MATHEMATICS OF OPERATIONS RESEARCH
Vol. 27, No. 1, February 2002, pp. 150–169
Printed in U.S.A.

SEMISMOOTH MATRIX-VALUED FUNCTIONS

DEFENG SUN AND JIE SUN

Matrix-valued functions play an important role in the development of algorithms for semidefinite programming problems. This paper studies generalized differential properties of such functions related to nonsmooth-smoothing Newton methods. The first part of this paper discusses basic properties such as the generalized derivative, Rademacher's theorem, $\mathcal{B}$-derivative, directional derivative, and semismoothness. The second part shows that the matrix absolute-value function, the matrix semidefinite-projection function, and the matrix projective residual function are strongly semismooth.

1. Introduction. Let $\mathcal{M}_{mn}$ and $\mathcal{M}_{pq}$ be the spaces of $m \times n$ and $p \times q$ matrices, respectively. Let $\mathcal{M}'$ be a subset of $\mathcal{M}_{mn}$. A *matrix-valued function* (*matrix function* for short) is a function that maps a matrix in $\mathcal{M}'$ to a matrix in $\mathcal{M}_{pq}$. We are particularly concerned with the case in which both $\mathcal{M}_{mn}$ and $\mathcal{M}_{pq}$ are real, symmetric, and of the same sizes, although many of the results can be established for the general case.

The differential properties of F are important in view of the recent research on semidefinite programs (SDP) and its generalization, the semidefinite complementarity problems (SDCP). For instance, it is shown (Tseng 1998) that the solution of SDP and SDCP can be reduced to solving a matrix equation $F(X) = 0$, where F is a certain matrix merit function. However, in order to develop Newton-type methods for such equations, some kind of semismoothness properties has to be established for F. Historically, vector-valued semismooth functions played a crucial role in constructing high-order nonsmooth and smoothing Newton methods (see e.g., Qi and Sun 1993, 2001) for nonlinear equations, complementarity problems, and variational inequality problems. There are computational evidences showing that those methods are competitive compared with interior point methods in solving complementarity and variational inequality problems (Burke and Xu 1998, Facchinei and Kanzow 1997). In particular, the recent paper of Chen and Tseng (1999) provided promising results in this type of methods for SDCP. In order to extend semismooth and smoothing Newton methods to SDP, SDCP, and other problems involving the cone of positive semidefinite matrices, the first goal in this direction naturally is to study the semismoothness of matrix functions. In our research we find that this task is not trivial and is often difficult due to the hardship of algebraically representing the constraint of positive semidefiniteness.

This paper is organized as follows. We study the Fréchet differentiability of matrix functions in §2 by using an isometry between vector spaces and matrix spaces and by invoking some results in matrix analysis. The semismoothness and related concepts are considered in §3. Then we show that the absolute-value, the semidefinite projection, and the projective residual matrix functions are strongly semismooth in §4.

Some notations to be used are as follows.

- Usually, calligraphic letters denote matrix sets. Capital letters represent matrices or matrix-valued functions. Lowercase letters are for vectors or vector-valued functions. Greek letters stand for scalars.

Received November 28, 1999; revised November 23, 2000, and June 1, 2001.
MSC 2000 subject classification. Primary: 65K05, 90C25, 90C33.
OR/MS subject classification. Primary: Mathematics/functions; secondary: programming/semidefinite.
Key words. Matrix functions, Newton's method, nonsmooth optimization, semidefinite programming.

0364-765X/02/2701/0150/$05.00
1526-5471 electronic ISSN, © 2002, INFORMS

- $\mathcal{S}^n$ is the space of real symmetric $n \times n$ matrices; $\mathcal{S}^n_+$ is the subset of $\mathcal{S}^n$ consisting of positive semidefinite matrices.
- $\mathcal{S}(n_1, \ldots, n_m)$ is the space of real symmetric $n \times n$ block-diagonal matrices with m blocks of sizes $n_1, \ldots, n_m$.
- $\mathcal{O}$ is the set of all $n \times n$ orthogonal matrices.
- For matrices $A, B \in \mathcal{S}^n$, we define

$$\langle A, B\rangle = \langle A, B\rangle_F = A \bullet B := \mathrm{Trace}\,(AB) = \sum_{i,j=1}^{n} A_{ij} B_{ij}.$$

- A superscript "T" represents the transpose of matrices and vectors.
- $\|A\|_F$, or simply $\|A\|$, is the Frobenius norm of matrix A : $\|A\|_F := \langle A, A\rangle^{1/2}$.
- We write $X \succ 0$ and $X \succeq 0$ if X is positive definite and positive semidefinite, respectively. For $X \succeq 0$, we denote its symmetric square root by $\sqrt{X}$ or $X^{1/2}$.
- **vec** (A) is a column vector whose entries come from A by stacking up columns of A, from the first to the nth column, on top of each other. The operator **mat** is the inverse of **vec**; i.e., **mat** $(\mathbf{vec}(A)) = A$.
- We write $X = O(\alpha)$ (respectively, $o(\alpha)$) if $\|X\|/|\alpha|$ is uniformly bounded (respectively, tends to zero) as $\alpha \to 0$.

2. The differential properties of matrix functions.

2.1. Differentiability of functions.

DEFINITION 2.1. Let $F : \mathcal{S}^n \to \mathcal{S}^n$ and let $X, H \in \mathcal{S}^n$. If $F'_X : \mathcal{S}^n \to \mathcal{S}^n$ is a linear operator that satisfies

$$\lim_{H\to 0} \frac{\|F(X+H) - F(X) - F'_X(H)\|}{\|H\|} = 0, \tag{1}$$

then F is said to be Fréchet differentiable ($\mathcal{F}$-differentiable) at X and F'_X is the $\mathcal{F}$-derivative of F at X.

EXAMPLE 2.2. Let $F(X) = AXA^T$ for any $X \in \mathcal{S}^n$, where $A \in \mathbb{R}^{n\times n}$. It is easy to verify that $\|F(X+H) - F(X) - AHA^T\|_F = 0$. Thus, $F'_X(H) = AHA^T$.

Since $\mathcal{S}^n$ and the inner product $\bullet$ form an Euclidean space, there is a natural isometry identifying $\mathcal{S}^n$ and $\mathbb{R}^\nu$ where $\nu := n(n+1)/2$. We could formally define it as follows.

DEFINITION 2.3. If $A \in \mathcal{S}^n$, then **svec**(A) is a column vector of length ν whose entries come from A by stacking up the lower half of A, including the diagonal entries, in the order of column 1, column 2,. . . , up to column n. The operator **smat** is the inverse of **svec**; i.e., **smat**(**svec** $(A)) = A$. The map **svec** and its inverse then define an isometry between $\mathcal{S}^n$ and $\mathbb{R}^\nu$ if the inner product of the latter space is defined as $\langle a, b\rangle_\nu := \langle \mathbf{smat}(a), \mathbf{smat}(b)\rangle_F$.

EXAMPLE 2.4.

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}, \quad \mathbf{svec}\,(A) = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \|\mathbf{svec}(A)\|_\nu = \sqrt{1^2 + 2(2)^2 + 3^2} = \sqrt{18}.$$

Under such an isometry, each matrix function $F : \mathcal{U} \to \mathcal{S}^n$ will induce a vector function $f : U = \mathbf{svec}(\mathcal{U}) \to \mathbb{R}^\nu$ by

$$f(x) = \mathbf{svec}[F(X)], \tag{2}$$

where $x := \mathbf{svec}(X)$, $X \in \mathcal{U}$. Then $F(X)$ and $f(x)$ will have the same topological properties. In particular, the following results are immediate.

THEOREM 2.5. (i) *F is* (*locally Lipschitz, respectively*) *continuous on* $\mathcal{U}$ *if and only if* f *is* (*locally Lipschitz, respectively*) *continuous on* U.

(ii) *If F has an $\mathcal{F}$-derivative F'_X at X, then f also has an $\mathcal{F}$-derivative f'_x defined similarly to* (1) *and vise versa. Moreover, f'_x is the Jacobian in the usual sense and there holds*

$$F'_X = \mathbf{smat} \circ J(x) \circ \mathbf{svec}\,, \tag{3}$$

where $\circ$ *represents the operator composition and* $J(x) := \frac{\partial f(x)}{\partial x}$.

