SIAM J. MATRIX ANAL. APPL.
Vol. 25, No. 3, pp. 766–783

SEMISMOOTHNESS OF SPECTRAL FUNCTIONS*

HOUDUO QI† AND XIAOQI YANG‡

Abstract. Any spectral function can be written as a composition of a symmetric function $f : \mathbb{R}^n \mapsto \mathbb{R}$ and the eigenvalue function $\lambda(\cdot) : \mathcal{S} \mapsto \mathbb{R}^n$, often denoted by $(f \circ \lambda)$, where $\mathcal{S}$ is the subspace of $n \times n$ symmetric matrices. In this paper, we present some nonsmooth analysis for such spectral functions. Our main results are (a) $(f \circ \lambda)$ is directionally differentiable if f is semidifferentiable, (b) $(f \circ \lambda)$ is LC^1 if and only if f is LC^1, and (c) $(f \circ \lambda)$ is SC^1 if and only if f is SC^1. Result (a) is complementary to a known (negative) fact that $(f \circ \lambda)$ might not be directionally differentiable if f is directionally differentiable only. Results (b) and (c) are particularly useful for the solution of LC^1 and SC^1 minimization problems which often can be solved by fast (generalized) Newton methods. Our analysis makes use of recent results on continuously differentiable spectral functions as well as on nonsmooth symmetric–matrix-valued functions.

Key words. symmetric function, spectral function, nonsmooth analysis, semismooth function

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

DOI. 10.1137/S0895479802417921

1. Introduction.

There has been growing interest in the variational analysis of spectral functions. This growing trend is probably due to the following reasons. On one hand, spectral functions have important applications to some fundamental problems in applied mathematics such as semidefinite programs and engineering problems. See a survey paper by Lewis and Overton [14] for many such applications. On the other hand, efficient nonsmooth analysis tools have only been available in the past few years; see the book by Rockafellar and Wets [26]. In this paper, we study some nonsmooth properties of spectral functions which have not been reported in the literature. Our study is inspired by recent progress on spectral functions [13, 15, 16] and progress on symmetric–matrix-valued functions [2, 27, 3, 28, 11].

Let $\mathcal{S}$ be the space of $n \times n$ real symmetric matrices endowed with the inner product $\langle X, Y \rangle := \mathrm{trace}(XY)$ for any $X, Y \in \mathcal{S}$. $\|X\|$ is the Frobenius norm of X. Let $\lambda(\cdot) : \mathcal{S} \to \mathbb{R}^n$ be the eigenvalue function such that $\lambda_i(X)$, $i = 1, \ldots, n$, yield eigenvalues of X for any $X \in \mathcal{S}$ and are patterned in nonincreasing order, i.e., $\lambda_1(X) \geq \cdots \geq \lambda_n(X)$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is symmetric on an open set $\Omega \subseteq \mathbb{R}^n$ if f is invariant under coordinate permutation, i.e.,

$$f(x) = f(Px) \quad \text{for any permutation matrix } P \text{ and any } x \in \Omega.$$

For simplicity, we assume that Ω is $\mathbb{R}^n$ in this paper (all results remain valid when restricted to some open symmetric set Ω). Formally, a *spectral function* is a composition of a symmetric function $f : \mathbb{R}^n \to \mathbb{R}$ and the eigenvalue function $\lambda(\cdot) : \mathcal{S} \to \mathbb{R}^n$;

*Received by the editors November 15, 2002; accepted for publication (in revised form) by M. L. Overton July 28, 2003; published electronically February 24, 2004.

http://www.siam.org/journals/simax/25-3/41792.html

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

‡Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (maqilq@polyu.edu.hk). The research of this author was supported by the Research Grants Council of the Hong Kong Special Administrative Region, China (project PolyU 5141/01E).

that is, the spectral function $(f \circ \lambda) : \mathcal{S} \to \mathbb{R}$ is given by

$$(f \circ \lambda)(X) := f(\lambda(X)), \qquad X \in \mathcal{S}.$$

For more explanation leading to this definition, see [16]. Typical spectral functions include the kth largest eigenvalue of a symmetric matrix [14, 15] and the Schatten p-norm of a symmetric matrix ($p \geq 1$).

It is well known that the eigenvalue function $\lambda(\cdot)$ is not everywhere differentiable. So it is natural to expect that the composite function $(f \circ \lambda)$ could be not everywhere differentiable no matter how smooth f is. It was therefore surprising when Lewis claimed in [13] that $(f \circ \lambda)$ is indeed (strictly) differentiable at $X \in \mathcal{S}$ if and only if f is (strictly) differentiable at $\lambda(X)$. Moreover, it is further proved in [16] that $(f \circ \lambda)$ is twice (continuously) differentiable at $X \in \mathcal{S}$ if and only if f is twice (continuously) differentiable at $\lambda(X)$. Those two results on derivatives play an important role in this paper. It is also known that $(f \circ \lambda)$ is convex if and only if f is convex [5]. Since the eigenvalue function is Lipschitz continuous, $(f \circ \lambda)$ is locally Lipschitzian if f is. Then the generalized gradient $\partial(f \circ \lambda)$ in the sense of Clarke [4] is well defined. A beautiful formula for calculating elements in $\partial(f \circ \lambda)$ can be found in [13]. Several other subgradients of $(f \circ \lambda)$ are studied in [15]; see also [8].

The above results show that $(f \circ \lambda)$ inherits smoothness properties from f. However, this is not the case for directional differentiability. The punctured hyperbola example constructed by Lewis [13] shows that $(f \circ \lambda)$ is not necessarily directionally differentiable if f is directionally differentiable only. We will show that a sufficient condition for directional differentiability of $(f \circ \lambda)$ at $X \in \mathcal{S}$ is the semidifferentiability of f at $\lambda(X)$ (see Proposition 3.2). This result suggests that f should have differentiability properties stronger than directional differentiability in order for $(f \circ \lambda)$ to inherit the same properties from f. In fact, we will show that $(f \circ \lambda)$ is $\min(1, \rho)$-order semismooth if and only if f is ρ-order semismooth (see Proposition 3.5), generalizing a recent result of Sun and Sun [28] which proves that the eigenvalue function is strongly semismooth. As mentioned earlier, $(f \circ \lambda)$ is (twice) differentiable if and only if f is (twice) differentiable. We are also interested in the case when f is an LC^1 function (also called a $C^{1,1}$ function in the literature), i.e., f is once continuously differentiable and its derivative function $\nabla f(\cdot)$ is locally Lipschitz. Another interesting case is when f is an SC^1 function, i.e., f is not only an LC^1 function, but also its derivative function is *semismooth*. For both cases, we will show that $(f \circ \lambda)$ is an LC^1 (respectively, SC^1) function (see Propositions 4.3 and 4.5). The importance of LC^1 and SC^1 functions is that they constitute a class of minimization problems which can be solved by Newton-type methods (see [6, 20, 22]) and by penalty-type methods (see [31, 30]).

The property of semismoothness, as introduced by Mifflin [17, 18] 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 [10, 21, 23]. Recent attention in research on semismoothness is on symmetric–matrix-valued functions which have important applications to semidefinite complementarity problems [29, 27, 2, 3, 28, 11]. Several important results have been established and inspired our research in this paper. For example, the absolute matrix-valued function $|X| := \sqrt{X^2}$, $X \in \mathcal{S}$, is strongly semismooth [27, 3]; the eigenvalue function $\lambda(\cdot)$ is strongly semismooth [28]. This latter result is found to be particularly useful in quadratic convergence analysis of Newton methods for inverse eigenvalue problems. Another useful result is a lemma of Chen and Tseng [2] about the locally upper Lipschitzian property of certain orthogonal matrices yielding the spectral decomposition of a symmetric matrix.

