SIAM J. OPTIM.
Vol. 13, No. 4, pp. 960–985

ANALYSIS OF NONSMOOTH SYMMETRIC-MATRIX-VALUED FUNCTIONS WITH APPLICATIONS TO SEMIDEFINITE COMPLEMENTARITY PROBLEMS*

XIN CHEN†, HOUDUO QI‡, AND PAUL TSENG§

Abstract. For any function f from $\mathbb{R}$ to $\mathbb{R}$, one can define a corresponding function on the space of $n \times n$ (block-diagonal) real symmetric matrices by applying f to the eigenvalues of the spectral decomposition. We show that this matrix-valued function inherits from f the properties of continuity, (local) Lipschitz continuity, directional differentiability, Fréchet differentiability, continuous differentiability, as well as (ρ-order) semismoothness. Our analysis uses results from nonsmooth analysis as well as perturbation theory for the spectral decomposition of symmetric matrices. We also apply our results to the semidefinite complementarity problem, addressing some basic issues in the analysis of smoothing/semismooth Newton methods for solving this problem.

Key words. symmetric-matrix-valued function, nonsmooth analysis, semismooth function, semidefinite complementarity problem

AMS subject classifications. 49M45, 90C25, 90C33

PII. S1052623400380584

1. Introduction.

Let $\mathcal{X}$ denote the space of $n \times n$ block-diagonal real matrices with m blocks of size $n_1, \ldots, n_m$, respectively (the blocks are fixed). Thus, $\mathcal{X}$ is closed under matrix addition $x + y$, multiplication xy, transposition x^T, and inversion x^{-1}, where $x, y \in \mathcal{X}$. We endow $\mathcal{X}$ with the inner product and norm

$$\langle x, y \rangle := \mathrm{tr}[x^T y], \qquad \|x\| := \sqrt{\langle x, x \rangle},$$

where $x, y \in \mathcal{X}$ and $\mathrm{tr}[\cdot]$ denotes the matrix trace, i.e., $\mathrm{tr}[x] = \sum_{i=1}^n x_{ii}$. [$\|x\|$ is the Frobenius norm of x and " := " means "define"]. Let $\mathcal{O}$ denote the set of $p \in \mathcal{X}$ that are orthogonal, i.e., $p^T = p^{-1}$. Let $\mathcal{S}$ denote the subspace comprising those $x \in \mathcal{X}$ that are symmetric, i.e., $x^T = x$. This is a subspace of $\mathbb{R}^{n \times n}$ of dimension $n_1(n_1 + 1)/2 + \cdots + n_m(n_m + 1)/2$.

For any $x \in \mathcal{S}$, its (repeated) eigenvalues $\lambda_1, \ldots, \lambda_n$ are real and it admits a spectral decomposition of the form

$$x = p \ \mathrm{diag}[\lambda_1, \ldots, \lambda_n] p^T \tag{1}$$

for some $p \in \mathcal{O}$, where $\mathrm{diag}[\lambda_1, \ldots, \lambda_n]$ denotes the $n \times n$ diagonal matrix with its ith diagonal entry λ_i. Then, for any function $f : \mathbb{R} \to \mathbb{R}$, we can define a corresponding function $f^{\square} : \mathcal{S} \to \mathcal{S}$ [1], [13] by

$$f^{\square}(x) := p \ \mathrm{diag}[f(\lambda_1), \ldots, f(\lambda_n)] p^T. \tag{2}$$

*Received by the editors November 7, 2000; accepted for publication (in revised form) July 12, 2002; published electronically March 5, 2003.
http://www.siam.org/journals/siopt/13-4/38058.html

†Operations Research Center, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Building E40-194, Cambridge, MA 02139 (xchen@mit.edu).

‡School of Mathematics, The University of New South Wales, Sydney, New South Wales 2052, Australia (hdqi@maths.unsw.edu.au). This author was supported by the Australian Research Council.

§Department of Mathematics, University of Washington, Seattle, WA 98195 (tseng@math.washington.edu). This author was supported by the National Science Foundation, grant CCR-9731273.

It is known that $f^{\square}(x)$ is well defined (independent of the ordering of $\lambda_1, \ldots, \lambda_n$ and the choice of p) and belongs to $\mathcal{S}$; see [1, Chap. V] and [13, sec. 6.2]. Moreover, a result of Daleckii and Krein showed that if f is continuously differentiable, then $f^{\square}$ is differentiable (in the Fréchet sense) and its Jacobian $\nabla f^{\square}(x)$ has a simple formula—see [1, Thm. V.3.3]; also see Proposition 4.3. In fact, in this case $f^{\square}$ is continuously differentiable—see [8, Lem. 4]; also see Proposition 4.4. Much of the studies on $f^{\square}$ has focused on conditions for it to be *operator monotone* or *operator convex*—see [1], [13], and the references cited in [1, pp. 150–151] for discussions. We note that [8] swaps p and p^T in (1)–(2), but this is only a difference in notation.

The above results show that $f^{\square}$ inherits smoothness properties from f. In this paper, we make an analogous study for properties associated with nonsmooth functions. In particular, we show that the properties of continuity, strict continuity, Lipschitz continuity, directional differentiability, differentiability, continuous differentiability, and (ρ-order) semismoothness are each inherited by $f^{\square}$ from f (see Propositions 4.1, 4.2, 4.3, 4.4, 4.6, 4.8, and 4.10). Our ρ-order semismoothness result generalizes a recent result of Sun and Sun [29] which considers the case of the absolute-value function $f(\xi) = |\xi|$ and shows that $f^{\square}(x) = (x^2)^{1/2}$ is strongly semismooth. In the case where $f = g'$ for some function g, our differentiability and continuous differentiability results can also be inferred from a recent work of Lewis and Sendov [19] on twice differentiability of spectral functions. Our proofs use a combination of results from matrix analysis and nonsmooth analysis—in particular, perturbation results for spectral decomposition [17, 28] and properties of the generalized gradient ∂f (in the Clarke sense) [9, 26], as well as a lemma from [29]. The property of semismoothness, as introduced by Mifflin [20] for functionals and scalar-valued functions and further extended by Qi and Sun [23] for vector-valued functions, is of particular interest due to the key role it plays in the superlinear convergence analysis of certain generalized Newton methods [14, 21, 23]. In section 5, we formulate the semidefinite complementarity problem (SDCP) as a nonsmooth equation

$$H(x, y) = 0,$$

where $H : \mathcal{S} \times \mathcal{S} \to \mathcal{S} \times \mathcal{S}$ is a certain semismooth function. This facilitates the development of nonsmooth Newton methods for solving the SDCP—a contrast to existing smoothing or differentiable merit function approaches [8, 27, 30, 32]. We show that H, together with the Chen–Mangasarian class of smoothing functions studied in [8], satisfies the Jacobian Consistence Property introduced in [6]. This paves a way for extending some smoothing methods for nonlinear complementarity problems (NCPs), such as those studied by Chen, Qi, and Sun [6] and later by Kanzow and Pieper [16], to the SDCP. Final remarks are given in section 6.

Our notations are, for the most part, consistent with those used in [8, 30]. If $F : \mathcal{S} \to \mathcal{S}$ is differentiable (in the Fréchet sense) at $x \in \mathcal{S}$, we denote by $\nabla F(x)$ the Jacobian of F at $x \in \mathcal{S}$, viewed as a linear mapping from $\mathcal{S}$ to $\mathcal{S}$. Throughout, $\|\cdot\|$ denotes the Frobenius norm for matrices and the 2-norm for vectors. For any linear mapping $M : \mathcal{S} \to \mathcal{S}$, we denote its operator norm $|||M||| := \max_{\|x\|=1} \|Mx\|$. For any $x \in \mathcal{S}$, we denote by x_{ij} the (i, j)th entry of x. We use $\circ$ to denote the Hardamard product, i.e.,

$$x \circ y = [x_{ij} y_{ij}]_{i,j=1}^n.$$

For any $x \in \mathcal{S}$ and scalar $\gamma > 0$, we denote the γ-ball around x by $\mathcal{B}(x, \gamma) := \{y \in \mathcal{S} \mid \|y - x\| \leq \gamma\}$. We write $z = O(\alpha)$ (respectively, $z = o(\alpha)$), with $\alpha \in \mathbb{R}$ and $z \in \mathcal{S}$, to mean $\|z\|/|\alpha|$ is uniformly bounded (respectively, tends to zero) as $\alpha \to 0$.

2. Basic properties. In this section, we review some basic properties of vector-valued functions. These properties are continuity, (local) Lipschitz continuity, directional differentiability, continuous differentiability, as well as (ρ-order) semismoothness. We note that $\mathcal{S}$ is a vector space of dimension $n_1(n_1+1)/2+\cdots+n_m(n_m+1)/2$, so these properties apply to the symmetric-matrix-valued function $f^{\square}$ defined by (1)–(2). In what follows, we consider a function/mapping $F : \mathbb{R}^k \to \mathbb{R}^\ell$.

We say F is continuous at $x \in \mathbb{R}^k$ if

$$F(y) \to F(x) \quad \text{as} \quad y \to x;$$

and F is continuous if F is continuous at every $x \in \mathbb{R}^k$. F is strictly continuous (also called "locally Lipschitz continuous") at $x \in \mathbb{R}^k$ [26, Chap. 9] if there exist scalars $\kappa > 0$ and $\delta > 0$ such that

$$\|F(y) - F(z)\| \le \kappa\|y - z\| \quad \forall y, z \in \mathbb{R}^k \text{ with } \|y - x\| \le \delta,\ \|z - x\| \le \delta;$$

and F is strictly continuous if F is strictly continuous at every $x \in \mathbb{R}^k$. If δ can be taken to be ∞, then F is Lipschitz continuous with Lipschitz constant κ. Define the function $\mathrm{lip}F : \mathbb{R}^k \to [0, \infty]$ by

$$\mathrm{lip}F(x) := \limsup_{\substack{y,z\to x\\ y\neq z}} \frac{\|F(y) - F(z)\|}{\|y - z\|}.$$

Then F is strictly continuous at x if and only if $\mathrm{lip}F(x)$ is finite.

We say F is directionally differentiable at $x \in \mathbb{R}^k$ if

$$F'(x; h) \ := \ \lim_{t\to 0^+} \frac{F(x + th) - F(x)}{t} \quad \text{exists} \qquad \forall h \in \mathbb{R}^k;$$

and F is directionally differentiable if F is directionally differentiable at every $x \in \mathbb{R}^k$. F is differentiable (in the Fréchet sense) at $x \in \mathbb{R}^k$ if there exists a linear mapping $\nabla F(x) : \mathbb{R}^k \to \mathbb{R}^\ell$ such that

$$F(x + h) - F(x) - \nabla F(x)h \ = \ o(\|h\|).$$

We say that F is continuously differentiable if F is differentiable at every $x \in \mathbb{R}^k$ and ∇F is continuous.

If F is strictly continuous, then F is almost everywhere differentiable by Rademacher's theorem—see [9] and [26, sec. 9J]. Then the generalized Jacobian $\partial F(x)$ of F at x (in the Clarke sense) can be defined as the convex hull of the generalized Jacobian $\partial_B F(x)$ (in the Bouligand sense), where

$$\partial_B F(x) := \left\{ \lim_{x^j\to x} \nabla F(x^j) \big| F \text{ is differentiable at } x^j \in \mathbb{R}^k \right\}.$$

In [26, Chap. 9], the case of $\ell = 1$ is considered and the notations "$\bar{\nabla}$" and "$\bar{\partial}$" are used instead of, respectively, "∂_B" and "∂."

Assume $F : \mathbb{R}^k \to \mathbb{R}^\ell$ is strictly continuous. We say F is semismooth at x if F is directionally differentiable at x and, for any $V \in \partial F(x + h)$, we have

$$F(x + h) - F(x) - Vh = o(\|h\|).$$

We say F is ρ-order semismooth at x ($0 < \rho < \infty$) if F is semismooth at x and, for any $V \in \partial F(x+h)$, we have

$$F(x+h) - F(x) - Vh = O(\|h\|^{1+\rho}).$$

We say F is semismooth (respectively, ρ-order semismooth) if F is semismooth (respectively, ρ-order semismooth) at every $x \in \mathbb{R}^k$. We say F is strongly semismooth if it is 1-order semismooth. Convex functions and piecewise continuously differentiable functions are examples of semismooth functions. The composition of two (respectively, ρ-order) semismooth functions is also a (respectively, ρ-order) semismooth function. The property of semismoothness plays an important role in nonsmooth Newton methods [23] as well as in some smoothing methods mentioned in the previous section. For extensive discussions of semismooth functions, see [10, 20, 23].

3. Perturbation results for symmetric matrices. In this section, we review some useful perturbation results for the spectral decomposition of real symmetric matrices. These results will be used in the next section to analyze properties of the symmetric-matrix-valued function $f^{\square}$ given by (1)–(2). The main sources of reference for the results are Chapter 2 of the book by Kato [17] and the book by Stewart and Sun [28].

Let $\mathcal{D}$ denote the space of $n \times n$ real diagonal matrices with nonincreasing diagonal entries. For each $x \in \mathcal{S}$, define the two sets of orthonormal eigenvectors of x by

$$\mathcal{O}_x := \{p \in \mathcal{O} |\ p^T x p \in \mathcal{D}\}, \qquad \tilde{\mathcal{O}}_x := \{p \in \mathcal{O} |\ p^T x p \text{ is diagonal }\}.$$

Clearly, $\mathcal{O}_x$ and $\tilde{\mathcal{O}}_x$ are nonempty for each $x \in \mathcal{S}$. The following key lemma, proved in [8, Lem. 3] using results from [28, pp. 92 and 250], shows that $\mathcal{O}_x$ is locally upper Lipschitzian with respect to x.

LEMMA 3.1. *For any $x \in \mathcal{S}$, there exist scalars $\eta > 0$ and $\epsilon > 0$ such that*

$$\min_{p \in \mathcal{O}_x} \|p - q\| \le \eta \|x - y\| \quad \forall\ y \in \mathcal{B}(x, \epsilon),\ \forall q \in \mathcal{O}_y. \tag{3}$$

We will also need the following perturbation result of Weyl for eigenvalues of symmetric matrices—see [1, p. 63] and [12, p. 367].

