Noname manuscript No.
(will be inserted by the editor)

Nonsmooth Composite Matrix Optimization: Strong Regularity, Constraint Nondegeneracy and Beyond

Ying Cui · Chao Ding

This version: July 30, 2019

Abstract The nonsmooth composite matrix optimization problem (CMatOP), in particular, the matrix norm minimization problem, is a generalization of the matrix conic programming problem with wide applications in numerical linear algebra, computational statistics and engineering. This paper is devoted to the characterization of the strong regularity for the CMatOP via the generalized strong second order sufficient condition and constraint nondegeneracy for problems with nonsmooth objective functions. The derived result supplements the existing characterization of the strong regularity for the constrained optimization problems with twice continuously differentiable data.

Keywords: matrix optimization, spectral functions, strong regularity, piecewise affine, strong second order sufficient condition, constraint nondegeneracy

AMS Class: 65K05, 90C25, 90C31

1 Introduction

Matrix conic programming is a class of optimization problems with matrix cone constraints, in particular, the positive semidefinite constraint. Being an extension of the classical nonlinear programming, this subject has now grown into a fruitful discipline in optimization, including deep and rich mathematical theory, a bunch of efficient and robust solvers [33,36,42,41] and a wide range of important applications in combinatorial optimization [5] and control theory [1].

A natural generalization of the matrix conic programming problem is the matrix norm minimization problem. Starting from the nuclear norm formulation of the low rank matrix completion problem [7,27], there is a growing list of algorithms and applications in such nonsmooth matrix optimization problems that also involve the spectral norm or the general matrix Ky Fan k-norm function [39,37,6] in the objective. Denote $\mathbb{R}^n$ as the real n-dimensional space and $\mathbb{S}^n$ as the set of all $n \times n$ symmetric matrices. A general form of the nonsmooth composite matrix optimization problems (CMatOPs) can be written as

$$\begin{aligned} &\underset{\mathbf{x}\in\mathbb{X}}{\text{minimize}} \quad \Phi(\mathbf{x}) \triangleq f(\mathbf{x}) + \phi \circ \lambda(g(\mathbf{x})) \\ &\text{subject to} \quad h(\mathbf{x}) = 0, \end{aligned} \tag{1}$$

where $\mathbb{X}$ and $\mathbb{Y}$ are two given finite dimensional Euclidean spaces, $f : \mathbb{X} \to \mathbb{R}$ is a twice continuously differentiable function, $g : \mathbb{X} \to \mathbb{S}^n$ and $h : \mathbb{X} \to \mathbb{Y}$ are twice continuously differentiable mappings, and $\phi : \mathbb{R}^n \to (-\infty, +\infty]$ is a symmetric function (i.e., for any $u \in \mathbb{R}^n$, $\phi(Pu) = \phi(u)$ for any $n \times n$ permutation matrix P). Here $\lambda(\bullet)$ denotes the vector of eigenvalues for a symmetric matrix with the

Y. Cui
Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA 90089
E-mail: yingcui@usc.edu

C. Ding
Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing, P.R. China.
E-mail: dingchao@amss.ac.cn

components being arranged in the non-increasing order. Obviously the function $\phi \circ \lambda$ only depends on the spectrum of a given matrix, and is thus called the spectral function in the literature. There is a one-to-one correspondence between the spectral function and the so-called orthogonal-invariant matrix function, i.e., the matrix function that is invariant under orthogonal similarity transformations [19]. Notice that if the function ϕ is taken to be the indicator function over the nonnegative orthant, the problem (1) reduces to the nonlinear semidefinite programming problem.

One fundamental concept in the sensitivity analysis and perturbation theory is the so-called *strong regularity*, which is originally introduced by Robinson [29] for generalized equations; see Section 4 for its definition. The Karush-Kuhn-Tucker (KKT) optimality condition of the conventional nonlinear programming problem with twice continuously differentiable data can be formulated as a special generalized equation, whose strong regularity at a KKT solution is known to be equivalent to the strong second order sufficient condition and the constraint nondegeneracy (when restricted to the nonlinear programming problem, the constraint nondegeneracy reduces to the linear independence constraint qualification) [29, 15]; see also [4, Proposition 5.38]. This result has been further generalized to the nonlinear semidefinite programming problems in [34]. For a general class of C^2-reducible problems where the constraint nondegeneracy holds, the strong regularity is further proved to be equivalent to the Lipschitzian full stability [23].

