

Twice Epi-Differentiability of Spectral Functions and its applications

CHAO DING¹ EBRAHIM SARABI² SHIWEI WANG³

Abstract. Second-order variational properties have been shown to play important theoretical and numerical roles for different classes of optimization problems. Among such properties, twice epi-differentiability has a special place because of its ubiquitous presence in various classes of extended-real-valued functions that are important for optimization problems. We provide a useful characterization of this property for spectral functions by demonstrating that it can be characterized via the same property of the symmetric part of the spectral representation of an eigenvalue function. Our approach allows us to bypass the rather restrictive convexity assumption, used in many recent works that targeted second-order variational properties of spectral functions. By this theoretical tool, several applications on the proto-differentiability of subgradient mappings, the directional differentiability of the proximal mapping of spectral functions are achieved. We finally use our established theory to study twice epi-differentiability of leading eigenvalue functions and practical regularization terms that have important applications in statistics and the robust PCA.

Key words. spectral functions, twice epi-differentiability, proto-differentiability, generalized twice differentiable, leading eigenvalue

Mathematics Subject Classification (2000) 49J52, 15A18, 49J53

1 Introduction

In this paper, we study the second-order variational properties of orthogonally invariant functions (also called spectral functions [14]). A mapping $g : \mathbf{S}^n \rightarrow \overline{\mathbf{R}} := [-\infty, \infty]$, is orthogonally invariant if

$$g(X) = g(U^\top XU), \text{ for all } X \in \mathbf{S}^n \text{ and for any } U \in \mathbf{O}^n,$$

where \mathbf{S}^n stands for the real vector space of $n \times n$ symmetric matrices and \mathbf{O}^n is the set of all real $n \times n$ orthogonal matrix. It is known (see [26]) that any spectral function admits the composite representation

$$g(X) = (\theta \circ \lambda)(X), \quad X \in \mathbf{S}^n, \tag{1.1} \quad \{\text{spec}\}$$

where $\theta : \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ is a symmetric (permutation-invariant) function on \mathbf{R}^n , i.e., for all $x \in \mathbf{R}^n$ and any $n \times n$ permutation matrix P , $\theta(x) = \theta(Px)$, where $\lambda(X) := (\lambda_1(X), \dots, \lambda_n(X))$ denotes the vector of eigenvalues of X arranged in nonincreasing order $\lambda_1(X) \geq \dots \geq \lambda_n(X)$. Spectral functions are widely used in areas such as neural networks, signal processing, robust PCA, and

¹State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China; School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China; Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China (dingchao@amss.ac.cn). Research of this author is supported in part by the National Key R&D Program of China (No. 2021YFA1000300, No. 2021YFA1000301), National Natural Science Foundation of China (No. 12531014), and CAS Project for Young Scientists in Basic Research (No. YSBR-034).

²Department of Mathematics, Miami University, Oxford, OH 45065, USA (sarabim@miamioh.edu). Research of this author is partially supported by the U.S. National Science Foundation under the grant DMS 2108546.

³Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China. This work is done while the author is a visiting scholar at the Institute of Operational Research and Analytics, National University of Singapore, Singapore (wangshiwei@amss.ac.cn).

graph theory, among others [11, 12, 15, 25, 39]. They also cover many classical models, including semidefinite programming problems (SDP) [52] and eigenvalue optimization [38]. Over the past three decades, extensive theoretical foundations have been developed for spectral functions; see, e.g., [8–10, 26, 28, 40].

While prior studies have delivered notable advances in first-order variational properties of spectral functions and algorithm design for particular classes of optimization problems such as CMatOP [3], the existing second-order variational theory of spectral functions remains limited in light of rapid progress in machine learning and statistical optimization. To narrow this gap, we aim to present a comprehensive analysis of twice epi-differentiability of spectral functions that are not necessarily convex. Twice epi-differentiability, introduced in [42], provides second-order approximations to the epigraphs of extended-real-valued functions. It also can be used to establish the second-order optimality conditions for different class of optimization problems [17]. On the algorithmic side, recent results in [20, 31, 33] highlight a fundamental role that twice epi-differentiability plays for establishing convergence of numerical algorithms including generalized Newton methods [35] and the augmented Lagrangian method [18] that are widely used in large-scale optimization.

Twice epi-differentiability has been explored for important classes of functions such as piecewise linear quadratic (cf. [43, Proposition 13.9]), C^2 cone reducible convex sets (cf. [31, Theorem 6.2]), and more generally parabolically regular functions (cf. [33, Theorem 3.8]). These results were extended to some classes of eigenvalue functions by Torik in [48, Theorem 3.2]), where he showed for the first time that some eigenvalue functions enjoy this property as well. His result, however, fell short of providing a systematic approach to characterizing twice epi-differentiability of eigenvalue functions through the same property of their symmetric parts, meaning the function θ in (1.1), a pattern which was pioneered by Davis in [5] for convexity of spectral functions and was remarkably extended to various first-order variational properties by Lewis; see [27]. It is worth mentioning that a similar pattern was demonstrated to hold for some second-order variational properties of spectral functions such as twice differentiability (cf. [28]), prox-regularity (cf. [7]), C^2 -cone reducibility (cf. [4]), and parabolic regularity (cf. [21, 34]).

This paper continues the path, initiated recently in [34], to conduct second-order variational analysis of spectral functions. We should add here that the latter work focused mainly on the characterization of parabolic regularity, which is a strictly stronger property than twice epi-differentiability, of spectral functions. Moreover, convexity was assumed in the main results of [34]. Our results have two major differences with [33]. First and foremost, we study a different second-order variational property than the one in [34]. Second, we improve some of the results therein that allow us to drop convexity in our developments in this paper. This opens the door to possibly remove convexity from the main assumptions in [34], which will be a subject for our future research. Below, we summarize our main contributions:

- We show that twice epi-differentiability of the spectral function g in (1.1) can be characterized via the same property of its symmetric part θ even when g is nonconvex. As a consequence of this characterization, we also calculate the second subderivative of spectral functions. A similar result is proven for twice semidifferentiability of differentiable spectral functions. Note that the principal challenge in the nonconvex case is that subgradients $\partial\theta$ need not have components in nonincreasing order, which complicates compatibility with the ordered eigenvalue map. These brand new results provide important insights in how to study second-order variational properties of nonconvex spectral functions.
- Based on these inheritance results, the proto-differentiability of subgradient mappings and the directional differentiability of the proximal mapping of spectral functions are characterized. Furthermore, we provide a characterization of the concept of generalized

twice differentiability of spectral functions, which was introduced recently in [44].

- As a consequence of our second order variational developments, we study variational properties of nonconvex spectral models, e.g., leading eigenvalue-based penalties and MCP penalties on eigenvalue values, which are widely used in modern statistics and machine learning yet lack rigorous analysis. We further propose a unified framework that treats such terms without requiring a case-by-case argument.

The remaining parts of this paper are organized as follows. In Section 2, we introduce the notation used throughout and reviews some preliminaries on eigenvalue and spectral functions. Section 3 presents a characterization of twice epi-differentiability of spectral functions via the same property of their symmetric parts. Several theoretical applications of this characterization are provided in Section 4. Section 5 illustrates some applications of our established theory in important examples, including the leading eigenvalues, MCP penalties on singular values, and several commonly used statistical regularizers. We conclude our paper in the final section.

2 Notations and Preliminaries

{sect02}

In this section, we mainly introduce some notations used throughout this paper, some preliminary knowledge on the eigenvalue function and spectral functions. Suppose \mathbf{X} and \mathbf{Y} are Euclidean spaces. For any set $C \subset \mathbf{X}$, we use δ_C to denote its indicator function. $\text{dist}(x, C)$ is defined as the distance between $x \in \mathbf{X}$ and set C . $\mathbf{R}^{m \times n}$ is the space of all real $n \times m$ matrices. For any $Z \in \mathbf{R}^{m \times n}$ and given two index sets $\alpha \subseteq \{1, \dots, m\}$, $\beta \subseteq \{1, \dots, n\}$, we use Z_α to denote the sub-matrix of Z obtained by removing all the columns of Z not in α . $Z_{\alpha\beta}$ denotes the $|\alpha| \times |\beta|$ sub-matrix of Z obtained by removing all the rows of Z not in α and all the columns of Z not in β , where $|\alpha|$ is cardinality of the index set α . In particular, we use Z_j to represent the j -th column of Z , $j = 1, \dots, n$. Z^\dagger is the Moore-Penrose generalized inverse of Z . \mathbf{N} stands for the set of natural numbers. For any vector $x = (x_1, \dots, x_n) \in \mathbf{R}^n$, denote by $\text{Diag}(x)$ the diagonal matrix whose i -th diagonal entry is x_i for any $i = 1, \dots, n$.

Given $X \in \mathbf{S}^n$, we use $\mu_1(X) > \dots > \mu_r(X)$ to denote the distinct eigenvalues of X and define the index sets

$$\alpha_m := \{i \in \{1, \dots, n\} \mid \lambda_i(X) = \mu_m(X)\}, \quad m = 1, \dots, r. \quad (2.1) \quad \{\text{index}\}$$

Let $\mathbf{O}^n(X)$ be the set of all orthogonal matrices U satisfying the eigenvalue decomposition of $X \in \mathbf{S}^n$, i.e.,

$$X = U \text{Diag}(\lambda(X)) U^\top. \quad (2.2) \quad \{\text{specdocom}\}$$

Define $\ell_i(X)$ for any $i \in \{1, \dots, n\}$ to be the number of eigenvalues of X that are equal to $\lambda_i(X)$ but are ranked before $\lambda_i(X)$ including $\lambda_i(X)$. This integer allows us to locate $\lambda_i(X)$ in the group of the eigenvalues of X as follows:

$$\lambda_1(X) \geq \dots \geq \lambda_{i-\ell_i(X)} > \lambda_{i-\ell_i(X)+1}(X) = \dots = \lambda_i(X) \geq \dots \geq \lambda_n(X). \quad (2.3) \quad \{\text{e11}\}$$

For simplicity, we drop X from $\ell_i(X)$ when the dependence of ℓ_i on X can be seen from the context. For any function $f : \mathbf{X} \rightarrow \mathbf{R}$, we use $f'(x; h)$ to denote its directional derivative at x for h (cf. [43, 7(17)]). The following first-order expansion of the eigenvalue function is given in [47, Proposition 1.4], which is of great importance in the derivation of the main result of this paper.

Proposition 2.1 (First-order expansion of eigenvalue functions). *Assume that $X \in \mathbf{S}^n$ has the eigenvalue decomposition (2.2) for some $U \in \mathbf{O}^n(X)$. Let $\mu_1 > \dots > \mu_r$ be distinct eigenvalues of X . Then, for any $H \in \mathbf{S}^n$ that $H \rightarrow 0$ and any $i \in \{1, \dots, n\}$, we have* {olemma}

$$\lambda_i(X + H) = \lambda_i(X) + \lambda_{\ell_i}(U_{\alpha_m}^\top H U_{\alpha_m} + U_{\alpha_m}^\top H(\mu_m I - X)^\dagger H U_{\alpha_m}) + O(\|H\|^3) \quad (2.4) \quad \{\text{secondexp1}\}$$

and

$$\lambda_i(X + H) = \lambda_i(X) + \lambda_{\ell_i}(U_{\alpha_m}^\top H U_{\alpha_m}) + O(\|H\|^2), \quad (2.5) \quad \{\text{fexpan}\}$$

where $m \in \{1, \dots, r\}$ with $i \in \alpha_m$. Thus, we have

$$\lambda'(X; H) = (\lambda(U_{\alpha_1}^\top H U_{\alpha_1}), \dots, \lambda(U_{\alpha_r}^\top H U_{\alpha_r})).$$

Next, we review some concepts from variational analysis that will be used extensively in this paper. Given a nonempty set $C \subset \mathbf{X}$ with $\bar{x} \in C$, the tangent cone $T_C(\bar{x})$ to C at \bar{x} is defined by

$$T_C(\bar{x}) = \{w \in \mathbf{X} \mid \exists t_k \downarrow 0, w_k \rightarrow w \text{ as } k \rightarrow \infty \text{ with } \bar{x} + t_k w_k \in C\}.$$

The regular normal cone of C at \bar{x} is defined as $\widehat{N}_C(\bar{x}) = T_C(\bar{x})^*$. The (limiting) normal cone $N_C(\bar{x})$ is defined as $\{v \in \mathbf{X} \mid \exists x_k \rightarrow \bar{x}, x_k \in C \text{ and } v_k \rightarrow v \text{ with } v_k \in \widehat{N}_C(x_k)\}$. The set of all subgradients of f at \bar{x} is defined as $\partial f(\bar{x}) = \{v \in \mathbf{X} \mid (v, -1) \in N_{\text{epi } f}(\bar{x}, f(\bar{x}))\}$ and is called the (limiting) subdifferential of f at \bar{x} . Similarly, we can define the regular subdifferential of f at \bar{x} , denoted by $\widehat{\partial} f(\bar{x})$. Below, we record a result, established in [27, Theorem 6], that provides a nice formula for the subdifferential of spectral functions. {subsp}

Proposition 2.2. *Assume that $\theta : \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ is a proper and lower semicontinuous (lsc) symmetric function. Then we have*

$$\partial(\theta \circ \lambda)(X) = \{U \text{Diag}(v) U^\top \mid v \in \partial\theta(\lambda(X)), U \in \mathbf{O}^n(X)\}.$$

A similar conclusion holds for the regular subdifferential of $\theta \circ \lambda$.

Given a function $f : \mathbf{X} \rightarrow \overline{\mathbf{R}}$, its domain is defined by $\text{dom } f = \{x \in \mathbf{X} \mid f(x) < \infty\}$. The function f is called locally Lipschitz continuous around \bar{x} relative to $C \subset \text{dom } f$ with constant $\ell \geq 0$ if $\bar{x} \in C$ with $f(\bar{x})$ finite and there exists a neighborhood U of \bar{x} such that

$$|f(x) - f(y)| \leq \ell \|x - y\| \quad \text{for all } x, y \in U \cap C.$$

Such a function is called locally Lipschitz continuous relative to C if it is locally Lipschitz continuous around every $\bar{x} \in C$ relative to C . Piecewise linear-quadratic functions (not necessarily convex) and an indicator function of a nonempty set are important examples of functions that are locally Lipschitz continuous relative to their domains.

The subderivative function of $f : \mathbf{X} \rightarrow \overline{\mathbf{R}}$ at \bar{x} with $f(\bar{x})$ finite, denoted by $\text{d}f(\bar{x}) : \mathbf{X} \rightarrow \overline{\mathbf{R}}$, is defined for any $\bar{w} \in \mathbf{X}$ by

$$\text{d}f(\bar{x})(\bar{w}) = \liminf_{\substack{t \downarrow 0 \\ w \rightarrow \bar{w}}} \frac{f(\bar{x} + tw) - f(\bar{x})}{t}. \quad (2.6) \quad \{\text{semi1}\}$$

Next, we review some preliminary results about the subderivative of the spectral functions. Denote \mathbf{P}^n as the set of all $n \times n$ permutation matrices. Recall that a function $\theta : \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ is called symmetric if for every $x \in \mathbf{R}^n$ and every $Q \in \mathbf{P}^n$, $\theta(Qx) = \theta(x)$ holds. It can be checked directly from (1.1) that θ can be expressed as a composite function in the form

$$\theta(x) = g(\text{Diag}(x)) \quad \text{for all } x \in \mathbf{R}^n. \quad (2.7) \quad \{\text{spec2}\}$$

Suppose X has the eigenvalue decomposition of (2.2) with r distinct eigenvalues. Denote $\mathbf{P}_X^n := \{Q = \text{Diag}(P_1, \dots, P_r) \mid P_m \in \mathbf{P}^{|\alpha_m|}, m = 1, \dots, r\}$. It can be checked directly that $\mathbf{P}_X^n \subset \mathbf{P}^n$. Moreover, if $Q \in \mathbf{P}_X^n$, we have $Q\lambda(X) = \lambda(X)$. If $\theta : \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ is a symmetric function and $X \in \mathbf{S}^n$ with $\theta(\lambda(X))$ finite, it is not hard to see that

$$d\theta(\lambda(X))(v) = d\theta(\lambda(X))(Qv) \quad (2.8) \quad \{\text{sub_sym}\}$$

for any $v \in \mathbf{R}^n$ and any $Q \in \mathbf{P}_X^n$.