LEMMA 3.2. *Let $\lambda_1 \ge \cdots \ge \lambda_n$ be the eigenvalues of any $x \in \mathcal{S}$ and $\mu_1 \ge \cdots \ge \mu_n$ be the eigenvalues of any $y \in \mathcal{S}$. Then*

$$|\lambda_i - \mu_i| \le \|x - y\| \quad \forall\ i = 1, \ldots, n.$$

Lastly, for our differential analysis, we need the following classical result [25, Thm. 1] showing that, for any $x \in \mathcal{S}$ and any $h \in \mathcal{S}$, the orthonormal eigenvectors of $x + th$ may be chosen to be analytic in t. As is remarked in [17, p. 122], the existence of such orthonormal eigenvectors depending smoothly on t is one of the most remarkable results in the analytic perturbation theory for symmetric operators.

LEMMA 3.3. *For any $x \in \mathcal{S}$ and any $h \in \mathcal{S}$, there exist $p(t) \in \tilde{\mathcal{O}}_{x+th}$, $t \in \mathbb{R}$, whose entries are power series in t, convergent in a neighborhood of $t = 0$.*

4. Continuity and differential properties of symmetric-matrix functions. In this section, we use the results from section 3 to show that if $f : \mathbb{R} \to \mathbb{R}$ has the property of continuity (respectively, strict continuity, Lipschitz continuity, directional differentiability, semismoothness, ρ-order semismoothness), then so does the symmetric-matrix-valued function $f^{\square}$ defined by (1)–(2). We begin with the continuity result below.

PROPOSITION 4.1. *For any $f : \mathbb{R} \to \mathbb{R}$, the following results hold:*

(a) *$f^{\square}$ is continuous at an $x \in \mathcal{S}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$ if and only if f is continuous at $\lambda_1, \ldots, \lambda_n$.*

(b) *$f^{\square}$ is continuous if and only if f is continuous.*

Proof. (a) Fix any $x \in \mathcal{S}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$. Assume without loss of generality that $\lambda_1 \geq \cdots \geq \lambda_n$.

Suppose f is continuous at $\lambda_1, \ldots, \lambda_n$. By Lemma 3.1, there exist $\eta > 0$ and $\epsilon > 0$ such that (3) holds. Then, for any $y \in \mathcal{B}(x, \epsilon)$ and any $q \in \mathcal{O}_y$, there exists $p \in \mathcal{O}_x$ satisfying

$$\|p - q\| \leq \eta \|x - y\|.$$

Moreover,

$$q^T y q = \text{diag}[\mu_1, \ldots, \mu_n], \qquad p^T x p = \text{diag}[\lambda_1, \ldots, \lambda_n],$$

where $\mu_1 \geq \cdots \geq \mu_n$ and $\lambda_1 \geq \cdots \geq \lambda_n$ are the eigenvalues of y and x, respectively. Since f is continuous and, by Lemma 3.2, $|\lambda_i - \mu_i| \leq \|x - y\|$ for all i, we have $f(\mu_i) \to f(\lambda_i)$ and $\|p - q\| \to 0$ as $y \to x$. Then (2) yields

$$\begin{aligned} f^{\square}(x) - f^{\square}(y) &= p \ \text{diag}[f(\lambda_1), \ldots, f(\lambda_n)]p^T - q \ \text{diag}[f(\mu_1), \ldots, f(\mu_n)]q^T \\ &= p \ \text{diag}[f(\lambda_1) - f(\mu_1), \ldots, f(\lambda_n) - f(\mu_n)]p^T \\ &\quad + (p - q)\text{diag}[f(\mu_1), \ldots, f(\mu_n)]p^T + q \ \text{diag}[f(\mu_1), \ldots, f(\mu_n)](p - q)^T \\ &\to 0 \qquad \text{as} \quad y \to x. \end{aligned}$$

Thus $f^{\square}$ is continuous at x.

Suppose instead $f^{\square}$ is continuous at x. Fix any $p \in \mathcal{O}_x$. Then for each $i \in \{1, \ldots, n\}$, $p \ \text{diag}[\lambda_1, \ldots, \mu_i, \ldots, \lambda_n]p^T \to x$ as $\mu_i \to \lambda_i$ so that $f^{\square}(p \ \text{diag}[\lambda_1, \ldots, \mu_i, \ldots, \lambda_n]p^T) \to f^{\square}(x)$ or, equivalently, $f(\mu_i) \to f(\lambda_i)$. Thus f is continuous at λ_i for $i = 1, \ldots, n$.

(b) is an immediate consequence of (a). □

For any $\lambda = (\lambda_1, \ldots, \lambda_n)^T \in \mathbb{R}^n$, any $h \in \mathcal{S}$, and any function $f : \mathbb{R} \to \mathbb{R}$ that is directionally differentiable at $\lambda_1, \ldots, \lambda_n$, we denote by $f^{[1]}(\lambda; h)$ the $n \times n$ symmetric matrix whose (i, j)th entry is

$$f^{[1]}(\lambda; h)_{ij} := \begin{cases} \dfrac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j} h_{ij} & \text{if } \lambda_i \neq \lambda_j, \\ f'(\lambda_i; h_{ij}) & \text{if } \lambda_i = \lambda_j. \end{cases} \tag{4}$$

By using Lemma 3.3, we have the directional differentiability result below.

PROPOSITION 4.2. *For any $f : \mathbb{R} \to \mathbb{R}$, the following results hold:*

(a) *$f^{\square}$ is directionally differentiable at an $x \in \mathcal{S}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$ if and only if f is directionally differentiable at $\lambda_1, \ldots, \lambda_n$. Moreover, for any nonzero $h \in \mathcal{S}$,*

$$(f^{\square})'(x; h) = p \ f^{[1]}(\lambda; p^T h p) \ p^T \tag{5}$$

for some $p \in \mathcal{O}$ such that $(p^T h p)_{ij} = 0$ whenever $\lambda_i = \lambda_j$ and $i \neq j$.

(b) *$f^{\square}$ is directionally differentiable if and only if f is directionally differentiable.*

Proof. (a) Fix any $x \in \mathcal{S}$. By Lemma 3.3, for any nonzero $h \in \mathcal{S}$ there exist $p(t) \in \tilde{\mathcal{O}}_{x(t)}$, $t \in \mathbb{R}$, whose entries are power series in t, convergent in a neighborhood $\mathcal{I}$ of $t = 0$, where $x(t) := x + th$. Then the corresponding eigenvalues

$$\lambda_i(t) := [p(t)^T x(t) p(t)]_{ii}, \quad i = 1, \ldots, n,$$

are also power series in t, convergent for $t \in \mathcal{I}$, and satisfy

$$x(t) = p(t)\mathrm{diag}[\lambda_1(t), \ldots, \lambda_n(t)]p(t)^T. \tag{6}$$

Multiplying both sides of (6) by $p(t)^T$ from the left and then differentiating both sides with respect to t using the product rule, we obtain

$$p'(t)^T x(t) + p(t)^T x'(t) = \Lambda'(t)p(t)^T + \Lambda(t)p'(t)^T,$$

where $\Lambda(t) := \mathrm{diag}[\lambda_1(t), \ldots, \lambda_n(t)]$ and $\Lambda'(t) := \mathrm{diag}[\lambda_1'(t), \ldots, \lambda_n'(t)]$. Multiplying both sides on the right by $p(t)$ and using $x'(t) = h$, we arrive at

$$\Lambda'(t) - \hat{h}(t) = \hat{p}(t)\Lambda(t) - \Lambda(t)\hat{p}(t),$$

where $\hat{h}(t) := p(t)^T hp(t)$ and $\hat{p}(t) := p'(t)^T p(t)$. This implies

$$\hat{h}(t)_{ii} = \lambda_i'(t), \qquad i = 1, \ldots, n, \tag{7}$$

$$\hat{h}(t)_{ij} = \hat{p}(t)_{ij}(\lambda_i(t) - \lambda_j(t)) \quad \forall i \neq j. \tag{8}$$

For simplicity, let

$$\begin{array}{llll} p := p(0), & p' := p'(0), & \hat{p} := \hat{p}(0), & \\ \lambda_i := \lambda_i(0), & \lambda_i' := \lambda_i'(0), & & i = 1, \ldots, n. \end{array}$$

Assume f is directionally differentiable at $\lambda_1, \ldots, \lambda_n$. Then we have from $\lambda_i(t) = \lambda_i + t\lambda_i' + o(t)$ and the positive homogeneity property of $f'(\lambda_i; \cdot)$ the expansions

$$p(t) = p + tp' + o(t) \quad \text{and} \quad f(\lambda_i(t)) = f(\lambda_i) + tf'(\lambda_i; \lambda_i') + o(t), \quad i = 1, \ldots, n.$$

Also, $p(\cdot)$ and $p'(\cdot)$ are continuous at $t = 0$ so that $\lim_{t \to 0} \hat{h}(t) = p^T hp$ and $\lim_{t \to 0} \hat{p}(t) = \hat{p}$. Using (2) and the above expansions, we then obtain

$$\begin{aligned} f^{\square}(x + th) &= p(t)\mathrm{diag}[f(\lambda_1(t)), \ldots, f(\lambda_n(t))]p(t)^T \\ &= p\ \mathrm{diag}[f(\lambda_1), \ldots, f(\lambda_n)]p^T + t\left(p\ \mathrm{diag}[f'(\lambda_1; \lambda_1'), \ldots, f'(\lambda_n; \lambda_n')]p^T\right) \\ &\quad + t\left(p'\mathrm{diag}[f(\lambda_1), \ldots, f(\lambda_n)]p^T + p\ \mathrm{diag}[f(\lambda_1), \ldots, f(\lambda_n)](p')^T\right) + o(t) \\ &= f^{\square}(x) + tp\ \mathrm{diag}[f'(\lambda_1; \lambda_1'), \ldots, f'(\lambda_n; \lambda_n')]p^T \\ &\quad + tp\left(\hat{p}^T\mathrm{diag}[f(\lambda_1), \ldots, f(\lambda_n)] + \mathrm{diag}[f(\lambda_1), \ldots, f(\lambda_n)]\hat{p}\right)p^T + o(t) \\ &= f^{\square}(x) + tp\ \mathrm{diag}[f'(\lambda_1; \lambda_1'), \ldots, f'(\lambda_n; \lambda_n')]p^T \\ &\quad + tp\,[(f(\lambda_i) - f(\lambda_j))\hat{p}_{ij}]_{i,j=1}^n\, p^T + o(t) \\ &= f^{\square}(x) + tp\ f^{[1]}(\lambda; p^T hp)\ p^T + o(t), \end{aligned} \tag{9}$$

where the fourth equality follows from $p(t)^T p(t) = I$ so that $p'(t)^T p(t) + p(t)^T p'(t) = 0$, implying $\hat{p}^T = -\hat{p}$; the last equality follows from (7) so that $\lambda_i' = \hat{h}(0)_{ii} = (p^T hp)_{ii}$ for $i = 1, \ldots, n$, and from (8) so that $\hat{p}_{ij} = (p^T hp)_{ij}/(\lambda_i - \lambda_j)$ whenever $\lambda_i \neq \lambda_j$ and $(p^T hp)_{ij} = 0$ whenever $\lambda_i = \lambda_j$ and $i \neq j$. It follows from (9) that

$$(f^{\square})'(x; h) = \lim_{t \to 0^+} \frac{f^{\square}(x + th) - f^{\square}(x)}{t} \;=\; p\ f^{[1]}(\lambda; p^T hp)\ p^T.$$

This proves (5).

Suppose instead $f^{\square}$ is directionally differentiable at x with eigenvalues $\lambda_1, \ldots, \lambda_n$. Fix any $p \in \mathcal{O}$ satisfying $x = p \text{ diag}[\lambda_1, \ldots, \lambda_n]p^T$. For each $i \in \{1, \ldots, n\}$ and each $d_i \in \mathbb{R}$, let $h := p \text{ diag}[0, \ldots, d_i, \ldots, 0]p^T$. Then, it is readily verified that $\text{diag}[0, \ldots, f'(\lambda_i; d_i), \ldots, 0] = p^T(f^{\square})'(x; h)p$, so $f'(\lambda_i; d_i)$ is well defined.

(b) is an immediate consequence of (a). ∎

We note that p in the formula for $(f^{\square})'(x; h)$ depends on h as well as x. In fact, the proof of Proposition 4.2 shows that a necessary condition for $p(t)$ to comprise orthonormal eigenvectors of $x + th$ that are differentiable at $t = 0$ is that $(p^T hp)_{ij} = 0$ whenever $\lambda_i = \lambda_j$ and $i \neq j$, where $p := p(0)$. In the case of $f(\cdot) = |\cdot|$, directional differentiability of $f^{\square}$ has been shown by Sun and Sun [29, Lem. 4.8]. In addition, they derived a formula for the directional derivative $(f^{\square})'(x; h)$ that also involves $p \in \mathcal{O}_x$ but with p *independent* of h.

For any $\lambda = (\lambda_1, \ldots, \lambda_n)^T \in \mathbb{R}^n$ and any function $f : \mathbb{R} \to \mathbb{R}$ that is differentiable at $\lambda_1, \ldots, \lambda_n$, we denote by $f^{[1]}(\lambda)$ the $n \times n$ symmetric matrix whose (i, j)th entry is

$$f^{[1]}(\lambda)_{ij} = \begin{cases} \dfrac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j} & \text{if } \lambda_i \neq \lambda_j, \\ f'(\lambda_i) & \text{if } \lambda_i = \lambda_j. \end{cases}$$

$f^{[1]}(\lambda)$ is called the first divided difference of f at λ [1, p. 123]. The next proposition, based on Lemmas 3.1, 3.2, and the proof idea for Proposition 4.10, characterizes when $f^{\square}$ is differentiable (in the Fréchet sense) at an $x \in \mathcal{S}$. This characterization will be needed for computing the generalized Jacobian of a strictly continuous $f^{\square}$ and for analyzing semismooth property of $f^{\square}$. We note that the proof idea of Proposition 4.2 cannot be used here because the $p(t)$ constructed in that proof depends on h. In particular, it is not known if $\|p''(t)\|$ is uniformly bounded in $\|h\|$.