Notation used in this paper is as follows: vectors in $\mathbb{R}^n$ are viewed as columns and capital letters such as X, Y, etc. always denote matrices in $\mathcal{S}$. For $X \in \mathcal{S}$, we denote by X_{ij} the (i,j)th entry of X. We use $\circ$ to denote the Hadamard product between two matrices, i.e.,

$$X \circ Y = [X_{ij} Y_{ij}]_{i,j=1}^n.$$

Let the operator diag : $\mathcal{S} \to \mathbb{R}^n$ be defined by $\mathrm{diag}[X] := (X_{11}, \ldots, X_{nn})^T$, while for $\mu \in \mathbb{R}^n$, $\mathrm{Diag}[\mu_1, \ldots, \mu_n]$ denotes the diagonal matrix with its ith diagonal entry μ_i. Sometimes we write $\mathrm{Diag}[\mu]$ instead of $\mathrm{Diag}[\mu_1, \ldots, \mu_n]$ for simplicity. Let $\mathcal{P}$ denote the set of all permutation matrices in $\mathbb{R}^{n\times n}$. For any given $\mu \in \mathbb{R}^n$, $\mathcal{P}_\mu$ denotes the stabilizer of μ defined by

$$\mathcal{P}_\mu := \{P \in \mathcal{P} | \; P\mu = \mu\}.$$

Throughout, $\|\cdot\|$ denotes the Frobenius norm for matrices and the 2-norm for vectors. For any linear mapping $\mathcal{L} : \mathcal{S} \to \mathcal{S}$, we define its operator norm $\||\mathcal{L}|\| := \max_{\|X\|=1} \|\mathcal{L}X\|$. For any $x \in \mathbb{R}^n$, $X \in \mathcal{S}$, and any scalar $\gamma > 0$, we denote the γ-ball around x in $\mathbb{R}^n$ and the γ-ball around X in $\mathcal{S}$, respectively, by

$$\mathcal{N}(x, \gamma) := \{y \in \mathbb{R}^n | \; \|y - x\| \le \gamma\},$$
$$\mathcal{B}(X, \gamma) := \{Y \in \mathcal{S} | \; \|Y - X\| \le \gamma\}.$$

For any $\mu \in \mathbb{R}^n$ and $P \in \mathcal{P}$, we will frequently use the following fact:

$$\mathrm{Diag}[P\mu] = P\mathrm{Diag}[\mu]P^T.$$

2. Miscellaneous. In this section, we review some basic concepts on continuity and differentiability of vector-valued functions in order to avoid confusion with other concepts not treated in this paper. Those concepts also apply to the spectral function $(f \circ \lambda)$ and its gradient map $\nabla(f \circ \lambda)$ (if it exists) since the symmetric matrix space $\mathcal{S}$ can be cast as a vector space of dimension $n(n+1)/2$. All those concepts except semismoothness and their equivalent characterizations can be found in the book [26]. We also list some perturbation results on symmetric matrices for later use.

2.1. Basic concepts. Consider the mapping $F : \mathbb{R}^k \mapsto \mathbb{R}^\ell$. We say F is continuous at $x \in \mathbb{R}^k$ if $F(y) \to F(x)$ as $y \to x$, 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

$$F'(x;h) \ := \ \lim_{\tau \to 0^+} \frac{F(x+\tau h) - F(x)}{\tau} \quad \text{exists} \qquad \forall h \in \mathbb{R}^k,$$

and F is directionally differentiable if F is directionally differentiable at every $x \in \mathbb{R}^k$. We say F is semidifferentiable at $x \in \mathbb{R}^k$ if the limit

$$\lim_{\substack{\tau \searrow 0 \\ \hat{h} \to h}} \frac{F(x+\tau \hat{h}) - F(x)}{\tau}$$

exists for every direction $h \in \mathbb{R}^n$. It is easy to see that the limit (if it exists) equals $F'(x;h)$. F is differentiable (in the Fréchet sense) at $x \in \mathbb{R}^k$ if there exists a linear mapping $\nabla F(x) : \mathbb{R}^k \mapsto \mathbb{R}^\ell$ 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 [4] and [26, Sec. 9J]. Then the generalized Jacobian $\partial F(x)$ of F at x (in the Clarke sense) is well defined.

Definition 2.1 (semismoothness). *Suppose that $F : \mathbb{R}^k \to \mathbb{R}^\ell$ is a strictly continuous function. F is said to be* semismooth *at $x \in \mathbb{R}^k$ if F is directionally differentiable at x and for any $V \in \partial F(x+h)$,*

$$F(x+h) - F(x) - Vh = o(\|h\|). \tag{1}$$

F is said to be ρ-order semismooth $(0 < \rho < \infty)$ at x if F is semismooth at x and

$$F(x+h) - F(x) - Vh = O(\|h\|^{1+\rho}). \tag{2}$$

In particular, F is called strongly semismooth at x if F is 1*-order semismooth at x.*

We say F is semismooth (respectively, ρ-order semismooth) if F is semismooth (respectively, ρ-order semismooth) at every $x \in \mathbb{R}^k$. 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 characterization below obtained by Sun and Sun [27, Thm. 3.7] provides a convenient way for proving ρ-order semismoothness and semismoothness as well. For more applications of this result, see [3, 28].

Lemma 2.2. *Suppose that $F : \mathbb{R}^k \to \mathbb{R}^\ell$ is strictly continuous and directionally differentiable in a neighborhood of x. Then for any $\rho \in (0, \infty)$ the following two statements are equivalent:*

(a) *for any $V \in \partial F(x+h)$,*

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

(b) *for any $h \in \mathbb{R}^k$ such that F is differentiable at $x+h$,*

$$F(x+h) - F(x) - \nabla F(x+h)h = O(\|h\|^{1+\rho}). \tag{3}$$

In particular, the following two statements are equivalent:

(c) *for any* $V \in \partial F(x+h)$,

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

(d) *for any* $h \in \mathbb{R}^k$ *such that* F *is differentiable at* $x+h$,

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

Finally we assume that $F : \mathbb{R}^k \mapsto \mathbb{R}^\ell$ is continuously differentiable. We say that F is an LC^1 function if ∇F is strictly continuous, and that F is an SC^1 function if F is an LC^1 function and ∇F is semismooth. For more discussion on LC^1 and SC^1 functions and their roles in superlinear convergence analysis of certain generalized Newton methods for some minimization problems, see [22, 20, 6]. We note that the LC^1 problem is also known as $C^{1,1}$ data in [9], where second-order analysis of the underlying function is conducted. For further development along this line, see [30, 31] and the references therein.

2.2. Perturbation results for symmetric matrices. In this subsection, 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 spectral function $(f \circ \lambda)$.

Let $\mathcal{O}$ denote the group of $n \times n$ real orthogonal matrices. For each $X \in \mathcal{S}$, define the set of orthogonal matrices giving the ordered spectral decomposition of X by

$$\mathcal{O}_X := \{P \in \mathcal{O} |\ P^T X P = \mathrm{Diag}[\lambda(X)]\}.$$

Clearly $\mathcal{O}_X$ is nonempty for each $X \in \mathcal{S}$. The following lemma, proved in [2, Lem. 3], gives a key perturbation result for eigenvectors of symmetric matrices. For a different yet simple proof of this lemma, see [28].

LEMMA 2.3. *For any* $X \in \mathcal{S}$, *there exist scalars* $\eta > 0$ *and* $\epsilon > 0$ *such that*

$$\min_{P \in \mathcal{O}_X} \|P - Q\| \le \eta \|X - Y\| \quad \forall\ Y \in \mathcal{B}(X, \epsilon),\ \forall Q \in \mathcal{O}_Y. \tag{4}$$

We will also need the following perturbation results of von Neumann [19]; see also [1].

LEMMA 2.4. *For any* $X, Y \in \mathcal{S}$, *we have*

$$\|\lambda(X) - \lambda(Y)\| \le \|X - Y\| \quad \textit{and} \quad |\lambda_i(X) - \lambda_i(Y)| \le \|X - Y\|_2 \quad \forall\ i = 1, \ldots, n,$$

where $\|\cdot\|_2$ *is the* 2*-norm.*

Last, 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 $X + \tau H$ may be chosen to be analytic in τ. As is remarked in [12, p. 122], the existence of such orthonormal eigenvectors depending smoothly on τ is one of the most remarkable results in the analytic perturbation theory for symmetric operators.

LEMMA 2.5. *For any* $X \in \mathcal{S}$ *and any* $H \in \mathcal{S}$, *there exist* $P(\tau) \in \mathcal{O}$, $\tau \in \mathbb{R}$, *whose entries are power series in* τ, *convergent in a neighborhood of* $\tau = 0$, *and* $P(\tau)^T (X + \tau H) P(\tau)$ *is diagonal.*

3. Directional differentiability and semismoothness of spectral functions. This section includes two main results. Proposition 3.2 says that the spectral function $(f \circ \lambda)$ is directionally differentiable if f is semidifferentiable. Without this condition, the punctured hyperbola example [13] shows that $(f \circ \lambda)$ is not necessarily directionally differentiable. Proposition 3.5 says that $(f \circ \lambda)$ inherits semismoothness from f. The following preliminary results, which shall be used from time to time in our proofs, are due to the symmetry of f. For example, parts (a), (c), and (d) of Lemma 3.1 follow from differentiating both sides of the equality $f(\mu) = f(P\mu)$ $(P \in \mathcal{P})$ and the chain rule. Part (b) is a direct consequence from the definition of semidifferentiability and the symmetry of f.