Recalling the index sets α_m , $m = 1, \dots, r$, from (2.1), we partition a vector $p \in \mathbf{R}^n$ into $(p_{\alpha_1}, \dots, p_{\alpha_r})$, where $p_{\alpha_m} \in \mathbf{R}^{|\alpha_m|}$ for any $m = 1, \dots, r$. Recall from [43, Definition 7.25] that an lsc function $f : \mathbf{X} \rightarrow \overline{\mathbf{R}}$ is said to be subdifferentially regular at $\bar{x} \in \mathbf{X}$ if $f(\bar{x})$ is finite and $\widehat{N}_{\text{epi } f}(\bar{x}, f(\bar{x})) = N_{\text{epi } f}(\bar{x}, f(\bar{x}))$. Denote by \mathbf{R}_\downarrow^n the set of all vectors (x_1, \dots, x_n) such that $x_1 \geq \dots \geq x_n$. The following result shows the explicit formula of the subderivative of spectral functions when they are subdifferentially regular. Note that a similar result was recently established [34, Theorem 3.5] for convex spectral functions.

Theorem 2.3 (Subderivatives of spectral functions). *Let $\theta : \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ be a symmetric function and let $X \in \mathbf{S}^n$ with $(\theta \circ \lambda)(X)$ finite. If θ is lsc and subdifferentially regular with $\partial\theta(\lambda(X)) \neq \emptyset$, then for all $H \in \mathbf{S}^n$ we have*

$$d(\theta \circ \lambda)(X)(H) = d\theta(\lambda(X))(\lambda'(X; H)). \quad (2.9) \quad \{\text{dchain}\}$$

Proof. The proof of the inequality “ \geq ” in (2.9) is exactly the same as that in [34, Theorem 3.5], so we omit it here for brevity. To justify the opposite inequality, we can also obtain for any $\varepsilon > 0$, we may choose $Y \in \partial(\theta \circ \lambda)(X)$ such that

$$d(\theta \circ \lambda)(X)(H) \leq \varepsilon + \langle Y, H \rangle$$

by following the same argument as in the proof of [34, Theorem 3.5]. Since $Y \in \partial(\theta \circ \lambda)(X)$, it follows from Proposition 2.2 that there are $U \in \mathbf{O}^n(X)$ and $v \in \partial\theta(\lambda(X))$ such that $Y = U\text{Diag}(v)U^\top$. Since $v = (v_{\alpha_1}, \dots, v_{\alpha_r})$, we can choose a permutation matrix $Q_m \in \mathbf{P}^{|\alpha_m|}$ such that $\tilde{v}_{\alpha_m} = Q_m v_{\alpha_m}$ with $\tilde{v}_{\alpha_m} \in \mathbf{R}_\downarrow^{|\alpha_m|}$ for any $m = 1, \dots, r$. Set $Q = \text{Diag}(Q_1, \dots, Q_r)$ and use Fan’s inequality [13] to obtain

$$\begin{aligned} d(\theta \circ \lambda)(X)(H) &\leq \varepsilon + \langle Y, H \rangle = \varepsilon + \langle \text{Diag}(v), U^\top H U \rangle = \varepsilon + \sum_{m=1}^r \langle \text{Diag}(v)_{\alpha_m \alpha_m}, U_{\alpha_m}^\top H U_{\alpha_m} \rangle \\ &\leq \varepsilon + \sum_{m=1}^r \langle \tilde{v}_{\alpha_m}, \lambda(U_{\alpha_m}^\top H U_{\alpha_m}) \rangle = \varepsilon + \sum_{m=1}^r \langle Q_m v_{\alpha_m}, \lambda(U_{\alpha_m}^\top H U_{\alpha_m}) \rangle \\ &= \varepsilon + \langle v, Q^\top \lambda'(X; H) \rangle \leq \varepsilon + d\theta(\lambda(X))(Q^\top \lambda'(X; H)) \\ &= \varepsilon + d\theta(\lambda(X))(\lambda'(X; H)), \end{aligned}$$

where the last inequality results from the fact that θ is subdifferentially regular, and $v \in \partial\theta(\lambda(X))$ and where the last equality results from (2.8). Letting $\varepsilon \downarrow 0$, we get the opposite inequality in (2.9). \square

3 Twice Epi-Differentiability of Spectral Functions

The main objective of this section is to provide a characterization of twice epi-differentiability of the spectral function g in (1.1) via the same property of the symmetric function θ therein.

To do so, we recall some second-order variational properties, important for our developments in this section. Given a function $f : \mathbf{X} \rightarrow \overline{\mathbf{R}}$ and $\bar{x} \in \mathbf{X}$ with $f(\bar{x})$ finite, denote the parametric family of second-order difference quotients for f at \bar{x} for $\bar{v} \in \mathbf{X}$ as

$$\Delta_t^2 f(\bar{x}, \bar{v})(w) = \frac{f(\bar{x} + tw) - f(\bar{x}) - t\langle \bar{v}, w \rangle}{\frac{1}{2}t^2} \quad \text{with } w \in \mathbf{X}, t > 0.$$

The second subderivative of f at \bar{x} for \bar{v} is defined by

$$d^2 f(\bar{x}, \bar{v})(w) = \liminf_{\substack{t \downarrow 0 \\ w' \rightarrow w}} \Delta_t^2 f(\bar{x}, \bar{v})(w'), \quad w \in \mathbf{X}.$$

Following [43, Definition 13.6], a function $f : \mathbf{X} \rightarrow \overline{\mathbf{R}}$ is said to be twice epi-differentiable at \bar{x} for $\bar{v} \in \mathbf{X}$, with $f(\bar{x})$ finite, if the sets $\text{epi } \Delta_t^2 f(\bar{x}, \bar{v})$ converge to $\text{epi } d^2 f(\bar{x}, \bar{v})$ as $t \downarrow 0$; see [43, Chapter 4] for the definition of a sequence of sets. The latter means by [43, Proposition 7.2] that for every sequence $t_k \downarrow 0$ and every $w \in \mathbf{X}$, there exists a sequence $w_k \rightarrow w$ such that

$$d^2 f(\bar{x}, \bar{v})(w) = \lim_{k \rightarrow \infty} \Delta_{t_k}^2 f(\bar{x}, \bar{v})(w_k). \quad (3.1) \quad \{\text{dedf}\}$$

The second subderivative was introduced in [41], where the concept of twice epi-differentiability for convex functions was studied for the first time. The importance of the second subderivative resides in the fact that it can characterize the quadratic growth condition for optimization problems; see [43, Theorem 13.24]. So, it is crucial for many applications to calculate it in terms of the initial data of an optimization problem. This task was carried out for major classes of functions including the convex piecewise linear-quadratic functions in [43, Proposition 13.9], the second-order cone in [32, Example 5.8], the cone of positive semidefinite symmetric matrices in [33, Example 3.7], and the augmented Lagrangian of constrained optimization problems in [32, Theorem 8.3]. The second subderivative can also be leveraged in the study of stability properties of optimization problems and generalized equations. To this end, it is vital to be able to calculate the second subderivative and characterize twice epi-differentiability for important classes of functions in order to study various stability properties of constrained and composite optimization problems.

We aim in this section to provide a unified approach that simultaneously answers both aforementioned aspects for spectral functions. We begin with the following result that gives a lower bound for the second subderivative of the spectral function in (1.1). We should add that a similar result was recently proved in [33, Proposition 5.3] for convex spectral functions. Here, we extend this result by removing the convexity restriction. While its proof resembles that of [33, Proposition 5.3], it requires careful modifications of the argument therein.

Proposition 3.1 (Lower estimate for second subderivatives). *Assume that $g : \mathbf{S}^n \rightarrow \overline{\mathbf{R}}$ is lsc and has the spectral representation in (1.1) and $Y \in \partial g(X)$. Let $\mu_1 > \dots > \mu_r$ be the distinct eigenvalues of X and $U \in \mathbf{O}^n(X)$. Then, for any $H \in \mathbf{S}^n$ we have*

$$d^2 g(X, Y)(H) \geq d^2 \theta(\lambda(X), v)(\lambda'(X; H)) + 2 \sum_{m=1}^r \langle \text{Diag}(y)_{\alpha_m \alpha_m}, U_{\alpha_m}^\top H (\mu_m I - X)^\dagger H U_{\alpha_m} \rangle, \quad (3.2) \quad \{\text{lbs1}\} \quad \{\text{lbs2}\}$$

where $\alpha_m, m = 1, \dots, r$, are defined in (2.1), and $y \in \partial \theta(\lambda(X))$ such that $Y = U \text{Diag}(y) U^\top$ and $v = (v_{\alpha_1}, \dots, v_{\alpha_r}) = Qy$ for some $Q \in \mathbf{P}_X^n$ with $v_{\alpha_m} \in \mathbf{R}_{\downarrow}^{|\alpha_m|}$ for any $m = 1, \dots, r$. Moreover, for any such permutation matrix Q , we always have

$$d^2 \theta(\lambda(X), v)(\lambda'(X; H)) = d^2 \theta(\lambda(X), y)(Q^\top \lambda'(X; H))$$

Proof. Since $Y \in \partial g(X)$, it follows from Proposition 2.2 that there are $U \in \mathbf{O}^n(X)$ and $y \in \partial\theta(\lambda(X))$ such that $Y = U\text{Diag}(y)U^\top$. Since $y = (y_{\alpha_1}, \dots, y_{\alpha_r})$, we can choose a permutation matrix $Q_m \in \mathbf{S}^{|\alpha_m|}$ such that $v_{\alpha_m} = Q_m y_{\alpha_m}$ with $v_{\alpha_m} \in \mathbf{R}_{\downarrow}^{|\alpha_m|}$ for any $m = 1, \dots, r$. Set $Q = \text{Diag}(Q_1, \dots, Q_r)$ and observe that

$$v = Qy \in Q\partial\theta(\lambda(X)) = \partial\theta(Q\lambda(X)) = \partial\theta(\lambda(X)). \quad (3.3) \quad \{\text{subj}\}$$

Let $H \in \mathbf{S}^n$ and pick sequences $H_k \rightarrow H$ and $t_k \downarrow 0$. Setting $\Delta_{t_k}\lambda(X)(H_k) := (\lambda(X + t_k H_k) - \lambda(X))/t_k$, we get

$$\begin{aligned} \Delta_{t_k}^2 g(X, Y)(H_k) &= \frac{\theta(\lambda(X + t_k H_k)) - \theta(\lambda(X)) - t_k \langle Y, H_k \rangle}{\frac{1}{2}t_k^2} \\ &= \frac{\theta(\lambda(X) + t_k \Delta_{t_k}\lambda(X)(H_k)) - \theta(\lambda(X)) - t_k \langle v, \Delta_{t_k}\lambda(X)(H_k) \rangle}{\frac{1}{2}t_k^2} \\ &\quad + \frac{\langle v, \Delta_{t_k}\lambda(X)(H_k) \rangle - \langle Y, H_k \rangle}{\frac{1}{2}t_k} \\ &= \Delta_{t_k}^2 \theta(\lambda(X), v)(\Delta_{t_k}\lambda(X)(H_k)) + \frac{\langle v, \Delta_{t_k}\lambda(X)(H_k) \rangle - \langle Y, H_k \rangle}{\frac{1}{2}t_k}. \end{aligned}$$

It follows from $Y = U\text{Diag}(y)U^\top$ that

$$\langle Y, H_k \rangle = \langle U\text{Diag}(y)U^\top, H_k \rangle = \langle \text{Diag}(y), U^\top H_k U \rangle = \sum_{m=1}^r \langle \text{Diag}(y)_{\alpha_m \alpha_m}, U_{\alpha_m}^\top H_k U_{\alpha_m} \rangle. \quad (3.4) \quad \{\text{sdcy}\}$$

On the other hand, it results from (2.4) and Fan's inequality [13] that

$$\begin{aligned} \langle v, \Delta_{t_k}\lambda(X)(H_k) \rangle &= \sum_{m=1}^r \sum_{j \in \alpha_m} \frac{v_j (\lambda_j(X + t_k H_k) - \lambda_j(X))}{t_k} \\ &= \sum_{m=1}^r \sum_{j \in \alpha_m} v_j \lambda_{\ell_j} (U_{\alpha_m}^\top H_k U_{\alpha_m} + t_k U_{\alpha_m}^\top H_k (\mu_m I - X)^\dagger H_k U_{\alpha_m}) + O(t_k^2) \\ &\geq \sum_{m=1}^r \langle \text{Diag}(y)_{\alpha_m \alpha_m}, U_{\alpha_m}^\top H_k U_{\alpha_m} + t_k U_{\alpha_m}^\top H_k (\mu_m I - X)^\dagger H_k U_{\alpha_m} \rangle + O(t_k^2). \end{aligned}$$

Combining this with (3.4) brings us to

$$\frac{\langle v, \Delta_{t_k}\lambda(X)(H_k) \rangle - \langle Y, H_k \rangle}{\frac{1}{2}t_k} \geq 2 \sum_{m=1}^r \langle \text{Diag}(y)_{\alpha_m \alpha_m}, U_{\alpha_m}^\top H_k (\mu_m I - X)^\dagger H_k U_{\alpha_m} \rangle + O(t_k).$$

This leads us to the estimate

$$\begin{aligned} \Delta_{t_k}^2 g(X, Y)(H_k) &\geq \Delta_{t_k}^2 \theta(\lambda(X), v)(\Delta_{t_k}\lambda(X)(H_k)) \\ &\quad + 2 \sum_{m=1}^r \langle \text{Diag}(y)_{\alpha_m \alpha_m}, U_{\alpha_m}^\top H_k (\mu_m I - X)^\dagger H_k U_{\alpha_m} \rangle + O(t_k), \end{aligned}$$

which in turn clearly justifies the lower estimate in (3.2) for the second subderivative of g at X for Y since $\Delta_{t_k}\lambda(X)(H_k) \rightarrow \lambda'(X; H)$ as $k \rightarrow \infty$.

To prove the last claim, pick $Q \in \mathbf{P}_X^n$ such that $v = (v_{\alpha_1}, \dots, v_{\alpha_r}) = Qy$ with $v_{\alpha_m} \in \mathbf{R}_{\downarrow}^{|\alpha_m|}$ for any $m = 1, \dots, r$ and observe that $Q^\top = Q^{-1} = \text{Diag}(P_1^{-1}, \dots, P_r^{-1}) \in \mathbf{P}_X^n$. Thus, we obtain for any $w \in \mathbf{R}^n$ that

$$\begin{aligned} d^2\theta(\lambda(X), v)(w) &= \liminf_{\substack{t \downarrow 0 \\ w' \rightarrow w}} \frac{\theta(\lambda(X) + tw') - \theta(\lambda(X)) - t\langle v, w' \rangle}{\frac{1}{2}t^2} \\ &= \liminf_{\substack{t \downarrow 0 \\ w' \rightarrow w}} \frac{\theta(\lambda(X) + tQ^\top w') - \theta(\lambda(X)) - t\langle y, Q^\top w' \rangle}{\frac{1}{2}t^2} \\ &\geq \liminf_{\substack{t \downarrow 0 \\ w' \rightarrow Q^\top w}} \frac{\theta(\lambda(X) + tw') - \theta(\lambda(X)) - t\langle y, w' \rangle}{t} = d^2\theta(\lambda(X), y)(Q^\top w). \end{aligned}$$

One can argue similarly to derive the opposite inequality $d^2\theta(\lambda(X), v)(w) \leq d^2\theta(\lambda(X), y)(Q^\top w)$ for any $w \in \mathbf{R}^n$, which confirms the claimed equality. \square

Note that when the spectral function g in Proposition 3.1 is convex, the vector y therein can be chosen as $\lambda(Y)$, which leads us to $v = y$, where v is taken from Proposition 3.1. In this case, the matrix Q will be the identity matrix and consequently, this result boils down to [33, Proposition 5.3].