EXAMPLE 2.6. Let

$$F(X) = X^2, \quad X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}, \quad H = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix}.$$

Then

$$f(x) = \begin{pmatrix} x_{11}^2 + x_{21}^2 \\ x_{21}x_{11} + x_{22}x_{21} \\ x_{21}^2 + x_{22}^2 \end{pmatrix}, \quad J(x) = \begin{pmatrix} 2x_{11} & 2x_{21} & 0 \\ x_{21} & x_{11} + x_{22} & x_{21} \\ 0 & 2x_{21} & 2x_{22} \end{pmatrix},$$

and

$$\begin{aligned} F'_X(H) &= \mathbf{smat} \circ J(x) \circ \mathbf{svec}\,(H) \\ &= \mathbf{smat} \circ \begin{pmatrix} 2x_{11} & 2x_{21} & 0 \\ x_{21} & x_{11} + x_{22} & x_{21} \\ 0 & 2x_{21} & 2x_{22} \end{pmatrix} \begin{pmatrix} h_{11} \\ h_{21} \\ h_{22} \end{pmatrix} \\ &= \mathbf{smat} \begin{pmatrix} 2x_{11}h_{11} + 2x_{21}h_{21} \\ x_{21}h_{11} + x_{11}h_{21} + x_{22}h_{21} + x_{21}h_{22} \\ 2x_{21}h_{21} + 2x_{22}h_{22} \end{pmatrix} \\ &= \begin{pmatrix} 2x_{11}h_{11} + 2x_{21}h_{21} & x_{21}h_{11} + x_{11}h_{21} + x_{22}h_{21} + x_{21}h_{22} \\ x_{21}h_{11} + x_{11}h_{21} + x_{22}h_{21} + x_{21}h_{22} & 2x_{21}h_{21} + 2x_{22}h_{22} \end{pmatrix} \\ &= XH + HX. \end{aligned}$$

There are two important implications of Theorem 2.5, in which the second claim is a direct extension of Rademacher's Theorem from vector functions to matrix functions.

COROLLARY 2.7. (i) *The $\mathcal{F}$-derivative of F at X is unique.*

(ii) *If F is locally Lipschitzian on $\mathcal{S}^n$, then F is almost everywhere $\mathcal{F}$-differentiable on $\mathcal{S}^n$.*

PROOF. The first claim is due to the uniqueness of Jacobian $J(x)$ and Theorem 2.5. The second is obtained by combining Theorem 2.5 and Rademacher's Theorem on R^ν. □

2.2. Extension to symmetric block-diagonal matrix functions. Some semidefinite complementarity problems particularly involve symmetric block-diagonal matrix functions. The results in the last subsection can be extended directly to those matrix functions. The proofs are straightforward.

DEFINITION 2.8. For any $X \in \mathcal{S}(n_1, \ldots, n_m)$ let $\mathbf{Vec}\,(X)$ be a column vector with

$$\mathbf{Vec}(X) := \left(\mathbf{vec}(X_1)^T, \ldots, \mathbf{vec}(X_m)^T\right)^T,$$

where $X_i, i = 1, \ldots, m$ is the ith block of X. The operator $\mathbf{Mat}$ is the inverse of $\mathbf{Vec}$; i.e., $\mathbf{Mat}[\mathbf{Vec}(X)] = X$. The operator $\mathbf{Svec}$ is defined by

$$\mathbf{Svec}(X) := \left(\mathbf{svec}(X_1)^T, \ldots, \mathbf{svec}(X_m)^T\right)^T$$

and $\mathbf{Smat}$ is the inverse of $\mathbf{Svec}$.

The matrix function $F : \mathcal{S}(n_1, \dots, n_m) \to \mathcal{S}(n_1, \dots, n_m)$ under **Svec** becomes a vector-valued function $f : \mathbb{R}^{\bar{\nu}} \to \mathbb{R}^{\bar{\nu}}$ defined by

$$f(x) := \mathbf{Svec}[F(X)], \tag{4}$$

where $x = \mathbf{Svec}(X)$ and $\bar{\nu} = \sum_{i=1}^{m} n_i(n_i + 1)/2$. Conversely, the vector valued f under **Smat** turns to be a matrix function $F : \mathcal{S}(n_1, \dots, n_m) \to \mathcal{S}(n_1, \dots, n_m)$ with

$$F(X) := \mathbf{Smat}\,[f(x)],$$

where $X = \mathbf{Smat}\,(x)$.

Parallel to Theorem 2.5 and Corollary 2.7, we have the following results.

THEOREM 2.9. 1. *F is (locally Lipschitz) continuous on $\mathcal{S}(n_1, \dots, n_m)$ if and only if f is (locally Lipschitz) continuous on $\mathbb{R}^{\bar{\nu}}$.*

2. *$F(X)$ is $\mathcal{F}$-differentiable at X if and only if $f(x)$ is differentiable at x in the usual Jacobian sense.*

3. *The $\mathcal{F}$-derivative of F at X is unique.*

4. *If F is locally Lipschitzian on $\mathcal{S}(n_1, \dots, n_m)$, then F is almost everywhere $\mathcal{F}$-differentiable on $\mathcal{S}(n_1, \dots, n_m)$.*

5. *Let $f : \mathbb{R}^{\bar{\nu}} \to \mathbb{R}^{\bar{\nu}}$ be defined and differentiable on an open subset U of $\mathbb{R}^{\bar{\nu}}$. Let $\mathcal{U} =$ $\mathbf{Smat}(U)$ and $x = \mathbf{Svec}(X)$. Then*

$$\frac{\partial f(x)}{\partial x} = J(x) \quad \textit{for } x \in U$$

if and only if $F'_X = \mathbf{Smat} \circ J(x) \circ \mathbf{Svec}$.

3. Generalized derivative, directional derivative, $\mathcal{B}$-derivative, and semismoothness of matrix functions. Suppose that $F : \mathcal{S}^n \to \mathcal{S}^n$ is a locally Lipschitzian matrix function and $f : \mathbb{R}^{\nu} \to \mathbb{R}^{\nu}$ is defined as in (2). According to Corollary 2.7, F is $\mathcal{F}$-differentiable almost everywhere on $\mathcal{S}^n$. Denote the set of points at which F is $\mathcal{F}$-differentiable by D_F.

DEFINITION 3.1. The generalized derivative of F at X, denoted by ∂F_X, is the set defined as the following:

$$\partial F_X := \mathrm{co}\left\{\lim F'_Z \mid Z \to X, Z \in D_F\right\}, \tag{5}$$

where “co” stands for the convex hull in the usual sense of convex analysis (Rockafellar 1970).

This definition is an extension of the generalized (Clarke) Jacobian $\partial f(x)$ (Clarke 1983):

$$\partial f(x) = \mathrm{co}\, \partial_B f(x) \quad \text{with } \partial_B f(x) = \{\lim J(z) \mid z \to x, z \in D_f\}, \tag{6}$$

where D_f is the set of differentiable points of f in $\mathbb{R}^{\nu}$ and $J(z)$ is the Jacobian of f at $z \in D_f$. It is easy to see by the differential correspondence that there is an isometry between $\partial f(x)$ and ∂F_X. Similar to (3), one can write

$$\partial F_X = \mathbf{smat} \circ \partial f(x) \circ \mathbf{svec}. \tag{7}$$

The following topological properties of ∂F_X can then be identified with the corresponding properties of $\partial f(x)$.

THEOREM 3.2. *Let $F : \mathcal{S}^n \to \mathcal{S}^n$ be locally Lipschitzian on an open neighborhood $\mathcal{U}$ of X.*

(a) *∂F_X is nonempty, convex, and compact subset of $(\mathcal{S}^n)^*$.*

(b) *The mapping* ∂F_X *is closed at* X; *that is, if* $Z \to X, W \in \partial F_Z, W \to V$, *then* $V \in \partial F_X$.

(c) ∂F *is upper-semicontinuous*: *For any* $\varepsilon > 0$ *there is* $\delta > 0$ *such that, for all* $Y \in X + \delta B$ (B *is the unit ball in* $\mathcal{S}^n$),

$$\partial F_Y \subset \partial F_X + \varepsilon B^*,$$

where B^* *is the unit ball in* $(\mathcal{S}^n)^*$.

(d) *If each entry* F_{ij} *of* $F(X)$ *is* α_{ij} *Lipschitzian at* $x (= \mathbf{svec}\,(X))$ *in the sense that*

$$\|F_{ij}(x) - F_{ij}(y)\| \le \alpha_{ij} \|x - y\|,$$

then $F(X)$ *is* $\alpha := \max(\alpha_{ij})$ *Lipschitzian at* X.

(e) *The mean value theorem holds*:

$$F(X) - F(Y) \in \mathrm{co}\big\{\partial F_Z(X - Y) | Z \in [X, Y]\big\} =: \partial F_{[X,Y]}, \tag{8}$$

where $[X, Y]$ *is the line segment connecting* X *and* Y.

The directional derivative and $\mathcal{B}$-derivative of a matrix function can be defined in a similar way to scalar or vector-valued functions.

Definition 3.3. Let $F : \mathcal{S}^n \to \mathcal{S}^n$ and $X, H \in \mathcal{S}^n$. The directional derivative of F at X along H is the following limit (if it exists at all):

$$F'(X; H) = \lim_{\tau \downarrow 0} \frac{F(X + \tau H) - F(X)}{\tau}.$$

F is said to be directionally differentiable at X if $F'(X; H)$ exists for all H.