PROPOSITION 4.3. *For any $f : \mathbb{R} \to \mathbb{R}$, the following results hold:*

(a) *$f^{\square}$ is differentiable at an $x \in \mathcal{S}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$ if and only if f is differentiable at $\lambda_1, \ldots, \lambda_n$. Moreover, $\nabla f^{\square}(x)$ is given by*

$$\nabla f^{\square}(x)h = p(f^{[1]}(\lambda) \circ (p^T hp))p^T \qquad \forall h \in \mathcal{S} \tag{10}$$

for any $p \in \mathcal{O}$ satisfying $x = p \text{ diag}[\lambda_1, \ldots, \lambda_n]p^T$, where $\lambda = (\lambda_1, \ldots, \lambda_n)^T$.

(b) *$f^{\square}$ is differentiable if and only if f is differentiable.*

Proof. (a) Fix any $x \in \mathcal{S}$ and let $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of x.

It is known [1] that the right-hand side of (10) is independent of the choice of $p \in \mathcal{O}$ satisfying $p^T xp = \text{diag}[\lambda_1, \ldots, \lambda_n]$. This can be seen by noting that any two such p are related by a right multiplication by a block diagonal $o \in \mathcal{O}$ whose diagonal blocks correspond to the distinct eigenvalues of x, while the entries of $f^{[1]}(\lambda)$ in each of these diagonal blocks, as well as in each of the off-diagonal blocks, are equal.

Suppose $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $\lambda_1, \ldots, \lambda_n$. We can without loss of generality assume that $\lambda_1 \geq \cdots \geq \lambda_n$. By Lemma 3.1, there exist scalars $\eta > 0$ and $\epsilon > 0$ such that (3) holds. We will show that, for any $h \in \mathcal{S}$ with $\|h\| \leq \epsilon$, there exists $p \in \mathcal{O}_x$ such that

$$f^{\square}(x + h) - f^{\square}(x) - p(c \circ (p^T hp))p^T = o(\|h\|), \tag{11}$$

where $c := f^{[1]}(\lambda)$ and $o(\cdot)$, $O(\cdot)$ depend on f and x only. This together with the independence of the third term on p would show that $f^{\square}$ is differentiable at x and $\nabla f^{\square}(x)$ is given by (10) for any $p \in \mathcal{O}$ satisfying $p^T xp = \text{diag}[\lambda_1, \ldots, \lambda_n]$. Let

$\mu_1 \geq \cdots \geq \mu_n$ denote the eigenvalues of $x+h$, and choose any $q \in \mathcal{O}_{x+h}$. Then, there exists $p \in \mathcal{O}_x$ satisfying

$$\|p-q\| \leq \eta\|h\|.$$

For simplicity, let r denote the left-hand side of (11), i.e.,

$$r := f^{\square}(x+h) - f^{\square}(x) - p(c \circ (p^T hp))p^T,$$

and denote $\tilde{r} = p^T rp$ and $\tilde{h} := p^T hp$. Then we have from (2) that

$$\tilde{r} = o^T bo - a - c \circ \tilde{h}, \tag{12}$$

where for simplicity we also denote $a := \operatorname{diag}[f(\lambda_1), \ldots, f(\lambda_n)]$, $b := \operatorname{diag}[f(\mu_1), \ldots, f(\mu_n)]$, and $o := q^T p$.

Since $\operatorname{diag}[\lambda_1, \ldots, \lambda_n] = p^T xp = o^T \operatorname{diag}[\mu_1, \ldots, \mu_n]o - \tilde{h}$, we have

$$\sum_{k=1}^{n} o_{ki}o_{kj}\mu_k - \tilde{h}_{ij} = \begin{cases} \lambda_i & \text{if } i = j; \\ 0 & \text{else,} \end{cases} \qquad i, j = 1, \ldots, n. \tag{13}$$

Since $o = q^T p = (q-p)^T p + I$ and $\|p-q\| \leq \eta\|h\|$, it follows that

$$o_{ij} = O(\|h\|) \quad \forall i \neq j. \tag{14}$$

Since $p, q \in \mathcal{O}$, we have $o \in \mathcal{O}$ so that $o^T o = I$. This implies

$$1 = o_{ii}^2 + \sum_{k \neq i} o_{ki}^2 = o_{ii}^2 + O(\|h\|^2), \quad i = 1, \ldots, n, \tag{15}$$

$$0 = o_{ii}o_{ij} + o_{ji}o_{jj} + \sum_{k \neq i,j} o_{ki}o_{kj} = o_{ii}o_{ij} + o_{ji}o_{jj} + O(\|h\|^2) \quad \forall i \neq j. \tag{16}$$

We now show that $\tilde{r} = o(\|h\|)$ which, by $\|r\| = \|\tilde{r}\|$, would prove (11). For any $i \in \{1, \ldots, n\}$, we have from (12) and (13) that

$$\begin{aligned}
\tilde{r}_{ii} &= \sum_{k=1}^{n} o_{ki}^2 f(\mu_k) - f(\lambda_i) - f'(\lambda_i)\tilde{h}_{ii} \\
&= \sum_{k=1}^{n} o_{ki}^2 f(\mu_k) - f(\lambda_i) - f'(\lambda_i)\left(-\lambda_i + \sum_{k=1}^{n} o_{ki}^2 \mu_k\right) \\
&= o_{ii}^2 f(\mu_i) - f(\lambda_i) - f'(\lambda_i)(-\lambda_i + o_{ii}^2 \mu_i) + O(\|h\|^2) \\
&= (1 + O(\|h\|^2))f(\mu_i) - f(\lambda_i) - f'(\lambda_i)(-\lambda_i + (1 + O(\|h\|^2))\mu_i) + O(\|h\|^2) \\
&= f(\mu_i) - f(\lambda_i) - f'(\lambda_i)(\mu_i - \lambda_i) + O(\|h\|^2),
\end{aligned}$$

where the third and fifth equalities use (14), (15), and the local boundedness of f. Since f is differentiable at $\lambda_1, \ldots, \lambda_n$ and Lemma 3.2 implies $|\mu_i - \lambda_i| \leq \|h\|$, the right-hand side is $o(\|h\|)$. For any $i, j \in \{1, \ldots, n\}$ with $i \neq j$, we have from (12) and (13) that

$$\begin{aligned}
\tilde{r}_{ij} &= \sum_{k=1}^{n} o_{ki}o_{kj}f(\mu_k) - c_{ij}\tilde{h}_{ij} \\
&= \sum_{k=1}^{n} o_{ki}o_{kj}f(\mu_k) - c_{ij}\sum_{k=1}^{n} o_{ki}o_{kj}\mu_k \\
&= o_{ii}o_{ij}f(\mu_i) + o_{ji}o_{jj}f(\mu_j) - c_{ij}(o_{ii}o_{ij}\mu_i + o_{ji}o_{jj}\mu_j) + O(\|h\|^2) \\
&= (o_{ii}o_{ij} + o_{ji}o_{jj})f(\mu_i) + o_{ji}o_{jj}(f(\mu_j) - f(\mu_i)) \\
&\quad - c_{ij}\left((o_{ii}o_{ij} + o_{ji}o_{jj})\mu_i + o_{ji}o_{jj}(\mu_j - \mu_i)\right) + O(\|h\|^2) \\
&= o_{ji}o_{jj}\left(f(\mu_j) - f(\mu_i) - c_{ij}(\mu_j - \mu_i)\right) + O(\|h\|^2),
\end{aligned}$$

where the third and fifth equalities use (14), (16), and the local boundedness of f. Thus, if $\lambda_i = \lambda_j$, the preceding relation together with (14) and $|\mu_i - \lambda_i| \le \|h\|$, $|\mu_j - \lambda_j| \le \|h\|$ and the continuity of f at λ_i yields

$$\tilde{r}_{ij} = o(\|h\|).$$

If $\lambda_i \ne \lambda_j$, then $c_{ij} = (f(\lambda_j) - f(\lambda_i))/(\lambda_j - \lambda_i)$ and the preceding relation yields

$$\begin{aligned}
\tilde{r}_{ij} &= o_{ji}o_{jj}\left(f(\mu_j) - f(\mu_i) - \frac{f(\lambda_j) - f(\lambda_i)}{\lambda_j - \lambda_i}(\mu_j - \mu_i)\right) + O(\|h\|^2) \\
&= o_{ji}o_{jj}\left(f(\mu_j) - f(\mu_i) - (f(\lambda_j) - f(\lambda_i))\left(1 + \frac{\mu_j - \mu_i - \lambda_j + \lambda_i}{\lambda_j - \lambda_i}\right)\right) + O(\|h\|^2).
\end{aligned}$$

This together with (14) and $|\mu_i - \lambda_i| \le \|h\|$, $|\mu_j - \lambda_j| \le \|h\|$ and the continuity of f at λ_i and λ_j yields $\tilde{r}_{ij} = o(\|h\|)$.

Suppose $f : \mathbb{R} \to \mathbb{R}$ is not differentiable at λ_i for some $i \in \{1, \dots, n\}$. Then, either f is not directionally differentiable at λ_i or, if it is, the right- and left-directional derivatives of f at λ_i are unequal. In either case, this means there exist two sequences of nonzero scalars t^ν and τ^ν, $\nu = 1, 2, \dots$, converging to zero, such that the limits

$$\lim_{\nu\to\infty} \frac{f(\lambda_i + t^\nu) - f(\lambda_i)}{t^\nu}, \qquad \lim_{\nu\to\infty} \frac{f(\lambda_i + \tau^\nu) - f(\lambda_i)}{\tau^\nu}$$

exist (possibly $-\infty$ or ∞) and either are unequal or are both equal to ∞ or are both equal to $-\infty$. Consider any $p \in \mathcal{O}$ satisfying $x = p\ \mathrm{diag}[\lambda_1, \dots, \lambda_n]p^T$. Then, letting $h = p\mathrm{diag}[0, \dots, 1, \dots, 0]p^T$ with the 1 being in the ith diagonal, we obtain that $x + th = p\mathrm{diag}[\lambda_1, \dots, \lambda_i + t, \dots, \lambda_n]p^T$ for all $t \in \mathbb{R}$ and hence

$$\begin{aligned}
\lim_{\nu\to\infty} \frac{f^{\square}(x + t^\nu h) - f^{\square}(x)}{t^\nu} &= p\ \mathrm{diag}\left[0, \dots, 0, \lim_{\nu\to\infty} \frac{f(\lambda_i + t^\nu) - f(\lambda_i)}{t^\nu}, 0, \dots, 0\right] p^T, \\
\lim_{\nu\to\infty} \frac{f^{\square}(x + \tau^\nu h) - f^{\square}(x)}{\tau^\nu} &= p\ \mathrm{diag}\left[0, \dots, 0, \lim_{\nu\to\infty} \frac{f(\lambda_i + \tau^\nu) - f(\lambda_i)}{\tau^\nu}, 0, \dots, 0\right] p^T.
\end{aligned}$$

It follows that these two limits either are unequal or are both nonfinite. Thus f is not differentiable at x.

(b) is an immediate consequence of (a). $\square$

Notice that the Jacobian formula (10) is independent of the choice of p and the ordering of $\lambda_1, \dots, \lambda_n$. This formula, together with the differentiability of $f^{\square}$, has been shown under the assumption that f is continuously differentiable—see Theorem V.3.3 and p. 150 of [1]. Proposition 4.3(b) improves on this result by assuming only that f is differentiable. After obtaining Proposition 4.3, we learned of a closely related recent result of Lewis and Sendov [19] on twice differentiability of spectral functions. In particular, in the case where $f = g'$ for some differentiable $g : \mathbb{R} \to \mathbb{R}$, applying Theorem 3.3 in [19] to the spectral function

$$x \mapsto g(\lambda_1) + \cdots + g(\lambda_n),$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $x \in \mathcal{S}$ in nonincreasing order, yields Proposition 4.3(a). For general f, however, Proposition 4.3(a) appears to be distinct from the results in [19]. In particular, for any $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$, there exists a function $f : \mathbb{R} \to \mathbb{R}$ that is differentiable at $\lambda_1, \ldots, \lambda_n$ and yet there is no differentiable function $g : \mathbb{R} \to \mathbb{R}$ satisfying $g' = f$. One such f is

$$f(\xi) := \begin{cases} (\xi - \lambda_1)^2 & \text{if } \xi \in \{\alpha_1, \alpha_2, \ldots\}; \\ 0 & \text{else,} \end{cases}$$

where $\alpha_1, \alpha_2, \ldots$ is any sequence of points in $\mathbb{R} \backslash \{\lambda_1, \ldots, \lambda_n\}$ converging to λ_1. Here f is differentiable at $\lambda_1, \ldots, \lambda_n$, but the range of f is not an interval, so f cannot be the derivative of a differentiable function. Specifically, a theorem of Darboux says that, for any open interval $\mathcal{I}$ containing a closed interval $[\alpha, \beta]$ and any differentiable $g : \mathcal{I} \to \mathbb{R}$, either $[g'(\alpha), g'(\beta)]$ or $[g'(\beta), g'(\alpha)]$ is a subset of $\{g'(\xi) | \alpha \le \xi \le \beta\}$. (This can be seen by defining, for each η strictly between $g'(\alpha)$ and $g'(\beta)$, the function $h(\xi) := g(\xi) - \eta\xi$. Then h is differentiable on $[\alpha, \beta]$ and $h'(\alpha) = g'(\alpha) - \eta$, $h'(\beta) = g'(\beta) - \eta$ have opposite signs. Thus, h has an extremum at some ξ^* in (α, β), implying $h'(\xi^*) = 0$ or, equivalently, $g'(\xi^*) = \eta$.) In fact, any function that coincides with f in a neighborhood of λ_1 cannot be the derivative of a differentiable function. Also, we speculate that the proof idea for Proposition 4.3(a) may be useful for second-or-higher order analysis of spectral functions.