Our next goal is to investigate when the inequality in (3.2) becomes equality. Such a task requires knowing more about the critical cone of spectral functions. Recall that the critical of a function $f : \mathbf{X} \rightarrow \overline{\mathbf{R}}$ at x for $v \in \partial f(x)$ is defined by

$$C_f(x, v) = \{w \in \mathbf{X} \mid df(x)(w) = \langle w, v \rangle\}.$$

Note that the critical cone of a function has a close relationship with its second subderivative. To see this, taking $(x, v) \in \text{gph } \partial f$ such that $d^2f(x, v)$ is proper, we can conclude from [43, Proposition 13.5] that $\text{dom } d^2f(x, v) \subset C_f(x, v)$. As shown later in this section, equality requires some additional assumptions; [33, Proposition 3.4] for more details.

The following proposition presents a characterization of the critical cone of the spectral function, and extends [34, Proposition 5.4], where a similar result was obtained for convex spectral functions. Since its proof is similar to that of [34, Proposition 5.4] with similar modifications illustrated in Proposition 3.1, we only provide a sketch of its proof.

Proposition 3.2 (critical cone of spectral functions). *Assume that $g : \mathbf{S}^n \rightarrow \overline{\mathbf{R}}$ is lsc and has the spectral representation in (1.1) and that $Y \in \partial g(X)$. Assume further that one of the following conditions is satisfied:*

- (a) g is subdifferentially regular;
- (b) θ is locally Lipschitz continuous at $\lambda(X)$ relative to its domain and $Y \in \widehat{\partial}g(X)$.

Then, we have $H \in C_g(X, Y)$ if and only if $\lambda'(X; H) \in C_\theta(\lambda(X), v)$ and the matrices $\text{Diag}(y)_{\alpha_m \alpha_m}$ and $U_{\alpha_m}^\top H U_{\alpha_m}$ have a simultaneous ordered spectral decomposition for any $m = 1, \dots, r$, where α_m taken from (2.1) and $U \in \mathbf{O}^n(X)$, where $Y = U \text{Diag}(y) U^\top$ for some $y \in \partial\theta(\lambda(X))$, and where v is taken from Proposition 3.1.

Proof. Suppose that either (a) or (b) holds. Since $Y \in \partial g(X)$, it follows from Proposition 2.2 that there is $U \in \mathbf{O}^n(X)$ and $y \in \partial\theta(\lambda(X))$ such that $Y = U \text{Diag}(y) U^\top$. Take the matrix Q

{crit}

from Proposition 3.1 and observe for any $H \in \mathbf{S}^n$ that

$$\begin{aligned} \langle Y, H \rangle &= \langle \text{Diag}(y), U^\top H U \rangle = \sum_{m=1}^r \langle \text{Diag}(y)_{\alpha_m \alpha_m}, U_{\alpha_m}^\top H U_{\alpha_m} \rangle \\ &\leq \sum_{m=1}^r \langle v_{\alpha_m}, \lambda(U_{\alpha_m}^\top H U_{\alpha_m}) \rangle = \langle v, \lambda'(X; H) \rangle \leq d\theta(\lambda(X))(\lambda'(X; H)) = dg(X)(H), \end{aligned}$$

where the last equality results from [33, Theorem 3.5] when (b) holds and from Theorem 2.3 when (a) is satisfied and where the last inequality follows from $v \in \widehat{\partial}\theta(\lambda(X))$ and [43, Exercise 8.4] in both cases. The rest of the proof can be done as that of [33, Proposition 5.4], which was argued for convex spectral functions. \square

The next result consists of some useful observations that will be utilized in our characterization of twice epi-differentiability of spectral functions.

Lemma 3.3. *Assume that $g : \mathbf{S}^n \rightarrow \overline{\mathbf{R}}$ is lsc and has the spectral representation in (1.1). If $(X, Y) \in \text{gph } \partial g$ and $Y = U \text{Diag}(y) U^\top$ for some $y \in \partial\theta(\lambda(X))$ and $U \in \mathbf{O}^n(X)$, then the following properties hold.*

{tpd_g}

- (a) *Twice epi-differentiability of θ at $\lambda(X)$ for y amounts to that of θ at $\lambda(X)$ for v , where v is defined in Proposition 3.1.*
- (b) *If g is twice epi-differentiable at X for Y , then it enjoys the same property at $\Lambda(X) = \text{Diag}(\lambda(X))$ for $\text{Diag}(y)$.*
- (c) *The inequality*

$$d^2g(\Lambda(X), \text{Diag}(y))(\text{Diag}(w)) \leq d^2\theta(\lambda(X), y)(w)$$

holds for any $w \in \mathbf{R}^n$.

Proof. To prove (a), take the block permutation matrix $Q \in \mathbf{P}_X^n$ from Proposition 3.1 such that $v = Qy$. Take $w \in \mathbf{R}^n$ and a sequence $t_k \downarrow 0$, and assume that θ is twice epi-differentiable at $\lambda(X)$ for y . So, we find a sequence $w_k \rightarrow Q^\top w$ such that $\Delta_{t_k}^2 \theta(\lambda(X), y)(w_k) \rightarrow d^2\theta(\lambda(X), y)(Q^\top w)$. It is not hard to see that

$$\Delta_{t_k}^2 \theta(\lambda(X), v)(Qw_k) = \Delta_{t_k}^2 \theta(\lambda(X), y)(w_k).$$

This brings us to

$$\begin{aligned} \lim_{k \rightarrow \infty} \Delta_{t_k}^2 \theta(\lambda(X), v)(Qw_k) &= \lim_{k \rightarrow \infty} \Delta_{t_k}^2 \theta(\lambda(X), y)(w_k) \\ &= d^2\theta(\lambda(X), y)(Q^\top w) = d^2\theta(\lambda(X), v)(w), \end{aligned}$$

where the last equality was justified at the end of the proof of Proposition 3.1. This demonstrates that θ is twice epi-differentiable at $\lambda(X)$ for v . The opposite implication can be justified similarly.

To prove (b), it can be checked directly from (2.7) that

$$\begin{cases} d^2g(\Lambda(X), \text{Diag}(y))(W) = d^2g(X, Y)(UWU^\top) & \text{and} \\ \Delta_{t_k}^2 g(X, Y)(W) = \Delta_{t_k}^2 g(\Lambda(X), \text{Diag}(y))(U^\top W U) \end{cases} \quad (3.5) \quad \{\mathbf{n_ted}\}$$

for any $W \in \mathbf{S}^n$ and any $U \in \mathbf{O}^n(X)$ such that $Y = U \text{Diag}(y) U^\top$. Take $H \in \mathbf{S}^n$ and a sequence $t_k \downarrow 0$. Since g is twice epi-differentiable at X for Y , there are sequences $H_k \rightarrow UH U^\top$ such that

$\Delta_{t_k}^2 g(X, Y)(H_k) \rightarrow d^2 g(X, Y)(UHU^\top)$. Set $\widehat{H}_k := U^\top H_k U$ and observe that $\Delta_{t_k}^2 g(X, Y)(H_k) = \Delta_{t_k}^2 g(\Lambda(X), \text{Diag}(y))(\widehat{H}_k)$. This, together with (3.5), leads us to

$$\begin{aligned} \lim_{k \rightarrow \infty} \Delta_{t_k}^2 g(\Lambda(X), \text{Diag}(y))(\widehat{H}_k) &= \lim_{k \rightarrow \infty} \Delta_{t_k}^2 g(X, Y)(H_k) \\ &= d^2 g(X, Y)(UHU^\top) = d^2 g(\Lambda(X), \text{Diag}(y))(H), \end{aligned}$$

which justifies (b).

Turning now to (c), we know from (2.7) that the symmetric function θ can be represented as the composite function $\theta = g \circ F$ with $F(x) := \text{Diag}(x)$ for any $x \in \mathbf{R}^n$. This allows us to use the extensive theory of variational properties of composite functions (cf. [31, 33, 43]) to justify (c). Note that since $y \in \partial\theta(\lambda(X))$, we deduce from [27, Theorem 5] that $\text{Diag}(y) \in \partial g(\text{Diag}(\lambda(X)))$. Using this and the fact that $\nabla F(\lambda(X))^* \text{Diag}(y) = y$, we can view $\text{Diag}(y)$ as a Lagrange multiplier associated with y for the composite representation of θ . Therefore, a similar argument as the proof of [43, Proposition 13.14] gives us the claimed inequality. \square

After these preparations, we are now in a position to present our main result in this section in which a characterization of twice epi-differentiability of spectral functions will be given via the same property of their symmetric parts. We should add here that twice epi-differentiability for some classes of eigenvalue functions, including the maximum and leading eigenvalue functions, was justified first by Torki in [48]. The latter, however, did not provide a similar characterization as the one below by relating it to the same property of the symmetric part of a spectral function.

Theorem 3.4. *Assume that $g : \mathbf{S}^n \rightarrow \overline{\mathbf{R}}$ is lsc and has the spectral representation in (1.1) and that θ is locally Lipschitz continuous relative to its domain. If $(X, Y) \in \text{gph } \widehat{\partial}g$ and $Y = U \text{Diag}(y) U^\top$ for some $y \in \widehat{\partial}\theta(\lambda(X))$ and $U \in \mathbf{O}^n(X)$, then g is twice epi-differentiable at X for Y if and only if θ is twice epi-differentiable at $\lambda(X)$ for y . Moreover, if one of these equivalent conditions holds, then the second subderivative of g at X for Y is calculated by*

$$d^2 g(X, Y)(H) = d^2 \theta(\lambda(X), v)(\lambda'(X; H)) + 2 \sum_{m=1}^r \langle \text{Diag}(y)_{\alpha_m \alpha_m}, U_{\alpha_m}^\top H (\mu_m I - X)^\dagger H U_{\alpha_m} \rangle, \quad (3.6)$$

for any $H \in \mathbf{S}^n$ such that $d^2 g(X, Y)(H) < \infty$, where v is defined in Proposition 3.1 and where α_m , $m = 1, \dots, r$, are defined in (2.1) and where $\mu_1 > \dots > \mu_r$ are the distinct eigenvalues of X .

Proof. Assume first that θ is twice epi-differentiable at $\lambda(X)$ for y and take $H \in \mathbf{S}^n$. Our goal is to demonstrate the validity of (3.1) for the following two cases: 1) $d^2 g(X, Y)(H) = \infty$; 2) $d^2 g(X, Y)(H) < \infty$. For the first case, take a sequence $t_k \downarrow 0$ and set $H_k := H$ for each k and observe that

$$\infty = d^2 g(X, Y)(H) \leq \liminf_{k \rightarrow \infty} \Delta_{t_k}^2 g(X, Y)(H) \leq \limsup_{k \rightarrow \infty} \Delta_{t_k}^2 g(X, Y)(H) \leq \infty = d^2 g(X, Y)(H).$$

This implies that $\lim_{k \rightarrow \infty} \Delta_{t_k}^2 g(X, Y)(H) = d^2 g(X, Y)(H)$, which confirms (3.1) for this case. Turning now to the second case, observe that given a sequence $t_k \downarrow 0$, it suffices to construct a sequence $H_k \rightarrow H$ such that

$$\lim_{k \rightarrow \infty} \Delta_{t_k}^2 g(X, Y)(H_k) = d^2 \theta(\lambda(X), v)(\lambda'(X; H)) + 2 \sum_{m=1}^r \langle \text{Diag}(y)_{\alpha_m \alpha_m}, U_{\alpha_m}^\top H (\mu_m I - X)^\dagger H U_{\alpha_m} \rangle,$$

where y and v are defined in Proposition 3.1. Indeed, doing so leads us via Proposition 3.1 to

$$\lim_{k \rightarrow \infty} \Delta_{t_k}^2 g(X, Y)(H_k) = d^2 g(X, Y)(H),$$

which proves (3.1) for the second case. According to Lemma 3.3(a), θ is twice epi-differentiable at $\lambda(X)$ for v . Thus, we find a sequence $\{w^k\}_{k \in \mathbf{N}}$, converging to $\lambda'(X; H)$, such that

$$\Delta_{t_k}^2 \theta(\lambda(X), v)(w^k) \rightarrow d^2 \theta(\lambda(X), v)(\lambda'(X; H)). \quad (3.7) \quad \{\text{seq1}\}$$

Note that w^k has an equivalent representation $w^k = (w_{\alpha_1}^k, \dots, w_{\alpha_r}^k)$ with $w_{\alpha_m}^k \in \mathbf{R}^{|\alpha_m|}$ for any $m = 1, \dots, r$. For each k and each $m \in \{1, \dots, r\}$, we can find a permutation matrices G_m^k such that $\tilde{w}_{\alpha_m}^k := G_m^k w_{\alpha_m}^k \in \mathbf{R}_{\downarrow}^{|\alpha_m|}$. Since the entries of G_m^k are either 0 or 1, we can assume by passing to a subsequence if necessary that for each $m \in \{1, \dots, r\}$ the sequence $\{G_m^k\}_{k \in \mathbf{N}}$ is constant. For any $m \in \{1, \dots, r\}$, set $G_m := G_m^k$ and define the block diagonal permutation matrix $G \in \mathbf{P}_X^n$ by $G := \text{Diag}(G_1, \dots, G_r)$. It follows from $w^k \rightarrow \lambda'(X; H)$ that $\tilde{w}^k \rightarrow G\lambda'(X; H)$. Since $\tilde{w}^k \in \mathbf{R}_{\downarrow}^{|\alpha_1|} \times \dots \times \mathbf{R}_{\downarrow}^{|\alpha_r|}$, we get $G\lambda'(X; H) \in \mathbf{R}_{\downarrow}^{|\alpha_1|} \times \dots \times \mathbf{R}_{\downarrow}^{|\alpha_r|}$. On the other hand, we already have

$$\lambda'(X; H) = (\lambda(U_{\alpha_1}^\top H U_{\alpha_1}), \dots, \lambda(U_{\alpha_r}^\top H U_{\alpha_r})) \in \mathbf{R}_{\downarrow}^{|\alpha_1|} \times \dots \times \mathbf{R}_{\downarrow}^{|\alpha_r|}. \quad (3.8) \quad \{\text{dir_eigen}\}$$

Combining these and remembering that G is a permutation matrix allows us to conclude that

$$G\lambda'(X; H) = \lambda'(X; H), \quad (3.9) \quad \{\text{eq:G}\}$$

which in turn implies that $\tilde{w}^k \rightarrow \lambda'(X; H)$. Moreover, since $G \in \mathbf{P}_X^n$ and θ is symmetric, we obtain for any k that

$$\theta(\lambda(X) + t_k w^k) = \theta(\lambda(X) + t_k \tilde{w}^k). \quad (3.10) \quad \{\text{th_1m}\}$$

By $d^2 g(X, Y)(H) < \infty$ and the lower estimate in (3.2), we get $d^2 \theta(\lambda(X), v)(\lambda'(X; H)) < \infty$, which coupled with (3.7) tells us that $\lambda(X) + t_k w^k \in \text{dom } \theta$ for all k sufficiently large. Combining it with (3.10), we get for all k sufficiently large that

$$\lambda(X) + t_k \tilde{w}^k \in \text{dom } \theta. \quad (3.11) \quad \{\text{th_1m2}\}$$

