseng [6] that the latter is true. Thus, for any $\varepsilon \neq 0$ and $X \in \mathcal{S}$, $JE(\varepsilon, X)$ is nonsingular. By Theorem 4.2, an infinite sequence $\{Z^k\}$ is generated by Algorithm 4.1 and each accumulation point Z^* of $\{Z^k\}$ is a solution of $E(Z) = 0$.

If $JF(\cdot)$ is Lipschitz continuous around X^* , then by Theorem 2.5 and a property on the strong semismoothness of a composite function (originally due to Fischer [10]; for the matrix version, see [32, Theorem 3.10]), we know that E is strongly semismooth at $(0, X^*)$. Furthermore, by Proposition 6.2, all $U \in \partial_B E(0, X^*)$ are nonsingular. Thus, by Theorem 4.2, the whole sequence $\{Z^k\}$ converges to Z^* , and (6.6) and (6.7) hold. \square

7. Conclusions. We have studied quadratic convergence of a squared smoothing Newton method for nonsmooth matrix equations. For the SDCP, the strong semismoothness of G , together with the positive definiteness of $JF(X^*)$ on the affine hull of $\mathcal{C}(X^* - F(X^*); S_+)$, implies that the proposed algorithm has quadratic rate of convergence without requiring the strict complementarity.

There are several possible directions to extend our work. One direction is to study the strong semismoothness of other smoothing functions used in [6] and then to improve the local analysis in [6]; another direction is to relax the nonsingularity condition on the Jacobians. It is also possible to use some regularization techniques, for example, the Tikhonov-type regularization, to get stronger global convergence results as has been done for vector-valued complementarity problems [8, 30].

Acknowledgments. The authors are grateful to the referees for their very constructive comments. In particular, the present proof of Lemma 2.4 was suggested by a referee.

REFERENCES

- [1] F. ALIZADEH, J. -P. HAEBERLY, AND M. L. OVERTON, *Complementarity and nondegeneracy in semidefinite programming*, Math. Programming, Ser. B, 77 (1997), pp. 111–128.
- [2] R. BHATIA, *Matrix Analysis*, Springer-Verlag, New York, 1997.
- [3] J. F. BONNANS AND A. SHAPIRO, *Perturbation Analysis of Optimization Problems*, Springer-Verlag, New York, 2000.