We next have the following continuous differentiability result based on [8, Lem. 4], which in turn was proven using Lemmas 3.1 and 3.2.

PROPOSITION 4.4. *For any $f : \mathbb{R} \to \mathbb{R}$, the matrix function $f^{\square}$ is continuously differentiable if and only if f is continuously differentiable.*

Proof. The "if" direction was proven in [8, Lem. 4]. To see the "only if" direction, suppose $f^{\square}$ is continuously differentiable. Then it follows from (10)and the definition of $f^{[1]}(\cdot)$ that $f'(\lambda_1)$ is well defined for all $\lambda_1 \in \mathbb{R}$. Moreover, $\nabla f^{\square}(\text{diag}[\lambda_1, 0, \ldots, 0])$ is continuous in λ_1 or, equivalently, $f'(\lambda_1)$ is continuous in λ_1. ∎

Similar to Proposition 4.3, it can be seen that, in the case where $f = g'$ for some differentiable g, Proposition 4.4 is a special case of Theorem 4.2 in [19]. We next have the following result of Rockafellar and Wets [26, Thm. 9.67] which we need to analyze strict continuity and Lipschitz continuity of $f^{\square}$.

LEMMA 4.5. *Suppose $f : \mathbb{R}^k \to \mathbb{R}$ is strictly continuous. Then there exist continuously differentiable functions $f^\nu : \mathbb{R}^k \to \mathbb{R}$, $\nu = 1, 2, \ldots$, converging uniformly to f on any compact set C in $\mathbb{R}^k$ and satisfying*

$$|||\nabla f^\nu(x)||| \le \sup_{x \in C} \; \text{lip} f(x) \qquad \forall x \in C, \; \forall \nu.$$

Lemma 4.5 is slightly different from the original version given in [26, Thm. 9.67]. In particular, the second part of Lemma 4.5 is not contained in [26, Thm. 9.67], but it is implicit in its proof. This second part is needed to show that strict continuity and Lipschitz continuity are inherited by $f^{\square}$ from f. We note that the proof idea of Proposition 4.1 cannot be used because eigenvectors do not behave in a (locally) Lipschitzian manner.

PROPOSITION 4.6. *For any $f : \mathbb{R} \to \mathbb{R}$, the following results hold:*

(a) *$f^{\square}$ is strictly continuous at an $x \in \mathcal{S}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$ if and only if f is strictly continuous at $\lambda_1, \ldots, \lambda_n$.*

(b) *$f^{\square}$ is strictly continuous if and only if f is strictly continuous.*

(c) *$f^{\square}$ is Lipschitz continuous with constant κ if and only if f is Lipschitz continuous with constant κ.*

Proof. (a) Fix any $x \in \mathcal{S}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$.

Suppose f is strictly continuous at $\lambda_1, \ldots, \lambda_n$. Then, there exist scalars $\kappa_i > 0$ and $\delta_i > 0$, $i = 1, \ldots, n$, such that

$$|f(\xi) - f(\zeta)| \le \kappa_i |\xi - \zeta| \quad \forall \xi, \zeta \in [\lambda_i - \delta_i, \lambda_i + \delta_i]$$

for all i. Let $\tilde{f} : \mathbb{R} \to \mathbb{R}$ be the function that coincides with f on

$$C := \bigcup_{i=1}^{n} [\lambda_i - \delta_i, \lambda_i + \delta_i]$$

and, on $\mathbb{R} \setminus C$, is defined by linearly extrapolating f at the boundary points of C. In other words, if $\xi < \zeta$ are two points in C such that $(\xi, \zeta) \subseteq \mathbb{R} \setminus C$, then $\tilde{f}(t\xi + (1-t)\zeta) = tf(\xi) + (1-t)f(\zeta)$ for all $t \in (0,1)$. If ξ is a point in C such that $(\xi, \infty) \subseteq \mathbb{R} \setminus C$, then $\tilde{f}(\zeta) = f(\xi)$ for all $\zeta > \xi$. Similarly, if ζ is a point in C such that $(-\infty, \zeta) \subseteq \mathbb{R} \setminus C$, then $\tilde{f}(\xi) = f(\zeta)$ for all $\xi < \zeta$. By definition, $\tilde{f}$ is Lipschitz continuous, so there exists a scalar $\kappa > 0$ such that $\operatorname{lip} f(\xi) \le \kappa$ for all $\xi \in \mathbb{R}$. Since C is compact, by Lemma 4.5, there exist continuously differentiable functions $f^{\nu} : \mathbb{R} \to \mathbb{R}$, $\nu = 1, 2, \ldots$, converging uniformly to $\tilde{f}$ and satisfying

$$|(f^{\nu})'(\xi)| \le \kappa \qquad \forall\, \xi \in C,\ \forall \nu. \tag{17}$$

Denote $\delta := \min_{i=1,\ldots,n} \delta_i$. By Lemma 3.2, C contains all the eigenvalues of $y \in \mathcal{B}(x, \delta)$. Moreover, for any $w \in \mathcal{B}(x, \delta)$, any $q \in \mathcal{O}$, and any $\mu = (\mu_1, \ldots, \mu_n)^T \in \mathbb{R}^n$ such that $w = q \operatorname{diag}[\mu_1, \ldots, \mu_n] q^T$, we have

$$\begin{aligned} \|(f^{\nu})^{\square}(w) - f^{\square}(w)\| &= \|q \operatorname{diag}[f^{\nu}(\mu_1), \ldots, f^{\nu}(\mu_n)] q^T - q \operatorname{diag}[f(\mu_1), \ldots, f(\mu_n)] q^T\| \\ &= \|\operatorname{diag}[f^{\nu}(\mu_1) - f(\mu_1), \ldots, f^{\nu}(\mu_n) - f(\mu_n)]\|, \end{aligned}$$

where the second equality uses $q^T q = I$ and properties of the Frobenius norm $\|\cdot\|$. Since $\{f^{\nu}\}_1^{\infty}$ converges uniformly to f on C, this shows that $\{(f^{\nu})^{\square}\}_1^{\infty}$ converges uniformly to $f^{\square}$ on $\mathcal{B}(x, \delta)$. Moreover, it follows from (10) that, for all $w \in \mathcal{B}(x, \delta)$ and all ν, we have

$$\begin{aligned} |||\nabla (f^{\nu})^{\square}(w)||| &= \sup_{\|h\|=1} \|\nabla (f^{\nu})^{\square}(w) h\| \\ &= \sup_{\|h\|=1} \|q((f^{\nu})^{[1]}(\mu) \circ (q^T h q)) q^T\| \\ &= \sup_{\|h\|=1} \|(f^{\nu})^{[1]}(\mu) \circ (q^T h q)\| \\ &\le \sup_{\|h\|=1} \kappa \|q^T h q\| \ = \ \kappa, \end{aligned} \tag{18}$$

where the first inequality uses (17). Fix any $y, z \in \mathcal{B}(x,\delta)$ with $y \neq z$. Since $\{(f^\nu)^\square\}_1^\infty$ converges uniformly to $f^\square$ on $\mathcal{B}(x,\delta)$, then for any $\epsilon > 0$ there exists an integer ν_0 such that for all $\nu \geq \nu_0$ we have

$$\|(f^\nu)^\square(w) - f^\square(w)\| \leq \epsilon\|y - z\| \quad \forall w \in \mathcal{B}(x,\delta).$$

Since f^ν is continuously differentiable, then Proposition 4.4 shows that $(f^\nu)^\square$ is continuously differentiable for all ν. Then, by (18) and the mean-value theorem for continuously differentiable functions, we have

$$\begin{aligned}
&\|f^\square(y) - f^\square(z)\| \\
&= \|f^\square(y) - (f^\nu)^\square(y) + (f^\nu)^\square(y) - (f^\nu)^\square(z) + (f^\nu)^\square(z) - f^\square(z)\| \\
&\leq \|f^\square(y) - (f^\nu)^\square(y)\| + \|(f^\nu)^\square(y) - (f^\nu)^\square(z)\| + \|(f^\nu)^\square(z) - f^\square(z)\| \\
&\leq 2\epsilon\|y - z\| + \|\int_0^1 \nabla(f^\nu)^\square(z + \tau(y - z))(y - z)d\tau\| \\
&\leq (\kappa + 2\epsilon)\|y - z\|.
\end{aligned}$$

Since $y, z \in \mathcal{B}(x,\delta)$ and ϵ is arbitrary, this yields

$$\|f^\square(y) - f^\square(z)\| \leq \kappa\|y - z\| \quad \forall y, z \in \mathcal{B}(x,\delta). \tag{19}$$

Thus $f^\square$ is strictly continuous at x.

Suppose instead that $f^\square$ is strictly continuous at x. Then, there exist scalars $\kappa > 0$ and $\delta > 0$ such that (19) holds. Choose any $p \in \mathcal{O}$ satisfying $x = p\,\mathrm{diag}[\lambda_1, \ldots, \lambda_n]p^T$. For any $i \in \{1, \ldots, n\}$ and any $\psi, \zeta \in [\lambda_i - \delta, \lambda_i + \delta]$, let

$$\begin{aligned}
y &:= p\,\mathrm{diag}[\lambda_1, \ldots, \lambda_{i-1}, \psi, \lambda_{i+1}, \ldots, \lambda_n]p^T, \\
z &:= p\,\mathrm{diag}[\lambda_1, \ldots, \lambda_{i-1}, \zeta, \lambda_{i+1}, \ldots, \lambda_n]p^T.
\end{aligned}$$

Then, $\|y - x\| = |\psi - \lambda_i| \leq \delta$ and $\|z - x\| = |\zeta - \lambda_i| \leq \delta$, so it follows from (2) and (19) that

$$\begin{aligned}
|f(\psi) - f(\zeta)| &= \|f^\square(y) - f^\square(z)\| \\
&\leq \kappa\|y - z\| \\
&= \kappa|\psi - \zeta|.
\end{aligned}$$

This shows that f is strictly continuous at λ_i for $i = 1, \ldots, n$.

(b) is an immediate consequence of (a).

(c) Suppose f is Lipschitz continuous with constant κ. Then $\mathrm{lip} f(\xi) \leq \kappa$ for all $\xi \in \mathbb{R}$. Fix any $x \in \mathcal{S}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$. For any scalar $\delta > 0$, define the compact set C in $\mathbb{R}$ by

$$C := \bigcup_{i=1}^n [\lambda_i - \delta, \lambda_i + \delta].$$

Then, as in the proof of (a), we obtain that (19) holds. Since the choice of $\delta > 0$ was arbitrary and κ is independent of δ, this implies

$$\|f^\square(y) - f^\square(z)\| \leq \kappa\|y - z\| \qquad \forall y, z \in \mathcal{S}.$$

Hence $f^\square$ is Lipschitz continuous with constant κ.

Suppose instead that $f^{\square}$ is Lipschitz continuous with constant $\kappa > 0$. Then, for any $\xi, \zeta \in \mathbb{R}$ we have

$$\begin{aligned}|f(\xi) - f(\zeta)| &= \|f^{\square}(\mathrm{diag}[\xi, 0, \dots, 0]) - f^{\square}(\mathrm{diag}[\zeta, 0, \dots, 0])\| \\ &\le \kappa \|\mathrm{diag}[\xi, 0, \dots, 0] - \mathrm{diag}[\zeta, 0, \dots, 0]\| \\ &= \kappa|\xi - \zeta|,\end{aligned}$$

so f is Lipschitz continuous with constant κ. ∎

Suppose $f : \mathbb{R} \to \mathbb{R}$ is strictly continuous. Then, by Proposition 4.6, $f^{\square}$ is strictly continuous. Hence $\partial_B f^{\square}(x)$ is well defined for all $x \in \mathcal{S}$. The following lemma studies the structure of this generalized Jacobian.

LEMMA 4.7. *Let $f : \mathbb{R} \to \mathbb{R}$ be strictly continuous. Then, for any $x \in \mathcal{S}$, the generalized Jacobian $\partial_B f^{\square}(x)$ is well defined and nonempty. Moreover, for any $V \in \partial_B f^{\square}(x)$, we have*

$$Vh = p((p^T hp) \circ c)p^T \qquad \forall h \in \mathcal{S} \tag{20}$$

for some $p \in \mathcal{O}_x$, $c \in \mathcal{S}$, and $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ satisfying $x = p\ \mathrm{diag}[\lambda_1, \dots, \lambda_n]p^T$ and

$$c_{ij} = \frac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j} \quad \textit{whenever } \lambda_i \ne \lambda_j, \qquad c_{ij} \in \partial f(\lambda_i) \quad \textit{whenever } \lambda_i = \lambda_j. \tag{21}$$

Proof. Fix any $V \in \partial_B f^{\square}(x)$. According to the definition of $\partial_B f^{\square}(x)$, there exists a sequence $\{x_k\} \subseteq \mathcal{S}$ converging to x such that f is differentiable at x_k for all k and $\lim_{k\to\infty} \nabla f^{\square}(x_k) = V$. Let $\lambda_1 \ge \cdots \ge \lambda_n$ and $\lambda_1^k \ge \cdots \ge \lambda_n^k$ be the eigenvalues of x and x_k, $k = 1, 2, \dots$, respectively. Choose any $p_k \in \mathcal{O}_{x_k}$. By Lemma 3.1, there exist η and $\tilde{p}_k \in \mathcal{O}_x$ satisfying

$$\|p_k - \tilde{p}_k\| \le \eta \|x - x_k\|$$

for all k sufficiently large. By passing to a subsequence if necessary, we assume that this holds for all k and that p_k converges. By Lemma 3.2, we have $\lambda_i^k \to \lambda_i$ for $i = 1, \dots, n$. Denote $\lambda^k = (\lambda_1^k, \dots, \lambda_n^k)^T$. Then we have from Proposition 4.3 that f is differentiable at $\lambda_1^k, \dots, \lambda_n^k$ and

$$\nabla f^{\square}(x_k)h = p_k((p_k^T hp_k) \circ c^k)p_k^T \qquad \forall h \in \mathcal{S}, \tag{22}$$

where we denote $c^k := f^{[1]}(\lambda^k)$. Thus,

$$c_{ij}^k \;=\; \begin{cases} (f(\lambda_i^k) - f(\lambda_j^k))/(\lambda_i^k - \lambda_j^k) & \text{if } \lambda_i^k \ne \lambda_j^k; \\ f'(\lambda_i^k) & \text{if } \lambda_i^k = \lambda_j^k. \end{cases} \tag{23}$$