Since $d^2 g(X, Y)(H) < \infty$, we deduce from [43, Proposition 13.5] that $dg(X)(H) \leq \langle H, Y \rangle$, which together with $Y \in \partial g(X)$ yields $H \in C_g(X, Y)$. We thus conclude from Proposition 3.2 that the matrices $\text{Diag}(y)_{\alpha_m \alpha_m}$ and $U_{\alpha_m}^\top H U_{\alpha_m}$ have a simultaneous ordered spectral decomposition for any $m = 1, \dots, r$. Thus, we find for each $m = 1, \dots, r$ an orthogonal matrix $P_m \in \mathbf{O}^{|\alpha_m|}(\text{Diag}(y)_{\alpha_m \alpha_m}) \cap \mathbf{O}^{|\alpha_m|}(U_{\alpha_m}^\top H U_{\alpha_m})$ such that

$$\text{Diag}(y)_{\alpha_m \alpha_m} = P_m \text{Diag}(v)_{\alpha_m \alpha_m} P_m^\top \quad \text{and} \quad U_{\alpha_m}^\top H U_{\alpha_m} = P_m \Lambda(U_{\alpha_m}^\top H U_{\alpha_m}) P_m^\top. \quad (3.12) \quad \{\text{spc_ref}\}$$

Take the matrix P_m , $m = 1, \dots, r$, and define the matrix

$$A_k := \text{Diag}(P_1 \text{Diag}(w_{\alpha_1}^k) P_1^\top, \dots, P_r \text{Diag}(w_{\alpha_r}^k) P_r^\top) + \widehat{H}, \quad k \in \mathbf{N}, \quad (3.13) \quad \{\text{mata}\}$$

and

$$D := \text{Diag}(U_{\alpha_1}^\top D_1 U_{\alpha_1}, \dots, U_{\alpha_r}^\top D_r U_{\alpha_r})$$

where

$$\widehat{H}_{ij} = \begin{cases} 0 & \text{if } i, j \in \alpha_m \text{ for any } m = 1, \dots, r, \\ \tilde{H}_{ij} & \text{otherwise,} \end{cases}$$

with $\tilde{H} := U^\top H U$ and $D_m := H(\mu_m I - X)^\dagger H$ for any $m = 1, \dots, r$, and then set

$$E_k := U(A_k - t_k D)U^\top, \quad k \in \mathbf{N}. \quad (3.14) \quad \{\text{what}\}$$

Using the second identity in (3.12) and (3.8), one can easily see for each $m = 1, \dots, r$ that

$$P_m \text{Diag}(w_{\alpha_m}^k) P_m^\top \rightarrow P_m \text{Diag}(\lambda(U_{\alpha_m}^\top H U_{\alpha_m})) P_m^\top = P_m \Lambda(U_{\alpha_m}^\top H U_{\alpha_m}) P_m^\top = U_{\alpha_m}^\top H U_{\alpha_m},$$

which in turn results in $E_k \rightarrow U^\top \tilde{H} U = H$. Since $\langle Y, U \hat{H} U^\top \rangle = \langle \text{Diag}(y), \hat{H} \rangle = 0$, the identity $U_{\alpha_m}^\top U A_k U^\top U_{\alpha_m} = P_m \text{Diag}(w_{\alpha_m}^k) P_m^\top$, $m = 1, \dots, r$, brings us to

$$\begin{aligned} \langle Y, U A_k U^\top \rangle &= \sum_{m=1}^r \langle \text{Diag}(y)_{\alpha_m \alpha_m}, U_{\alpha_m}^\top (U A_k U^\top) U_{\alpha_m} \rangle + \langle Y, U \hat{H} U^\top \rangle \\ &= \sum_{m=1}^r \langle \text{Diag}(y)_{\alpha_m \alpha_m}, P_m \text{Diag}(w_{\alpha_m}^k) P_m^\top \rangle = \sum_{m=1}^r \langle P_m^\top \text{Diag}(y)_{\alpha_m \alpha_m} P_m, \text{Diag}(w_{\alpha_m}^k) \rangle \\ &= \sum_{m=1}^r \langle \text{Diag}(v_{\alpha_m}), \text{Diag}(w_{\alpha_m}^k) \rangle = \langle v, w^k \rangle, \end{aligned}$$

where the forth equality results from the first relationship in (3.12). Thus, we obtain

$$\begin{aligned} \frac{\langle v, w^k \rangle - \langle Y, E_k \rangle}{\frac{1}{2} t_k} &= \frac{\langle v, w^k \rangle - \langle Y, U A_k U^\top \rangle}{\frac{1}{2} t_k} + 2 \langle U \text{Diag}(y) U^\top, U D U^\top \rangle \\ &= 2 \sum_{m=1}^r \langle \text{Diag}(y)_{\alpha_m \alpha_m}, U_{\alpha_m}^\top (U D_m U^\top) U_{\alpha_m} \rangle \\ &= 2 \sum_{m=1}^r \langle \text{Diag}(y)_{\alpha_m \alpha_m}, U_{\alpha_m}^\top H (\mu_m I - X)^\dagger H U_{\alpha_m} \rangle. \end{aligned} \quad (3.15) \quad \{\text{ted_g2}\}$$

Take $m \in \{1, \dots, r\}$ and $i \in \alpha_m$. Appealing now to (2.4), we can conclude that

$$\begin{aligned} &\lambda_i(X + t_k E_k) - \lambda_i(X) \\ &= \lambda_{\ell_i}(t_k P_m \text{Diag}(w_{\alpha_m}^k) P_m^\top - t_k^2 (U_{\alpha_m}^\top D_m U_{\alpha_m} - U_{\alpha_m}^\top U A_k U^\top (\mu_m I - X)^\dagger U A_k U^\top U_{\alpha_m})) + o(t_k^2) + O(t_k^3). \end{aligned}$$

As argued above, we have $U A_k U^\top \rightarrow H$, which ensures that

$$t_k^2 (U_{\alpha_m}^\top D_m U_{\alpha_m} - U_{\alpha_m}^\top U A_k U^\top (\mu_m I - X)^\dagger U A_k U^\top U_{\alpha_m}) = o(t_k^2).$$

Thus, we arrive at

$$\lambda_i(X + t_k E_k) - \lambda_i(X) = \lambda_{\ell_i}(t_k P_m \text{Diag}(w_{\alpha_m}^k) P_m^\top + o(t_k^2)) + O(t_k^3).$$

This, coupled with the Lipschitz continuity of eigenvalues, leads us for each $m \in \{1, \dots, r\}$ to

$$\begin{aligned} \lambda_{\alpha_m}(X + t_k E_k) - \lambda_{\alpha_m}(X) - t_k \tilde{w}_{\alpha_m}^k &= t_k \lambda(P_m \text{Diag}(w_{\alpha_m}^k) P_m^\top) + o(t_k^2) - t_k \tilde{w}_{\alpha_m}^k \\ &= t_k \tilde{w}_{\alpha_m}^k + o(t_k^2) - t_k \tilde{w}_{\alpha_m}^k = o(t_k^2). \end{aligned}$$

Thus, we arrive at

$$\lambda(X + t_k E_k) - \lambda(X) - t_k \tilde{w}^k = o(t_k^2). \quad (3.16) \quad \{\text{taylor_1}\}$$

Note that we can not guarantee that $X + t_k E_k \in \text{dom } g$. To resolve this issue, we can utilize [7, Proposition 2.3] to obtain for any k that

$$\text{dist}(X + t_k E_k, \text{dom } g) = \text{dist}(\lambda(X + t_k E_k), \text{dom } \theta),$$

which together with (3.16) and (3.11) brings us to the relationships

$$\begin{aligned}
\text{dist}\left(E_k, \frac{\text{dom } g - X}{t_k}\right) &= \frac{1}{t_k} \text{dist}(\lambda(X + t_k E_k), \text{dom } \theta) \\
&\leq \frac{1}{t_k} \|\lambda(X) + t_k \tilde{w}^k + o(t_k^2) - \lambda(X) - t_k \tilde{w}^k\| \\
&= o(t_k) \text{ for all } k \in \mathbf{N}.
\end{aligned} \tag{3.17} \quad \{\text{eq:indom}\}$$

Thus, for each k sufficiently large, we can find a matrix $H_k \in \mathbf{S}^n$ such that $X + t_k H_k \in \text{dom } g$ and $H_k - E_k = o(t_k)$. Using this observation, one can argue that

$$\begin{aligned}
\Delta_{t_k}^2 g(X, Y)(H_k) &= \frac{\theta(\lambda(X + t_k H_k)) - \theta(\lambda(X)) - t_k \langle Y, H_k \rangle}{\frac{1}{2} t_k^2} \\
&= \Delta_{t_k}^2 \theta(\lambda(X), v)(w^k) + \frac{\langle v, w^k \rangle - \langle Y, H_k \rangle}{\frac{1}{2} t_k} + \frac{\theta(\lambda(X + t_k H_k)) - \theta(\lambda(X) + t_k w^k)}{\frac{1}{2} t_k^2} \\
&= \Delta_{t_k}^2 \theta(\lambda(X), v)(w^k) + \frac{\langle v, w^k \rangle - \langle Y, E_k \rangle}{\frac{1}{2} t_k} + \frac{o(t_k)}{t_k} + \frac{\theta(\lambda(X + t_k H_k)) - \theta(\lambda(X) + t_k \tilde{w}^k)}{\frac{1}{2} t_k^2},
\end{aligned} \tag{3.18} \quad \{\text{ted_g}\}$$

where the last equality results from (3.10). Using the fact that θ is locally Lipschitz continuous at $\lambda(X)$ with respect to its domain, we find $\ell \geq 0$ such that for all k sufficiently large we have

$$\begin{aligned}
|\theta(\lambda(X + t_k H_k)) - \theta(\lambda(X) + t_k \tilde{w}^k)| &\leq \ell \|\lambda(X + t_k H_k) - \lambda(X) - t_k \tilde{w}^k\| \\
&\leq \ell \|\lambda(X + t_k E_k) - \lambda(X) - t_k \tilde{w}^k\| + o(t_k^2) \\
&= o(t_k^2),
\end{aligned} \tag{3.19} \quad \{\text{eq:liph}\}$$

where the last inequality results from $H_k - E_k = o(t_k)$ and where the last equality comes from (3.16). Passing to the limit in (3.18) and using (3.7) and (3.15) lead us to

$$\Delta_{t_k}^2 g(X, Y)(H_k) \rightarrow d^2 \theta(\lambda(X), v)(\lambda'(X; H)) + 2 \sum_{m=1}^r \langle \text{Diag}(y)_{\alpha_m \alpha_m}, U_{\alpha_m}^\top H (\mu_m I - X)^\dagger H U_{\alpha_m} \rangle, \tag{3.20} \quad \{\text{eq:kinfy}\}$$

which ensures (3.1) for the second case and consequently justifies twice epi-differentiability of g at X for Y . Moreover, it confirms the claimed formula for the second subderivative of g at X for Y .

Turning now to the proof of the opposite implication, assume that g is twice epi-differentiable at X for Y . We again consider the following two cases: 1) $d^2 \theta(\lambda(X), y)(h) = \infty$; 2) $d^2 \theta(\lambda(X), y)(h) < \infty$. For the first case, its verification can use a similar argument as the proof of the first case in the previous implication. To ensure twice epi-differentiability of θ at $\lambda(X)$ for y in the second case, observe first via Lemma 3.3(b) that g is twice epi-differentiable at $\Lambda(X)$ for $\text{Diag}(y)$. Pick any $h \in \mathbb{R}^n$ and denote $H = \text{Diag}(h) \in \mathbf{S}^n$. It follows from twice epi-differentiability of g at $\Lambda(X)$ for $\text{Diag}(y)$ that for any $t_k \downarrow 0$, there exists a sequence $\{H_k\}_{k \in \mathbf{N}}$, converging to H , such that

$$\Delta_{t_k}^2 g(\Lambda(X), \text{Diag}(y))(H_k) \rightarrow d^2 g(\Lambda(X), \text{Diag}(y))(H).$$

Suppose $h^\downarrow = (\tilde{h}_{\alpha_1}, \dots, \tilde{h}_{\alpha_r}) = Qh$ with $\tilde{h}_{\alpha_m} := Q_m h_{\alpha_m} \in \mathbf{R}_\downarrow^{|\alpha_m|}$, $Q = \text{Diag}(Q_1, \dots, Q_r)$, $Q_m \in \mathbf{P}^{|\alpha_m|}$. Let $V^k = \text{Diag}(V_1^k, \dots, V_r^k)$ with $V_m^k \in \mathbf{O}^{|\alpha_m|}((H_k)_{\alpha_m \alpha_m})$. It can be checked directly that

$$\begin{aligned}
g(\Lambda(X) + t_k H_k) &= g(\Lambda(X) + t_k \text{Diag}((H_k)_{\alpha_1 \alpha_1}, \dots, (H_k)_{\alpha_r \alpha_r}) + t_k \text{OD}(H_k)) \\
&= g(QV^k \Lambda(X)(V^k)^\top Q^\top + t_k Q \text{Diag}(\Lambda((H_k)_{\alpha_1 \alpha_1}), \dots, \Lambda((H_k)_{\alpha_r \alpha_r})) Q^\top + t_k QV^k \text{OD}(H_k)(V^k)^\top Q^\top) \\
&= g(\Lambda(X) + t_k Q \text{Diag}(\Lambda((H_k)_{\alpha_1 \alpha_1}), \dots, \Lambda((H_k)_{\alpha_r \alpha_r})) Q^\top + t_k QV^k \text{OD}(H_k)(V^k)^\top Q^\top),
\end{aligned}$$

where $\text{OD}(H_k) := H_k - \text{Diag}((H_k)_{\alpha_1\alpha_1}, \dots, (H_k)_{\alpha_r\alpha_r})$ denotes the off block diagonal part of H_k and where the last equality comes from the block diagonal structures of $\Lambda(X)$, Q , and V_k . Let

$$\widehat{H}_k := Q\text{Diag}(\Lambda((H_k)_{\alpha_1\alpha_1}), \dots, \Lambda((H_k)_{\alpha_r\alpha_r}))Q^\top + QV^k\text{OD}(H_k)(V^k)^\top Q^\top.$$

It can be checked directly that

$$d^2g(\Lambda(X), \text{Diag}(y))(H) = \lim_{k \rightarrow \infty} \Delta_{t_k}^2 g(\Lambda(X), \text{Diag}(y))(\widehat{H}_k) = \lim_{k \rightarrow \infty} \Delta_{t_k}^2 g(\Lambda(X), \text{Diag}(y))(H_k). \quad \{\text{eq:p4}\}$$

Let h_k be such that $(h_k)_{\alpha_m} = Q_m \lambda((\widehat{H}_k)_{\alpha_m\alpha_m} + t_k(\widehat{H}_k)_{\alpha_m}^\top (\mu_m I - \Lambda(X))^\dagger (\widehat{H}_k)_{\alpha_m})$ for any $m \in \{1, \dots, r\}$. Since $\widehat{H}_k \rightarrow \text{Diag}(h)$, we know for all $m = 1, \dots, r$ that

$$(\widehat{H}_k)_{\alpha_m}^\top (\mu_m I - \Lambda(X))^\dagger (\widehat{H}_k)_{\alpha_m} \rightarrow 0 \quad \{\text{eq:15}\}$$

as $k \rightarrow \infty$, which implies that $h_k \rightarrow h$. It follows from the definition of h_k that for k sufficiently large,

$$\lambda(\Lambda(X) + t_k \widehat{H}_k) - \lambda(\Lambda(X) + t_k \text{Diag}(h_k)) = \lambda(\Lambda(X) + t_k \widehat{H}_k) - \lambda(\Lambda(X)) - t_k Q h_k.$$

Moreover, using (2.4) again brings us to

$$\lambda_{\alpha_m}(\Lambda(X) + t_k \widehat{H}_k) - \lambda_{\alpha_m}(\Lambda(X)) = \lambda_{\alpha_m}(t_k(\widehat{H}_k)_{\alpha_m\alpha_m} + t_k^2(\widehat{H}_k)_{\alpha_m}^\top (\mu_m I - \Lambda(X))^\dagger (\widehat{H}_k)_{\alpha_m}) + o(t_k^2).$$

Combining the above two equalities, we arrive at

$$\lambda(\Lambda(X) + t_k \widehat{H}_k) - \lambda(\Lambda(X) + t_k \text{Diag}(h_k)) = o(t_k^2). \quad \{\text{eq:p2}\}$$