- [4] S. BOYD, L. EL GHAOU, E. FERON, AND V. BALAKRISHNAN, *Linear Matrix Inequalities in System and Control Theory*, SIAM Studies in Applied Mathematics 15, Philadelphia, 1994.
- [5] X. CHEN, H. QI, AND P. TSENG, *Analysis of nonsmooth symmetric-matrix-valued functions with applications to semidefinite complementarity problems*, SIAM J. Optim., 13 (2003), pp. 960–985.
- [6] X. CHEN AND P. TSENG, *Non-interior continuation methods for solving semidefinite complementarity problems*, Math. Program., 95 (2003), pp. 431–474.
- [7] B. FARES, D. NOLL, AND P. APKARIAN, *Robust control via sequential semidefinite programming*, SIAM J. Control Optim., 40 (2002), pp. 1791–1820.
- [8] F. FACCHINEI AND C. KANZOW, *Beyond monotonicity in regularization methods for nonlinear complementarity problems*, SIAM J. Control Optim., 37 (1999), pp. 1150–1161.
- [9] F. FACCHINEI AND J. S. PANG, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Vols. I and II, Springer-Verlag, New York, 2003.
- [10] A. FISCHER, *Solution of monotone complementarity problems with locally Lipschitzian functions*, Math. Programming, 76 (1997), pp. 513–532.
- [11] M. FUKUSHIMA AND L. QI, EDS., *Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999.
- [12] M. S. GOWDA AND T. PARTHASARATHY, *Complementarity forms of theorems of Lyapunov and Stein, and related results*, Linear Algebra Appl., 320 (2000), pp. 131–144.
- [13] M. S. GOWDA AND Y. SONG, *On semidefinite linear complementarity problems*, Math. Program., 88 (2000), pp. 575–587.
- [14] F. JARRE, *An interior method for nonconvex semidefinite programs*, Optim. Eng., 1 (2000), pp. 347–372.
- [15] C. KANZOW AND C. NAGEL, *Semidefinite programs: New search directions, smoothing-type methods, and numerical results*, SIAM J. Optim., 13 (2002), pp. 1–23.
- [16] M. KOJIMA, M. SHIDA, AND S. SHINDOH, *Local convergence of predictor-corrector infeasible-interior-point algorithm for SDPs and SDLCPs*, Math. Programming, 80 (1998), pp. 129–160.
- [17] M. KOJIMA, M. SHIDA, AND S. SHINDOH, *A predictor-corrector interior-point algorithm for the semidefinite linear complementarity problem using the Alizadeh–Haeberly–Overton search direction*, SIAM J. Optim., 9 (1999), pp. 444–465.
- [18] M. KOJIMA, M. SHIDA, AND S. SHINDOH, *Search directions in the SDP and the monotone SDLCP: Generalization and inexact computation*, Math. Program., 85 (1999), pp. 51–80.
- [19] M. KOJIMA, S. SHINDO, AND S. HARA, *Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices*, SIAM J. Optim., 7 (1997), pp. 86–125.
- [20] F. LEIBFRIITZ AND E. M. E. MOSTAFA, *An interior point constrained trust region method for a special class of nonlinear semidefinite programming problems*, SIAM J. Optim., 12 (2002), pp. 1048–1074.
- [21] R. D. C. MONTEIRO AND J. S. PANG, *On two interior-point mappings for nonlinear semidefinite complementarity problems*, Math. Oper. Res., 23 (1998), pp. 39–60.
- [22] R. D. C. MONTEIRO AND J.-S. PANG, *A potential reduction Newton method for constrained equations*, SIAM J. Optim., 9 (1999), pp. 729–754.
- [23] Y. NESTEROV AND A. NEMIROVSKII, *Interior Point Polynomial Algorithms in Convex Programming*, SIAM Studies in Applied Mathematics 13, Philadelphia, 1994.
- [24] J. S. PANG, D. SUN, AND J. SUN, *Semismooth homeomorphisms and strong stability of semidefinite and Lorentz complementarity problems*, Math. Oper. Res., 28 (2003), pp. 39–63.
- [25] L. QI, *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res., 18 (1993), pp. 227–244.
- [26] L. QI, D. SUN, AND G. ZHOU, *A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities*, Math. Program., 87 (2000), pp. 1–35.
- [27] L. QI AND J. SUN, *A nonsmooth version of Newton’s method*, Math. Programming, 58 (1993), pp. 353–367.
- [28] R. T. ROCKAFELLAR AND R. J.-B. WETS, *Variational Analysis*, Springer-Verlag, Berlin, 1998.
- [29] A. SHAPIRO, *First and second order analysis of nonlinear semidefinite programs*, Math. Programming, Ser. B, 77 (1997) pp. 301–320.
- [30] D. SUN, *A regularization Newton method for solving nonlinear complementarity problems*, Appl. Math. Optim., 40 (1999), pp. 315–339.
- [31] D. SUN AND L. QI, *Solving variational inequality problems via smoothing-nonsmooth reformulations*, J. Comput. Appl. Math., 129 (2001), pp. 37–62.

- [32] D. SUN AND J. SUN, *Semismooth matrix valued functions*, Math. Oper. Res., 27 (2002), pp. 150–169.
- [33] D. SUN AND J. SUN, *Strong semismoothness of eigenvalues of symmetric matrices and its application to inverse eigenvalue problems*, SIAM J. Numer. Anal., 40 (2003), pp. 2352–2367.
- [34] M. J. TODD, K. C. TOH, AND R. H. TÜTÜNCÜ, *On the Nesterov–Todd direction in semidefinite programming*, SIAM J. Optim., 8 (1998), pp. 769–796.
- [35] P. TSENG, *Merit functions for semidefinite complementarity problems*, Math. Programming, 83 (1998), pp. 159–185.
- [36] P. TSENG, *Convergent infeasible interior-point trust-region methods for constrained minimization*, SIAM J. Optim., 13 (2002), pp. 432–469.
- [37] G. ZHOU, D. SUN, AND L. QI, *Numerical experiments for a class of squared smoothing Newton methods for box constrained variational inequality problems*, in Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, M. Fukushima and L. Qi, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999, pp. 421–441.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.