Since f is strictly continuous, then $\{c_{ij}^k\}$ is bounded for all i, j. By passing to a subsequence if necessary, we can assume that $\{c_{ij}^k\}$ converges to some $c_{ij} \in \mathbb{R}$ for all i, j. For each i, we have

$$c_{ii}^k = f'(\lambda_i^k) \to c_{ii} \in \partial_B f(\lambda_i).$$

For each $i \ne j$ such that $\lambda_i \ne \lambda_j$, we have $\lambda_i^k \ne \lambda_j^k$ for all k sufficiently large and hence

$$c_{ij}^k = \frac{f(\lambda_i^k) - f(\lambda_j^k)}{\lambda_i^k - \lambda_j^k} \to c_{ij} = \frac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j}.$$

For each $i \neq j$ such that $\lambda_i = \lambda_j$, if $\lambda_i^k = \lambda_j^k$ for k along some subsequence, then

$$c_{ij}^k = f'(\lambda_i^k) \to c_{ii} \in \partial_B f(\lambda_i) \subseteq \partial f(\lambda_i);$$

if $\lambda_i^k \neq \lambda_j^k$ for k along some subsequence, then a mean-value theorem of Lebourg [9, Proposition 2.3.7], [26, Thm. 10.48] yields

$$c_{ij}^k = \frac{f(\lambda_i^k) - f(\lambda_j^k)}{\lambda_i^k - \lambda_j^k} \in \partial f(\hat{\lambda}_{ij}^k)$$

for some $\hat{\lambda}_{ij}^k$ in the interval between λ_i^k and λ_j^k. Since f is strictly continuous so that ∂f is upper semicontinuous [9, Proposition 2.1.5] or, equivalently, outer semicontinuous [26, Proposition 8.7], this together with $\hat{\lambda}_{ij}^k \to \lambda_i = \lambda_j$ implies the limit of $\{c_{ij}^k\}$ belongs to $\partial f(\lambda_i)$. Thus, taking limits on both sides of (22) and using the above results, we obtain (20) and (21) for some $p \in \mathcal{O}_x$ and $c \in \mathcal{S}$, which are the limit of $\{p_k\}$ and $\{f^{[1]}(\lambda^k)\}$, respectively. This proves the lemma. ∎

Lemma 4.7 does not, however, provide a characterization of $\partial_B f^{\square}$. It is an open question whether such a (tractable) characterization can be found for any strictly continuous f. In the special case where f is piecewise continuously differentiable (e.g., $f(\cdot) = |\cdot|$) and, more generally, where the directional derivative of f has a one-sided continuity property, a simple characterization of $\partial_B f^{\square}$ can be found as we show below. In what follows we denote the right- and left-directional derivative of $f : \mathbb{R} \to \mathbb{R}$ by

$$f'_+(\xi) := \lim_{\zeta \to \xi^+} \frac{f(\zeta) - f(\xi)}{\zeta - \xi}, \qquad f'_-(\xi) := \lim_{\zeta \to \xi^-} \frac{f(\zeta) - f(\xi)}{\zeta - \xi}.$$

PROPOSITION 4.8. *Let $f : \mathbb{R} \to \mathbb{R}$ be a strictly continuous and directionally differentiable function with the property that*

$$\lim_{\substack{\zeta,\nu \to \xi^\sigma \\ \zeta \neq \nu}} \frac{f(\zeta) - f(\nu)}{\zeta - \nu} = \lim_{\substack{\zeta \to \xi^\sigma \\ \zeta \in D_f}} f'(\zeta) = f'_\sigma(\xi) \quad \forall \xi \in \mathbb{R},\ \sigma \in \{-,+\}, \tag{24}$$

where $D_f := \{\xi \in \mathbb{R} | f \text{ is differentiable at } \xi\}$. Then, for any $x \in \mathcal{S}$, we have that $V \in \partial_B f^{\square}(x)$ if and only if V has the form (20) *for some $p \in \mathcal{O}_x$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ satisfying $x = p \operatorname{diag}[\lambda_1, \ldots, \lambda_n] p^T$ and c has the form*

$$c_{ij} = \begin{cases} (f(\lambda_i) - f(\lambda_j))/(\lambda_i - \lambda_j) & \text{if } \lambda_i \neq \lambda_j, \\ f'_{\sigma_i}(\lambda_i) & \text{if } \lambda_i = \lambda_j \text{ and } i \in \alpha_l, j \in \beta \cup \alpha_\nu \text{ for some } \\ & l < \nu, \\ f'_{\sigma_j}(\lambda_j) & \text{if } \lambda_i = \lambda_j \text{ and } i \in \beta \cup \alpha_l, j \in \alpha_\nu \text{ for some } \\ & l > \nu, \\ (\omega_i f'_{\sigma_i}(\lambda_i) + \omega_j f'_{\sigma_j}(\lambda_j))/(\omega_i + \omega_j) & \text{if } \lambda_i = \lambda_j \text{ and } i, j \in \alpha_l \text{ for some } l, \\ f'(\lambda_i) & \text{if } \lambda_i = \lambda_j \text{ and } i, j \in \beta \end{cases} \tag{25}$$

for some partition $\alpha_1, \ldots, \alpha_\ell, \beta$ of $\{1, \ldots, n\}$ ($\ell \geq 0$) and some $\sigma_i \in \{-,+\}$ and $\omega_i \in (0, \infty)$ for $i \in \alpha_1 \cup \cdots \cup \alpha_\ell$. (Implicit in (25) *is the differentiability of f at λ_i, $i \in \beta$.)*

Proof. Consider any $V \in \partial_B f^{\square}(x)$. By Lemma 4.7 and its proof, V has the form (20) for some $p \in \mathcal{O}_x$ and $\lambda_1 \geq \cdots \geq \lambda_n$ satisfying $x = p \operatorname{diag}[\lambda_1, \ldots, \lambda_n] p^T$ and with

c being the cluster point of c^k given by (23), $k = 1, 2, \ldots$ for some $\lambda^k = (\lambda_1^k, \ldots, \lambda_n^k)^T$ converging to $\lambda = (\lambda_1, \ldots, \lambda_n)^T$. Moreover, f is differentiable at $\lambda_1^k, \ldots, \lambda_n^k$ for all k. By passing to a subsequence if necessary, we can assume that, for each $i \in \{1, \ldots, n\}$, either (i) $\lambda_i^k > \lambda_i$ for all k or (ii) $\lambda_i^k < \lambda_i$ for all k or (iii) $\lambda_i^k = \lambda_i$ for all k. Denote

$$\beta := \{i \in \{1, \ldots, n\} | \text{case (iii) holds for } i\}.$$

By further passing to a subsequence if necessary, we can assume that, for each $i, j \in \{1, \ldots, n\} \setminus \beta$,

$$\frac{|\lambda_i^k - \lambda_i|}{|\lambda_j^k - \lambda_j|} \text{ has a limit } \rho_{ij} \in [0, \infty] \text{ as } k \to \infty.$$

Then, $\{1, \ldots, n\} \setminus \beta$ may be partitioned into disjoint subsets $\alpha_1, \ldots, \alpha_\ell$ for some $\ell \geq 0$ such that

$$\begin{array}{ll} \rho_{ij} \in (0, \infty) & \text{whenever } i, j \in \alpha_l \text{ for some } l, \\ \rho_{ij} = \infty & \text{whenever } i \in \alpha_l, j \in \alpha_\nu \text{ for some } l < \nu. \end{array}$$

Moreover, for each $l \in \{1, \ldots, \ell\}$ and each $i \in \alpha_l$, the quantity

$$\omega_i^k := |\lambda_i^k - \lambda_i| / \left(\sum_{j \in \alpha_l} |\lambda_j^k - \lambda_j| \right)$$

converges to a positive limit, which we denote by ω_i. For each $i \in \{1, \ldots, n\} \setminus \beta$, set $\sigma_i = +$ if case (i) holds for i and set $\sigma_i = -$ if case (ii) holds for i. We now verify that c has the form (25). For any $i, j \in \{1, \ldots, n\}$ with $\lambda_i \neq \lambda_j$, this follows from (21). For any $i, j \in \{1, \ldots, n\}$ with $\lambda_i = \lambda_j$, we consider the following disjoint cases.

Case 1. Suppose $i \in \alpha_l$ and $j \in \alpha_\nu$ for some $l, \nu \in \{1, \ldots, \ell\}$ and $\sigma_i = \sigma_j = +$. Then $\lambda_i^k > \lambda_i$ and $\lambda_j^k > \lambda_i$ for all k. If $l = \nu$, it follows from (23) and (24) that

$$c_{ij}^k \to f'_+(\lambda_i) = (\omega_i f'_{\sigma_i}(\lambda_i) + \omega_j f'_{\sigma_j}(\lambda_j))/(\omega_i + \omega_j) = c_{ij},$$

where the last equality uses (25). If $l < \nu$, a similar argument shows that

$$c_{ij}^k \to f'_+(\lambda_i) = f'_{\sigma_i}(\lambda_i) = c_{ij}.$$

The remaining subcase of $l > \nu$ can be treated analogously.

Case 2. Suppose $i \in \alpha_l$ and $j \in \alpha_\nu$ for some $l, \nu \in \{1, \ldots, \ell\}$ and $\sigma_i = +, \sigma_j = -$. Then $\lambda_i^k > \lambda_i$ and $\lambda_j^k < \lambda_i$ for all k. If $l = \nu$, it follows from (23) and (24) that

$$\begin{aligned} c_{ij}^k &= \frac{f(\lambda_i^k) - f(\lambda_j^k)}{\lambda_i^k - \lambda_j^k} \\ &= \frac{\omega_i^k}{\omega_i^k + \omega_j^k} \frac{f(\lambda_i^k) - f(\lambda_i)}{\lambda_i^k - \lambda_i} + \frac{\omega_j^k}{\omega_i^k + \omega_j^k} \frac{f(\lambda_j^k) - f(\lambda_i)}{\lambda_j^k - \lambda_i} \\ &\to \frac{\omega_i}{\omega_i + \omega_j} f'_+(\lambda_i) + \frac{\omega_j}{\omega_i + \omega_j} f'_-(\lambda_j) \\ &= (\omega_i f'_{\sigma_i}(\lambda_i) + \omega_j f'_{\sigma_j}(\lambda_j))/(\omega_i + \omega_j) \\ &= c_{ij}, \end{aligned}$$

where the last equality uses (25). If $l < \nu$, a similar argument together with $\rho_{ij} = \infty$ shows that

$$\begin{aligned} c_{ij}^k &= \frac{|\lambda_i^k - \lambda_i|}{|\lambda_i^k - \lambda_i| + |\lambda_j^k - \lambda_j|} \frac{f(\lambda_i^k) - f(\lambda_i)}{\lambda_i^k - \lambda_i} + \frac{|\lambda_j^k - \lambda_j|}{|\lambda_i^k - \lambda_i| + |\lambda_j^k - \lambda_j|} \frac{f(\lambda_j^k) - f(\lambda_i)}{\lambda_j^k - \lambda_i} \\ &\to f'_+(\lambda_i) \\ &= c_{ij}. \end{aligned}$$

The remaining subcase of $l > \nu$ can be treated analogously.

Case 3. Suppose $i \in \alpha_l$ and $j \in \beta$ for some $l \in \{1, \dots, \ell\}$ and $\sigma_i = +$. Then $\lambda_i^k > \lambda_i$ and $\lambda_j^k = \lambda_i$ for all k. It follows from (23) and (24) that

$$c_{ij}^k = \frac{f(\lambda_i^k) - f(\lambda_i)}{\lambda_i^k - \lambda_i} \to f'_+(\lambda_i) = c_{ij}.$$

Case 4. Suppose $i, j \in \beta$. Then $\lambda_i^k = \lambda_j^k = \lambda_i$ for all k and it follows from (23) that f is differentiable at λ_i, $i \in \beta$, and

$$c_{ij}^k = f'(\lambda_i) = c_{ij}.$$

Case 5. Suppose $i \in \alpha_l$ and $j \in \alpha_\nu$ for some $l, \nu \in \{1, \dots, \ell\}$ and $\sigma_i = \sigma_j = -$. This case is analogous to Case 1.

Case 6. Suppose $i \in \alpha_l$ and $j \in \beta$ for some $l \in \{1, \dots, \ell\}$ and $\sigma_i = -$. This case is analogous to Case 3.