Since $h \in \text{dom } d^2\theta(\lambda(X), y)$, we can use Lemma 3.3(c) and (3.21) to ensure that $\Lambda(X) + t_k \widehat{H}_k \in \text{dom } g$ for k sufficiently large. Consequently, using [7, Proposition 2.3], (3.23), and θ being symmetric, we get for any k that

$$\begin{aligned} 0 &= \text{dist}(\Lambda(X) + t_k \widehat{H}_k, \text{dom } g) = \text{dist}(\lambda(\Lambda(X) + t_k \widehat{H}_k), \text{dom } \theta) \\ &= \text{dist}(\lambda(\Lambda(X) + t_k \text{Diag}(h_k)), \text{dom } \theta) + o(t_k^2) = \text{dist}(\lambda(X) + t_k h_k, \text{dom } \theta) + o(t_k^2). \end{aligned}$$

Thus, for each k sufficiently large, we can find $l_k \in \mathbf{R}^n$ such that $\lambda(X) + t_k l_k \in \text{dom } \theta$ and $h_k - l_k = o(t_k)$. Simple calculation then leads us to

$$\begin{aligned} \Delta_{t_k}^2 \theta(\lambda(X), y)(l_k) &= \frac{\theta(\lambda(X) + t_k l_k) - \theta(\lambda(X)) - t_k \langle y, l_k \rangle}{\frac{1}{2} t_k^2} \\ &= \frac{g(\Lambda(X) + t_k \text{Diag}(l_k)) - g(\Lambda(X)) - t_k \langle \text{Diag}(y), \text{Diag}(h_k) \rangle}{\frac{1}{2} t_k^2} + \frac{o(t_k)}{t_k} \\ &= \frac{g(\Lambda(X) + t_k \widehat{H}_k) - \theta(\lambda(X)) - t_k \langle \text{Diag}(y), \widehat{H}_k \rangle}{\frac{1}{2} t_k^2} + \frac{g(\Lambda(X) + t_k \text{Diag}(l_k)) - g(\Lambda(X) + t_k \widehat{H}_k)}{\frac{1}{2} t_k^2} \\ &\quad + \frac{\langle \text{Diag}(y), \widehat{H}_k \rangle - \langle y, h_k \rangle}{\frac{1}{2} t_k} + o(1). \end{aligned} \quad \{\text{eq:qtheta}\}$$

Since θ is locally Lipschitz continuous at $\lambda(X)$ with respect to its domain, we obtain

$$|g(\Lambda(X) + t_k \text{Diag}(h_k)) - g(\Lambda(X) + t_k \widehat{H}_k)| \leq l \|\lambda(\Lambda(X) + t_k \widehat{H}_k) - \lambda(\Lambda(X) + t_k \text{Diag}(h_k))\|,$$

which together with (3.23) yeilds

$$g(\Lambda(X) + t_k \text{Diag}(h_k)) - g(\Lambda(X) + t_k \widehat{H}_k) = o(t_k^2). \quad \{\text{eq:p3}\}$$

Moreover, we have

$$\frac{\langle \text{Diag}(y), \widehat{H}_k \rangle - \langle y, h_k \rangle}{\frac{1}{2}t_k} = \frac{\langle Qy, \text{diag}(\lambda((H_k)_{\alpha_1\alpha_1}), \dots, \lambda((H_k)_{\alpha_r\alpha_r})) \rangle - \langle Qy, Qh_k \rangle}{\frac{1}{2}t_k} \quad (3.26) \quad \{\text{eq:p5}\}$$

Combining (3.22) with the definition of h_k , we can conclude for any sufficiently large k that

$$\begin{aligned} & \|\lambda((H_k)_{\alpha_m\alpha_m}) - \lambda((\widehat{H}_k)_{\alpha_m\alpha_m} + t_k(\widehat{H}_k)_{\alpha_m}^\top (\mu_m I - \Lambda(X))^\dagger (\widehat{H}_k)_{\alpha_m})\| \\ &= \|\lambda((\widehat{H}_k)_{\alpha_m\alpha_m}) - \lambda((\widehat{H}_k)_{\alpha_m\alpha_m} + t_k(\widehat{H}_k)_{\alpha_m}^\top (\mu_m I - \Lambda(X))^\dagger (\widehat{H}_k)_{\alpha_m})\| \\ &\leq \|t_k(\widehat{H}_k)_{\alpha_m}^\top (\mu_m I - \Lambda(X))^\dagger (\widehat{H}_k)_{\alpha_m}\| = o(t_k), \end{aligned}$$

which, coupled with (3.26), tell us that

$$\frac{\langle \text{Diag}(y), \widehat{H}_k \rangle - \langle y, h_k \rangle}{\frac{1}{2}t_k} = o(1).$$

Combining this with (3.24), (3.21), and (3.25) implies that

$$\Delta_{t_k}^2 \theta(\lambda(X), y)(l_k) \rightarrow d^2 g(\Lambda(X), \text{Diag}(y))(\text{Diag}(h)),$$

which together with (3.2) confirms that θ is twice epi-differentiable at $\lambda(X)$ for y and hence completes the proof. \square

Theorem 3.4 provides a practical characterization of twice epi-differentiability of spectral functions and can be viewed as a far-reaching extension of [48, Theorem 3.1] in which this property was justified for some classes of eigenvalue functions. Note that Lipschitz continuity of θ with respect to its domain, assumed in Theorem 3.4, doesn't impose a major restriction in applications and often holds automatically.

Remark 3.5. Note that a characterization of directionally differentiability of the proximal mapping of spectral functions can be found in [10, Theorem 3]. It is well-known that if the spectral function g in (1.1) is lsc and convex, its proximal mapping can be calculated by

$$\text{prox}_g(X) := \text{argmin}_{W \in \mathbf{S}^n} \{g(W) + \frac{1}{2}\|W - X\|^2\} = U \text{Diag}(\text{prox}_\theta(\lambda(X)))U^\top, \quad (3.27) \quad \{\text{prox_con}\}$$

where $U \in \mathbf{O}^n(X)$. Given an lsc and convex spectral function g with representation (1.1), we infer from [10, Theorem 3] that prox_g is directionally differentiable at $X \in \mathbf{S}^n$ if and only if prox_θ is directionally differentiable at $\lambda(X)$. It also follows from [43, Exercise 13.45] that twice epi-differentiability of g at X for $Y \in \partial g(X)$ amounts to directional differentiability of prox_g at $X + Y$. One may wonder whether the combination of these two results can give us the equivalence established in Theorem 3.4 for convex spectral functions. To explore such a possibility, take $(X, Y) \in \text{gph } \partial g$ and assume that g is twice epi-differentiable at X for Y . According to [43, Exercise 13.45], the proximal mapping prox_g is directionally differentiable at $X + Y$. Employing now [10, Theorem 3] implies that prox_θ is directionally differentiable at $\lambda(X + Y)$. Appealing again to [43, Exercise 13.45] indicates that θ is twice epi-differentiable at z for $\lambda(X + Y) - z$ with $z = \text{prox}_\theta(\lambda(X + Y))$. It is not hard to see from (3.27) and the spectral representation of X that $z = \lambda(X)$. On the other hand, since $Y \in \partial g(X)$ and g is convex, there exists $U \in \mathbf{O}^n$ such that $X = U\Lambda(X)U^\top$ and $Y = U\Lambda(Y)U^\top$, which indicate that

$$X + Y = U\Lambda(X)U^\top + U\Lambda(Y)U^\top = U(\Lambda(X) + \Lambda(Y))U^\top.$$

Thus we have $\lambda(X + Y) = \lambda(X) + \lambda(Y)$. Combining these tells us that θ is twice epi-differentiable at $\lambda(X)$ for $\lambda(Y)$. A similar argument can prove the opposite implication, which demonstrates that for convex spectral functions, Theorem 3.4 can be concluded from [10, Theorem 3].

The formula for the second subderivative, established in Theorem 3.4, deals with the domain of this object and so one may wonder what can be said about this set. Under the hypotheses of Theorem 3.4, we can conclude from [43, Proposition 13.5] that if $d^2g(X, Y)$ is a proper function, then we always have the inclusion

$$\text{dom } d^2g(X, Y) := \{H \in \mathbf{S}^n \mid d^2g(X, Y)(H) < \infty\} \subset C_g(X, Y). \quad (3.28) \quad \{\text{dom_ssub}\}$$

Note that equality happens to be valid in this inclusion in many applications. To elaborate more on this, recall that for a function $f : \mathbf{X} \rightarrow \overline{\mathbf{R}}$, and $\bar{x} \in \mathbf{X}$ with $f(\bar{x})$ finite and $w \in \mathbf{X}$ with $df(\bar{x})(w)$ finite, the parabolic subderivative (see [43, Definition 13.59]) of f at \bar{x} for w with respect to z is defined by

$$d^2f(\bar{x})(w \mid z) = \liminf_{\substack{t \searrow 0 \\ z' \rightarrow z}} \frac{f(\bar{x} + tw + \frac{1}{2}t^2z') - f(\bar{x}) - tdf(\bar{x})(w)}{\frac{1}{2}t^2}.$$

It was shown in [33, Proposition 3.4] that if Y is a proximal subgradient in the sense of [43, Definition 8.45] and if $\text{dom } d^2g(X)(H \mid \cdot) \neq \emptyset$ for all $H \in C_g(X, Y)$, the inclusion in (3.28) becomes equality. One may ask when the condition $\text{dom } d^2g(X)(H \mid \cdot) \neq \emptyset$ holds. Indeed, this condition does hold when the function g is parabolically epi-differentiable at X for any H in the sense of [43, Definition 13.59]. The latter condition was extensively studied recently in [34, Theorem 4.7] for spectral functions. Roughly speaking, it was shown that the latter property holds for the spectral function g in (1.1) whenever the symmetric function θ in (1.1) enjoys this property. The latter automatically holds for many symmetric functions that appear in applications. We record below a result that provides sufficient conditions for ensuring equality in (3.28).

Proposition 3.6. *Suppose that g has the spectral representation in (1.1) and Y is a proximal subgradient of g at X and that θ is locally Lipschitz continuous with respect to its domain. If θ is parabolically epi-differentiable at $\lambda(X)$ for $\lambda'(X; H)$ for any $H \in T_{\text{dom } g}(X)$, then we have $\text{dom } d^2g(X, Y) = C_g(X, Y)$.* \{\text{prop:dom_ssub}\}

Proof. It follows from [34, Theorem 4.7(a)] that g is parabolically epi-differentiable at X for any $H \in T_{\text{dom } g}(X)$. Since $C_g(X, Y) \subset T_{\text{dom } g}(X)$, the discussion above gives us the claimed equality via [33, Proposition 3.4]. \square

4 Applications of Twice Epi-Differentiability

This section is devoted to establishing several second-order variational properties of spectral functions as an immediate consequence of Theorem 3.4. We begin by leveraging the latter result to obtain a characterization of a generalized twice differentiability of functions, recently introduced in [44, Definition 4.1]. To present its definition, recall that a function $\varphi : \mathbf{X} \rightarrow \overline{\mathbf{R}}$ is called a generalized quadratic form (GQF) on \mathbf{X} if $\text{dom } \varphi$ is a linear subspace of \mathbf{X} and there exists a linear symmetric operator L (i.e. $\langle Lx, y \rangle = \langle x, Ly \rangle$ for any $x, y \in \text{dom } \varphi$) from $\text{dom } \varphi$ to \mathbf{X} such that φ has a representation of form

$$\varphi(x) = \langle Lx, x \rangle \quad \text{for all } x \in \text{dom } \varphi.$$

A function $f : \mathbf{X} \rightarrow \overline{\mathbf{R}}$ is called generalized twice differentiable at x for $v \in \partial f(x)$ if the second subderivative $d^2f(x, v)$ is a GQF and twice epi-differentiable at x for v .

{gted}

Theorem 4.1. *Suppose that all the hypothesis of Proposition 3.6 hold and that g is subdifferentially regular and $Y = U\text{Diag}(y)U^\top$ for some $y \in \widehat{\partial}\theta(\lambda(X))$ and $U \in \mathbf{O}^n(X)$. Then g is generalized twice differentiable at X for Y if and only if θ has the same property at $\lambda(X)$ for y .*

Proof. Assume first that θ is generalized twice differentiable at $\lambda(X)$ for y . By Theorem 3.4, g is twice epi-differentiable at X for Y . Moreover, since $d^2\theta(\lambda(X), y)$ is a proper function and a GQF, it follows from (3.6) that $d^2g(X, Y)$ is quadratic for any $H \in \text{dom } d^2g(X, Y) = C_g(X, Y)$. What remains is to show that $C_g(X, Y)$ is a linear subspace of \mathbf{S}^n . By assumptions and [33, Proposition 3.4], we have $\text{dom } d^2\theta(\lambda(X), y) = C_\theta(\lambda(X), y)$ is a linear subspace of \mathbf{R}^n . Since θ is subdifferentially regular, we conclude from [19, Proposition 3.3] that $C_\theta(\lambda(X), y) = N_{\partial\theta(\lambda(X))}(y)$. Thus, we infer from a well-known fact from convex analysis that $y \in \text{ri } \partial\theta(\lambda(X))$. By [6, Proposition 3.19], we have $Y \in \text{ri } \partial g(X)$. This, together with the identities $\text{dom } d^2g(X, Y) = C_g(X, Y) = N_{\partial g(X)}(Y)$, which result from Proposition 3.6 and [19, Proposition 3.3], tells us that $C_g(X, Y)$ is a linear subspace.

Assume now that g is generalized twice differentiable at X for Y . It follows from Theorem 3.4 that θ is twice epi-differentiable at $\lambda(X)$ for y . It remains to show that $d^2\theta(\lambda(X), y)$ is a GQF. To do so, we infer from the second part of the proof of Theorem 3.4 that

$$d^2\theta(\lambda(X), y)(w) = d^2g(\text{Diag}(\lambda(X)), \text{Diag}(y))(\text{Diag}(w)) \quad (4.29) \quad \{\text{ssub_th}\}$$

for any $w \in \mathbf{R}^n$. On the other hand, Lemma 3.3(b), tells us that g is twice epi-differentiable at $\text{Diag}(\lambda(X))$ for $\text{Diag}(y)$. Moreover, the proof of the latter result indicates that

$$d^2g(\text{Diag}(\lambda(X)), \text{Diag}(y))(H) = d^2g(X, Y)(UHU^\top)$$

for any $H \in \mathbf{S}^n$. Combining these implies that g is generalized twice differentiable at $\text{Diag}(\lambda(X))$ for $\text{Diag}(y)$. This, along with (4.29), confirms that θ is generalized twice differentiable at $\lambda(X)$ for y and so completes the proof. \square

Corollary 4.2. *Suppose that g has the spectral representation in (1.1) and θ therein is a polyhedral function. Then, if $Y \in \text{ri } \partial g(X)$, then g is generalized twice differentiable at X for Y .*

Proof. It results from [43, Exercise 13.61] that θ is parabolically epi-differentiable at $\lambda(X)$ for any $y \in \partial\theta(\lambda(X))$. Moreover, we know from [6, Proposition 3.19] and $Y \in \text{ri } \partial g(X)$ that there exists $y \in \text{ri } \theta(\lambda(X))$ such that $Y = U\text{Diag}(y)U^\top$ and $U \in \mathbf{O}^n(X)$. According to [43, Proposition 13.9], θ is twice epi-differentiable at $\lambda(X)$ for y and $d^2\theta(\lambda(X), y) = \delta_{C_\theta(\lambda(X), y)}$. Combining this, $y \in \text{ri } \theta(\lambda(X))$, and Corollary 4.1 ensures that g is generalized twice differentiable at X for Y . \square

We proceed with exploring twice differentiability of spectral functions. Note that it was proven in [28, Theorem 3.3] that the spectral function g from (1.1) is twice differentiable at $X \in \mathbf{S}^n$ if and only if θ enjoys this property at $\lambda(X)$. It is tempting to ask whether our approach can be used to derive the latter result. We show below that this can be achieved when the spectral function g is prox-regular. To this end, recall that a function $f : \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ is called semidifferentiable at \bar{x} if it is finite at \bar{x} and the “lim inf” in (2.6) is a “lim.” Furthermore, f is said to be twice semidifferentiable at \bar{x} if it is semidifferentiable at \bar{x} and the “lim inf” in

$$d^2f(\bar{x})(w) := \liminf_{\substack{t \downarrow 0 \\ w' \rightarrow w}} \frac{f(\bar{x} + tw') - f(\bar{x}) - tdf(\bar{x})(w')}{\frac{1}{2}t^2}, \quad w \in \mathbf{X} \quad (4.30) \quad \{\text{ssud}\}$$

is a “lim.” In such a case, we refer to the left-hand side of (4.30) as the second semiderivative of f at \bar{x} . It is known that the twice semidifferentiability has serious limitations to handle

nonsmoothness in variational analysis; see [43, p. 590] for a detailed discussion. Below, we provide a characterization of twice semidifferentiability of differentiable spectral functions.