Corollary 3.4. *If* F *is* $\mathcal{F}$*-differentiable at* X, *then for any* $H \in \mathcal{S}^n$,

$$F'(X; H) = F_X(H).$$

Definition 3.5. Let $F : \mathcal{S}^n \to \mathcal{S}^n$ and $X \in \mathcal{S}^n$. Then F is said to be $\mathcal{B}$-differentiable at X if there exists a function $BF(X) : \mathcal{S}^n \to \mathcal{S}^n$, called the $\mathcal{B}$-derivative of F at X, which is positively homogeneous of degree 1 (i.e., $BF(X)(\tau H) = \tau BF(X)(H)$ for all $H \in \mathcal{S}^n$ and all $\tau \ge 0$), such that for any $H \in \mathcal{S}^n$ and $H \to 0$,

$$F(X + H) - F(X) - BF(X)(H) = o(\|H\|). \tag{9}$$

It can easily be seen that there is also an isometry identifying the $\mathcal{B}$-differentiable of f at x with $\mathcal{B}$-derivative of F at X.

Shapiro (1990) showed that a vector-valued locally Lipschitzian function $f : R^\nu \to R^\nu$ is $\mathcal{B}$-differentiable at x if and only if it is directionally differentiable at x. Therefore, the $\mathcal{B}$-differentiability of matrix-valued locally Lipschitzian functions is also equivalent to their directional differentiability.

In the case in which a function $f : \mathbb{R}^\nu \to \mathbb{R}^\nu$ is not differentiable, but locally Lipschitzian, Mifflin (1977) and Qi and Sun (1993) introduced the concepts of semismoothness and p-order semismoothness for vector-valued function f. The definition of semismoothness implies that the function is directionally differentiable (hence is $\mathcal{B}$-differentiable according to Shapiro 1990) and that (Proposition 1 of Pang and Qi 1993) there exists $\varepsilon > 0$ such that

$$\sup_{v \in \partial f(x+h)} \|f(x + h) - f(x) - vh\| = o(\|h\|) \ \forall \ \|h\| < \varepsilon. \tag{10}$$

In the case of p-order semismooth the term $o(\|h\|)$ is replaced by $O(\|h\|^{1+p})$.

Equation (10) plays an important role in developing semismooth Newton methods and smoothing methods for nonlinear equation systems and nonlinear programming problems. We now extend the definition of semismoothness to matrix functions in terms of (10).

DEFINITION 3.6. Suppose that $F : \mathcal{S}^n \to \mathcal{S}^n$ is a locally Lipschitzian matrix-valued function. F is said to be semismooth at $X \in \mathcal{S}^n$ if F is directionally differentiable at X and for any $V \in \partial F_{X+H}$ and $H \to 0$,

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

F is said to be p-order ($0 < p < \infty$) semismooth at X if F is semismooth at X and

$$F(X+H) - F(X) - V(H) = O(\|H\|^{1+p}). \tag{11}$$

In particular, F is called strongly semismooth at X if F is first-order semismooth at X.

We find it is convenient to establish an equivalent definition for the analysis of matrix functions.

THEOREM 3.7. *Suppose that $F : \mathcal{S}^n \to \mathcal{S}^n$ is locally Lipschitzian and directionally differentiable in a neighborhood of X. Then for any $p \in (0, \infty)$ the following two statements are equivalent*:

(a) *for any $V \in \partial F(X+H), H \to 0$,*

$$F(X+H) - F(X) - V(H) = O(\|H\|^{1+p});$$

(b) *for any $X + H \in D_F, H \to 0$,*

$$F(X+H) - F(X) - F'(X+H; H) = O(\|H\|^{1+p}).$$

PROOF. From the differential correspondence between F and f the conclusion of this theorem is valid if one can show that the following two statements are equivalent:

(a′) for any $v \in \partial f(x+h), h \to 0$,

$$f(x+h) - f(x) - vh = O(\|h\|^{1+p});$$

(b′) for any $x + h \in D_f, h \to 0$,

$$f(x+h) - f(x) - f'(x+h; h) = O(\|h\|^{1+p}).$$

(a′)⇒(b′) is obvious.

Next we prove (b′)⇒(a′): Assume by contradiction that (b′) holds while (a′) does not hold. Then, there exists a sequence $\{h^i\}_{i=1}^{\infty}$ ($h^i \neq 0$) converging to 0 and a corresponding generalized Jacobian sequence $v_i \in \partial f(x+h^i)$ such that

$$\frac{\|f(x+h^i) - f(x) - v_i h^i\|}{\|h^i\|^{1+p}} \to \infty, \quad i \to \infty. \tag{12}$$

By the Carathéodory theorem, any $v_i \in \partial f(x+h^i)$ can be expressed as

$$v_i = \sum_{j=1}^{\nu^2+1} \lambda_{i_j} v_{i_j}, \tag{13}$$

where

$$\lambda_{i_j} \in [0, 1], \quad \sum_{j=1}^{\nu^2+1} \lambda_{i_j} = 1 \tag{14}$$

and $v_{i_j} \in \partial_B f(x+h^i)$. For each $v_{i_j} \in \partial_B f(x+h^i)$, by the definition of $\partial_B f(x+h^i)$, there exists $y^{i_j} \in D_f$ such that

$$\|y^{i_j} - (x+h^i)\| \le \|h^i\|^{1+p} \tag{15}$$

and

$$\|v_{i_j} - f'(y^{i_j})\| \le \|h^i\|^p. \tag{16}$$

By (13)–(16), we obtain

$$\begin{aligned}
&\|f(x+h^i)-f(x)-v_i h^i\| \\
&\quad \le \sum_{j=1}^{\nu^2+1} \lambda_{i_j} \|f(x+h^i)-f(x)-f'(y^{i_j})h^i\| + \sum_{j=1}^{\nu^2+1} \lambda_{i_j} \|[v_{i_j} - f'(y^{i_j})]h^i\| \\
&\quad \le \sum_{j=1}^{\nu^2+1} \lambda_{i_j} \|f(x+h^i)-f(x)-f'(y^{i_j})h^i\| + O(\|h^i\|^{1+p}) \\
&\quad = \sum_{j=1}^{\nu^2+1} \lambda_{i_j} \|f(x+h^i)-f(x)-f'(y^{i_j})[(y^{i_j}-x)+(x+h^i-y^{i_j})]\| \\
&\qquad + O(\|h^i\|^{1+p}) \\
&\quad \le \sum_{j=1}^{\nu^2+1} \lambda_{i_j} \|f(x+h^i)-f(x)-f'(y^{i_j})(y^{i_j}-x)\| \\
&\qquad + \sum_{j=1}^{\nu^2+1} \lambda_{i_j} \|f'(y^{i_j})(x+h^i-y^{i_j})\| + O(\|h^i\|^{1+p}) \\
&\quad \le \sum_{j=1}^{\nu^2+1} \lambda_{i_j} \|f(x+h^i)-f(x)-f'(y^{i_j})(y^{i_j}-x)\| + O(\|h^i\|^{1+p}),
\end{aligned} \tag{17}$$

where in the last inequality we used the fact that $\{f'(y^{i_j})\}$ are uniformly bounded because of the local Lipschitzian property of f. Relations (17), (14), and (15), together with the Lipschitzian continuity of f, imply that

$$\|f(x+h^i)-f(x)-v_i h^i\| \le \sum_{j=1}^{\nu^2+1} \lambda_{i_j} \|f(y^{i_j})-f(x)-f'(y^{i_j})(y^{i_j}-x)\| + O(\|h^i\|^{1+p}). \tag{18}$$

Thus, by (b′), (14), and (15), from (18) we obtain

$$\|f(x+h^i)-f(x)-v_i h^i\| \le O(\|y^{i_j}-x\|^{1+p}) + O(\|h^i\|^{1+p}) = O(\|h^i\|^{1+p}),$$

which contradicts (12). This contradiction shows that (b′) ⇒ (a′). □

REMARK. Note that the result of Theorem 3.7 is also true if $O(\|H\|^{1+p})$ is replaced by $o(\|H\|)$.

We list below some properties of semismooth matrix functions involving components and composition.

THEOREM 3.8. *Suppose that $F : \mathcal{S}^n \to \mathcal{S}^n$ is a locally Lipschitzian function. Let $p \in (0,\infty)$. If each entry of F, viewed as $F_{kj} : \mathcal{S}^n \to \mathrm{diag}(\mathbb{R}, 0, \ldots, 0) \subset \mathcal{S}^n$, is (p-order semismooth) semismooth at X, then F is (p-order semismooth) semismooth at X.*

PROOF. For $X+H \in D_F, H \to 0$,

$$\begin{aligned}&\|F(X+H)-F(X)-F'(X+H;H)\| \\ &\quad \le \sum_{j,k=1}^{n} |[F(X+H)]_{kj}-[F(X)]_{kj}-[(F)_{kj}]'(X+H;H)|,\end{aligned}$$

which completes the proof. □

The following result on semismoothness is due to Mifflin (1977) and the result on p-order semismoothness is essentially due to Fischer (1997).