Conversely, suppose that V has the form (20) for some $p \in \mathcal{O}_x$ and $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ satisfying $x = p \operatorname{diag}[\lambda_1, \dots, \lambda_n] p^T$ and c has the form (25) for some partition $\alpha_1, \dots, \alpha_\ell, \beta$ of $\{1, \dots, n\}$ ($\ell \geq 0$) and some $\sigma_i \in \{-, +\}$ and $\omega_i \in (0, \infty)$ for $i \in \alpha_1 \cup \dots \cup \alpha_\ell$. For each $i \in \beta$, set $d_i^k := 0$ for $k = 1, 2, \dots$. For each $i \in \alpha_l$, $l \in \{1, \dots, \ell\}$, let $\delta_i^k = \omega_i (1/2)^{kl}$ if $\sigma_i = +$ and let $\delta_i^k = -\omega_i (1/2)^{kl}$ if $\sigma_i = -$, $k = 1, 2, \dots$. Since f is strictly continuous, by Rademacher's theorem (see [26, Thm. 9.60]), D_f is dense in $\mathbb{R}$. Thus, for each $i \in \alpha_1 \cup \dots \cup \alpha_\ell$ and each index k, there exists $d_i^k \in \mathbb{R}$ satisfying

$$\lambda_i + d_i^k \in D_f \quad \text{and} \quad |d_i^k - \delta_i^k| \leq |\delta_i^k|^2.$$

Let $\lambda_i^k := \lambda_i + d_i^k$ for all i. Then, by Proposition 4.3, $f^{\square}$ is differentiable at

$$x^k := p \operatorname{diag}[\lambda_1^k, \dots, \lambda_n^k] p^T$$

for all k and

$$\nabla f^{\square}(x^k) h = p(c^k \circ (p^T h p)) p^T \quad \forall h \in \mathcal{S},$$

where c^k is given by (23). Also, the definition of $d_1^k, \dots, d_n^k$ yields

$$d_i^k \to 0 \ \forall i, \quad \frac{|d_i^k|}{|d_j^k|} \to \frac{\omega_i}{\omega_j} \ \forall i, j \in \alpha_l, \ l = 1, \dots, \ell, \quad \frac{|d_i^k|}{|d_j^k|} \to \infty \ \forall i \in \alpha_l, j \in \alpha_\nu, l < \nu,$$

and $\sigma_i = +$ implies $d_i^k > 0$ for all k and $\sigma_i = -$ implies $d_i^k < 0$ for all k. Then, it is straightforward to verify that $x^k \to x$ and $c^k \to c$, implying

$$\nabla f^{\square}(x^k) h \to p(c \circ (p^T h p)) p^T = V h \quad \forall h \in \mathcal{S}.$$

This shows that $V \in \partial_B f^{\square}(x)$. ∎

Notice that a V of the form (20) is invertible if and only all entries of c are nonzero. Also, notice that the p in the formula (20) depends on V; i.e., two elements of $\partial_B f^{\square}(x)$ may have different p in their formulas. Thus $\partial f^{\square}(x)$, being the convex hull of $\partial_B f^{\square}(x)$, has a rather complicated structure.

The following lemma, proven by Sun and Sun [29, Thm. 3.6] using the definition of generalized Jacobian,[1] enables one to study the semismooth property of $f^{\square}$ by examining only those points $x \in \mathcal{S}$ where $f^{\square}$ is differentiable and thus work only with the Jacobian of $f^{\square}$, rather than the generalized Jacobian.

LEMMA 4.9. *Suppose $F : \mathcal{S} \to \mathcal{S}$ is strictly continuous and directionally differentiable in a neighborhood of $x \in \mathcal{S}$. Then, for any $0 < \rho < \infty$, the following two statements (where $O(\cdot)$ depends on F and x only) are equivalent:*

(a) *For any $h \in \mathcal{S}$ and any $V \in \partial F(x+h)$,*

$$F(x+h) - F(x) - Vh = o(\|h\|) \quad (\textit{respectively, } O(\|h\|^{1+\rho})).$$

(b) *For any $h \in \mathcal{S}$ such that F is differentiable at $x+h$,*

$$F(x+h) - F(x) - \nabla F(x+h)h = o(\|h\|) \quad (\textit{respectively, } O(\|h\|^{1+\rho})).$$

By using Lemmas 3.1, 3.2, and 4.9 and Propositions 4.2, 4.3, and 4.6, we are now ready to state and prove the last result of this section. The proof is motivated by and in some sense generalizes the proof of Lemma 4.12 in [29], though it is also simpler. The proof idea was also used for proving Proposition 4.3, with the main difference being that here $x + h$ is diagonalized rather than x.

PROPOSITION 4.10. *For any $f : \mathbb{R} \to \mathbb{R}$, the matrix function $f^{\square}$ is semismooth if and only if f is semismooth. If $f : \mathbb{R} \to \mathbb{R}$ is ρ-order semismooth $(0 < \rho < \infty)$, then $f^{\square}$ is $\min\{1, \rho\}$-order semismooth.*

Proof. Suppose f is semismooth. Then f is strictly continuous and directionally differentiable. By Propositions 4.2 and 4.6, $f^{\square}$ is strictly continuous and directionally differentiable. Let $\mathcal{D} := \{x \in \mathcal{S} | f^{\square} \text{ is differentiable at } x\}$.

Fix any $x \in \mathcal{S}$ and let $\lambda_1 \geq \cdots \geq \lambda_n$ denote the eigenvalues of x. By Lemma 3.1, there exist scalars $\eta > 0$ and $\epsilon > 0$ such that (3) holds. By taking ϵ smaller if necessary, we can assume that $\epsilon < (\lambda_i - \lambda_{i+1})/2$ whenever $\lambda_i \neq \lambda_{i+1}$. We will show that, for any $h \in \mathcal{S}$ with $x + h \in \mathcal{D}$ and $\|h\| \leq \epsilon$, we have

$$f^{\square}(x+h) - f^{\square}(x) - \nabla f^{\square}(x+h)h = o(\|h\|), \tag{26}$$

where $o(\cdot)$ and $O(\cdot)$ depend on f and x only. Then, it follows from Lemma 4.9 that $f^{\square}$ is semismooth at x. Since the choice of $x \in \mathcal{S}$ was arbitrary, $f^{\square}$ is semismooth. Let $\mu_1 \geq \cdots \geq \mu_n$ denote the eigenvalues of $x + h$, and choose any $q \in \mathcal{O}_{x+h}$. Then, there exists $p \in \mathcal{O}_x$ satisfying

$$\|p - q\| \leq \eta \|h\|.$$

For simplicity, let r denote the left-hand side of (26), i.e.,

$$r := f^{\square}(x+h) - f^{\square}(x) - \nabla f^{\square}(x+h)h,$$

[1]Sun and Sun did not consider the case of $o(\|h\|)$, but their argument readily applies to this case.

and denote $\tilde{r} = q^T r q$ and $\tilde{h} := q^T h q$. Since $x + h \in \mathcal{D}$, Proposition 4.3 implies f is differentiable at $\mu_1, \ldots, \mu_n$. Then we have from (2) and (10) that

$$\tilde{r} = b - o^T a o - c \circ \tilde{h}, \tag{27}$$

where for simplicity we also denote $a := \mathrm{diag}[f(\lambda_1), \ldots, f(\lambda_n)]$, $b := \mathrm{diag}[f(\mu_1), \ldots, f(\mu_n)]$, $c := f^{[1]}(\mu)$, and $o := p^T q$.

Since $\mathrm{diag}[\mu_1, \ldots, \mu_n] = q^T(x + h)q = o^T \mathrm{diag}[\lambda_1, \ldots, \lambda_n] o + \tilde{h}$, we have

$$\sum_{k=1}^{n} o_{ki} o_{kj} \lambda_k + \tilde{h}_{ij} = \begin{cases} \mu_i & \text{if } i = j, \\ 0 & \text{else,} \end{cases} \qquad i, j = 1, \ldots, n. \tag{28}$$

Since $o = p^T q = (p - q)^T q + I$ and $\|p - q\| \le \eta \|h\|$, it follows that

$$o_{ij} = O(\|h\|) \quad \forall i \ne j. \tag{29}$$

Since $p, q \in \mathcal{O}$, we have $o \in \mathcal{O}$ so that $o^T o = I$. This implies

$$1 = o_{ii}^2 + \sum_{k \ne i} o_{ki}^2 = o_{ii}^2 + O(\|h\|^2), \quad i = 1, \ldots, n, \tag{30}$$

$$0 = o_{ii} o_{ij} + o_{ji} o_{jj} + \sum_{k \ne i,j} o_{ki} o_{kj} = o_{ii} o_{ij} + o_{ji} o_{jj} + O(\|h\|^2) \quad \forall i \ne j. \tag{31}$$

We now show that $\tilde{r} = o(\|h\|)$ which, by $\|r\| = \|\tilde{r}\|$, would prove (26). For any $i \in \{1, \ldots, n\}$, we have from (27) and (28) that

$$\begin{aligned}
\tilde{r}_{ii} &= f(\mu_i) - \sum_{k=1}^{n} o_{ki}^2 f(\lambda_k) - f'(\mu_i) \tilde{h}_{ii} \\
&= f(\mu_i) - \sum_{k=1}^{n} o_{ki}^2 f(\lambda_k) - f'(\mu_i) \left(\mu_i - \sum_{k=1}^{n} o_{ki}^2 \lambda_k \right) \\
&= f(\mu_i) - o_{ii}^2 f(\lambda_i) - f'(\mu_i)(\mu_i - o_{ii}^2 \lambda_i) + O(\|h\|^2) \\
&= f(\mu_i) - (1 + O(\|h\|^2)) f(\lambda_i) - f'(\mu_i)(\mu_i - (1 + O(\|h\|^2))\lambda_i) + O(\|h\|^2) \\
&= f(\mu_i) - f(\lambda_i) - f'(\mu_i)(\mu_i - \lambda_i) + O(\|h\|^2),
\end{aligned}$$

where the third and fifth equalities use (29), (30), and the local boundedness of f and f'. Since f is semismooth and Lemma 3.2 implies $|\mu_i - \lambda_i| \le \|h\|$, then clearly the right-hand side is of $o(\|h\|)$. For any $i, j \in \{1, \ldots, n\}$ with $i \ne j$, we have from (27) and (28) that

$$\begin{aligned}
\tilde{r}_{ij} &= -\sum_{k=1}^{n} o_{ki} o_{kj} f(\lambda_k) - c_{ij} \tilde{h}_{ij} \\
&= -\sum_{k=1}^{n} o_{ki} o_{kj} f(\lambda_k) + c_{ij} \sum_{k=1}^{n} o_{ki} o_{kj} \lambda_k \\
&= -(o_{ii} o_{ij} f(\lambda_i) + o_{ji} o_{jj} f(\lambda_j)) + c_{ij}(o_{ii} o_{ij} \lambda_i + o_{ji} o_{jj} \lambda_j) + O(\|h\|^2) \\
&= -\left((o_{ii} o_{ij} + o_{ji} o_{jj}) f(\lambda_i) + o_{ji} o_{jj} (f(\lambda_j) - f(\lambda_i)) \right) \\
&\quad + c_{ij} \left((o_{ii} o_{ij} + o_{ji} o_{jj}) \lambda_i + o_{ji} o_{jj} (\lambda_j - \lambda_i) \right) + O(\|h\|^2) \\
&= -o_{ji} o_{jj} \left(f(\lambda_j) - f(\lambda_i) - c_{ij} (\lambda_j - \lambda_i) \right) + O(\|h\|^2),
\end{aligned}$$

where the third and fifth equalities use (29), (31), and the local boundedness of f and f'. Thus, if $\lambda_i = \lambda_j$, the preceding relation yields

$$\tilde{r}_{ij} = O(\|h\|^2).$$

If $\lambda_i \neq \lambda_j$, then Lemma 3.2 implies $|\mu_i - \lambda_i| \leq \|h\|$ and $|\mu_j - \lambda_j| \leq \|h\|$ so that $|\mu_i - \mu_j| = |\lambda_i - \lambda_j - (\lambda_i - \mu_i) + (\lambda_j - \mu_j)| \geq |\lambda_i - \lambda_j| - 2\|h\| > 2\epsilon - 2\|h\| \geq 0$. Hence $\mu_i \neq \mu_j$, so $c_{ij} = (f(\mu_j) - f(\mu_i))/(\mu_j - \mu_i)$ and the preceding relation yields

$$\begin{aligned}\tilde{r}_{ij} &= -o_{ji}o_{jj}\left(f(\lambda_j) - f(\lambda_i) - \frac{f(\mu_j) - f(\mu_i)}{\mu_j - \mu_i}(\lambda_j - \lambda_i)\right) + O(\|h\|^2) \\ &= -o_{ji}o_{jj}\left(f(\lambda_j) - f(\lambda_i) - (f(\mu_j) - f(\mu_i))\left(1 + \frac{\lambda_j - \lambda_i - \mu_j + \mu_i}{\mu_j - \mu_i}\right)\right) + O(\|h\|^2) \\ &= O(\|h\|^2),\end{aligned}$$

where the last equality uses (29) and the strict continuity of f at λ_i, λ_j, so that $f(\mu_i) - f(\lambda_i) = O(|\mu_i - \lambda_i|) = O(\|h\|)$ and $f(\mu_j) - f(\lambda_j) = O(|\mu_j - \lambda_j|) = O(\|h\|)$.

Suppose f is ρ-order semismooth ($0 < \rho < \infty$). Then the preceding argument shows that $\tilde{r}_{ii} = O(\max\{\|h\|^{1+\rho}, \|h\|^2\}) = O(\|h\|^{1+\min\{1,\rho\}})$ for all i while we still have $\tilde{r}_{ij} = O(\|h\|^2)$ for all $i \neq j$. This shows that $f^{\square}$ is $\min\{1, \rho\}$-order semismooth at x. Since the choice of $x \in \mathcal{S}$ was arbitrary, $f^{\square}$ is $\min\{1, \rho\}$-order semismooth.

Suppose $f^{\square}$ is semismooth. Then $f^{\square}$ is strictly continuous and directionally differentiable. By Propositions 4.2 and 4.6, f is strictly continuous and directionally differentiable. For any $\xi \in \mathbb{R}$ and any $\eta \in \mathbb{R}$ such that f is differentiable at $\xi+\eta$, Proposition 4.3 yields that $f^{\square}$ is differentiable at $x + h$, where we denote $x := \operatorname{diag}[\xi, \ldots, \xi] = \xi I$ and $h := \operatorname{diag}[\eta, \ldots, \eta] = \eta I$. Since $f^{\square}$ is semismooth, it follows from Lemma 4.9 that

$$f^{\square}(x + h) - f^{\square}(x) - \nabla f^{\square}(x + h)h = o(\|h\|),$$

which, by (2) and (10), is equivalent to

$$f(\xi + \eta) - f(\xi) - f'(\xi + \eta)\eta = o(|\eta|).$$

Then Lemma 4.9 yields that f is semismooth. $\square$

We note that for each of the preceding global results there is a corresponding local result. This can be seen from our proofs where, in order to show that a global property of f is inherited by $f^{\square}$, we first show that this property is locally inherited from f by $f^{\square}$. For example, we can show the following local analogue of Proposition 4.4: If $f : \mathbb{R} \to \mathbb{R}$ is continuously differentiable at each of the eigenvalues of $x \in \mathcal{S}$, then $f^{\square}$ is continuously differentiable at x and $\nabla f^{\square}(x)$ is given by (10).