{prop:sed}

Theorem 4.3. *Suppose that θ is locally Lipschitz continuous around $\lambda(X)$, differentiable at $\lambda(X)$. Then the spectral function g from (1.1) is twice semidifferentiable at X if and only if θ is twice semidifferentiable at $\lambda(X)$. If one of these equivalent conditions holds, then the second semiderivative of g at X can be calculated for any $H \in \mathbf{S}^n$ by*

$$d^2g(X)(H) = d^2\theta(\lambda(X))(\lambda'(X; H)) + 2 \sum_{m=1}^r \langle \text{Diag}(y)_{\alpha_m \alpha_m}, U_{\alpha_m}^\top H(\mu_m I - X)^\dagger H U_{\alpha_m} \rangle,$$

where α_m , $m = 1, \dots, r$, are defined in (2.1), $U \in \mathbf{O}^n(X)$, and $y = \nabla\theta(\lambda(X))$ and where $\mu_1 > \dots > \mu_r$ are the distinct eigenvalues of X . Moreover, $d^2g(X)(H) = d^2g(X, \nabla g(X))(H)$ for any $H \in \mathbf{S}^n$.

Proof. Assume first that θ is twice semidifferentiable at $\lambda(X)$. To prove twice semidifferentiability of g at X , take $H \in \mathbf{S}^n$ and pick sequences $t_k \downarrow 0$ and $H_k \rightarrow H$, and observe that

$$\begin{aligned} \Delta_{t_k}^2 g(X, \nabla g(X))(H_k) &= \frac{\theta(\lambda(X + t_k H_k)) - \theta(\lambda(X)) - t_k \langle \nabla g(X), H_k \rangle}{\frac{1}{2} t_k^2} = \Delta_{t_k}^2 \theta(\lambda(X), y)(w^k) \\ &\quad + \frac{\langle y, w^k \rangle - \langle \nabla g(X), H_k \rangle}{\frac{1}{2} t_k} + \frac{\theta(\lambda(X + t_k H_k)) - \theta(\lambda(X) + t_k w^k)}{\frac{1}{2} t_k^2}, \end{aligned}$$

where $\nabla g(X) = U^\top \text{Diag}(y)U$, $y = \nabla\theta(\lambda(X))$, and

$$(w^k)_{\alpha_m} = \lambda(U_{\alpha_m}^\top H_k U_{\alpha_m} + t_k U_{\alpha_m}^\top H(\mu_m I - X)^\dagger H U_{\alpha_m}) \quad \text{for all } m = 1, \dots, r,$$

with the index sets α_m taken from (2.1). It can be checked directly that $w^k \rightarrow \lambda'(X, H)$. Since θ is differentiable at $\lambda(X)$, we get $\partial\theta(\lambda(X)) = \{y\}$. Because θ is symmetric, we conclude for any permutation matrices P satisfying $\lambda(X) = P\lambda(X)$ that $P y \in \widehat{\partial}\theta(\lambda(X))$, which leads us to $P y = y$. Thus, we can assume without loss of generality that $y \in \mathbf{R}_{\downarrow}^{|\alpha_1|} \times \dots \times \mathbf{R}_{\downarrow}^{|\alpha_r|}$. This implies that for any $m = 1, \dots, r$, there exists $u_m \in \mathbf{R}$ such that $y_{\alpha_m} = u_m \mathbf{1}_{|\alpha_m|}$, where $\mathbf{1}_{|\alpha_m|}$ is a vector in $\mathbf{R}^{|\alpha_m|}$ with all entries being 1. Hence, the matrices $\text{Diag}(y_{\alpha_m})$ and $U_{\alpha_m}^\top H_k U_{\alpha_m} + t_k U_{\alpha_m}^\top H(\mu_m I - X)^\dagger H U_{\alpha_m}$ are simultaneously diagonalizable for any $m = 1, \dots, r$. Appealing to Fan's inequality from [13] leads us to

$$\begin{aligned} &\frac{\langle y, w^k \rangle - \langle \nabla g(X), H_k \rangle}{\frac{1}{2} t_k} \\ &= \frac{\sum_{m=1}^r \langle y_{\alpha_m}, \lambda(U_{\alpha_m}^\top H_k U_{\alpha_m} + t_k U_{\alpha_m}^\top H(\mu_m I - X)^\dagger H U_{\alpha_m}) \rangle - \langle \nabla g(X), H_k \rangle}{\frac{1}{2} t_k} \\ &= \frac{\sum_{m=1}^r \langle \text{Diag}(y_{\alpha_m}), U_{\alpha_m}^\top H_k U_{\alpha_m} + t_k U_{\alpha_m}^\top H(\mu_m I - X)^\dagger H U_{\alpha_m} \rangle - \langle \nabla g(X), H_k \rangle}{\frac{1}{2} t_k} \\ &= \frac{\langle \text{Diag}(y), U^\top H_k U \rangle - \langle \nabla g(X), H_k \rangle}{\frac{1}{2} t_k} + 2 \sum_{m=1}^r \langle y_{\alpha_m}, \lambda(U_{\alpha_m}^\top H(\mu_m I - X)^\dagger H U_{\alpha_m}) \rangle \\ &= \frac{\langle y, \lambda'(X, H_k) \rangle - \langle \text{Diag}(y), U^\top H_k U \rangle}{\frac{1}{2} t_k} + 2 \sum_{m=1}^r \langle \text{Diag}(y)_{\alpha_m \alpha_m}, U_{\alpha_m}^\top H(\mu_m I - X)^\dagger H U_{\alpha_m} \rangle \\ &= 2 \sum_{m=1}^r \langle \text{Diag}(y)_{\alpha_m \alpha_m}, U_{\alpha_m}^\top H(\mu_m I - X)^\dagger H U_{\alpha_m} \rangle. \end{aligned}$$

On the other hand, we can obtain from (2.4) that

$$\lambda(X + t_k H_k) - \lambda(X) - t_k w^k = o(t_k^2).$$

Thus, it follows from the local Lipschitz continuity of θ around $\lambda(X)$ and [47, Theorem 1.5], and applying a similar argument as the proof of the estimate in (3.19) that

$$\frac{\theta(\lambda(X + t_k H_k)) - \theta(\lambda(X) + t_k w^k)}{\frac{1}{2}t_k^2} \rightarrow 0.$$

Since θ is twice semidifferentiable at $\lambda(X)$ and differentiable at $\lambda(X)$, we arrive at

$$\lim_{k \rightarrow \infty} \Delta_{t_k}^2 \theta(\lambda(X), y) = d^2 \theta(\lambda(X))(H) + 2 \sum_{m=1}^r \langle \text{Diag}(y)_{\alpha_m \alpha_m}, U_{\alpha_m}^\top H (\mu_m I - X)^\dagger H U_{\alpha_m} \rangle.$$

Recall that the sequences H_k and t_k were taken arbitrarily. Combining these tells us that g is twice semidifferentiable at X and gives us the claimed formula for the second semiderivative of g at X . The last claim about the equivalence of the second subderivative and second semiderivative of g follows from the fact that g is differentiable and $\text{dg}(X)(H) = \langle \nabla g(X), H \rangle$.

Finally, if g is twice semidifferentiable at X , it follows from the composite representation of θ in (2.7) and [32, Proposition 8.2(i)] that θ is twice semidifferentiable at $\lambda(X)$. \square

The result above opens the door to characterize twice differentiability of spectral functions that are prox-regular. To do so, we say that f has a quadratic expansion at \bar{x} if f is differentiable at \bar{x} and there exists a linear mapping $A : \mathbf{X} \rightarrow \mathbf{X}$ such that

$$f(x) = f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \frac{1}{2} \langle A(x - \bar{x}), x - \bar{x} \rangle + o(\|x - \bar{x}\|^2).$$

Recall also that a function $f : \mathbf{X} \rightarrow \overline{\mathbf{R}}$ is called prox-regular at \bar{x} for \bar{v} if f is finite at \bar{x} and locally lsc around \bar{x} with $\bar{v} \in \partial f(\bar{x})$, and there exist constants $\varepsilon > 0$ and $r \geq 0$ such that

$$\begin{cases} f(x') \geq f(x) + \langle v, x' - x \rangle - \frac{r}{2} \|x' - x\|^2 & \text{for all } x' \in \mathbf{B}_\varepsilon(\bar{x}) \\ \text{whenever } (x, v) \in (\text{gph } \partial f) \cap \mathbf{B}_\varepsilon(\bar{x}, \bar{v}) & \text{with } f(x) < f(\bar{x}) + \varepsilon. \end{cases} \quad (4.31) \quad \{\text{prox}\}$$

The function f is called prox-regular at \bar{x} if it enjoys this property at \bar{x} for any $\bar{v} \in \partial f(\bar{x})$. We are now in a position to provide a characterization of twice differentiability of spectral functions that are prox-regular. We should note that this was obtained in [28, Theorem 3.3] without assuming prox-regularity via a different approach. We, however, derive it as an immediate conclusion of our second-order variational analysis of spectral functions. It remains as an open question how to drop prox-regularity in the following result using our approach.

Corollary 4.4. *Assume that $g : \mathbf{S}^n \rightarrow \overline{\mathbf{R}}$ has the spectral representation in (1.1) and that g is prox-regular at $\lambda(X)$. Then, g is twice differentiable at X if and only if θ is twice differentiable at $\lambda(X)$. If one of these equivalent conditions holds, the second derivative of g at X can be calculated for any $H \in \mathbf{S}^n$ by*

$$\langle \nabla^2 g(X) H, H \rangle = \langle \nabla^2 \theta(\lambda(X)) \lambda'(X; H), \lambda'(X; H) \rangle + 2 \sum_{m=1}^r \langle \text{Diag}(y)_{\alpha_m \alpha_m}, U_{\alpha_m}^\top H (\mu_m I - X)^\dagger H U_{\alpha_m} \rangle,$$

where α_m , $m = 1, \dots, r$, are defined in (2.1), $U \in \mathbf{O}^n(X)$, and $y = \nabla \theta(\lambda(X))$ and where $\mu_1 > \dots > \mu_r$ are the distinct eigenvalues of X .

Proof. If g is twice differentiable, then it follows from the composite representation θ in (2.7) that θ is twice differentiable at $\lambda(X)$. Observe that in this case prox-regularity of g is not required.

Assume now that θ is twice differentiable at $\lambda(X)$. This requires that θ be differentiable around $\lambda(X)$ which implies that its gradient must be locally bounded around this point and thus it is locally Lipschitz continuous around this point. It results then from [26, Theorem 1.1] that g is locally differentiable around X . Moreover, twice differentiability of θ at $\lambda(X)$ implies twice semidifferentiability of θ at $\lambda(X)$ and $d^2\theta(\lambda(X))(w) = \langle \nabla^2\theta(\lambda(X))w, w \rangle$ for any $w \in \mathbf{R}^n$. By Theorem 4.3, g is twice semidifferentiable at X . Appealing now to [43, Example 13.8] confirms that g has a quadratic expansion at X . Since g is prox-regular at X , we deduce from [43, Corollary 13.42] that ∇g is differentiable, which amounts to saying that g is twice differentiable at X . \square

Our next result provides a characterization of subdifferential continuity for spectral functions. Recall that a function $f : \mathbf{X} \rightarrow \overline{\mathbf{R}}$ is called subdifferentially continuous at \bar{x} for $\bar{v} \in \partial f(\bar{x})$ if the convergence $(x_k, v_k) \rightarrow (\bar{x}, \bar{v})$ with $v_k \in \partial f(x_k)$ yields $f(x_k) \rightarrow f(\bar{x})$ as $k \rightarrow \infty$. When this property holds for all subgradients $v \in \partial f(\bar{x})$, we call f subdifferentially continuous at \bar{x} . For spectral functions, it was shown in [7, Theorem 4.2] that the spectral function g from (1.1) is prox-regular at X if and only if θ enjoys this property at $\lambda(X)$. A similar observation can be made for subdifferential continuity as shown below.

Proposition 4.5. *Assume that $g : \mathbf{S}^n \rightarrow \overline{\mathbf{R}}$ has the spectral representation in (1.1). Then, g is subdifferentially continuous at X if and only if θ is subdifferentially continuous at $\lambda(X)$.*

Proof. Suppose first that θ is subdifferentially continuous at $\lambda(X)$. To show that g is subdifferentially continuous at X , take $Y \in \partial g(X)$. If $(X_k, Y_k) \rightarrow (X, Y)$ with $Y_k \in \partial g(X_k)$, we know that there exists y_k such that $y_k \in \partial\theta(\lambda(X_k))$ and $\lambda(X_k) \rightarrow \lambda(X)$. It is known from [7, Lemma 14] that there exists a permutation matrix P such that $y_k \rightarrow Py$ with $Py \in \partial\theta(\lambda(X))$. Since θ is subdifferentially continuous at $\lambda(X)$, we know that $\theta(\lambda(X_k)) \rightarrow \theta(\lambda(X))$, which implies $g(X_k) \rightarrow g(X)$. Thus, g is subdifferentially continuous at X . To show the opposite implication, assume that g is subdifferentially continuous at X . We claim that g is subdifferentially continuous at $\Lambda(X) = \text{Diag}(\lambda(X))$. To see this, take $Y \in \partial g(\Lambda(X))$ and $U \in \mathbf{O}^n(X)$, and assume that $(X_k, Y_k) \rightarrow (\Lambda(X), Y)$ with $Y_k \in \partial g(X_k)$. In this case, it is not hard to see that

$$UYU^\top \in \partial g(X) \quad \text{and} \quad UY_kU^\top \in \partial g(UX_kU^\top).$$

for any $U \in \mathbf{O}^n(X)$. Since $(UX_kU^\top, UY_kU^\top) \rightarrow (X, UYU^\top)$ and since g is subdifferentially continuous at X , we get

$$g(X_k) = g(UX_kU^\top) \rightarrow g(X) = g(\Lambda(X)),$$

which implies that g is subdifferentially continuous at $\Lambda(X)$. To show that θ is subdifferentially continuous at $\lambda(X)$, take $y \in \partial\theta(\lambda(X))$ and $(x_k, y_k) \rightarrow (\lambda(X), y)$ with $y_k \in \partial\theta(x_k)$. It follows from [26, Theorem 1.4] that

$$\text{Diag}(y) \in \partial g(\text{Diag}(x)) \quad \text{and} \quad \text{Diag}(y_k) \in \partial g(\text{Diag}(x_k)).$$

Since $(\text{Diag}(x_k), \text{Diag}(y_k)) \rightarrow (\Lambda(X), \text{Diag}(y))$ and since g is subdifferentially continuous at $\Lambda(X)$, we arrive via (2.7) at

$$\theta(x_k) = g(\text{Diag}(x_k)) \rightarrow g(\Lambda(X)) = \theta(\lambda(X)),$$

which confirms that θ is subdifferentially continuous at $\lambda(X)$ and hence ends the proof. \square

Next, we are going to use our characterization in Theorem 3.4 to study proto-differentiability of subgradient mappings of spectral functions. To achieve our goal, recall that the graphical derivative of a set-valued mapping $F : \mathbf{X} \rightrightarrows \mathbf{Y}$ at \bar{x} for $\bar{y} \in F(\bar{x})$, denoted by $DF(\bar{x}, \bar{y})$, is a set-valued mapping from \mathbf{X} into \mathbf{Y} , defined by

$$\text{gph } DF(\bar{x}, \bar{y}) = T_{\text{gph } F}(\bar{x}, \bar{y}).$$

If, in addition, for any $\eta \in DF(\bar{x}, \bar{y})(w)$ and any choice of $t_k \downarrow 0$, there exist sequences $w_k \rightarrow w$ and $\eta_k \rightarrow \eta$ with $\bar{y} + t_k \eta_k \in F(\bar{x} + t_k w_k)$, then F is said to be proto-differentiable at \bar{x} for \bar{y} ; see [43, page 331] for more details. One of the main sources of proto-differentiability in variational analysis is subgradient mappings of various classes of functions.