LEMMA 3.1. *Suppose $f : \mathbb{R}^n \mapsto \mathbb{R}$ is symmetric. Then we have the following results:*

(a) *f is directionally differentiable at $\mu \in \mathbb{R}^n$ along $h \in \mathbb{R}^n$ if and only if f is directionally differentiable at $P\mu$ along Ph for any $P \in \mathcal{P}$.*

(b) *f is semidifferentiable at $\mu \in \mathbb{R}^n$ if and only if f is semidifferentiable at $P\mu$ for any $P \in \mathcal{P}$.*

(c) *f is differentiable at $\mu \in \mathbb{R}^n$ if and only if f is differentiable at $P\mu$ for any $P \in \mathcal{P}$. In particular, $\nabla f(P\mu) = P\nabla f(\mu)$. Moreover, if $P \in \mathcal{P}_\mu$, then $\nabla f(\mu) = P\nabla f(\mu)$. Consequently, $(\nabla f(\mu))_i = (\nabla f(\mu))_j$ if $\mu_i = \mu_j$ for some $i, j \in \{1, \ldots, n\}$.*

(d) *f is twice differentiable at $\mu \in \mathbb{R}^n$ if and only if f is twice differentiable at $P\mu$ for any $P \in \mathcal{P}$. In this case we have $\nabla^2 f(P\mu) = P\nabla^2 f(\mu)P^T$.*

The next result states that under the condition of semidifferentiability the directional differentiability of f is inherited by the spectral function $(f \circ \lambda)$. Without this condition, this result is no longer valid as the punctured hyperbola example in [13, p. 587] illustrates.

PROPOSITION 3.2. *Let $X \in \mathcal{S}$ be given. The following results hold.*

(a) *Suppose that f is semidifferentiable at $\lambda(X)$. Then $(f \circ \lambda)$ is directionally differentiable at X.*

(b) *Conversely, if $(f \circ \lambda)$ is directionally differentiable at X, then f is directionally differentiable at $\lambda(X)$.*

(c) *Suppose that f is both strictly continuous and directionally differentiable at $\lambda(X)$. Then $(f \circ \lambda)$ is directionally differentiable at X.*

Proof. (a) Let $H \in \mathcal{S}$ and define

$$X(\tau) = X + \tau H, \quad \tau \in \mathbb{R}.$$

Then by Lemma 2.5 there exists $P(\tau) \in \mathcal{O}$, $\tau \in \mathbb{R}$, whose entries are power series in τ, convergent in a neighborhood $\mathcal{I}$ of $\tau = 0$, and $P^T(\tau)X(\tau)P(\tau)$ is diagonal. Consequently the corresponding eigenvalues

$$\mu_i(\tau) = [P^T(\tau)X(\tau)P(\tau)]_{ii}, \quad i = 1, \ldots, n,$$

are also power series in τ, convergent for $\tau \in \mathcal{I}$. Denote

$$\mu(\tau) := (\mu_1(\tau), \ldots, \mu_n(\tau))^T.$$

Then we have the expansion

$$\mu(\tau) = \mu(0) + \tau\mu'(0) + o(\tau). \tag{5}$$

The fact that the elements of $\mu(\tau)$ are eigenvalues of $X(\tau)$ yields

$$\lim_{\tau \searrow 0} \frac{(f \circ \lambda)(X + \tau H) - (f \circ \lambda)(X)}{\tau}$$

$$= \lim_{\tau \searrow 0} \frac{f(\mu(\tau)) - f(\mu(0))}{\tau}$$
$$= \lim_{\tau \searrow 0} \frac{f(\mu(0) + \tau\mu'(0) + o(\tau)) - f(\mu(0))}{\tau}$$
$$= f'(\mu(0); \mu'(0)),$$

where the last equality uses the semidifferentiability of f at $\lambda(X)$. This proves that $(f \circ \lambda)$ is directionally differentiable at X.

(b) The proof of this part is standard and follows by restricting the spectral function to the subspace of diagonal matrices and application of Lemma 3.1(a).

(c) This part follows directly from (a) since the strict continuity and directional differentiability of f at $\lambda(X)$ imply the semidifferentiability of f at $\lambda(X)$. ∎

The sufficient condition of semidifferentiability in Proposition 3.2(a) cannot be replaced by directional differentiability in general. However, it can be so if f has the separable form

$$f(x) = g(x_1) + \cdots + g(x_n), \tag{6}$$

where $g : \mathbb{R} \to \mathbb{R}$ is directionally differentiable. The proof is simple by noticing in the preceding argument for (a) that

$$f(\mu(\tau)) = \sum_{i=1}^{n} g(\mu_i(\tau)) = \sum_{i=1}^{n} (g(\mu_i(0) + \tau g'(\mu_i(0); \mu_i'(0)) + o(\tau))).$$

Hence for this special case we have

$$(f \circ \lambda)'(X; H) = \sum_{i=1}^{n} g'(\mu_i(0); \mu_i'(0)).$$

The next result on differentiability of spectral functions will be used in our analysis of semismoothness (Proposition 3.5) and LC^1 property (Proposition 4.3) of spectral functions.

Lemma 3.3 (see [13, Thm. 1.1 and Cor. 2.5]). *Let $X \in \mathcal{S}$. $(f \circ \lambda)$ is differentiable at X if and only if f is differentiable at $\lambda(X)$. In this case the gradient of $(f \circ \lambda)$ at X is*

$$\nabla(f \circ \lambda)(X) = V \mathrm{Diag}[\nabla f(\mu)] V^T \tag{7}$$

for any orthogonal matrix $V \in \mathcal{O}$ and $\mu \in \mathbb{R}^n$ satisfying $X = V\mathrm{Diag}[\mu]V^T$.

The result below shows that semismoothness implies semidifferentiability.

Lemma 3.4. *Let $F : \mathbb{R}^k \mapsto \mathbb{R}^\ell$ and $x \in \mathbb{R}^k$. Suppose that F is semismooth at x. Then F is semidifferentiable at x.*

Proof. An equivalent characterization of semismoothness of F at x is that the limit

$$\lim_{\substack{\hat{h} \to h \\ \tau \searrow 0}} \left\{ V\hat{h} \mid V \in \partial F(x + \tau\hat{h}) \right\} \tag{8}$$

exists for any $h \in \mathbb{R}^k$ and equals $F'(x; h)$; see [23]. Let $h \in \mathbb{R}^k$ be given. For any $\tau \searrow 0$ and $\hat{h} \to h$, choose any element $V \in \partial F(x + \tau\hat{h})$; we then have

$$\lim_{\substack{\hat{h} \to h \\ \tau \searrow 0}} \left(F(x + \tau\hat{h}) - F(x) \right) / \tau = \lim_{\substack{\hat{h} \to h \\ \tau \searrow 0}} \left(F(x + \tau\hat{h}) - F(x) - \tau V\hat{h} + \tau V\hat{h} \right) / \tau$$

$$= \lim_{\substack{\hat{h}\to h\\ \tau\searrow 0}} o(\tau\|\hat{h}\|)/\tau + \lim_{\substack{\hat{h}\to h\\ \tau\searrow 0}} V\hat{h} = F'(x;h).$$

Hence F is semidifferentiable at x. □

The converse of the above result is not true, i.e., a semidifferentiable function is not necessarily semismooth. For example, let $F : \mathbb{R}^n \mapsto \mathbb{R}$ be defined by

$$F(x) := \begin{cases} \|x\|^2 \sin(\frac{1}{\|x\|}) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

This function is locally Lipschitzian, differentiable everywhere, smooth everywhere except at the origin, and semidifferentiable at 0. But it is not semismooth at 0 [24].

Now we present the second main result in this section. The sufficient part says that the spectral function $(f \circ \lambda)$ inherits semismoothness from f, which can also be obtained by using a recent result of Sun and Sun [28] that the eigenvalue function $\lambda(\cdot)$ is strongly semismooth and the fact that compositions of ρ-order semismooth functions are ρ-order semismooth [7]. However, we include a different proof here because it is direct and suggests a proof technique in analyzing SC^1 property of spectral functions in the next section.

PROPOSITION 3.5. *For any symmetric function $f : \mathbb{R}^n \mapsto \mathbb{R}$, the spectral function $(f \circ \lambda)$ is semismooth if and only if f is semismooth. If f is ρ-order semismooth $(0 < \rho < \infty)$, then $(f \circ \lambda)$ is $\min\{1, \rho\}$-order semismooth.*

Proof. Suppose f is semismooth. Then f is strictly continuous and semidifferentiable (Lemma 3.4). Hence $(f\circ\lambda)$ is strictly continuous and directionally differentiable (Lemma 3.2). Let $\mathcal{D} := \{X \in \mathcal{S} |\ (f \circ \lambda) \text{ is differentiable at } X\}$.