LEMMA 3.9. *Let $p \in (0,\infty)$. Suppose that $f : \mathbb{R}^{\nu} \to \mathbb{R}^{\nu_1}$ is (p-order semismooth) semismooth at $x \in \mathbb{R}^{\nu}$ and $g : \mathbb{R}^{\nu_1} \to \mathbb{R}^{\nu_2}$ is (p-order semismooth) semismooth at $f(x)$. Then, the composite function $h := g \circ f$ is (p-order semismooth) semismooth at x.*

The above lemma implies

THEOREM 3.10. *Let $p \in (0,\infty)$. Suppose $F : \mathcal{S}^n \to \mathcal{S}^d$ is (p-order semismooth) semismooth at X and $G : \mathcal{S}^d \to \mathcal{S}^t$ is (p-order semismooth) semismooth at $F(X)$. Then the composite function $G \circ F$ is (p-order semismooth) semismooth at X.*

PROOF. Define two vector-valued functions g and h by

$$g(y) := \mathbf{svec}[G(Y)],$$

where $y := \mathbf{svec}(Y)$, $Y \in \mathcal{S}^d$ and

$$h(x) := \mathbf{svec}\,[G(F(X))],$$

where $x := \mathbf{svec}(X)$, $X \in \mathcal{S}^n$. Then, by Lemma 3.9 and the relationship $\mathbf{smat} \circ (g \circ f) \circ \mathbf{svec} = G \circ F$, the above result follows directly. □

We conclude this section by noting that all the results obtained in this section can be analogously extended to block-diagonal matrix functions.

4. Some strongly semismooth matrix functions. The projective residual function $r : \mathbb{R}^{\nu} \to \mathbb{R}^{\nu}$

$$r(x) := g(x) - \max[0, g(x) - f(x)], \tag{19}$$

is often used in continuation methods, QP-free methods, smoothing methods, and merit function methods; see Qi and Sun (2001) for its applications. It is also closely related to other merit functions. We will show that the corresponding matrix projective residual function is locally Lipschitzian and strongly semismooth.

Let $\mathcal{S}(n_1,\dots,n_m)_+$ be the cone of symmetric block-diagonal positive semidefinite matrices. Define the *semidefinite projection matrix function* $[\cdot]_+ : \mathcal{S}(n_1,\dots,n_m) \to \mathcal{S}(n_1,\dots,n_m)_+$ as the matrix-valued function that satisfies:

(1) $[X]_+ \in \mathcal{S}(n_1,\dots,n_m)_+$;

(2) $\|X-[X]_+\| \le \|X-Z\|\ \forall\, Z \in \mathcal{S}(n_1,\dots,n_m)_+$.

The following proposition means that we can focus our discussion on matrix projection from $\mathcal{S}^n$ to $\mathcal{S}^n_+$ rather than that from $\mathcal{S}(n_1,\dots,n_m)$ to $\mathcal{S}(n_1,\dots,n_m)_+$.

PROPOSITION 4.1. *For any $X \in \mathcal{S}(n_1,\dots,n_m)$, we have*

$$[X]_+ = \begin{pmatrix} [X_1]_+ & & \\ & \ddots & \\ & & [X_m]_+ \end{pmatrix},$$

where X_i is the i-th block matrix of X.

PROOF. Straightforward. $\square$

Since $(\mathcal{S}^n, \langle\cdot,\cdot\rangle)$ is a Hilbert space, we have the following proposition (Zarantonello 1971).

PROPOSITION 4.2. *For the positive semidefinite projection function* $[\cdot]_+$, *we have*
(a) *for any* $X \in \mathcal{S}^n$ *and* $Z \in \mathcal{S}^n_+$, $Z = [X]_+$ *if and only if* $\langle X - Z, Y - Z\rangle \le 0\ \forall\ Y \in \mathcal{S}^n_+$.
(b) $\|[Y]_+ - [X]_+\| \le \|Y - X\|\ \forall\ Y, X \in \mathcal{S}^n$.

Proposition 4.2(b) says that $[\cdot]_+$ is globally Lipschitz continuous with Lipschitz constant 1.

DEFINITION 4.3. Let $F, G : \mathcal{S}(n_1, \ldots, n_m) \to \mathcal{S}(n_1, \ldots, n_m)$. The matrix projective residual function is defined as

$$R(X) := G(X) - [G(X) - F(X)]_+, \quad X \in \mathcal{S}(n_1, \ldots, n_m).$$

It is well known (Tseng 1998, Zarantonello 1971) that the root of the above function satisfies $G(X) \succeq 0, F(X) \succeq 0$, and $\langle G(X), F(X)\rangle = 0$. Thus, one can obtain a solution to SDCP by solving $R(X) = 0$. Similarly to Proposition 4.1, we could focus on residual functions from $\mathcal{S}^n$ to $\mathcal{S}^n$ only. The strong semismoothness of such functions depends on the strong semismoothness of $[\cdot]_+$ according to Theorem 3.10 if G and F are strongly semismooth.

Note that function $[\cdot]_+$ can be expressed as the optimal solution of the parametric SDP

$$[X]_+ = \arg\min\{\|X - Y\| \mid Y \succeq 0\}.$$

Hence $[\cdot]_+$ is quite different from the vector function $\max(0, \cdot)$. The vector case is always strongly semismooth due to the piecewise linear structure of the function max. However, the matrix case does not have a piecewise linear structure. This poses great difficulty in proving the (strong) semismoothness of $[\cdot]_+$.

We further reduce the problem to the strong semismoothness of the so-called absolute-value function.

DEFINITION 4.4. Let $A \in \mathcal{S}(n_1, \ldots, n_m)$ be positive semidefinite. Then there exists a unique symmetric positive semidefinite matrix $B \in \mathcal{S}(n_1, \ldots, n_m)$ such that $B^2 = A$. We call B the square root of A and denote it by $B = \sqrt{A}$. The matrix absolute value of A is defined as $|A| := \sqrt{A^2}$.

The following proposition says that the three matrix functions—the absolute-value, the semidefinite-projection, and the projective residual functions—will be all strongly semismooth if one of them is.

PROPOSITION 4.5. *There hold* $|X| = [X]_+ + [-X]_+$ *and* $[X]_+ = [X + |X|]/2$.

PROOF. Tseng (1998) showed that for any $X \in \mathcal{S}^n$,

$$[X]_+ = P\,\mathrm{diag}\big[\max(0, \lambda_1), \ldots, \max(0, \lambda_n)\big]P^T,$$

where $P \in \mathcal{O}$ and $\lambda_1, \ldots, \lambda_n \in R$ satisfy $X = P\mathrm{diag}\,[\lambda_1, \ldots, \lambda_n]P^T$. By definition we have

$$|X| = P\,\mathrm{diag}[|\lambda_1|, \ldots, |\lambda_n|]P^T.$$

The proposition follows. $\square$

For any matrix $X \in \mathcal{S}^n$, there exists $P \in \mathcal{O}$ such that

$$X = P\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}P^T, \tag{20}$$

where D is a nonsingular matrix. Let m denote the rank of X and let

$$K := \{1, \ldots, m\}, \qquad J := \{1, \ldots, n\} \backslash K.$$

Let $C := |D|$ and define the linear operator L_C by

$$L_C(Z) := CZ + ZC, \qquad Z \in \mathcal{S}^m$$

and L_C^{-1} its inverse operator (see Tseng 1998 for more discussions on the linear operator L_C and its inverse).