Theorem 4.6. *Assume that $g : \mathbf{S}^n \rightarrow \overline{\mathbf{R}}$ has the spectral representation in (1.1) and that θ is locally Lipschitz continuous relative to its domain and $(X, Y) \in \text{gph } \partial g$. Suppose that g is prox-regular and subdifferentially continuous at X . Then ∂g is proto-differentiable at X for Y if and only if $\partial \theta$ is proto-differentiable at $\lambda(X)$ for y , where $y \in \partial \theta(\lambda(X))$ such that $Y = U \text{Diag}(y) U^\top$ for some $U \in \mathbf{O}^n(X)$.*

Proof. Our plan is to combine Theorem 3.4 and [43, Theorem 13.40]. To this end, recall that since g is prox-regular at X for Y , then $Y \in \widehat{\partial} g(X)$. Assume that θ is proto-differentiable at $\lambda(X)$ for y . It follows from [43, Theorem 13.40] that θ is twice epi-differentiable at $\lambda(X)$ for y . Appealing now to Theorem 3.4 tells us that g is twice epi-differentiable at X for Y , which together with [43, Theorem 13.40] confirms that g is proto-differentiable at X for Y . The opposite claim can be proven similarly using Theorem 3.4. \square

We end this section with a characterization of directional differentiability of the proximal mapping of spectral functions. This property shows great potential in establishing local superlinear convergence of semismooth Newton method and forward backward envelop acceleration [24]. Recall that proximal mapping of a function $f : \mathbf{X} \rightarrow \overline{\mathbf{R}}$ for a parameter value $\gamma > 0$, denoted by $\text{prox}_{\gamma f}$, is defined by

$$\text{prox}_{\gamma f}(x) = \underset{w \in \mathbf{X}}{\text{argmin}} \left\{ f(w) + \frac{1}{2\gamma} \|w - x\|^2 \right\}.$$

Recall from [43, Exercise 1.24] that a function $f : \mathbf{X} \rightarrow \overline{\mathbf{R}}$ is called prox-bounded if the function $f + \alpha \|\cdot\|^2$ is bounded from below on \mathbf{X} for some $\alpha \in \mathbf{R}$.

Corollary 4.7. *Assume that $g : \mathbf{S}^n \rightarrow \overline{\mathbf{R}}$ has the spectral representation in (1.1) and that θ is locally Lipschitz continuous relative to its domain and $(X, Y) \in \text{gph } \partial g$. Suppose that g is prox-regular and subdifferentially continuous at X and that g is prox-bounded. Then the following statements are equivalent.*

- (a) *For any $\gamma > 0$ sufficiently small, the proximal mapping $\text{prox}_{\gamma g}$ is directionally differentiable at $X + \gamma Y$.*
- (b) *For any $\gamma > 0$ sufficiently small, the proximal mapping $\text{prox}_{\gamma \theta}$ is directionally differentiable at $\lambda(X) + \gamma y$, where*

Proof. According to [43, Exercise 13.45], twice epi-differentiability of g at X for Y amounts to directional differentiability of $\text{prox}_{\gamma g}$ at $X + \gamma Y$ for any $\gamma > 0$ sufficiently small. The same observation holds for θ . Since twice epi-differentiability of g and θ are equivalent, it follows from the discussion above that (a) and (b) are equivalent. \square

As mentioned in Remark 3.5, when g is convex, Corollary 4.7 can be obtained from [10, Theorem 3]. The above result, however, goes far beyond convexity and covers a large class of spectral functions that are not necessarily convex.

{coro:prox-d-d}

5 Illustrated Examples

{secExp}

In this section, we are going to apply our second-order variational theory for several important classes of spectral functions to illustrate its efficacy in the study of important practical optimization problems.

5.1 Second-Order Variational Properties of Leading Eigenvalues

Our first goal is to explore the second-order variational properties of the leading eigenvalue functions by using theoretical results in Section 3. Recall from (2.3) that the eigenvalue $\lambda_{i-\ell_i(X)+1}(X)$, ranking first in the group of eigenvalues equal to $\lambda_i(X)$ is called a *leading* eigenvalue of X . To express this class of functions in the form of the spectral functions in (1.1), define the i -th order statistic function ϕ_i by $\phi_i(x) := \lambda_i(\text{Diag}(x)) = i$ -th largest element of $\{x_1, \dots, x_n\}$. It is not hard to see that all the leading eigenvalue functions have the spectral representation

$$g(X) := \phi_i(\lambda(X)) \quad \text{with either } i = 1 \text{ or } \phi_{i-1}(\lambda(X)) > \phi_i(\lambda(X))$$

for any $X \in \mathbf{S}^n$. The leading eigenvalue function has important applications in various fields. For example, the largest eigenvalue is usually used in the principal component analysis for dimension reduction [23]. The Fiedler value, which is defined as the second smallest eigenvalue, measures graph connectivity and is applied for image segmentation or community detection in social networks [22]. Other applications includes neural network [36, 37], signal processing [45]; see [1, 12, 15, 39] for more applications. Such a representation was well studied in [27, Section 9] through which some first-order variational properties of this function were explored. Our main objective is to demonstrate that our developments in previous sections allow us to characterize second-order variational properties of the leading eigenvalues. We should add here that some of these properties were studied before in [48] using a different approach and without appealing to the spectral representation of the leading eigenvalue functions. We begin by recording some of the first- and second-order properties of the i -th order statistic function.

{stat}

Proposition 5.1. *Suppose that $x \in \mathbf{R}^n$ and ϕ_i is the i -th order statistic function ϕ_i where either $i = 1$ or $\phi_{i-1}(x) > \phi_i(x)$. Then the following properties are satisfied.*

- (a) *For any $z = (z_1, \dots, z_n)$ sufficiently close to $x = (x_1, \dots, x_n)$, we have $\phi_i(z) = \max\{z_i \mid i \in I(x)\}$, where $I(x) = \{j \in \{1, \dots, n\} \mid x_j = \phi_i(x)\}$.*
- (b) *ϕ_i is locally Lipschitz continuous and subdifferentially regular at x . Moreover, we have $d\phi_i(x)(w) = \max\{w_i \mid i \in I(x)\}$ for any $w = (w_1, \dots, w_n)$ and $\partial\phi_i(x) = \text{conv}\{e_i \mid i \in I(x)\}$.*
- (c) *The function ϕ_i is twice epi-differentiable at x for any $v \in \partial\phi_i(x)$.*
- (d) *The function ϕ_i is prox-regular and subdifferentially continuous at z for any $v \in \partial\phi_i(z)$.*

Proof. The claim in (a) falls directly from the definition of ϕ_i . The claims in Part (b) are immediate consequences of (a); see also [27, Theorem 9]. The formula for $\partial\phi_i(x)$ results from the equivalent local representation of ϕ_i from (a) and the well-known formula for the pointwise maximum of convex functions; see also [27, Proposition 6]. The formula for the subderivative of ϕ_i results from [43, Exercise 8.31]. Part (c) follows from (a) and [43, Example 13.10]. Finally, (d) falls out directly from the local representation of ϕ_i as a polyhedral function together with [43, Example 13.30]. \square

Using the properties of the i -th order statistic function ϕ_i , listed above, we first calculate the subderivative of these functions.

Corollary 5.2. *Suppose that $X \in \mathbf{S}^n$ and that $\mu_1 > \dots > \mu_r$ are the distinct eigenvalues of X . Then, for any $i \in \{1, \dots, r\}$, the subderivative of the leading eigenvalue function λ_i can be calculated by*

$$\lambda'_i(X, H) = d\lambda_i(X)(H) = \lambda_1(U_{\alpha_i}^\top H U_{\alpha_i}) \quad (5.32) \quad \{\text{subd_lead}\}$$

for any $H \in \mathbf{S}^n$, where the index sets α_i comes from (2.1). Moreover, λ_i is subdifferentially regular at X and its subdifferential at this point has a representation in the form

$$\partial\lambda_i(X) = \{U_{\alpha_i} \text{Diag}(y)_{\alpha_i \alpha_i} U_{\alpha_i}^\top \mid U \in \mathbf{O}^n(X), y \in \partial\phi_i(\lambda(X))\}. \quad (5.33) \quad \{\text{subd_lead2}\}$$

Proof. It is easy to see that for any $i \in \{1, \dots, r\}$, the leading eigenvalue function λ_i has the spectral representation as (1.1) with θ being the i -th order statistic function ϕ_i , where either $i = 1$ or $\phi_{i-1}(x) > \phi_i(x)$. To justify (5.32), it follows from Theorem 2.3 that

$$\begin{aligned} d\lambda_i(X)(H) &= d\phi_i(\lambda(X))(\lambda'(X, H)) \\ &= \max\{\lambda_s(U_{\alpha_i}^\top H U_{\alpha_i}) \mid s \in \alpha_i\} = \lambda_1(U_{\alpha_i}^\top H U_{\alpha_i}). \end{aligned}$$

By Proposition 2.2, the subdifferential of λ_i at X has a representation of the form

$$\partial\lambda_i(X) = \{U \text{Diag}(y) U^\top \mid U \in \mathbf{O}^n(X), y \in \partial\phi_i(\lambda(X))\}.$$

By Proposition 5.1(b), $y \in \partial\phi_i(\lambda(X))$ amounts to $y \in \text{conv}\{e_i \mid i \in \alpha_i\}$. Thus, we get

$$Y = U \text{Diag}(\lambda(Y)) U^\top = \sum_{i=1}^n \lambda_i(Y) U_i U_i^\top = \sum_{s \in \alpha_i} \lambda_s(Y) U_s U_s^\top = U_{\alpha_i} \Lambda(Y)_{\alpha_i \alpha_i} U_{\alpha_i}^\top.$$

Combining these confirms (5.33). Finally subdifferential regularity of λ_i at X was taken from [27, Corollary 4] along with the fact that in the spectral representation of λ_i , the symmetric function $\theta = \phi_i$ is subdifferentially regular according to Proposition 5.1(b). \square

Next, we characterize the twice epi-differentiability of the leading eigenvalues. It is worth mentioning that although Proposition 5.1(a) shows that ϕ_i can be written locally as a polyhedral function, that local representation is not symmetric. Thus, existing results in [34] are not applicable here.

Theorem 5.3. *Suppose that $X \in \mathbf{S}^n$ and that $\mu_1 > \dots > \mu_r$ are the distinct eigenvalues of X . Then, for any $i \in \{1, \dots, r\}$, the leading eigenvalue function λ_i is twice epi-differentiable at X for any $Y \in \partial\lambda_i(X)$ and*

$$d^2\lambda_i(X, Y)(H) = \delta_{C_g(X, Y)}(H) + 2\langle Y, H(\mu_i I - X)^\dagger H \rangle \quad \text{for all } H \in \mathbf{S}^n,$$

where $U \in \mathbf{O}^n(X)$ such that $Y = U \text{Diag}(y) U^\top$ for some $y \in \partial\phi_i(\lambda(X))$.

Proof. It is easy to see that for any $i \in \{1, \dots, r\}$, the leading eigenvalue function λ_i has the spectral representation as (1.1) with θ being the i -th order statistic function ϕ_i , where either $i = 1$ or $\phi_{i-1}(x) > \phi_i(x)$. It follows Proposition 5.1(b) that λ_i is locally Lipschitz continuous and subdifferentially regular at X . Appealing now to Proposition 5.1(c) and Theorem 3.4 confirms that λ_i is twice epi-differentiable at X for Y . To justify the claimed formula for the second subderivative of λ_i , we proceed with two Claims.

Claim 1: We have $\text{dom } d^2\lambda_i(X, Y) = C_{\lambda_i}(X, Y)$.

To justify this claim, we can conclude from Proposition 5.1(a) and [34, Theorem 4.7(a)] that λ_i is parabolically epi-differentiable at X for any $H \in C_{\lambda_i}(X, Y)$ in the sense of [43,

Definition 13.59]. This, coupled with Proposition 3.6., ensures the claim about the domain of $d^2\lambda_i(X, Y)$.

Claim 2: The second subderivative of λ_i at X for Y has a representation of the form

$$d^2g(X, Y)(H) = \delta_{C_g(X, Y)}(H) + 2 \sum_{s=1}^r \langle \Lambda(Y)_{\alpha_s \alpha_s}, U_{\alpha_s}^\top H(\mu_s I - X)^\dagger H U_{\alpha_s} \rangle$$

for any $H \in \mathbf{S}^n$.

To verify this claim, it follows from $Y \in \partial\lambda_i(X)$ and Proposition 5.3 that $Y = U_{\alpha_i} \text{Diag}(y)_{\alpha_i \alpha_i} U_{\alpha_i}^\top$ for some $y \in \partial\phi_i(\lambda(X))$. Recall from Claim 1 that $\text{dom } d^2\lambda_i(X, Y) = C_{\lambda_i}(X, Y)$. Take $H \in \text{dom } d^2g(X, Y)$ and conclude that $d^2g(X, Y)(H) < \infty$. Using the formula for the second subderivative from Theorem 3.4 then leads us to $\lambda'(X; H) \in \text{dom } d^2\theta(\lambda(X), y)$. On the other hand, we know from Proposition 5.1(a) that $d^2\theta(\lambda(X), y) = d^2h(\lambda(X), y) = \delta_{C_\theta(\lambda(X), y)}$, where $h(z) := \max\{z_i \mid i \in I(\lambda(X))\}$ for any $z = (z_1, \dots, z_n)$, and where the last equality results from h being a polyhedral function together with [43, Proposition 13.9]. Therefore, we get $d^2\theta(\lambda(X), y)(\lambda'(X; H)) = 0$. Combining this with the formula for the second subderivative from Theorem 3.4 confirms that for any $H \in \text{dom } d^2g(X, Y)$, we have

$$d^2g(X, Y)(H) = 2 \sum_{s=1}^r \langle \Lambda(Y)_{\alpha_s \alpha_s}, U_{\alpha_s}^\top H(\mu_s I - X)^\dagger H U_{\alpha_s} \rangle.$$

If $H \notin \text{dom } d^2g(X, Y)$, we conclude from the equality $\text{dom } d^2\lambda_i(X, Y) = C_{\lambda_i}(X, Y)$ that $d^2\lambda_i(X, Y)(H) = \infty$, which proves the claimed formula for $d^2\lambda_i(X, Y)$ in this claim.

To finish the proof, recall that $Y = U_{\alpha_i} \text{Diag}(y)_{\alpha_i \alpha_i} U_{\alpha_i}^\top$ for some $y \in \partial\phi_i(\lambda(X))$. This implies that $\Lambda(Y)_{\alpha_s \alpha_s} = 0$ for all $s \in \{1, \dots, r\} \setminus \{\alpha_i\}$. Combining these and Claim 2, we arrive at

$$\begin{aligned} d^2g(X, Y)(H) &= \delta_{C_g(X, Y)}(H) + 2 \langle \Lambda(Y)_{\alpha_i \alpha_i}, U_{\alpha_i}^\top H(\mu_i I - X)^\dagger H U_{\alpha_i} \rangle \\ &= \delta_{C_g(X, Y)}(H) + 2 \langle Y, H(\mu_i I - X)^\dagger H \rangle \end{aligned}$$

for any $H \in \mathbf{S}^n$, which completes the proof. \square

It is worth mentioning here that twice epi-differentiability of the leading eigenvalue functions was studied before in [48, Theorem 3.2]. Here, we obtain it via a different approach, which is much simpler than deriving it from the definition directly.