5. Applications to the SDCP. In this section, we consider the semidefinite complementarity problem (SDCP), which is to find, for a given function $F : \mathcal{S} \to \mathcal{S}$, an $(x, y) \in \mathcal{S} \times \mathcal{S}$ satisfying

$$x \in \mathcal{S}_+, \quad y \in \mathcal{S}_+, \quad \langle x, y \rangle = 0, \quad F(x) - y = 0, \tag{32}$$

where $\mathcal{S}_+$ denotes the convex cone comprising those $x \in \mathcal{S}$ that are positive semidefinite. We assume that F is continuously differentiable. The SDCP includes as a special case the nonlinear complementarity problem (NCP), where $n_1 = \cdots = n_m = 1$. It is also connected to eigenvalue optimization [18]. There has been much interest in the numerical solution of the SDCP (32) using, e.g., the interior-point approach [27], the merit function approach [30, 32], and the noninterior smoothing approach [8] (also see references therein). We will consider a related approach of reformulating the SDCP as a semismooth equation and then, by applying the results of section 4, study issues relevant to the design and analysis of smoothing Newton methods based on this reformulation.

It is known [30, Proposition 2.1] that $(x, y) \in \mathcal{S} \times \mathcal{S}$ solves the SDCP if and only if it solves the equations

$$H(x,y) := \begin{pmatrix} x - [x-y]_+ \\ F(x) - y \end{pmatrix} = 0, \tag{33}$$

where $[\cdot]_+ : \mathcal{S} \to \mathcal{S}_+$ denotes the nearest-point projection onto $\mathcal{S}_+$, i.e.,

$$[x]_+ := \arg\min\{\|x - y\| \mid y \in \mathcal{S}_+\}.$$

The function H is nonsmooth due to the nonsmoothness of the matrix projection operator $[\cdot]_+$. However, it was shown by Sun and Sun [29] that $[\cdot]_+$ is strongly semismooth, so that H is semismooth. We will see that this result also follows from Proposition 4.10 and, in particular, $f^{\square}(\cdot) = [\cdot]_+$ with $f(\cdot) = \max\{0, \cdot\}$ (Proposition 5.2).

There have been many smoothing methods proposed for solving semismooth equation reformulation of the NCP—see [2, 3, 4, 5, 6, 7, 11, 16, 22, 24] and references therein. These methods are based on making accurate smooth approximation of the semismooth equations. In particular, the smoothing method studied by Chen, Qi, and Sun [6] and later studied by Kanzow and Pieper [16] have an accuracy criterion called the Jacobian Consistence Property. We will verify this property with respect to a class of smoothing functions H_μ for H, as proposed by Chen and Mangasarian [4, 5] for the case of the linear program (LP) and the NCP and recently extended in [8] to the SDCP. This property, together with semismoothness of H, allows the development of methods of the form

$$(x^{k+1}, y^{k+1}) = (x^k, y^k) - t_k \nabla H_{\mu_k}(x^k, y^k)^{-1} H(x^k, y^k), \quad k = 0, 1, \ldots,$$

with $t_k > 0$ and $\mu_k \downarrow 0$ suitably chosen, that achieve both global convergence and local superlinear convergence, assuming nonsingularity of all $V \in \partial H(x,y)$ locally; see [6, Thm. 3.2]. Such methods have the advantage of requiring only one linear equation solve per iteration, in contrast to the two (or more) linear equation solves required by other smoothing methods having similar global and local convergence properties. Thus, our study paves the way for extending methods of the above form from the NCP to the SDCP. This, for example, would improve on the methods of [8, 15] which require two linear equation solves per iteration.

Let $\mathcal{CM}$ denote the class of convex continuously differentiable functions $g : \mathbb{R} \to \mathbb{R}$ with the properties

$$\lim_{\tau \to -\infty} g(\tau) = 0, \quad \lim_{\tau \to \infty} g(\tau) - \tau = 0, \quad \text{and} \quad 0 < g'(\tau) < 1 \ \forall \tau \in \mathbb{R}.$$

Two typical examples of g are the so-called *CHKS function* $g(\tau) = ((\tau^2+4)^{1/2} + \tau)/2$ and the *neural network function* $g(\tau) = \ln(e^\tau + 1)$. For any $g \in \mathcal{CM}$, consider the following smooth approximation of $x - [x-y]_+$, as proposed by Chen and Mangasarian [4, 5] for the case of the LP and the NCP:

$$\phi_\mu(x,y) := x - \mu g^{\square}((x-y)/\mu), \quad \mu > 0. \tag{34}$$

It was shown in [8, Lem. 1] that the limit $\lim_{\mu\to 0}\phi_\mu(x,y)$ exists and is equal to $x-[x-y]_+$. Moreover, one has [8, Cor. 1]

$$\|\phi_\mu(x,y)-(x-[x-y]_+)\| \le \sqrt{n}g(0)\mu, \tag{35}$$

and ϕ_μ is continuously differentiable for any $\mu > 0$ [8, Lem. 2]. Hence a smooth approximation of $H(x,y)$ is

$$H_\mu(x,y) := \begin{pmatrix} \phi_\mu(x,y) \\ F(x)-y \end{pmatrix}, \qquad \mu > 0. \tag{36}$$

We say that H_μ has the Jacobian Consistence Property relative to H if there exists a constant $\kappa > 0$ such that, for any $(x,y)\in\mathcal{S}\times\mathcal{S}$, we have (i)

$$\|H_\mu(x,y)-H(x,y)\| \le \kappa\mu \qquad \forall \mu > 0 \tag{37}$$

and (ii)

$$\lim_{\mu\to 0^+} \operatorname{dist}(\nabla H_\mu(x,y), \partial H(x,y)) = 0; \tag{38}$$

i.e., the distance between $\nabla H_\mu(x,y)$ and the set $\partial H(x,y)$ approaches zero as μ is decreased to zero. Here, we denote $\operatorname{dist}(L,\mathcal{M}) := \inf_{M\in\mathcal{M}} |||L-M|||$ for any linear mapping $L:\mathcal{S}\times\mathcal{S}\to\mathcal{S}\times\mathcal{S}$ and any nonempty collection $\mathcal{M}$ of linear mappings from $\mathcal{S}\times\mathcal{S}$ to $\mathcal{S}\times\mathcal{S}$. Also, for any $(x,y)\in\mathcal{S}\times\mathcal{S}$, we define $\|(x,y)\| = \sqrt{\|x\|^2+\|y\|^2}$. We show below that H is semismooth and H_μ has the Jacobian Consistence Property relative to H. These results facilitate the extension of the smoothing Newton methods of Chen, Qi, and Sun [6] for the NCP, later studied by Kanzow and Pieper [16], to the SDCP. Such methods are promising. For example, a smoothing method of [8], based on (34) and (36) with g being the CHKS function, is comparable to primal-dual interior-point methods in terms of the number of iterations to solve benchmark semidefinite programs with relative infeasibility and duality gap below $3\cdot 10^{-9}$. As with interior-point methods and barrier/penalty methods, the smoothing parameter μ needs to be small to obtain an accurate solution and, as μ becomes smaller, $\nabla H_\mu(x,y)$ can become more ill-conditioned. Thus, such smoothing methods could have difficulty achieving solution accuracy much greater than 10^{-9}.

We begin with the following lemma showing that the Jacobian Consistence Property is inherited by $f^{\square}$ and its smooth approximations from f and its smooth approximations.

LEMMA 5.1. *Let $f:\mathbb{R}\to\mathbb{R}$ be a strictly continuous function. Let $f_\mu:\mathbb{R}\to\mathbb{R}$, $\mu > 0$, be differentiable functions such that there exists a scalar constant $\kappa > 0$ for which*

$$|f_\mu(\zeta)-f(\zeta)| \le \kappa\mu \quad \forall \mu > 0, \tag{39}$$

$$\lim_{\mu\to 0^+} \operatorname{dist}(f'_\mu(\zeta), \partial f(\zeta)) = 0 \tag{40}$$

for all $\zeta\in\mathbb{R}$. Then, for any $z\in\mathcal{S}$, we have

$$\|f^{\square}_\mu(z)-f^{\square}(z)\| \le \sqrt{n}\kappa\mu \quad \forall \mu > 0, \tag{41}$$

$$\lim_{\mu\to 0^+} \operatorname{dist}(\nabla f^{\square}_\mu(z), \partial f^{\square}(z)) = 0. \tag{42}$$

Proof. Fix any $z \in \mathcal{S}$. Consider any $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ and any $p \in \mathcal{O}$ satisfying $z = p \operatorname{diag}[\lambda_1, \ldots, \lambda_n] p^T$.

By (1) and (2), we have

$$\begin{aligned} \|f_\mu^{\square}(z) - f^{\square}(z)\| &= \|p^T f_\mu^{\square}(z) p - p^T f^{\square}(z) p\| \\ &= \|\operatorname{diag}[f_\mu(\lambda_1) - f(\lambda_1), \ldots, f_\mu(\lambda_n) - f(\lambda_n)]\| \\ &\le \sqrt{n}\kappa\mu, \end{aligned}$$

where the last inequality uses (39). This proves (41).

We now prove (42). For any $\mu > 0$, since f_μ is differentiable, then Proposition 4.3 yields that $f_\mu^{\square}$ is differentiable and

$$\nabla f_\mu^{\square}(z) h = p(c_\mu \circ (p^T h p)) p^T \qquad \forall h \in \mathcal{S}, \tag{43}$$

where $c_\mu := f_\mu^{[1]}(\lambda)$ and $\lambda := (\lambda_1, \ldots, \lambda_n)^T$. Let $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_m$ denote the *distinct* eigenvalues of z and denote $\mathcal{I}_k := \{i \in \{1, \ldots, n\} | \lambda_i = \tilde{\lambda}_k\}$, $k = 1, \ldots, m$. We have

$$(c_\mu)_{ij} = \begin{cases} (f_\mu(\tilde{\lambda}_k) - f_\mu(\tilde{\lambda}_\ell))/(\tilde{\lambda}_k - \tilde{\lambda}_\ell) & \text{if } i \in \mathcal{I}_k,\ j \in \mathcal{I}_\ell \text{ for some } k \neq \ell, \\ f_\mu'(\tilde{\lambda}_k) & \text{if } i, j \in \mathcal{I}_k \text{ for some } k. \end{cases} \tag{44}$$

By (39) and (40), for each $\epsilon > 0$ there exists $\delta > 0$ such that for each $\mu \in (0, \delta)$ we have

$$|f_\mu(\tilde{\lambda}_k) - f(\tilde{\lambda}_k)| < \epsilon \quad \text{and} \quad |f_\mu'(\tilde{\lambda}_k) - v_k| < \epsilon, \qquad k = 1, \ldots, m, \tag{45}$$

for some $v_k \in \partial f(\tilde{\lambda}_k)$ depending on μ. Letting $c \in \mathcal{S}$ denote the symmetric matrix whose (i,j)th entry is

$$c_{ij} := \begin{cases} (f(\tilde{\lambda}_k) - f(\tilde{\lambda}_\ell))/(\tilde{\lambda}_k - \tilde{\lambda}_\ell) & \text{if } i \in \mathcal{I}_k,\ j \in \mathcal{I}_\ell \text{ for some } k \neq \ell, \\ v_k & \text{if } i, j \in \mathcal{I}_k \text{ for some } k, \end{cases} \tag{46}$$

we then obtain from (39), (44), (45), and (46) that

$$|(c_\mu)_{ij} - c_{ij}| < \epsilon\beta \qquad \forall i, j = 1, \ldots, n, \tag{47}$$

where $\beta > 0$ is a scalar independent of μ and ϵ. Define the linear mapping $V : \mathcal{S} \to \mathcal{S}$ by

$$Vh := p(c \circ (p^T h p)) p^T \qquad \forall h \in \mathcal{S}. \tag{48}$$

Then V depends on μ and, by (43) and (47), we have

$$|||\nabla f_\mu^{\square}(z) - V||| = \sup_{\|h\|=1} \|\nabla f_\mu^{\square}(z) h - Vh\| = \sup_{\|h\|=1} \|(c_\mu - c) \circ (p^T h p)\| < \epsilon\beta.$$

Thus $|||\nabla f_\mu^{\square}(z) - V||| \to 0$ as $\mu \to 0^+$. We now show that V belongs to $\partial f^{\square}(z)$. For each $k \in \{1, \ldots, m\}$, since $v_k \in \partial f(\tilde{\lambda}_k)$, there exist integer $\tau_k \ge 1$ and $\upsilon_k[\nu] \in \partial_B f(\tilde{\lambda}_k)$ and $\omega_k[\nu] \in (0, \infty)$, $\nu = 1, \ldots, \tau_k$, satisfying

$$\sum_{\nu=1}^{\tau_k} \omega_k[\nu] = 1, \qquad \sum_{\nu=1}^{\tau_k} \omega_k[\nu]\ \upsilon_k[\nu] = v_k.$$

Then, it is straightforward to verify that

$$\sum_{\nu_1=1}^{\tau_1}\cdots\sum_{\nu_m=1}^{\tau_m}\left(\prod_{k=1}^{m}\omega_k[\nu_k]\right)=1,\qquad \sum_{\nu_1=1}^{\tau_1}\cdots\sum_{\nu_m=1}^{\tau_m}\left(\prod_{k=1}^{m}\omega_k[\nu_k]\right)c[\nu_1,\ldots,\nu_m]=c,$$

where $c[\nu_1,\ldots,\nu_m]\in\mathcal{S}$ denotes the symmetric matrix whose (i,j)th entry is

$$c[\nu_1,\ldots,\nu_m]_{ij}:=\begin{cases}(f(\tilde\lambda_k)-f(\tilde\lambda_\ell))/(\tilde\lambda_k-\tilde\lambda_\ell) & \text{if } i\in\mathcal{I}_k,\ j\in\mathcal{I}_\ell \text{ for some } k\neq\ell,\\ \upsilon_k[\nu_k] & \text{if } i,j\in\mathcal{I}_k \text{ for some } k.\end{cases}$$

We now show that the linear mapping $V[\nu_1,\ldots,\nu_m]:\mathcal{S}\to\mathcal{S}$ defined by

$$V[\nu_1,\ldots,\nu_m]h:=p(c[\nu_1,\ldots,\nu_m]\circ(p^Thp))p^T\qquad\forall h\in\mathcal{S}$$

belongs to $\partial_B f^{\square}(z)$. For each $k\in\{1,\ldots,m\}$, since $\upsilon_k[\nu_k]\in\partial_B f(\tilde\lambda_k)$, there exist $\tilde\lambda_{kl}\in\mathbb{R}$, $l=1,2,\ldots$, such that f is differentiable at $\tilde\lambda_{kl}$ for all l and $\tilde\lambda_{kl}\to\tilde\lambda_k$ and $f'(\tilde\lambda_{kl})\to\upsilon_k[\nu_k]$ as $l\to\infty$. Then, letting

$$z_l:=p\ \mathrm{diag}[\lambda_{1l},\ldots,\lambda_{nl}]p^T\qquad\text{with}\qquad \lambda_{il}:=\tilde\lambda_{kl}\quad\forall i\in\mathcal{I}_k,\ k=1,\ldots,m,$$

for $l=1,2,\ldots$, we have from Proposition 4.3 that $f^{\square}$ is differentiable at z_l. Moreover, as $l\to\infty$, we have $z_l\to z$ and

$$\begin{aligned}|||\nabla f^{\square}(z_l)-V[\nu_1,\ldots,\nu_m]|||&=\sup_{\|h\|=1}\|\nabla f^{\square}(z_l)h-V[\nu_1,\ldots,\nu_m]h\|\\&=\sup_{\|h\|=1}\|(f^{[1]}(\lambda_{1l},\ldots,\lambda_{nl})-c[\nu_1,\ldots,\nu_m])\circ(p^Thp)\|\ \to\ 0.\end{aligned}$$

Hence $V[\nu_1,\ldots,\nu_m]\in\partial_B f(z)$. ◻

By using Lemma 5.1 together with Proposition 4.10, we can now establish the main result of this section. Part (a) of this result was already shown in [29]. Here we show that it also follows from Proposition 4.10.