THEOREM 4.6. *Denote $|Y|$ by $F(Y)$, $Y \in \mathcal{S}^n$. If $X \in \mathcal{S}^n$ is nonsingular, then F'_X exists and satisfies*

$$F'_X(H) = L_{|X|}^{-1}(XH + HX). \tag{21}$$

Moreover, for $H \to 0$, $H \in \mathcal{S}^n$, we have

$$F(X+H) - F(X) - F'_X(H) = O(\|H\|^2). \tag{22}$$

PROOF. Let

$$\begin{aligned} T &:= F(X+H) - F(X) - F'_X(H), \\ \Delta F &:= F(X+H) - F(X). \end{aligned}$$

We shall substitute (21) for $F'_X(H)$ to verify that $T = O(\|H\|^2)$, which will prove both (21) and (22). Notice that from (21) we have

$$|X|F'_X(H) + F'_X(H)|X| = XH + HX, \qquad H \in \mathcal{S}^n. \tag{23}$$

From the definition of T,

$$\begin{aligned} (|X+H| + |X|)T &= (X+H)^2 - X^2 - |X+H||X| + |X||X+H| - (|X+H| + |X|)F'_X(H) \\ &= (X+H)^2 - X^2 - \Delta F|X| + |X|\Delta F - (|X+H| + |X|)F'_X(H). \end{aligned}$$

That is,

$$[2|X| + O(\|H\|)]T = XH + HX + H^2 - \Delta F|X| + |X|\Delta F - [2|X| + O(\|H\|)]F'_X(H),$$

which implies that

$$2|X|\Delta F - 2|X|F'_X(H) = XH + HX - \Delta F|X| + |X|\Delta F - 2|X|F'_X(H) + O(\|H\|^2) \tag{24}$$

because from the definitions of $F'_X(H)$ and ΔF it is clear that

$$\|F'_X(H)\| = O(\|H\|) \quad \text{and} \quad \|\Delta F\| = O(\|H\|).$$

Hence, from (23) and (24) we have

$$L_{|X|}(\Delta F - F'_X(H)) = O(\|H\|^2).$$

By observing that $L_{|X|}$ is invertible because $|X|$ is positive definite, we have

$$T = \Delta F - F'_X(H) = O(\|H\|^2).$$

The proof is complete. $\square$

THEOREM 4.7. *Let $F : \mathscr{S}^n \to \mathscr{S}^n$ be defined by $F(Y) := |Y|$, $Y \in \mathscr{S}^n$. Then F is directionally differentiable at any $X \in \mathscr{S}^n$ and for any $H \in \mathscr{S}^n$,*

$$F'(X;H) = P\begin{pmatrix} L_C^{-1}[D\widetilde{H}_{KK} + \widetilde{H}_{KK}D] & C^{-1}D\widetilde{H}_{KJ} \\ \widetilde{H}_{KJ}^T DC^{-1} & |\widetilde{H}_{JJ}| \end{pmatrix} P^T,$$

where D is a nonsingular diagonal matrix defined in (20), *$C := |D|$, and $\widetilde{H} = P^T HP$.*

PROOF. For any $H \in S$ and $\tau \in [0, \infty)$, let

$$\Delta(\tau) := F(X + \tau H) - F(X)$$

and

$$\widetilde{\Delta}(\tau) := P^T \Delta(\tau) P.$$

Then,

$$\begin{aligned}\widetilde{\Delta}(\tau) &= P^T F(X+\tau H)P - P^T F(X)P \\ &= |P^T(X+\tau H)P| - |P^T XP| \\ &= \sqrt{[P^T XP + \tau P^T HP]^2} - \sqrt{(P^T XP)^2}, \\ &= \sqrt{[\widetilde{X} + \tau \widetilde{H}]^2} - \sqrt{\widetilde{X}^2},\end{aligned}$$

where $\widetilde{X} := P^T XP$. Define

$$\widetilde{C} := \sqrt{\widetilde{X}^2} = \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix}.$$

Then,

$$\widetilde{\Delta}(\tau) = \sqrt{\widetilde{C}^2 + \widetilde{W}} - \widetilde{C},$$

where

$$\widetilde{W} = \tau \widetilde{X}\widetilde{H} + \tau \widetilde{H}\widetilde{X} + \tau^2 \widetilde{H}^2.$$

After simple computations we have

$$\widetilde{W} = \tau \begin{pmatrix} D\widetilde{H}_{KK} + \widetilde{H}_{KK}D & D\widetilde{H}_{KJ} \\ \widetilde{H}_{KJ}^T D & 0 \end{pmatrix} + \begin{pmatrix} O(\tau^2) & O(\tau^2) \\ O(\tau^2) & \tau^2[\widetilde{H}_{KJ}^T \widetilde{H}_{KJ} + \widetilde{H}_{JJ}^2] \end{pmatrix}. \tag{25}$$

By Lemma 6.2 in Tseng (1998), we have

$$\widetilde{\Delta}(\tau)_{KK} = L_C^{-1}(\widetilde{W}_{KK}) + o(\|\widetilde{W}\|), \tag{26}$$

$$\widetilde{\Delta}(\tau)_{KJ} = C^{-1}\widetilde{W}_{KJ} + o(\|\widetilde{W}\|), \tag{27}$$

and

$$\widetilde{W}_{JJ} = \widetilde{\Delta}(\tau)_{KJ}^T \widetilde{\Delta}(\tau)_{KJ} + \widetilde{\Delta}(\tau)_{JJ}^2. \tag{28}$$

Thus,

$$\widetilde{\Delta}(\tau)_{KJ} = \tau C^{-1} D\widetilde{H}_{KJ} + o(\tau). \tag{29}$$

Let

$$E := C^{-1}D.$$

Then

$$E_{ij} = \begin{cases} 1 & \text{or } -1 \text{ if } j = i, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, from (29),

$$\widetilde{\Delta}(\tau)_{KJ}^T \widetilde{\Delta}(\tau)_{KJ} = \tau^2 \widetilde{H}_{KJ}^T E^2 \widetilde{H}_{KJ} + o(\tau^2) = \tau^2 \widetilde{H}_{KJ}^T \widetilde{H}_{KJ} + o(\tau^2). \tag{30}$$

According to (26) and (25),

$$\widetilde{\Delta}(\tau)_{KK} = \tau L_C^{-1}(D\widetilde{H}_{KK} + \widetilde{H}_{KK}D) + o(\tau). \tag{31}$$

Since

$$\widetilde{W}_{JJ} = \tau^2[\widetilde{H}_{KJ}^T \widetilde{H}_{KJ} + \widetilde{H}_{JJ}^2],$$

from (28) and(30), we obtain

$$\widetilde{\Delta}(\tau)_{JJ}^2 = \tau^2 \widetilde{H}_{JJ}^2 + o(\tau^2). \tag{32}$$

Furthermore, since $\widetilde{\Delta}(\tau)_{JJ}$ is positive semidefinite (see the definition of $\widetilde{\Delta}(\tau)$), from (32), $\widetilde{\Delta}(\tau)_{JJ}$ satisfies

$$\widetilde{\Delta}(\tau)_{JJ} = \tau\sqrt{\widetilde{H}_{JJ}^2 + o(1)}. \tag{33}$$

Then from (31), (29), and (33) we obtain

$$\lim_{\tau\downarrow 0} \frac{\widetilde{\Delta}(\tau)}{\tau} = \begin{pmatrix} L_C^{-1}[D\widetilde{H}_{KK} + \widetilde{H}_{KK}D] & C^{-1}D\widetilde{H}_{KJ} \\ \widetilde{H}_{KJ}^T D C^{-1} & |\widetilde{H}_{JJ}| \end{pmatrix},$$

which proves the theorem. □

Theorem 4.7 says that the matrix absolute-value function is directionally differentiable with the explicit formula given for the directional derivative. Theorem 4.6 says that the function is $\mathscr{F}$-differentiable at all nonsingular X. The following theorem says that the $\mathscr{F}$-differentiability is lost elsewhere.

THEOREM 4.8. *Let $F : \mathscr{S}^n \to \mathscr{S}^n$ be defined by $F(Y) := |Y|$, $Y \in \mathscr{S}^n$. Then for any singular matrix $X \in S$, F is not $\mathscr{F}$-differentiable at X.*

PROOF. By Theorem 4.7, F is directionally differentiable at X. By contradiction, suppose that F is $\mathscr{F}$-differentiable at X. Then $F'(X;\cdot) = F'_X(\cdot)$ is a linear operator. But, let $H \in S^n$ be such that

$$\widetilde{H} := P^T H P = \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}$$

and $G = -H$, where I is the identity matrix in $\mathbb{R}^{r\times r}$ and $r = n - \operatorname{rank}(X) \geq 1$. Then, by Theorem 4.7, we should have

$$\begin{aligned} 0 &= F'(X; G+H) = F'_X(G+H) = F'_X(G) + F'_X(H) \\ &= F'(X; G) + F'(X; H) = 2P\begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} P^T, \end{aligned}$$

which is a contradiction. This contradiction shows that F is not $\mathscr{F}$-differentiable at X. The proof is complete. □

An interesting implication of Theorems 4.6 and 4.8 is that the Lebesgue measure of singular symmetric matrices is zero. This is almost intuitively obvious and is a useful fact in our analysis later.