5.2 Second-Order Variational Properties of Statistic Regularizer

We continue to study second-order variational properties of some important regularization terms, used often in statistics to analyze eigenvalue optimization problems such as the largest eigenvalue gaps. This again demonstrates the great potential of our established theory in practical applications.

Example 5.4 (largest eigenvalue gap). Suppose that $X \in \mathbf{S}^n$. The largest eigenvalue gap of X is defined by

$$g(X) = \max \{ \lambda_k(X) - \lambda_{k+1}(X) \mid k = 1, \dots, n \}.$$

The largest eigenvalue gap of the Laplacian matrix is frequently used in statistics to choose the number k of clusters in spectral clustering; see [30] and [49, Section 8.3] for more details. It is also widely used in quantum mechanics to determine the mixing time of quantum walks [46]. It is easy to see that g has the spectral representation in (1.1) with

$$\theta(x) := \max \{ \langle x, e^{i, i+1} \rangle \mid i = 1, \dots, n-1 \},$$

where $e^{i,i+1} \in \mathbf{R}^n$ is the vector whose i -th component is 1, $i + 1$ -th component is -1 and the remaining components are 0. Note that θ is the fussed graphical Lasso regularization term and its proximal mapping was computed in [51]. Since θ is a polyhedral function, it follows from [43, Proposition 13.9] that such θ is twice epi-differentiable at any $x \in \mathbf{R}^n$ for any $y \in \partial\theta(x)$. Using Theorem 3.4, we conclude that g is also twice epi-differentiable at X for any $Y \in \partial g(X)$. It is also worth mentioning that our argument above can be applied to any polyhedral function θ . This tells us a similar conclusion can be reached about twice epi-differentiability of a wide range of functions appearing in different optimization problems such as positive semidefinite programming problems and the fastest mixing Markov chain problem [2]. In such a case, according to Corollary 4.7, prox_g is directionally differentiable, which allows us to apply fast algorithms like the generalized Newton method and forward-backward envelop acceleration [24] to solve the optimization problem with such a regularizer g .

Example 5.5 (Minimax Concave Penalty (MCP)). The MCP function was introduced in [50, equation (2.1)] for the sparse regression. A frequently used MCP function (cf. [29, 50]) has the form

$$\phi(t) = \begin{cases} c|t| - \frac{t^2}{2a} & \text{if } |t| \leq ac, \\ \frac{ac^2}{2} & \text{if } |t| > ac, \end{cases}$$

where $a > 1$ and $c > 0$ are two given constants and $t \in \mathbf{R}$. It is usually regarded as a smoothing function for the l_0 -norm [16, 50]. It can also be composed with an eigenvalue function to form a spectral function in applications. For instance, in the robust PCA (cf. [29]), it is common to consider the spectral function $g(X) = \sum_{i=1}^n \phi(\lambda_i(X))$. Indeed, all the penalty functions in [29, Table 1] can be used in a similar way to construct a spectral function to deal with the low rank optimization problems. Define the function $\theta : \mathbf{R}^n \rightarrow \mathbf{R}$ by

$$\theta(x_1, \dots, x_n) = \phi(x_1) + \dots + \phi(x_n).$$

Clearly, θ is symmetric and $g = \theta \circ \lambda$, which tells us that g is a spectral function. It is easy to see that θ is differentiable at any $x = (x_1, \dots, x_n) \in \mathbf{R}^n$ when $x_i \neq 0$ for all $i = 1, \dots, n$. At $x = 0$, θ enjoys the following properties:

- θ is Lipschitz continuous around 0 and $\partial\phi(0) = [-c, c]$. This, along with [43, Proposition 10.5], shows that $\partial\theta(0) = [-c, c]^n$.
- Since ϕ can be equivalently expressed as $\phi(t) = \max\{ct - \frac{t^2}{2a}, -ct - \frac{t^2}{2a}\}$ locally around 0, it is fully amenable at $0 \in \mathbf{R}$ in the sense of [43, Definition 10.23]. According to [43, Example 10.26], θ enjoys this property at $0 \in \mathbf{R}^n$ as well and therefore is prox-regular and subdifferential continuous at 0.
- By [43, Example 7.28], ϕ is subdifferentially regular at 0. It follows then from [43, Proposition 10.5] that θ is subdifferentially regular at $0 \in \mathbf{R}^n$. Employing now [27, Corollary 4], we can conclude that g is also subdifferentially regular at $0 \in \mathbf{S}^n$.
- It follows from [43, Corollary 13.15] that θ is twice epi-differentiable at $0 \in \mathbf{R}$ for any $v \in \partial\phi(0)$ and

$$d^2\phi(0, v)(w) = \delta_{C_\phi(0, v)}(w) - \frac{w^2}{a},$$

where the last equality comes from [43, Example 13.16]. Note that the critical cone $C_\phi(0, v)$ can be efficiently calculated using the representation $C_\phi(0, v) = N_{\partial\phi(0)}(v)$. Because of the

definition of θ , the latter tells us that θ is twice epi-differentiable at $0 \in \mathbf{R}^n$ for any $y = (y_1, \dots, y_n) \in \partial\theta(0)$ and

$$d^2\theta(0, y)(w) = d^2\phi(0, y_1)(w_1) + \dots + d^2\phi(0, y_n)(w_n),$$

for any $w = (w_1, \dots, w_n) \in \mathbf{R}^n$.

Combining these properties and Theorem 3.4 tells us that the spectral function g is twice epi-differentiable at $X := 0$ for any $Y \in \partial g(X)$. A similar argument can be made for any $x = (x_1, \dots, x_n) \in \mathbf{R}^n$ when one of $x_i = 0$. Note that for any $x \in \mathbf{R}$, it is easy to see that for sufficiently small γ that

$$\text{prox}_{\gamma\phi}(x) = \begin{cases} 0, & x \leq \gamma c, \\ \text{sgn}(x) \cdot \frac{a}{a-\gamma}(|x| - \gamma c) & \gamma c < x \leq ac, \\ x & x > ac. \end{cases}$$

It can be checked directly that $\text{prox}_{\gamma\theta}(x) = (\text{prox}_{\gamma\phi}(x_1), \dots, \text{prox}_{\gamma\phi}(x_n))$. Since ϕ is prox-bounded, we derive from Corollary 4.7 that $\text{prox}_{\gamma g}$ is directionally differentiable, which can be used to design efficient methods for solving eigenvalue optimization problems using the generalized Newton method in [24].

6 Conclusion

In this paper, we explore the second-order variational properties of a spectral function whose symmetric part is not necessarily convex. We prove twice epi-differentiability of spectral functions amounts to that of their symmetric parts. By applying this theoretical equivalence as a tool, we characterize several variational properties of spectral functions including generalized twice differentiability, twice semidifferentiability, proto-differentiability, and directional differentiability of their proximal mappings. As an immediate consequence, we obtain twice epi-differentiability of the leading eigenvalue functions and the largest eigenvalue gap. The established theory can be leveraged to solve practical optimization problems using the generalized Newton method or forward-backward envelop acceleration method, which is a subject for our future research. It is very likely that our second-order analysis in this paper can be extended to singular value spectral functions and it is a subject of our ongoing research.

References

- [1] Anderson, G. W., Guionnet, A., & Zeitouni, O. (2010). An introduction to random matrices (No. 118). Cambridge university press.
- [2] Boyd, S., Diaconis, P. and Xiao, L. (2004). Fastest mixing Markov chain on a graph. *SIAM Review*, 46(4), 667–689.
- [3] Cui Y, Ding C (2019) Nonsmooth composite matrix optimization: strong Regularity, constraint nondegeneracy and beyond. Preprint. arXiv:1907.13253.
- [4] Cui Y, Ding C, Zhao XY (2019) Quadratic growth conditions for convex matrix optimization problems associated with spectral functions, *SIAM J. Optim.* 27:2332–2355.
- [5] Davis C (1957) All convex invariant functions of hermitian matrices. *Arch. Math.* 8:276–278.
- [6] A. Daniilidis, D. Drusvyatskiy, and A.S. Lewis: *Orthogonal invariance and identifiability*, *SIAM J. Matrix Anal. Appl.* 35/2 (2014) 580–598.
- [7] Daniilidis A, Lewis AS, Malick J, Sendov H (2008) Prox-regularity of spectral functions and spectral sets. *J. Convex Anal.* 15(3):547–560.

- [8] Ding C, (2012) An introduction to a class of matrix optimization problems. PhD Thesis, National University of Singapore.
- [9] Ding, C., Sun, D., Sun, J., & Toh, K. C. (2020). Spectral operators of matrices: semismoothness and characterizations of the generalized Jacobian. *SIAM Journal on Optimization*, 30(1), 630-659.
- [10] Ding, C., Sun, D., Sun, J., & Toh, K. C. (2018). Spectral operators of matrices. *Mathematical Programming*, 168(1), 509-531.
- [11] Drusvyatskiy D, Paquette C (2018) Variational analysis of spectral functions simplified. *J. Convex Anal.* 25(1):119-134.
- [12] Erdős, L., & Yau, H. T. (2017). A dynamical approach to random matrix theory (Vol. 28). American Mathematical Soc..
- [13] K. FAN, *On a theorem of Weyl concerning eigenvalues of linear transformations*, Proceedings of the National Academy of Sciences of U.S.A. 35 (1949): 652–655.
- [14] Friedland, S. (1981). Convex spectral functions. *Linear and multilinear algebra*, 9(4), 299-316.
- [15] Gavish, M., & Donoho, D. L. (2014). The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Transactions on Information Theory*, 60(8), 5040-5053.
- [16] Gu, Y., Jin, J., & Mei, S. (2009). l_0 norm constraint LMS algorithm for sparse system identification. *IEEE Signal Processing Letters*, 16(9), 774-777.
- [17] Hang NTV, Sarabi ME (2021) Local convergence analysis of augmented Lagrangian methods for piecewise linear-quadratic composite optimization problems. *SIAM J. Optim.* 31(4): 2665–2694.
- [18] Hang NTV, Sarabi ME (2021) Convergence of Augmented Lagrangian Methods for Composite Optimization Problems. To appear at *Math. Oper. Res.*
- [19] Hang, N. T., Sarabi, M. E. (2023). Smoothness of Subgradient Mappings and Its Applications in Parametric Optimization, arxiv: 2311.06026
- [20] Hang, N. T., Sarabi, M. E. (2024). A chain rule for strict twice epi-differentiability and its applications. *SIAM Journal on Optimization*, 34(1), 918-945.
- [21] He, J., Kan, C., & Song, W. (2024). Twice Epi-Differentiability of Orthogonally Invariant Matrix Functions and Application. arXiv preprint arXiv:2412.09898.
- [22] Jiang, S., Lu, J. A., Zhou, J., & Dai, Q. (2024). Fiedler value: The cumulated dynamical contribution value of all edges in a complex network. *Physical Review E*, 109(5), 054301.
- [23] Jolliffe, I. (2011). Principal component analysis. In *International encyclopedia of statistical science* (pp. 1094-1096). Springer, Berlin, Heidelberg.
- [24] Duy Khanh, P., Mordukhovich, B., & Phat, V. T. (2023). A generalized Newton method for subgradient systems. *Mathematics of Operations Research*, 48(4), 1811-1845.
- [25] Kotlov, A., Lovász, L., & Vempala, S. (1997). The Colin de Verdiere number and sphere representations of a graph. *Combinatorica*, 17(4), 483-521.
- [26] Lewis, A. S. (1996). Derivatives of spectral functions. *Mathematics of Operations Research*, 21(3), 576-588.
- [27] Lewis AS (1999) Nonsmooth analysis of eigenvalues. *Math. Program.* 84:1–24.
- [28] Lewis AS, Sendov HS (2001) Twice differentiable spectral functions. *SIAM J. Matrix Anal. Appl.*, 23:368–386.
- [29] Lu C, Tang J, Yan S, et al. Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm[J]. *IEEE Transactions on Image Processing*, 2015, 25(2): 829-839.
- [30] Meilă M, Shortreed S, Xu L. Regularized spectral learning[C]//*International Workshop on Artificial Intelligence and Statistics*. PMLR, 2005: 230-237.
- [31] Mohammadi A, Mordukhovich BS, Sarabi ME (2021) Variational analysis of composite models with applications to continuous optimization. *Math. Oper. Res.* 47: 397–426.
- [32] Mohammadi A, Mordukhovich BS, Sarabi ME (2021) Parabolic regularity in geometric variational analysis. *Trans. Amer. Soc.* 374:1711–1763.
- [33] Mohammadi A, Sarabi ME (2020) Twice epi-differentiability of extended real-valued functions with applications in composite optimization. *SIAM J. Optim.* 30:2379–2409.
- [34] Mohammadi A, Sarabi ME (2023) Parabolic Regularity of Spectral Functions. arXiv:2301.04240
- [35] Mordukhovich, B. S., & Sarabi, M. E. (2021). Generalized Newton algorithms for tilt-stable minimizers in nonsmooth optimization. *SIAM Journal on Optimization*, 31(2), 1184-1214.

- [36] Nt, H., & Maehara, T. (2019). Revisiting graph neural networks: All we have is low-pass filters. arXiv preprint arXiv:1905.09550.
- [37] Oono, K., & Suzuki, T. (2019). Graph neural networks exponentially lose expressive power for node classification. arXiv preprint arXiv:1905.10947.
- [38] Overton, M. L., & Womersley, R. S. (1993). Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices. *Mathematical Programming*, 62(1), 321-357.
- [39] Pennington, J., & Worah, P. (2018). The spectrum of the fisher information matrix of a single-hidden-layer neural network. *Advances in neural information processing systems*, 31.
- [40] Qi, H., & Yang, X. (2003). Semismoothness of spectral functions. *SIAM journal on matrix analysis and applications*, 25(3), 766-783.
- [41] R. T. ROCKAFELLAR, *Maximal monotone relations and the second derivatives of nonsmooth functions*, *Ann. Inst. H. Poincaré Analyse Non Linéaire* 2 (1985), 167–184.
- [42] Rockafellar, R. T. (1990). Generalized second derivatives of convex functions and saddle functions. *Transactions of the American Mathematical Society*, 322(1), 51-77.
- [43] Rockafellar RT, Wets RJ-B (1998) *Variational Analysis* (Springer, Berlin).
- [44] Derivative tests for prox-regularity and the modulus of convexity, to appear in *J. Convex Anal.*, (2025).
- [45] Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., & Vandergheynst, P. (2013). The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE signal processing magazine*, 30(3), 83-98.
- [46] Temme, K., Osborne, T. J., Vollbrecht, K. G., Poulin, D., & Verstraete, F. (2011). Quantum metropolis sampling. *Nature*, 471(7336), 87-90.
- [47] Torik M (2001) Second-order directional derivatives of all eigenvalues of a symmetric matrix. *Nonlinear Anal.* 46:1133–1150.
- [48] Torik M (1998) First- and second-order epi-differentiability in eigenvalue optimization. *J. Math. Anal. Appl.* 234:391–416.
- [49] Von Luxburg U. A tutorial on spectral clustering[J]. *Statistics and computing*, 2007, 17: 395-416.
- [50] Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* 38(2): 894-942
- [51] Zhang, N., Zhang, Y., Sun, D., & Toh, K. C. (2021). An efficient linearly convergent regularized proximal point algorithm for fused multiple graphical lasso problems. *SIAM Journal on Mathematics of Data Science*, 3(2), 524-543.
- [52] Zhao, X. Y., Sun, D., & Toh, K. C. (2010). A Newton-CG augmented Lagrangian method for semidefinite programming. *SIAM Journal on Optimization*, 20(4), 1737-1765.