Fix any $X \in \mathcal{S}$. By Lemma 2.3, there exist scalars $\eta > 0$ and $\epsilon > 0$ such that (4) holds. We will show that, for any $H \in \mathcal{S}$ with $X + H \in \mathcal{D}$ and $\|H\| \leq \epsilon$, we have

$$(9) \qquad (f \circ \lambda)(X + H) - (f \circ \lambda)(X) - \langle \nabla(f \circ \lambda)(X + H), H\rangle = o(\|H\|),$$

where $o(\cdot)$ and $O(\cdot)$ depend on f and X only. Then it follows from Lemma 2.2 that $(f \circ \lambda)$ is semismooth at X. Since the choice of $X \in \mathcal{S}$ was arbitrary, $(f \circ \lambda)$ is semismooth. Now choose any $Q \in \mathcal{O}_{X+H}$. Then Lemma 2.3 implies that there exists $P \in \mathcal{O}_X$ satisfying

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

For simplicity, let r denote the left-hand side of (9), i.e.,

$$r := (f \circ \lambda)(X + H) - (f \circ \lambda)(X) - \langle \nabla(f \circ \lambda)(X + H), H\rangle.$$

We also let

$$\Delta_1 := f(\lambda(X + H)) - f(\lambda(X)) - \langle \nabla f(\lambda(X + H)), \lambda(X + H) - \lambda(X)\rangle$$

and

$$\Delta_2 := \langle \nabla f(\lambda(X + H)), \lambda(X + H) - \lambda(X) - \mathrm{diag}[Q^T HQ]\rangle.$$

Since $(f \circ \lambda)$ is differentiable at $X + H \in \mathcal{D}$, it follows from Lemma 3.3 that f is differentiable at $\lambda(X + H)$. Hence, Δ_1 and Δ_2 are well defined. Note that $\nabla f(\lambda(X + H))$ and $\lambda(X + H) - \lambda(X)$ are column vectors. We write their inner product in the form of $\langle \cdot, \cdot \rangle$ rather than $x^T y$ for $x, y \in \mathbb{R}^n$ in order to be consistent with the inner product in $\mathcal{S}$. Using the gradient formula (7), we then have

$$\begin{aligned} &\langle \nabla(f \circ \lambda)(X+H), H\rangle = \langle Q\mathrm{Diag}[\nabla f(\lambda(X+H))]Q^T, H\rangle \\ &= \langle \mathrm{Diag}[\nabla f(\lambda(X+H))], Q^T H Q\rangle = \langle \nabla f(\lambda(X+H)), \mathrm{diag}[Q^T H Q]\rangle, \end{aligned}$$

yielding

$$r = \Delta_1 + \Delta_2.$$

Since f is semismooth at $\lambda(X)$ and $\lambda(X+H) \to \lambda(X)$ as $\|H\| \to 0$, it follows from Lemmas 2.2 and 2.4 that

$$\Delta_1 = o(\|\lambda(X+H) - \lambda(X)\|) = o(\|H\|).$$

It remains to show $\Delta_2 = o(\|H\|)$ in order to show $r = o(\|H\|)$. Let $\tilde{H} := Q^T H Q$ and $O := P^T Q$. For simplicity, we let $\mu := \lambda(X+H)$ and $\beta := \lambda(X)$. Since

$$\mathrm{Diag}[\mu] = Q^T(X+H)Q = O^T \mathrm{Diag}[\beta] O + \tilde{H},$$

we have

$$\sum_{k=1}^{n} O_{ki} O_{kj} \beta_k + \tilde{H}_{ij} = \begin{cases} \mu_i & \text{if } i = j, \\ 0 & \text{else}, \quad i, j = 1, \ldots, n. \end{cases} \tag{10}$$

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

$$O_{ij} = O(\|H\|) \quad \text{for } i \neq j. \tag{11}$$

Since $P, Q \in \mathcal{O}$, we have $O \in \mathcal{O}$ so that $O^T O = I$. This and (11) imply

$$1 = O_{ii}^2 + \sum_{k \neq i} O_{ki}^2 = O_{ii}^2 + O(\|H\|^2), \quad i = 1, \ldots, n. \tag{12}$$

Then, for $i = 1, \ldots, n$, the relations (10)–(12) yield

$$\begin{aligned} &\mu_i - \beta_i - (Q^T H Q)_{ii} = \sum_{k=1}^{n} O_{ki}^2 \beta_k + \tilde{H}_{ii} - \beta_i - \tilde{H}_{ii} \\ &= O_{ii}^2 \beta_i + \sum_{k \neq i} O_{ki}^2 \beta_k - \beta_i = \beta_i - \beta_i + O(\|H\|^2) = O(\|H\|^2). \end{aligned}$$

Hence we have

$$\lambda(X+H) - \lambda(X) - \mathrm{diag}[Q^T H Q] = O(\|H\|^2),$$

which in turn implies $\Delta_2 = O(\|H\|^2)$. This proves that $(f \circ \lambda)$ is semismooth.

Suppose that f is ρ-order semismooth at $\lambda(X)$ $(0 < \rho < \infty)$. Then the preceding argument shows that

$$r = \Delta_1 + \Delta_2 = O(\|H\|^{1+\rho}) + O(\|H\|^2) = O(\|H\|^{1+\min\{1,\rho\}}).$$

This shows that $(f \circ \lambda)$ is $\min\{1, \rho\}$-order semismooth at X.

Suppose now $(f \circ \lambda)$ is semismooth. Then $(f \circ \lambda)$ is directionally differentiable and strictly continuous. By Proposition 3.2, f is directionally differentiable. It is well known that $(f \circ \lambda)$ is strictly continuous if and only if f is. Then the semismoothness of f follows from restricting the spectral function $(f \circ \lambda)$ to the subspace of diagonal matrices and using the property of semismoothness of $(f \circ \lambda)$ and Lemma 2.2. $\square$

4. LC^1 and SC^1 spectral functions. The purpose of this section is to show that the spectral function $(f \circ \lambda)$ inherits LC^1 and SC^1 properties from f. To establish those properties, we need two more known results. One is a result of Rockafellar and Wets saying that any Lipschitz function has a uniform approximation of a sequence of continuously differentiable functions (on compact domain). The other is a result of Lewis and Sendov on twice continuously differentiable spectral functions.

LEMMA 4.1 (see [26, Thm. 9.67]). *Given $f : \mathbb{R}^n \to \mathbb{R}$, and Ω is an open subset in $\mathbb{R}^n$. If f is strictly continuous on Ω, then there exist functions $f^\nu : \mathbb{R}^n \to \mathbb{R}$, $\nu = 1, 2, \dots$, continuously differentiable and converging uniformly to f on any compact set contained in Ω. Moreover, if f is an LC^1 function on Ω, then there are twice continuously differentiable functions f^ν such that $\{\nabla f^\nu\}$ converge uniformly to ∇f on any compact set C contained in Ω, and*

$$\left|\left\|\nabla^2 f^\nu(x)\right\|\right| \leq \operatorname{lip} \nabla f(x) \quad \forall \nu. \tag{13}$$

If f is symmetric, then the smooth approximants $\{f^\nu\}$ can also be selected to be symmetric.

In fact, [26, Thm. 9.67] contains only the first part of Lemma 4.1. But the second part can be obtained from its proof. To see this, let $\psi^\nu : \mathbb{R}^n \to \mathbb{R}$, $\nu = 1, 2, \dots$, be nonnegative, measurable, and bounded with $\int_{\mathbb{R}^n} \psi^\nu(z)dz = 1$, and the sets $\mathbb{B}^\nu := \{z \in \mathbb{R}^n |\ \psi^\nu(z) > 0\}$ form a bounded sequence that converges to $\{0\}$. Let C be a compact set contained in Ω. We assume $\mathbb{B}^\nu + C \subseteq \Omega$. Define

$$f^\nu(x) := \int_{\mathbb{R}^n} f(x - z)\psi^\nu(z)dz = \int_{\mathbb{B}^\nu} f(x - z)\psi^\nu(z)dz.$$

We observe that for $x \in \Omega$

$$\nabla f^\nu(x) = \int_{\mathbb{B}^\nu} \nabla f(x - z)\psi^\nu(z)dz.$$

Then the proof argument of [26, Thm. 9.67] can be applied to the functions ∇f^ν and ∇f, establishing that $\{\nabla f^\nu\}$ converge uniformly to ∇f on any compact set C contained in Ω and (13) holds. Suppose f is symmetric. We further assume that the measurable functions $\{\psi^\nu\}$ are symmetric; it follows from the symmetry of f and ψ^ν that for any $P \in \mathcal{P}$

$$f^\nu(Px) = \int_{\mathbb{R}^n} f(Px - z)\psi^\nu(z)dz = \int_{\mathbb{R}^n} f(x - P^T z)\psi^\nu(P^T z)d(P^T z) = f^\nu(x),$$

i.e., $\{f^\nu\}$ are also symmetric.