COROLLARY 4.9. *Matrices in $\mathcal{S}^n$ are almost everywhere nonsingular.*

PROOF. By Theorems 4.6 and 4.8, $F(Y) := |Y|,\ Y \in \mathcal{S}^n$ is $\mathcal{F}$-differentiable at $X \in \mathcal{S}^n$ if and only if X is nonsingular. Since F is Lipschitzian, F is $\mathcal{F}$-differentiable almost everywhere in $\mathcal{S}^n$ by Theorem 2.7. Thus, matrices in $\mathcal{S}^n$ are almost everywhere nonsingular. □

LEMMA 4.10. *Let $F : \mathcal{S}^n \to \mathcal{S}^n$ be defined by $F(Y) = |Y|,\ Y \in S$. Then for any nonsingular matrix $H \in \mathcal{S}^n$, $F'_H(\cdot)$ exists and satisfies*

$$F(H) - F(0) - F'_H(H) = 0. \tag{34}$$

PROOF. By Theorem 4.6, for any nonsingular $H \in \mathcal{S}^n$, $F'_H(\cdot)$ exists and satisfies

$$F'_H(H)|H| + |H|F'_H(H) = 2H^2,$$

which implies that $F'_H(H) = |H| = F(H)$ because $F'_H(H)$ is the unique $Z \in \mathcal{S}^n$ satisfying $Z|H| + |H|Z = 2H^2$ and $Z = |H|$ satisfies this equation. This completes the proof by observing $F(0) = 0$. □

The following Weyl's perturbation result for eigenvalues of symmetric matrices (see Bhatia 1997, Corollary III.2.6) is useful in proving the strong semismoothness of the absolute value function.

LEMMA 4.11. *Let $\sigma_1 \geq \cdots \geq \sigma_n$ be the eigenvalues of any $A \in \mathcal{S}^n$ and $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of $A + H$ for any $H \in \mathcal{S}^n$. Then*

$$\max_{1 \leq i \leq n} |\sigma_i - \lambda_i| \leq \|H\|.$$

We now prove a main lemma that, together with Theorem 3.7, leads to the strong semismoothness of the absolute-value function.

LEMMA 4.12. *Let $F : \mathcal{S}^n \to \mathcal{S}^n$ be defined by $F(Y) := |Y|,\ Y \in \mathcal{S}^n$. Assume that $X \in \mathcal{S}^n, X \neq 0$. Then, for any $H \in \mathcal{S}^n$ such that $X + H$ is nonsingular, $F(\cdot)$ is $\mathcal{F}$-differentiable at $X + H$ and*

$$F(X + H) - F(X) - F'_{X+H}(H) = O(\|H\|^2). \tag{35}$$

PROOF. For any $H \in \mathcal{S}^n$ such that $X + H$ is nonsingular, denote

$$\Delta T := F(X + H) - F(X) - F'_{X+H}(H). \tag{36}$$

Then,

$$|X + H|\Delta T = (X + H)^2 - |X + H||X| - |X + H|F'_{X+H}(H) \tag{37}$$

and

$$\Delta T|X + H| = (X + H)^2 - |X||X + H| - F'_{X+H}(H)|X + H|. \tag{38}$$

By (37), (38), and Theorem 4.6, we have

$$\begin{aligned} |X + H|\Delta T + \Delta T|X + H| &= 2(X + H)^2 - [|X + H||X| + |X||X + H|] \\ &\quad - [|X + H|F'_{X+H}(H) + F'_{X+H}(H)|X + H|] \\ &= 2(X + H)^2 - [|X + H||X| + |X||X + H|] \\ &\quad - [(X + H)H + H(X + H)] \\ &= (X + H)X + X(X + H) - [|X + H||X| + |X||X + H|]. \end{aligned} \tag{39}$$

Since $0 \neq X$, there exists an orthogonal matrix $Q \in \mathcal{O}$ and a diagonal matrix Σ such that

$$Q^T X Q = \Sigma = \operatorname{diag}[\sigma_1, \ldots, \sigma_m, 0, \ldots, 0],$$

where $\sigma_1 \geq \cdots \geq \sigma_m$, $\sigma_i \neq 0$, $i = 1, \ldots, m$ and $m \in \{1, \ldots, n-1\}$. Let

$$\widetilde{H} = Q^T H Q.$$

Then, from (39), we have

$$\begin{aligned} &|\Sigma + \widetilde{H}| Q^T \Delta T Q + Q^T \Delta T Q |\Sigma + \widetilde{H}| \\ &\quad = (\Sigma + \widetilde{H})\Sigma + \Sigma(\Sigma + \widetilde{H}) - \big[|\Sigma + \widetilde{H}||\Sigma| + |\Sigma||\Sigma + \widetilde{H}|\big]. \end{aligned} \tag{40}$$

It is noted that if $X + H \in \mathcal{S}^n$ is nonsingular, then $\Sigma + \widetilde{H} = Q^T(X+H)Q$ is also nonsingular and $H \to 0 \Longleftrightarrow \widetilde{H} \to 0$. Thus, for any $H \in \mathcal{S}^n$ such that $X + H$ is nonsingular, there exists an orthogonal matrix $P \in \mathcal{O}$ (depending on H) and a nonsingular diagonal matrix Λ (depending on H) such that

$$P^T(\Sigma + \widetilde{H})P = \Lambda = \operatorname{diag}[\lambda_1, \cdots, \lambda_m, \lambda_{m+1}, \ldots, \lambda_n], \tag{41}$$

where

$$\lambda_1 \geq \cdots \geq \lambda_m, \quad \min_{1 \leq i \leq m} |\lambda_i| \geq \min_{m+1 \leq i \leq n} |\lambda_i|, \quad \text{and} \quad \lambda_i \neq 0, \ i = 1, 2, \ldots, n.$$

Then, by Lemma 4.11,

$$\lambda_i = \sigma_i + O(\|\widetilde{H}\|), \qquad i = 1, \ldots, m \tag{42}$$

and

$$\lambda_i = O(\|\widetilde{H}\|), \qquad i = m+1, \ldots, n. \tag{43}$$

Equations (40) and (41) imply

$$|\Lambda| \Delta\widetilde{T} + \Delta\widetilde{T}|\Lambda| = \Lambda(P^T \Sigma P) + (P^T \Sigma P)\Lambda - \big[|\Lambda| P^T |\Sigma| P + P^T |\Sigma| P |\Lambda|\big], \tag{44}$$

where

$$\Delta\widetilde{T} := P^T Q^T \Delta T Q P. \tag{45}$$

Therefore, from (44) we have in fact obtained for $i, j = 1, \ldots, n$ that

$$\Delta\widetilde{T}_{ij} = \frac{\lambda_i + \lambda_j}{|\lambda_i| + |\lambda_j|} U_{ij} - V_{ij}, \tag{46}$$

where $U = P^T \Sigma P$ and $V = P^T |\Sigma| P$. Define a vector $s \in \mathbb{R}^m$ by

$$s_i = \begin{cases} 1 & \text{if } \sigma_i > 0, \\ -1 & \text{otherwise,} \end{cases} \qquad i = 1, \ldots, m.$$

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

$$U_{ij} = \sum_{k=1}^{m} P_{ki} P_{kj} \sigma_k \tag{47}$$

and

$$V_{ij} = \sum_{k=1}^{m} P_{ki} P_{kj} \sigma_k s_k, \tag{48}$$

where P_{ij} is the (i, j)-th entry of P. Hence, for $i, j = 1, \ldots, n$,

$$\Delta \widetilde{T}_{ij} = \frac{\lambda_i + \lambda_j}{|\lambda_i| + |\lambda_j|} \left(\sum_{k=1}^{m} P_{ki} P_{kj} \sigma_k \right) - \sum_{k=1}^{m} P_{ki} P_{kj} \sigma_k s_k. \tag{49}$$

Next, we establish a relationship between P and $\widetilde{H}$. Suppose q is a nonnegative integer such that

$$\sigma_1 \geq \cdots \geq \sigma_q > 0 > \sigma_{q+1} \geq \cdots \geq \sigma_m.$$

Then, from (41), we have $P\Lambda = (\Sigma + \widetilde{H})P$, which implies for $i = 1, \ldots, n$,

$$\lambda_i \begin{pmatrix} P_{1i} \\ \vdots \\ P_{qi} \\ \vdots \\ P_{mi} \\ \vdots \\ P_{ni} \end{pmatrix} = \begin{pmatrix} (\sigma_1 + \widetilde{H}_{11}) P_{1i} + \cdots + \widetilde{H}_{1q} P_{qi} + \cdots + \widetilde{H}_{1m} P_{mi} + \cdots + \widetilde{H}_{1n} P_{ni} \\ \vdots \\ \widetilde{H}_{1q} P_{1i} + \cdots + (\sigma_q + \widetilde{H}_{qq}) P_{qi} + \cdots + \widetilde{H}_{qm} P_{mi} + \cdots + \widetilde{H}_{qn} P_{ni} \\ \vdots \\ \widetilde{H}_{1m} P_{1i} + \cdots + \widetilde{H}_{qm} P_{qi} + \cdots + (\sigma_m + \widetilde{H}_{mm}) P_{mi} + \cdots + \widetilde{H}_{mn} P_{ni} \\ \vdots \\ \widetilde{H}_{1n} P_{1i} + \cdots + \widetilde{H}_{qn} P_{qi} + \cdots + \widetilde{H}_{mn} P_{mi} + \cdots + \widetilde{H}_{nn} P_{ni} \end{pmatrix}. \tag{50}$$