PROPOSITION 5.2. *For the functions H and H_μ defined by* (33) *and* (36) *with $g\in\mathcal{CM}$, respectively, the following results hold.*

(a) *H is semismooth. If F is ρ-order semismooth* ($0<\rho<\infty$)*, then H is $\min\{1,\rho\}$-order semismooth.*

(b) *H_μ has the Jacobian Consistence Property relative to H.*

Proof. Let

$$f(\zeta):=\max\{0,\zeta\},\qquad f_\mu(\zeta):=\mu g(\zeta/\mu)\qquad\forall\zeta\in\mathbb{R}. \tag{49}$$

(a) It was shown in [30, Lem. 2.1] that

$$f^{\square}(z)=[z]_+\qquad\forall z\in\mathcal{S}.$$

Also, it is well known that f is piecewise linear on $\mathbb{R}$ and hence f is strongly semismooth. Then, by Proposition 4.10, $f^{\square}$ is strongly semismooth. It is known that the composition of two ρ-order semismooth functions is also ρ-order semismooth [10, Thm. 19]. Hence the composite function $(x,y)\mapsto f^{\square}(x-y)=[x-y]_+$ is strongly semismooth. Since F is semismooth, then H is semismooth. If F is ρ-order semismooth ($0<\rho<\infty$), then H is $\min\{1,\rho\}$-order semismooth.

(b) It can be seen from (33), (35), and (36) that (37) is satisfied with $\kappa := \sqrt{n}g(0)$. Alternatively, this can be deduced by applying Lemma 5.1 and using (49). We now prove (38). It is readily seen from (49) and properties of g (see, e.g., [31]) that

$$\lim_{\mu\to 0^+} f'_\mu(\zeta) = \lim_{\mu\to 0^+} g'(\zeta/\mu) = \begin{cases} g'(0) & \text{if } \zeta = 0, \\ 1 & \text{if } \zeta > 0, \\ 0 & \text{if } \zeta < 0, \end{cases} \qquad \partial f(\zeta) = \begin{cases} [-1,1] & \text{if } \zeta = 0, \\ \{1\} & \text{if } \zeta > 0, \\ \{0\} & \text{if } \zeta < 0. \end{cases}$$

Since $g'(0) \in (0,1)$, this shows that (40) holds for all $\zeta \in \mathbb{R}$. Thus, by Lemma 5.1, (42) holds for all $z \in \mathcal{S}$. Fix any $x, y \in \mathcal{S}$. It can be seen from (33) and $f^{\square}(\cdot) = [\cdot]_+$ that

$$B \in \partial H(x,y) \quad \text{if and only if} \quad B = \begin{bmatrix} I - V & V \\ \nabla F(x) & -I \end{bmatrix} \text{ for some } V \in \partial f^{\square}(x-y).$$

Also, we have from (34) and (36) that

$$\nabla H_\mu(x,y) = \begin{bmatrix} I - \nabla f^{\square}_\mu(x-y) & \nabla f^{\square}_\mu(x-y) \\ \nabla F(x) & -I \end{bmatrix}.$$

Thus

$$\begin{aligned} \operatorname{dist}(\nabla H_\mu(x,y), \partial H(x,y)) &= \min_{V \in \partial f^{\square}(x-y)} \left\{ \max_{\|(u,v)\|=1} \|(\nabla f^{\square}_\mu(x-y) - V)(u-v)\| \right\} \\ &\le \sqrt{2}\ \operatorname{dist}(\nabla f^{\square}_\mu(x-y), \partial f^{\square}(x-y)) \\ &\to 0 \qquad \text{as } \ \mu \to 0^+, \end{aligned}$$

where the last relation follows from (42) with $z = x - y$. This verifies (38). ∎

We note that, for the particular choice (49) of f and f_μ, we can obtain an explicit formula for c given by (46) and directly verify that V given by (48) belongs to $\partial f^{\square}(z)$. Specifically, for any $z \in \mathcal{S}$ and any $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ and $p \in \mathcal{O}$ satisfying $z = p\ \mathrm{diag}[\lambda_1, \ldots, \lambda_n]p^T$, define the three index sets

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

Upon taking $\mu \to 0^+$ in (44) and using (49) and properties of g [31], we obtain in the limit that the (i,j)th entry of c is given by

$$c_{ij} = \lim_{\mu\to 0^+} (c_\mu)_{ij} = \begin{cases} 1 & \text{if } \ i,j \in \alpha, \\ 1 & \text{if } \ i \in \alpha, j \in \beta \ \text{ or } \ i \in \beta, j \in \alpha, \\ \lambda_i/(\lambda_i - \lambda_j) & \text{if } \ i \in \alpha, j \in \gamma, \\ \lambda_j/(\lambda_j - \lambda_i) & \text{if } \ i \in \gamma, j \in \alpha, \\ g'(0) & \text{if } \ i,j \in \beta, \\ 0 & \text{else}. \end{cases} \tag{50}$$

To see that V given by (48) belongs to $\partial f^{\square}(z)$, let ϵ_l, $l = 1, 2, \ldots$, be any sequence of positive scalars converging to 0, and define for $\sigma = -1, 1$ and $l = 1, 2, \ldots$ the symmetric matrix

$$z_l[\sigma] := z + \sigma\epsilon_l p\ \mathrm{diag}[d_1, \ldots, d_n]p^T, \qquad \text{with} \quad d_i := \begin{cases} 1 & \text{if } i \in \beta, \\ 0 & \text{else}. \end{cases}$$

For each $\sigma \in \{-1, 1\}$, it can be seen that the eigenvalues of $z_l[\sigma]$ are $\lambda_{il}[\sigma] := \lambda_i + \sigma\epsilon_l d_i$, $i = 1, \ldots, n$, which are nonzero for all l sufficiently large. Thus, f is differentiable at $\lambda_{il}[\sigma]$, $i = 1, \ldots, n$, for all l sufficiently large. Hence, by Proposition 4.3, $f^{\square}$ is differentiable at $z_l[\sigma]$ for all l sufficiently large and

$$\nabla f^{\square}(z_l[\sigma])h = p(c_l[\sigma] \circ (p^T hp))p^T \qquad \forall h \in \mathcal{S},$$

where $c_l[\sigma] := f^{[1]}(\lambda_{1l}[\sigma], \ldots, \lambda_{nl}[\sigma]) \in \mathcal{S}$. Using (49), it can be seen that, as $l \to \infty$, $z_l[\sigma] \to z$ and $c_l[\sigma]$ converges entrywise to $c[\sigma]$ whose (i,j)th entry is

$$(c[\sigma])_{ij} := \begin{cases} 1 & \text{if } i, j \in \alpha, \\ 1 & \text{if } i \in \alpha, j \in \beta \text{ or } i \in \beta, j \in \alpha, \\ \lambda_i/(\lambda_i - \lambda_j) & \text{if } i \in \alpha, j \in \gamma, \\ \lambda_j/(\lambda_j - \lambda_i) & \text{if } i \in \gamma, j \in \alpha, \\ \max\{0, \sigma\} & \text{if } i, j \in \beta, \\ 0 & \text{else.} \end{cases} \tag{51}$$

Hence $\nabla f^{\square}(z_l[\sigma])$ converges in operator norm to $V[\sigma] : \mathcal{S} \to \mathcal{S}$ defined by

$$V[\sigma]h := p(c[\sigma] \circ (p^T hp))p^T \qquad \forall h \in \mathcal{S}.$$

By the definition of $\partial_B f^{\square}(z)$, we see that $V[\sigma] \in \partial_B f^{\square}(z)$. Moreover, (50) and (51) show that $c = g'(0)c[-1] + (1 - g'(0))c[1]$, and hence $V = g'(0)V[-1] + (1 - g'(0))V[1]$. This shows that $V \in \partial f^{\square}(z)$.

6. Final remarks. In this paper, we studied various continuity and differentiability properties of a class of symmetric-matrix-valued functions, which are natural extensions of real-valued functions to matrix-valued functions. Using these properties, we reformulated the SDCP as a semismooth equation based on the matrix projection operator $[\cdot]_+$. We verified the Jacobian Consistence Property for the reformulated semismooth equation and its smooth approximation based on a class of smoothing functions proposed by Chen and Mangasarian [4, 5] for the LP and NCP and extended in [8] to the SDCP. This result facilitates the extension of the smoothing method studied in [6] and [16] for the NCP to the SDCP. We stress that, apart from the Jacobian Consistence Property, there are other important issues in extending the smoothing method of [6] to the SDCP. One of them is the solvability of the smoothing Newton equations. We leave this issue for future research.

REFERENCES

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

[2] B. Chen and X. Chen, *A global and local superlinear continuation-smoothing method for P_0 and R_0 NCP or monotone NCP*, SIAM J. Optim., 9 (1999), pp. 624–645.

[3] B. Chen and P.T. Harker, *Smoothing approximations to nonlinear complementarity problems*, SIAM J. Optim., 7 (1997), pp. 403–420.

[4] C. Chen and O.L. Mangasarian, *Smoothing methods for convex inequalities and linear complementarity problems*, Math. Programming, 71 (1995), pp. 51–69.

[5] C. Chen and O.L. Mangasarian, *A class of smoothing functions for nonlinear and mixed complementarity problems*, Comput. Optim. Appl., 5 (1996), pp. 97–138.

[6] X. Chen, L. Qi, and D. Sun, *Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities*, Math. Comp., 67 (1998), pp. 519–540.

[7] X. Chen and Y. Ye, *On homotopy-smoothing methods for box-constrained variational inequalities*, SIAM J. Control Optim., 37 (1999), pp. 589–616.

[8] X. Chen and P. Tseng, *Non-interior continuation methods for solving semidefinite complementarity problems*, Math. Programming, to appear.

[9] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.

[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, Boston, 1999.

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

[13] R.A. Horn and C.R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1991.

[14] H. Jiang and D. Ralph, *Global and local superlinear convergence analysis of Newton-type methods for semismooth equations with smooth least squares*, in Reformulation—Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, M. Fukushima and L. Qi, eds., Kluwer Academic Publishers, Boston, 1999, pp. 181–209.

[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] C. Kanzow and H. Pieper, *Jacobian smoothing methods for nonlinear complementarity problems*, SIAM J. Optim., 9 (1999), pp. 342–373.

[17] T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag, Berlin, 1984.

[18] A.S. Lewis and M.L. Overton, *Eigenvalue optimization*, Acta Numer., 5 (1996), pp. 149–190.

[19] A.S. Lewis and H.S. Sendov, *Twice differentiable spectral functions*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 368–386.

[20] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optim., 15 (1977), pp. 959–972.

[21] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res., 18 (1993), pp. 227–244.

[22] H.-D. Qi, *A regularized smoothing Newton method for box constrained variational inequality problems with P_0-functions*, SIAM J. Optim., 10 (1999), pp. 315–330.

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

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

[25] F. Rellich, *Perturbation Theory of Eigenvalue Problems*, Gordon and Breach, New York, 1969.

[26] R.T. Rockafellar and R.J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.

[27] M. Shida and S. Shindoh, *Monotone Semidefinite Complementarity Problems*, Research Report 312, Department of Mathematical and Computer Sciences, Tokyo Institute of Technology, Tokyo, 1996.

[28] G.W. Stewart and J.-G. Sun, *Matrix Perturbation Theory*, Academic Press, New York, 1990.

[29] D. Sun and J. Sun, *Semismooth matrix valued functions*, Math. Oper. Res., 27 (2002), pp. 150–169.

[30] P. Tseng, *Merit functions for semi-definite complementarity problems*, Math. Programming, 83 (1998), pp. 159–185.

[31] P. Tseng, *Analysis of a non-interior continuation method based on Chen-Mangasarian smoothing functions for complementarity problems*, in Reformulation—Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, M. Fukushima and L. Qi, eds., Kluwer Academic Publishers, Boston, 1999, pp. 381–404.

[32] N. Yamashita and M. Fukushima, *A new merit function and a descent method for semidefinite complementarity problems*, in Reformulation—Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, M. Fukushima and L. Qi, eds., Kluwer Academic Publishers, Boston, 1999, pp. 405–420.