To present Lewis and Sendov's result, we suppose the symmetric function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is twice differentiable at some points. Letting $\mu \in \mathbb{R}^n$ be such a point, we define a matrix map $\mathcal{A}(\cdot)$ mapping μ to an $n \times n$ matrix:

$$(\mathcal{A}(\mu))_{ij} := \begin{cases} 0 & \text{if } \ i = j, \\ \left(\nabla^2 f(\mu)\right)_{ii} - \left(\nabla^2 f(\mu)\right)_{ij} & \text{if } \ i \neq j \text{ and } \mu_i = \mu_j, \\ \dfrac{(\nabla f(\mu))_i - (\nabla f(\mu))_j}{\mu_i - \mu_j} & \text{else.} \end{cases} \tag{14}$$

The following results are [16, Thms. 3.3, 4.2].

LEMMA 4.2. *For any $X \in \mathcal{S}$, $(f \circ \lambda)$ is twice (continuously) differentiable at X if and only if f is twice (continuously) differentiable at $\lambda(X)$. Moreover, in this case the Hessian of the spectral function at X is*

$$\nabla^2(f \circ \lambda)(X)[H] = U(\mathrm{Diag}[\nabla^2 f(\lambda(X))\,\mathrm{diag}[\tilde{H}]] + \mathcal{A}(\lambda(X)) \circ \tilde{H})U^T, \ \forall H \in \mathcal{S}, \tag{15}$$

where U is any orthogonal matrix in $\mathcal{O}_X$ and $\tilde{H} = U^T H U$.

If f is twice continuously differentiable in a neighborhood of $\lambda(X)$, say, $\mathcal{N}(\lambda(X), \epsilon)$ for some $\epsilon > 0$, and

$$\|\nabla^2 f(\mu)\| \le \kappa \tag{16}$$

for any μ in this neighborhood for some $\kappa > 0$, then, according to Lemmas 2.4 and 4.2, $(f \circ \lambda)$ is twice continuously differentiable in the neighborhood $\mathcal{B}(X, \epsilon)$ of X and for any Y in this neighborhood

$$\begin{aligned}
|||\nabla^2(f \circ \lambda)(Y)||| &= \sup_{\|H\|=1} \|\nabla^2(f \circ \lambda)(Y)(H)\| \\
&= \sup_{\|H\|=1} \|U(\mathrm{Diag}[\nabla^2 f(\lambda(Y))\,\mathrm{diag}[U^T H U]] + \mathcal{A}(\lambda(Y)) \circ (U^T H U))U^T\| \\
&= \sup_{\|H\|=1} \|\mathrm{Diag}[\nabla^2 f(\lambda(Y))\,\mathrm{diag}[U^T H U]] + \mathcal{A}(\lambda(Y)) \circ (U^T H U)\| \\
&\le \sup_{\|H\|=1} \|\mathrm{Diag}[\nabla^2 f(\lambda(Y))\,\mathrm{diag}[U^T H U]]\| + \sup_{\|H\|=1} \|\mathcal{A}(\lambda(Y)) \circ (U^T H U)\| \\
&\le \bar{\kappa} \sup_{\|H\|=1} \|U^T H U\| = \bar{\kappa}, \qquad (17)
\end{aligned}$$

for some $\bar{\kappa} > 0$ which depends only on κ. Here we use the facts $\lambda(Y) \in \mathcal{N}(\lambda(X), \epsilon)$, (16), and the twice continuous differentiability of f.

Now we present our first main result on LC^1 spectral functions.

PROPOSITION 4.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable in an open set $\Omega \subseteq \mathbb{R}^n$. Let $X \in \mathcal{S}$ with $\lambda(X) \in \Omega$. The following results hold.*

(a) *$\nabla(f \circ \lambda)$ is strictly continuous at X if and only if ∇f is strictly continuous at $\lambda(X)$.*

(b) *$(f \circ \lambda)$ is an LC^1 function in $\mathcal{S}$ if and only if f is an LC^1 function in $\mathbb{R}^n$.*

Proof. For any $\epsilon > 0$ such that $\mathcal{N}(\lambda(X), \epsilon) \subset \Omega$, it is noted that f is differentiable at every point in $\mathcal{N}(\lambda(X), \epsilon)$ and $(f \circ \lambda)$ is also differentiable at every point in $\mathcal{B}(X, \epsilon)$ by Lemmas 2.4 and 3.3.

(a) Suppose that ∇f is strictly continuous at $\lambda(X)$. Then there exist scalars $\kappa > 0$ and $\delta > 0$ such that

$$\|\nabla f(y) - \nabla f(z)\| \le \kappa \|y - z\| \qquad \forall\, y, z \in \mathcal{N}(\lambda(X), \delta) \subset \Omega.$$

We note that $\lambda(Y) \in \mathcal{N}(\lambda(X), \delta)$ for any $Y \in \mathcal{B}(X, \delta)$. By letting $C := \mathcal{N}(\lambda(X), \delta)$ in Lemma 4.1, there exists a sequence of twice continuously differentiable and symmetric functions $f^\nu : \mathbb{R}^n \to \mathbb{R}$, $\nu = 1, 2, \ldots$, satisfying that ∇f^ν converges uniformly to ∇f on C and

$$\|\nabla^2 f^\nu(\xi)\| \le \kappa \qquad \forall\, \xi \in C, \ \forall \nu. \tag{18}$$

By Lemma 4.2, we know that each $(f^\nu \circ \lambda)$ is twice continuously differentiable. Letting $Y \in \mathcal{B}(X, \delta)$, it follows from Lemma 3.3 that for any $P \in \mathcal{O}_Y$ we have

$$\begin{aligned}
\|\nabla(f^\nu \circ \lambda)(Y) - \nabla(f \circ \lambda)(Y)\| &= \|P\mathrm{Diag}[\nabla f^\nu(\lambda(Y))]P^T - P\mathrm{Diag}[\nabla f(\lambda(Y))]P^T\| \\
&= \|\mathrm{Diag}[\nabla f^\nu(\lambda(Y)) - \nabla f(\lambda(Y))]\|,
\end{aligned}$$

where we use $PP^T = I$ and the properties of the Frobenius norm. Since $\{\nabla f^\nu\}_1^\infty$ converge uniformly to ∇f on C, this shows that $\{\nabla(f^\nu \circ \lambda)\}_1^\infty$ converge uniformly to $\nabla(f \circ \lambda)$ on $\mathcal{B}(X, \delta)$. Moreover, by repeating arguments for (17) to the function f^ν (noting that $\{f^\nu\}_1^\infty$ are twice continuously differentiable with the bound of (18)), we have for any $Y \in \mathcal{B}(X, \delta)$,

$$|||\nabla^2(f^\nu \circ \lambda)(Y)||| \le \bar{\kappa} \qquad \forall \nu, \tag{19}$$

for some $\bar{\kappa} > 0$, depending only on κ. Fix any $Y, Z \in \mathcal{B}(X, \delta)$ with $Y \neq Z$. Since $\{\nabla(f^\nu \circ \lambda)\}_1^\infty$ converges uniformly to $\nabla(f \circ \lambda)$ on $\mathcal{B}(X, \delta)$, for any $\epsilon > 0$ there exists an integer $\nu_1 > 0$ such that for all $\nu > \nu_1$ we have

$$\|\nabla(f^\nu \circ \lambda)(W) - \nabla(f \circ \lambda)(W)\| \le \epsilon\|Y - Z\| \qquad \forall W \in \mathcal{B}(X, \delta).$$

Then by (19) and the mean value theorem for continuously differentiable functions, we have

$$\begin{aligned}
&\|\nabla(f \circ \lambda)(Y) - \nabla(f \circ \lambda)(Z)\| \\
= &\|\nabla(f \circ \lambda)(Y) - \nabla(f^\nu \circ \lambda)(Y) + \nabla(f^\nu \circ \lambda)(Y) - \nabla(f^\nu \circ \lambda)(Z) \\
&+\nabla(f^\nu \circ \lambda)(Z) - \nabla(f \circ \lambda)(Z)\| \\
\le &\|\nabla(f \circ \lambda)(Y) - \nabla(f^\nu \circ \lambda)(Y)\| + \|\nabla(f^\nu \circ \lambda)(Y) - \nabla(f^\nu \circ \lambda)(Z)\| \\
&+\|\nabla(f^\nu \circ \lambda)(Z) - \nabla(f \circ \lambda)(Z)\| \\
\le &2\epsilon\|Y - Z\| + \|\int_0^1 \nabla^2(f^\nu \circ \lambda)(Y + \tau(Y - Z))(Y - Z)d\tau\| \\
\le &(\bar{\kappa} + 2\epsilon)\|Y - Z\| \quad \forall \ \nu > \nu_1.
\end{aligned}$$

Since $Y, Z \in \mathcal{B}(X, \delta)$ and ϵ are arbitrary, and by letting $\nu \to \infty$, this yields

$$\|\nabla(f \circ \lambda)(Y) - \nabla(f \circ \lambda)(Z)\| \le \bar{\kappa}\|Y - Z\| \quad \forall\, Y, Z \in \mathcal{B}(X, \delta).$$

Thus $\nabla(f \circ \lambda)$ is strictly continuous at X.