Since P in an orthogonal matrix, for $i, j = 1, \ldots, n$,

$$\|P_{\cdot i}\|_2 = 1, \quad P_{\cdot i}^T P_{\cdot j} = 0, \qquad j \neq i, \tag{51}$$

where for any $k = 1, \ldots, n$, $P_{\cdot k}$ denotes the k-th column of P. Thus, in terms of (42), (43), (50), and (51), we have

$$\begin{pmatrix} P_{(q+1)i} \\ \vdots \\ P_{ni} \end{pmatrix} = \begin{pmatrix} O(\|\widetilde{H}\|) \\ \vdots \\ O(\|\widetilde{H}\|) \end{pmatrix}, \qquad i = 1, \ldots, q, \tag{52}$$

$$\begin{pmatrix} P_{1i} \\ \vdots \\ P_{qi} \\ P_{(m+1)i} \\ \vdots \\ P_{ni} \end{pmatrix} = \begin{pmatrix} O(\|\widetilde{H}\|) \\ \vdots \\ O(\|\widetilde{H}\|) \\ O(\|\widetilde{H}\|) \\ \vdots \\ O(\|\widetilde{H}\|) \end{pmatrix}, \qquad i = q+1, \cdots, m \tag{53}$$

and

$$\begin{pmatrix} P_{1i} \\ \vdots \\ P_{mi} \end{pmatrix} = \begin{pmatrix} O(\|\widetilde{H}\|) \\ \vdots \\ O(\|\widetilde{H}\|) \end{pmatrix}, \qquad i = m+1, \ldots, n. \tag{54}$$

Then for $i = 1, \dots, m$, (50) and (52)–(54) give

$$\lambda_i \begin{pmatrix} P_{1i} \\ \vdots \\ P_{qi} \\ \vdots \\ P_{mi} \\ \vdots \\ P_{ni} \end{pmatrix} = \begin{pmatrix} (\sigma_1 + \widetilde{H}_{11})P_{1i} + \cdots + \widetilde{H}_{1q}P_{qi} + \cdots + \widetilde{H}_{1m}P_{mi} + O(\|\widetilde{H}\|^2) \\ \vdots \\ \widetilde{H}_{1q}P_{1i} + \cdots + (\sigma_q + \widetilde{H}_{qq})P_{qi} + \cdots + \widetilde{H}_{qm}P_{mi} + O(\|\widetilde{H}\|^2) \\ \vdots \\ \widetilde{H}_{1m}P_{1i} + \cdots + \widetilde{H}_{qm}P_{qi} + \cdots + (\sigma_m + \widetilde{H}_{mm})P_{mi} + O(\|\widetilde{H}\|^2) \\ \vdots \\ O(\|\widetilde{H}\|) \end{pmatrix}. \tag{55}$$

In the following we will prove for $i, j = 1, \dots, n$ the following equation

$$\Delta \widetilde{T}_{ij} = O(\|\widetilde{H}\|^2) \tag{56}$$

holds. We analyze this in eight different cases.

Case 1. $i, j = 1, \dots, q$. Then, by (49) and (52)–(54),

$$\Delta \widetilde{T}_{ij} = 2 \sum_{k=q+1}^{m} P_{ki}P_{kj}\sigma_k = O(\|\widetilde{H}\|^2).$$

Hence, (56) holds for this case.

Case 2. $i = q+1, \dots, m$ and $j = 1, \dots, q$. Then, again, by (49),

$$\Delta \widetilde{T}_{ij} = \frac{\lambda_i + \lambda_j}{|\lambda_i| + |\lambda_j|} \left(\sum_{k=1}^{m} P_{ki}P_{kj}\sigma_k \right) - \sum_{k=1}^{q} P_{ki}P_{kj}\sigma_k + \sum_{k=q+1}^{m} P_{ki}P_{kj}\sigma_k. \tag{57}$$

Now we assume

$$\sigma_1 \geq \cdots \geq \sigma_{j-s-1} > \sigma_{j-s} = \cdots = \sigma_j = \cdots = \sigma_{j+t} > \sigma_{j+t+1} \geq \cdots \geq \sigma_q > 0$$

and

$$0 > \sigma_{q+1} \geq \cdots \geq \sigma_{i-a-1} > \sigma_{i-a} = \cdots = \sigma_i = \cdots = \sigma_{i+b} > \sigma_{i+b+1} \geq \cdots \geq \sigma_m$$

for some nonnegative integers s, t, a, and b.

Then, by (42) and (55),

$$\begin{pmatrix} P_{1j} \\ \vdots \\ P_{(j-s-1)j} \\ P_{(j+t+1)j} \\ \vdots \\ P_{mj} \end{pmatrix} = \begin{pmatrix} O(\|\widetilde{H}\|) \\ \vdots \\ O(\|\widetilde{H}\|) \\ O(\|\widetilde{H}\|) \\ \vdots \\ O(\|\widetilde{H}\|) \end{pmatrix} \tag{58}$$

and

$$\begin{pmatrix} P_{1i} \\ \vdots \\ P_{(i-a-1)i} \\ P_{(i+b+1)i} \\ \vdots \\ P_{mi} \end{pmatrix} = \begin{pmatrix} O(\|\widetilde{H}\|) \\ \vdots \\ O(\|\widetilde{H}\|) \\ O(\|\widetilde{H}\|) \\ \vdots \\ O(\|\widetilde{H}\|) \end{pmatrix}. \tag{59}$$

Then, by (52)–(54) and (57)–(59), we have

(60)
$$\begin{aligned}\Delta\widetilde{T}_{ij} = \frac{\lambda_i+\lambda_j}{|\lambda_i|+|\lambda_j|}\big[&P_{(j-s)i}P_{(j-s)j}\sigma_j+\cdots+P_{ji}P_{jj}\sigma_j+\cdots+P_{(j+t)i}P_{(j+t)j}\sigma_j\\ &+P_{(i-a)i}P_{(i-a)j}\sigma_i+\cdots+P_{ii}P_{ij}\sigma_i+\cdots+P_{(i+b)i}P_{(i+b)j}\sigma_i+O(\|\widetilde{H}\|^2)\big]\\ -\big[&P_{(j-s)i}P_{(j-s)j}\sigma_j+\cdots+P_{ji}P_{jj}\sigma_j+\cdots+P_{(j+t)i}P_{(j+t)j}\sigma_j\\ &-P_{(i-a)i}P_{(i-a)j}\sigma_i-\cdots-P_{ii}P_{ij}\sigma_i-\cdots-P_{(i+b)i}P_{(i+b)j}\sigma_i+O(\|\widetilde{H}\|^2)\big].\end{aligned}$$

Let

$$\eta_1 := P_{(j-s)i}P_{(j-s)j}+\cdots+P_{ji}P_{jj}+\cdots+P_{(j+t)i}P_{(j+t)j} \tag{61}$$

and

$$\eta_2 = P_{(i-a)i}P_{(i-a)j}+\cdots+P_{ii}P_{ij}+\cdots+P_{(i+b)i}P_{(i+b)j}. \tag{62}$$

Then, by (42), (60) becomes

$$\begin{aligned}\Delta\widetilde{T}_{ij} &= \frac{\lambda_i+\lambda_j}{|\lambda_i|+|\lambda_j|}[\sigma_j\eta_1+\sigma_i\eta_2+O(\|\widetilde{H}\|)^2]-[\sigma_j\eta_1-\sigma_i\eta_2+O(\|\widetilde{H}\|)^2]\\ &= \frac{2}{\lambda_j-\lambda_i}\big[\lambda_i\sigma_j\eta_1+\lambda_j\sigma_i\eta_2\big]+O(\|\widetilde{H}\|)^2,\end{aligned} \tag{63}$$

where we used the fact that when $H\to 0$, $\lambda_j\to\sigma_j>0$, and $\lambda_i\to\sigma_i<0$.