Suppose instead that $\nabla(f \circ \lambda)$ is strictly continuous at X. Then the strict continuity of f follows from restricting $(f \circ \lambda)$ to the subspace of diagonal matrices and using formula (7).

(b) is an immediate consequence of (a) by choosing $\Omega = \mathbb{R}^n$. ☐

In addition to the LC^1 property, another prerequisite for being an SC^1 function is the directional differentiability of the gradient map. The following result concerns this prerequisite.

PROPOSITION 4.4. *Suppose f is differentiable on an open set $\Omega \subseteq \mathbb{R}^n$. Let $X \in \mathcal{S}$ and $\lambda(X) \in \Omega$. Then the following results hold.*

(a) *$\nabla(f \circ \lambda)$ is directionally differentiable at X provided that ∇f is semidifferentiable at $\lambda(X)$.*

(b) *∇f is directionally differentiable at $\lambda(X)$ if $\nabla(f \circ \lambda)$ is directionally differentiable at X.*

Proof. We emphasize again that for any $\epsilon > 0$ such that $\mathcal{N}(\lambda(X), \epsilon) \subset \Omega$, $(f \circ \lambda)$ is differentiable at every point in $\mathcal{B}(X, \epsilon)$. Fix such an ϵ. In the following, we will consider point $X + \tau H$ for $\tau \in \mathbb{R}$ and $H \in \mathcal{S}$. Then $X + \tau H \in \mathcal{B}(X, \epsilon)$ for all small $|\tau|$. Hence, f and $(f \circ \lambda)$ are differentiable at $\lambda(X + \tau H)$ and $X + \tau H$, respectively, for all small $|\tau|$.

(a) Let $H \in \mathcal{S}$, and define $X(\tau) = X + \tau H$, $\tau \in \mathbb{R}$. Then by Lemma 2.5 there exists $P(\tau) \in \mathcal{O}$, $\tau \in \mathbb{R}$, whose entries are power series in τ convergent in a neighborhood $\mathcal{I}$ of $\tau = 0$, and $P^T(\tau)X(\tau)P(\tau)$ is diagonal. Then the corresponding eigenvalues

$$\mu_i(\tau) := [P^T(\tau)X(\tau)P(\tau)]_{ii}, \qquad i = 1, \dots, n,$$

are also power series in τ, convergent for $\tau \in \mathcal{I}$. Denote $\mu(\tau) := (\mu_1(\tau), \dots, \mu_n(\tau))^T$. Then we have the expansions

$$\mu(\tau) = \mu(0) + \tau\mu'(0) + o(\tau) \quad \text{and} \quad P(\tau) = P(0) + \tau P'(0) + o(\tau).$$

We note that $\mu(0) = Q\lambda(X)$ for some $Q \in \mathcal{P}$. Hence ∇f is semidifferentiable at $\mu(0)$ by Lemma 3.1(b). In particular, we have

$$\nabla f(\mu(\tau)) = \nabla f(\mu(0) + \tau\mu'(0) + o(\tau)) = \nabla f(\mu(0)) + \tau(\nabla f)'(\mu(0); \mu'(0)) + o(\tau).$$

Then from those expansions above and the formula (7) we have

$$\begin{aligned}
&\nabla(f \circ \lambda)(X + \tau H) - \nabla(f \circ \lambda)(X) \\
&= P(\tau)\mathrm{Diag}[\nabla f(\mu(\tau))]P^T(\tau) - P(0)\mathrm{Diag}[\nabla f(\mu(0))]P^T(0) \\
&= \tau\big(P(0)\mathrm{Diag}[(\nabla f)'(\mu(0); \mu'(0))]P^T(0) + P(0)\mathrm{Diag}[\nabla f(\mu(0))](P'(0))^T \\
&\quad + P'(0)\mathrm{Diag}[\nabla f(\mu(0))]P(0)^T\big) + o(\tau).
\end{aligned}$$

Hence

$$\begin{aligned}
&\lim_{\tau \searrow 0} \left(\nabla(f \circ \lambda)(X + \tau H) - \nabla(f \circ \lambda)(X)\right)/\tau \\
&= P(0)\mathrm{Diag}[(\nabla f)'(\mu(0); \mu'(0))]P^T(0) + P(0)\mathrm{Diag}[\nabla f(\mu(0))](P'(0))^T \\
&\quad + P'(0)\mathrm{Diag}[\nabla f(\mu(0))]P^T(0).
\end{aligned}$$

This implies that the directional derivative $(\nabla(f \circ \lambda))'(X; H)$ is well defined.

(b) Suppose now that $\nabla(f \circ \lambda)$ is directionally differentiable at X. Then the directional differentiability of f follows again from restricting $(f \circ \lambda)$ to the subspace of diagonal matrices and using formula (7). ∎

Our last main result is on SC^1 property of spectral functions.

Proposition 4.5. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable on an open set Ω in $\mathbb{R}^n$. Let $X \in \mathcal{S}$ with $\lambda(X) \in \Omega$. Then the following results hold.*

(a) *∇f is semismooth at $\lambda(X)$ if and only if $\nabla(f \circ \lambda)$ is semismooth at X. If ∇f is ρ-order semismooth at $\lambda(X)$ $(0 < \rho < \infty)$, then $\nabla(f \circ \lambda)$ is $\min\{1, \rho\}$-order semismooth at X.*

(b) *$(f \circ \lambda)$ is an SC^1 function in $\mathcal{S}$ if and only if f is an SC^1 function in $\mathbb{R}^n$.*

Proof. First we note that there exist $\eta > 0$ and $\epsilon > 0$ such that both f and $(f \circ \lambda)$ are differentiable in $\mathcal{N}(\lambda(X), \epsilon)$ and $\mathcal{B}(X, \epsilon)$, respectively, and Lemma 2.3 holds for all $Y \in \mathcal{B}(X, \epsilon)$. For simplicity, we let $F(\cdot) = \nabla f(\cdot)$.

(a) Suppose F is semismooth at $\lambda(X)$. Then F is semidifferentiable, strictly continuous, and directionally differentiable at $\lambda(X)$. By Propositions 4.4 and 4.3, $\nabla(f \circ \lambda)$ is directionally differentiable at X and locally Lipschitz continuous in a neighborhood $\mathcal{B}(X, \delta)$ for some $\delta \le \epsilon$. Let $\mathcal{D} := \{Y \in \mathcal{B}(X, \delta) |\ \nabla(f \circ \lambda)$ is differentiable at $Y\}$ and $\lambda := \lambda(X)$. 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

$$\nabla(f\circ\lambda)(X+H) - \nabla(f\circ\lambda)(X) - \nabla^2(f\circ\lambda)(X+H)H = o(\|H\|). \tag{20}$$

Then, it follows from Lemma 2.2 that $\nabla(f\circ\lambda)$ is semismooth at X. Let $\mu := \lambda(X+H)$, and choose any $Q \in \mathcal{O}_{X+H}$; then there exists $P \in \mathcal{O}_X$ satisfying

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

Since $X+H \in \mathcal{D}$, Lemma 4.2 implies ∇f is differentiable at μ. For simplicity, let R denote the left-hand side of (20); then we have from (7) and (15) that

$$\begin{aligned} R &:= \nabla(f\circ\lambda)(X+H) - \nabla(f\circ\lambda)(X) - \nabla^2(f\circ\lambda)(X+H)H \\ &= Q\mathrm{Diag}[\nabla f(\mu)]Q^T - P\mathrm{Diag}[\nabla f(\lambda)]P^T \\ &\quad - Q(\mathrm{Diag}[\nabla^2 f(\mu)\,\mathrm{diag}[Q^T HQ]] + \mathcal{A}(\mu)\circ(Q^T HQ))Q^T. \end{aligned}$$

Once again for simplicity, we let

$$\tilde{R} := Q^T RQ,\ \ \tilde{H} := Q^T HQ,\ \ A := \mathrm{Diag}[F(\mu)],\ \ B := \mathrm{Diag}[F(\lambda)],\ \ D := P^T Q,\ \ C := \mathcal{A}(\mu).$$