According to (55) and the fact that $\sigma_{j-s}=\cdots=\sigma_{j+t}=\sigma_j$ and $\sigma_{i-a}=\cdots=\sigma_{i+b}=\sigma_i$, we have

$$\lambda_j\begin{pmatrix}P_{(j-s)j}\\ \vdots\\ P_{jj}\\ \vdots\\ P_{(j+t)j}\end{pmatrix}=\begin{pmatrix}\sigma_jP_{(j-s)j}+O(\|\widetilde{H}\|)\\ \vdots\\ \sigma_jP_{jj}+O(\|\widetilde{H}\|)\\ \vdots\\ \sigma_jP_{(j+t)j}+O(\|\widetilde{H}\|)\end{pmatrix} \tag{64}$$

and

$$\lambda_i\begin{pmatrix}P_{(i-a)i}\\ \vdots\\ P_{ii}\\ \vdots\\ P_{(i+b)i}\end{pmatrix}=\begin{pmatrix}\sigma_iP_{(i-a)i}+O(\|\widetilde{H}\|)\\ \vdots\\ \sigma_iP_{ii}+O(\|\widetilde{H}\|)\\ \vdots\\ \sigma_iP_{(i+b)i}+O(\|\widetilde{H}\|)\end{pmatrix}. \tag{65}$$

Equations (61), (62), (52)–(54), (58), (59), and (51) imply

$$\begin{aligned}(P_{(j-s)j})^2+\cdots+(P_{(j+t)j})^2 &= 1+O(\|\widetilde{H}\|^2),\\ (P_{(i-a)i})^2+\cdots+(P_{(i+b)i})^2 &= 1+O(\|\widetilde{H}\|^2)\end{aligned} \tag{66}$$

and

$$\eta_1=O(\|\widetilde{H}\|),\quad \eta_2=O(\|\widetilde{H}\|),\quad \eta_1+\eta_2=O(\|\widetilde{H}\|^2). \tag{67}$$

Therefore, by (63), (67), and (42), we obtain

$$\begin{aligned}\Delta\widetilde{T}_{ij} &= \frac{2}{\sigma_j - \sigma_i + O(\|\widetilde{H}\|)}\big[\sigma_i\sigma_j(\eta_1+\eta_2) + O(\|\widetilde{H}\|)\eta_1 + O(\|\widetilde{H}\|)\eta_2\big] \\ &= \frac{2}{\sigma_j - \sigma_i + O(\|\widetilde{H}\|)}\big[\sigma_i\sigma_j O(\|\widetilde{H}\|^2) + O(\|\widetilde{H}\|)\eta_1 + O(\|\widetilde{H}\|)\eta_2\big] \\ &= O(\|\widetilde{H}\|^2),\end{aligned}$$

which proves (56).

Case 3. $i = m+1, \ldots, n$ and $j = 1, \ldots, q$. Then, by (49),

$$\Delta\widetilde{T}_{ij} = \frac{\lambda_i - |\lambda_i|}{|\lambda_i| + \lambda_j}\left(\sum_{k=1}^{q} P_{ki}P_{kj}\sigma_k\right) + \frac{2\lambda_j + \lambda_i + |\lambda_i|}{|\lambda_i| + \lambda_j}\left(\sum_{k=q+1}^{m} P_{ki}P_{kj}\sigma_k\right),$$

which, together with (52)–(54), implies

$$\Delta\widetilde{T}_{ij} = \frac{\lambda_i - |\lambda_i|}{|\lambda_i| + \lambda_j}\left(\sum_{k=1}^{q} P_{ki}P_{kj}\sigma_k\right) + O(\|\widetilde{H}\|^2). \tag{68}$$

By Equations (54) and (43), we have

$$\lambda_i\left(\sum_{k=1}^{q} P_{ki}P_{kj}\sigma_k\right) = O(\|\widetilde{H}\|^2),$$

which, together with (68), proves (56).

Case 4. $i = 1, \ldots, q$ and $j = q+1, \ldots, m$. By the symmetry of $\Delta\widetilde{T}$ and Case 2, (56) holds.

Case 5. $i, j = q+1, \ldots, m$. Similar to the discussion in Case 1, we can prove (56). We omit the detail here.

Case 6. $i = m+1, \ldots, n$ and $j = q+1, \ldots, m$. Similar to the discussion in Case 3, we can prove (56). Again, we omit the detail here.

Case 7. $i = 1, \ldots, m$ and $j = m+1, \ldots, n$. Due to the symmetric property of $\Delta\widetilde{T}$ we can prove (56) according to Cases 3 and 6.

Case 8. $i, j = m+1, \ldots, n$. Then, by (53), (54), and (49), we have

$$\Delta\widetilde{T}_{ij} = \frac{\lambda_i + \lambda_j}{|\lambda_i| + |\lambda_j|}[O(\|\widetilde{H}\|^2)] - [O(\|\widetilde{H}\|^2)] = O(\|\widetilde{H}\|^2),$$

which proves (56).

Overall, we have proved that (56) holds for $i, j = 1, \ldots, n$. Since $\widetilde{H} = QHQ^T$, we have in fact proved

$$\|\Delta\widetilde{T}\| = O(\|\widetilde{H}\|^2), \tag{69}$$

which, together with (45), proves (35). $\square$

THEOREM 4.13. *Let $F : \mathcal{S}^n \to \mathcal{S}^n$ be defined by $F(Y) = |Y|,\ Y \in \mathcal{S}^n$. Then F is strongly semismooth at any $X \in \mathcal{S}^n$.*

PROOF. By Theorem 4.7, F is directionally differentiable at X. On the other hand, in terms of Lemmas 4.10 and 4.12,

$$F(X+H)-F(X)-F'_{X+H}(H)=O(\|H\|^2)\ \forall\ X+H\in D_F,\ H\to 0.$$

Thus, according to Theorem 3.7, F is semismooth at $X\in\mathcal{S}^n$.

By Theorem 3.7, Lemmas 4.10 and 4.12, and the semismoothness of F, F is strongly semismooth at any $X\in\mathcal{S}^n$. $\square$

COROLLARY 4.14. *Let $F:\mathcal{S}(n_1,\ldots,n_m)\to\mathcal{S}(n_1,\ldots,n_m)$ be defined by $F(Y)=|Y|$, $Y\in\mathcal{S}(n_1,\ldots,n_m)$. Then F is strongly semismooth at any $X\in\mathcal{S}(n_1,\ldots,n_m)$.*

COROLLARY 4.15. *The function of semidefinite projection is strongly semismooth on $\mathcal{S}(n_1,\ldots,n_m)$.*

COROLLARY 4.16. *The function of projective residual defined on $\mathcal{S}(n_1,\ldots,n_m)$ is strongly semismooth at $X\in\mathcal{S}(n_1,\ldots,n_m)$ if both F and G are strongly semismooth at $X\in\mathcal{S}(n_1,\ldots,n_m)$.*

PROOF. These are direct implications of Theorems 3.10 and 4.13. $\square$

5. Final remarks. We have discussed some important differential properties of matrix-valued functions. In particular, we have proved that the semidefinite-projection function, the absolute-value function, and the projective residual function are strongly semismooth on $\mathcal{S}(n_1,\ldots,n_m)$. To use these properties to design semismooth Newton methods and smoothing Newton methods for matrix-valued equations and the SDCP is a logical step in future research.

Acknowledgments. This research was partially supported by the Australian Research Council, the Singapore-MIT Alliance, and Grants R314000-026112 and R314000-009112 from the National University of Singapore.

References

Bhatia, R. 1997. *Matrix Analysis*. Springer-Verlag, New York.

Burke, J., S. Xu. 1998. The global linear convergence of a non-interior path-following algorithm for linear complementarity problem. *Math. Oper. Res.* **23** 719–734.

Chen, X., P. Tseng. 1999. Non-interior continuation methods for solving semidefinite complementarity problems. Tech. Report, Dept. of Mathematics, University of Washington, Seattle, WA.

Clarke, F. 1983. *Optimization and Nonsmooth Analysis*. John Wiley and Sons, New York.

Facchinei, F., C. Kanzow. 1997. A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems. *Math. Programming* **76** 493–512.

Fischer, A. 1997. Solution of monotone complementarity problems with locally Lipschitzian functions. *Math. Programming* **76** 513–532.

Mifflin, R. 1977. Semismooth and semiconvex functions in constrained optimization. *SIAM J. Control Optim.* **15** 957–972.

Pang, J. S., L. Qi. 1993. Nonsmooth equations: Motivation and algorithms. *SIAM J. Optim.* **3** 443–465.

Qi, L., D. Sun. 2001. Nonsmooth and smoothing methods for NCP and VI. C. Floudas, P. Pardalos, eds. *Encyclopedia of Optimization*. Kluwer Academic Publishers, Norwell, MA, 100–104.

———, J. Sun. 1993. A nonsmooth version of Newton's method. *Math. Programming* **58** 353–367.

Rockafellar, R. T. 1970. *Convex Analysis*. Princeton University Press, Princeton, NJ.

Shapiro, A. 1990. On concepts of directional differentiability. *J. Optim. Theory Appl.* **66** 477–487.

Tseng, P. 1998. Merit functions for semidefinite complementarity problems. *Math. Programming* **83** 159–185.
Zarantonello, E. H. 1971. Projections on convex sets in Hilbert space and spectral theory. E. H. Zarantonello, ed. *Contributions to Nonlinear Functional Analysis*. Academic Press, New York, 237–424.

D. Sun: Department of Mathematics, National University of Singapore, Republic of Singapore; e-mail: matsundf@nus.edu.sg

J. Sun: Faculty of Business Administration and Singapore-MIT Alliance, National University of Singapore, Republic of Singapore; e-mail: jsun@nus.edu.sg