Consequently we have

$$\tilde{R} = A - D^T BD - \mathrm{Diag}[\nabla F(\mu)\,\mathrm{diag}[\tilde{H}]] - C\circ\tilde{H}. \tag{21}$$

Since $\mathrm{Diag}[\mu_1,\ldots,\mu_n] = Q^T(X+H)Q = D^T\mathrm{Diag}[\lambda_1,\ldots,\lambda_n]D + \tilde{H}$, we have

$$\sum_{k=1}^n D_{ki}D_{kj}\lambda_k + \tilde{H}_{ij} = \begin{cases} \mu_i & \text{if } i=j, \\ 0 & \text{else,} \end{cases} \qquad i,j = 1,\ldots,n. \tag{22}$$

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

$$D_{ij} = O(\|H\|) \quad \forall i \neq j. \tag{23}$$

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

$$1 = D_{ii}^2 + \sum_{k\neq i} D_{ki}^2 = D_{ii}^2 + O(\|H\|^2), \quad i = 1,\ldots,n, \tag{24}$$

$$0 = D_{ii}D_{ij} + D_{ji}D_{jj} + \sum_{k\neq i,j} D_{ki}D_{kj} = D_{ii}D_{ij} + D_{ji}D_{jj} + O(\|H\|^2) \quad \forall i \neq j. \tag{25}$$

We now show that $\tilde{R} = o(\|H\|)$, which, by $\|R\| = \|\tilde{R}\|$, would prove (20). For any $i \in \{1,\ldots,n\}$, we have

$$\begin{aligned} \tilde{R}_{ii} &\overset{(21)}{=} F_i(\mu) - \sum_{k=1}^n D_{ki}^2 F_k(\lambda) - \sum_{j=1}^n ((\nabla F(\mu))_{ij}\tilde{H}_{jj}) \\ &\overset{(22)}{=} F_i(\mu) - \sum_{k=1}^n D_{ki}^2 F_k(\lambda) - \sum_{j=1}^n \left((\nabla F(\mu))_{ij}\left(\mu_j - \sum_{k=1}^n D_{kj}^2\lambda_k\right)\right) \\ &\overset{(23)}{=} F_i(\mu) - D_{ii}^2 F_i(\lambda) - \sum_{j=1}^n (\nabla F(\mu))_{ij}(\mu_j - D_{jj}^2\lambda_j) + O(\|H\|^2) \end{aligned}$$

$$
\begin{aligned}
&\overset{(24)}{=} F_i(\mu) - (1 + O(\|H\|^2))F_i(\lambda) \\
&\quad - \sum_{j=1}^{n}((\nabla F(\mu))_{ij}(\mu_j - (1 + O(\|H\|^2))\lambda_j)) + O(\|H\|^2) \\
&= F_i(\mu) - F_i(\lambda) - \sum_{j=1}^{n}(\nabla F(\mu))_{ij}(\mu_j - \lambda_j) + O(\|H\|^2) \\
&= F_i(\mu) - F_i(\lambda) - (\nabla F_i(\mu))^T(\mu - \lambda) + O(\|H\|^2),
\end{aligned}
$$

where we use local boundedness of F and ∇F. Since F is semismooth at λ, each of its components is also semismooth at λ. Lemma 2.4 implies that $\|\lambda - \mu\| \leq \|H\|$. Then clearly the right-hand side of the preceding relation is $o(\|H\|)$. For any $i, j \in \{1, \dots, n\}$ with $i \neq j$, we have

$$
\begin{aligned}
\tilde{R}_{ij} &\overset{(21)}{=} -\sum_{k=1}^{n} D_{ki}D_{kj}F_k(\lambda) - C_{ij}\tilde{H}_{ij} \\
&\overset{(22)}{=} -\sum_{k=1}^{n} D_{ki}D_{kj}F_k(\lambda) + C_{ij}\sum_{k=1}^{n} D_{ki}D_{kj}\lambda_k \\
&\overset{(23)}{=} -(D_{ii}D_{ij}F_i(\lambda) + D_{ji}D_{jj}F_j(\lambda)) + C_{ij}(D_{ii}D_{ij}\lambda_i + D_{ji}D_{jj}\lambda_j) + O(\|H\|^2) \\
&= -\left((D_{ii}D_{ij} + D_{ji}D_{jj})F_i(\lambda) + D_{ji}D_{jj}(F_j(\lambda) - F_i(\lambda))\right) \\
&\quad + C_{ij}\left((D_{ii}D_{ij} + D_{ji}D_{jj})\lambda_i + D_{ji}D_{jj}(\lambda_j - \lambda_i)\right) + O(\|H\|^2) \\
&\overset{(25)}{=} -D_{ji}D_{jj}\left(F_j(\lambda) - F_i(\lambda) - C_{ij}(\lambda_j - \lambda_i)\right) + O(\|H\|^2).
\end{aligned}
$$

Thus, if $\lambda_i = \lambda_j$, Lemma 3.1(c) implies that $F_i(\lambda) = F_j(\lambda)$, which with the preceding relation, yields

$$\tilde{R}_{ij} = O(\|H\|^2).$$

If $\lambda_i \neq \lambda_j$, then Lemma 2.4 implies $\|\mu - \lambda\| \leq \|H\|$, $|\mu_i - \lambda_i| \leq \|H\|$, 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 $C_{ij} = (F_j(\mu) - F_i(\mu))/(\mu_j - \mu_i)$ and the preceding relation yield

$$\tilde{R}_{ij} \;=\; -D_{ji}D_{jj}\left(F_j(\lambda) - F_i(\lambda) - \frac{F_j(\mu) - F_i(\mu)}{\mu_j - \mu_i}(\lambda_j - \lambda_i)\right) + O(\|H\|^2) \;=\; O(\|H\|^2),$$

where the second equality uses (23) and the strict continuity of F_i and F_j at λ, so that $F_i(\mu) - F_i(\lambda) = O(\|\mu - \lambda\|) = O(\|H\|)$ and $F_j(\mu) - F_j(\lambda) = O(\|\mu - \lambda\|) = O(\|H\|)$.

Suppose F is ρ-order semismooth at $\lambda(X)$ $(0 < \rho < \infty)$. 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 $\nabla(f \circ \lambda)$ is $\min\{1, \rho\}$-order semismooth at X.

Suppose $\nabla(f \circ \lambda)$ is semismooth at X. Then $\nabla(f \circ \lambda)$ is strictly continuous and directionally differentiable at X. By Propositions 4.3 and 4.4, $F := \nabla f$ is strictly continuous and directionally differentiable at $\lambda(X)$. For any $h \in \mathbb{R}^n$ such that F is differentiable at $\lambda(X) + h$, i.e., f is twice differentiable at $\lambda(X) + h$, let $H := Q\mathrm{Diag}[h]Q^T$ for some $Q \in \mathcal{O}_X$. Then there exists $P \in \mathcal{P}$ such that $P(\lambda(X) + h) =$

$\lambda(X+H)$. Lemma 3.1(d) implies that f is twice differentiable at $\lambda(X+H)$. In turn, Lemma 4.2 yields that $\nabla(f\circ\lambda)$ is twice differentiable at $X+H$. We note that

$$Q^T(X+H)Q = \mathrm{Diag}[\lambda(X)+h] = \mathrm{Diag}[P^T\lambda(X+H)] = P^T\mathrm{Diag}[\lambda(X+H)]P,$$

which is equivalent to

$$X+H = QP^T\mathrm{Diag}[\lambda(X+H)]PQ^T = U\mathrm{Diag}[\lambda(X+H)]U^T,$$

where $U := QP^T$, and hence $U\in\mathcal{O}$ since $Q,P\in\mathcal{O}$. For simplicity, let $\mu := \lambda(X+H)$; then we have

$$\begin{aligned} U^THU &= PQ^TQ\mathrm{Diag}[h]Q^TQP^T \\ &= P\mathrm{Diag}[h]P^T = \mathrm{Diag}[Ph] \qquad (\text{using } \ P\in\mathcal{P}) \end{aligned}$$

and

$$\begin{aligned} &\mathrm{Diag}[\nabla^2 f(\mu)\,\mathrm{diag}[U^THU]] = \mathrm{Diag}[\nabla^2 f(\mu)Ph] \\ &= \mathrm{Diag}[\nabla^2 f(P(\lambda(X)+h))Ph] \qquad (\text{using } \ \mu = P(\lambda(X)+h)) \\ &= \mathrm{Diag}[P\nabla^2 f(\lambda(X)+h)P^TPh] \qquad (\text{using Lemma 3.1(d)}) \\ &= \mathrm{Diag}[P\nabla^2 f(\lambda(X)+h)h] \\ &= P\mathrm{Diag}[\nabla^2 f(\lambda(X)+h)h]P^T \qquad (\text{using } \ P\in\mathcal{P}). \end{aligned} \tag{26}$$

Since $\nabla(f\circ\lambda)$ is semismooth at X, it follows from Lemma 2.2 that

$$R := \nabla(f\circ\lambda)(X+H) - \nabla(f\circ\lambda)(X) - \nabla^2(f\circ\lambda)(X+H)H = o(\|H\|),$$

which, by (7), (15), and (26), is equivalent to

$$\begin{aligned} R &= Q\mathrm{Diag}[\nabla f(\lambda(X)+h)]Q^T - Q\mathrm{Diag}[\nabla f(\lambda(X))]Q^T \\ &\quad - U(\mathrm{Diag}[\nabla^2 f(\mu)\,\mathrm{diag}[U^THU]] + \mathcal{A}(\mu)\circ(U^THU))U^T \\ &= Q\mathrm{Diag}[\nabla f(\lambda(X)+h) - \nabla f(\lambda(X)) - \nabla^2 f(\lambda(X)+h)h]Q^T. \end{aligned}$$

The second equality uses $\mathcal{A}(\mu)\circ(U^THU) = \mathcal{A}(\mu)\circ\mathrm{Diag}[Ph] = 0$. We then have

$$\tilde{R} := Q^TRQ = \mathrm{Diag}[\nabla f(\lambda(X)+h) - \nabla f(\lambda(X)) - \nabla^2 f(\lambda(X)+h)h].$$

Since $\|\tilde{R}\| = \|R\|$ and $\|H\| = \|h\|$, the preceding relation means by noting $F = \nabla f$

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

This proves that ∇f is semismooth at $\lambda(X)$.

(b) is an immediate consequence of (a) since the choice of X is arbitrary, Ω can be chosen as $\mathbb{R}^n$. $\square$

Remarks. In the special case where $f:\mathbb{R}^n\mapsto\mathbb{R}$ takes the form (6) and $g(\cdot):\mathbb{R}\mapsto\mathbb{R}$ is differentiable, according to Lemma 3.3 we have

$$\nabla(f\circ\lambda)(X) = U\mathrm{Diag}[g'(\lambda_1),\ldots,g'(\lambda_n)]U^T, \tag{27}$$

where $U\in\mathcal{O}_X$ and $\lambda := \lambda(X)$. Associated with this f, we define a symmetric–matrix-valued function $f^\square:\mathcal{S}\mapsto\mathcal{S}$ by

$$f^\square(X) = U\mathrm{Diag}[g'(\mu_1),\ldots,g'(\mu_n)]U^T$$

for any $U \in \mathcal{O}$ satisfying $X = U\text{Diag}[\mu_1, \ldots, \mu_n]U^T$. It is pointed out in [3] that for this special case

$$f^{\square}(X) = \nabla(f \circ \lambda)(X).$$

Among many results on continuity, differentiability, and nonsmoothness obtained in [3] is the semismoothness of $f^{\square}$. It is proved [3, Prop. 4.10] that $f^{\square}(\cdot)$ is semismooth if and only if $g'(\cdot)$ is semismooth. In other words, for this special case, the SC^1 result (Proposition 4.5) follows from [3, Prop. 4.10]. But for general cases other than (6), we do not have such direct consequences. Nevertheless, the proof here is inspired by [3, Prop. 4.10]. We would also like to point out that the treatment in [3] goes beyond this special case. In fact, given a real function of one dimension $f : \mathbb{R} \mapsto \mathbb{R}$, the symmetric–matrix-valued function defined in [3] is

$$f^{\square}(X) := U\text{Diag}[f(\mu_1), \ldots, f(\mu_n)]U^T,$$

where $U \in \mathcal{O}$ satisfying $X = U\text{Diag}[\mu_1, \ldots, \mu_n]U^T$. There are examples where f cannot be derivative of another real function.

5. An example. As an example, we consider the positive trace function $F : \mathcal{S} \mapsto \mathbb{R}$ by

$$F(X) := (\max\{0, \text{trace}(X)\})^2 \qquad \forall X \in \mathcal{S}.$$

Obviously, $F(X) = (f \circ \lambda)(X)$ with $f : \mathbb{R}^n \mapsto \mathbb{R}$ defined by

$$f(x) := \left(\max\left\{0, \sum x_i\right\}\right)^2 \qquad \forall x \in \mathbb{R}^n.$$

It is known that $f(\cdot)$ is continuously differentiable, and its derivative map is strongly semismooth. Hence, we can conclude that $F(\cdot)$ is continuously differentiable [13], and moreover, it is an SC^1 function (Proposition 4.5).

Acknowledgments. We thank the associate editor M. Overton for his suggestion of the paper title, and the two referees for their comments which improved the presentation of the paper significantly. In particular, [19] was brought to our attention by one referee and Proposition 3.2(c) was suggested to us by the other. We also thank Defeng Sun for his comments on this result.

REFERENCES

[1] R. Bhatia, *Matrix Analysis*, Springer-Verlag, New York, 1997.
[2] X. Chen and P. Tseng, *Non-interior continuation methods for solving semidefinite complementarity problems*, Math. Program., 95 (2003), pp. 431–474.
[3] X. Chen, H.D. Qi, and P. Tseng, *Analysis of nonsmooth symmetric-matrix-valued functions with applications to semidefinite complementarity problems*, SIAM J. Optim., 13 (2003), pp. 960–985.
[4] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
[5] C. Davis, *All convex invariant functions of hermitian matrices*, Arch. Mathematik, 8 (1957), pp. 276–278.
[6] F. Facchinei, *Minimization of SC^1 functions and the Maratos effect*, Oper. Res. Lett., 17 (1995), pp. 131–137.
[7] A. Fischer, *Solution of monotone complementarity problems with locally Lipschitzian functions*, Math. Program., 76 (1997), pp. 513–532.
[8] J.-B. Hiriart-Urruty and A.S. Lewis, *The Clarke and Michel–Penot subdifferentials of the eigenvalues of a symmetric matrix*, Comput. Optim. Appl., 13 (1999), pp. 13–23.

[9] J.-B. Hiriart-Urruty, J.J. Strodiot, and V. Hien Nguyen, *Generalized Hessian matrix and second-order optimality conditions for problems with $C^{1,1}$ data*, Appl. Math. Optim., 11 (1984), pp. 43–56.

[10] 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, Appl. Optim. 22, M. Fukushima and L. Qi, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999, pp. 181–209.

[11] C. Kanzow and C. Nagel, *Semidefinite programs: New search directions, smoothing-type methods, and numerical results*, SIAM J. Optim., 13 (2002), pp. 1–23.

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

[13] A.S. Lewis, *Derivatives of spectral functions*, Math. Oper. Res., 21 (1996), pp. 576–588.

[14] A.S. Lewis and M.L. Overton, *Eigenvalue optimization*, in Acta Numerica, Acta Numer. 5, Cambridge University Press, Cambridge, UK, 1996, pp. 149–190.

[15] A.S. Lewis, *Nonsmooth analysis of eigenvalues*, Math. Program., 84 (1999), pp. 1–24.

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

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

[18] R. Mifflin, *An algorithm for constrained optimization with semismooth functions*, Math. Oper. Res., 2 (1977), pp. 191–207.

[19] J. von Neumann, *Some matrix inequalities and metrization of matric space*, Tomsk University Review, 1 (1937), pp. 286–300. Also in Collected Works Vol. IV: Continuous Geometry and Other Topics, A. H. Taub, ed., Pergamon Press, Oxford, UK, 1962, pp. 205–218.

[20] J.-S. Pang and L. Qi, *A globally convergent Newton method for convex SC^1 minimization problems*, J. Optim. Theory Appl., 85 (1995), pp. 633–648.

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

[22] L. Qi, *Superlinearly convergent approximate Newton methods for LC^1 optimization problems*, Math. Programming, 64 (1994), pp. 277–294.

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

[24] L. Qi and P. Tseng, *An analysis of piecewise smooth functions and almost smooth functions*, Math. Oper. Res., submitted.

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

[28] D. Sun and J. Sun, *Strong semismoothness of eigenvalues of symmetric matrices and its application to inverse eigenvalue problems*, SIAM J. Numer. Anal., 40 (2003), pp. 2352–2367.

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

[30] X.Q. Yang and X.X. Huang, *A nonlinear Lagrangian approach to constrained optimization problems*, SIAM J. Optim., 11 (2001), pp. 1119–1144.

[31] X.Q. Yang and V. Jeyakumar, *Generalized second-order directional derivatives and optimization with $C^{1,1}$ functions*, Optimization, 26 (1992), pp. 165–185.