The non-polyhedrality of the matrix cone distinguishes the nature of the matrix conic programming from the classical nonlinear programming, where the constraints of the latter problems are given by finitely representable equalities and inequalities. As one can expect, such a distinction is carried to the sensitivity analysis of the CMatOPs and makes it a worthwhile effort for a deep investigation. In contrast to the relative long history of the theoretical research of the matrix conic programming, the sensitivity analysis of the nonsmooth composite matrix optimization program still stays at an early stage. It turns out that one can rewrite (1) via its epigraphical formulation

$$\begin{array}{ll} \underset{X\in\mathbb{X},\, t\in\mathbb{R}}{\text{minimize}} & f(X)+t \\ \text{subject to} & h(X)=0, \quad (g(X),t)\in \operatorname{epi}(\phi\circ\lambda), \end{array} \tag{2}$$

which transforms the original problem into a matrix conic programming problem. However, such a transformation itself is inadequate for drawing the whole picture of the sensitivity results of the CMatOPs, with the following two reasons. One, besides that of the semidefinite programming, the characterization of the strong regularity for other matrix conic programming problems, such as the one involves the epigraph of the Ky Fan k-norm cone, is in fact unknown. Two, the reformulation in (2) lifts the original problem from $\mathbb{X}$ to $\mathbb{X}\times\mathbb{R}$. Even if given the answer raised by the first point, it still needs the effort to bring those characterization back to the space $\mathbb{X}$.

In this paper, we characterize the strong regularity of the solution to the KKT system of (1) for the case where ϕ is piecewise affine. To accomplish this task, we study nonsmooth counterparts of the second order sufficient condition and the constraint nondegeneracy via the second order variational analysis of the spectral functions. The adopted approach is a departure from [23] that based on the second order subdifferential of the extended value function $\Phi(X)+\delta_Q(h(X))$, where $\delta_Q(h(\bullet))$ is the indicator function of $h(\bullet)$ over Q, i.e., $\delta_Q(h(X))$ equals to 0 if $h(X)\in Q$ and $+\infty$ otherwise. With a main focus on the characterization of the full stability in the above mentioned reference, the resulting equivalent conditions involves the limiting coderivative of $\operatorname{epi}(\phi\circ\lambda)$, whose calculation itself might be complicated.

The rest of the paper is organized as follows. Section 2 summarizes some useful variational properties of eigenvalues and piecewise affine functions. In Section 3, we investigate the properties of proximal mappings associate with spectral functions that are important to the subsequent analysis. The main result of this paper on the characterization of the strong regularity for the CMatOPs is presented in Section 4. An example of CMatOPs involving the largest eigenvalue of a symmetric matrix is used to illustrate the derived results in Section 5. We conclude our paper in the final section.

Unless otherwise specified in the paper, we use plain small Latin letters (e.g., x and y) to represent scalars, small Latin letters in boldface (e.g., $\mathbf{x}$) to represent vectors and capital Latin letters (e.g., X) to represent matrices. We also use Greek letters (e.g., α, β and ι) to denote the index sets and blackboard bold letters (e.g., $\mathbb{R}^n$ and $\mathbb{O}^n$) to denote spaces or sets. For $X\in\mathbb{R}^{n\times n}$, $\operatorname{diag}(X)$ denotes the column vector consisting of all the diagonal entries of X being arranged from the first to the last. For $\mathbf{x}\in\mathbb{R}^n$, $\operatorname{Diag}(\mathbf{x})$ denotes the $n\times n$ diagonal matrix whose i-th diagonal entry is $\mathbf{x}_i$ for $i=1,\ldots,n$. We write $\mathbb{O}^n$ as the set of all $n\times n$ orthogonal matrices.

2 Preliminaries and background results

2.1 Variational analysis of the eigenvalues

Let $X \in \mathbb{S}^n$ be an arbitrary symmetric matrix. Suppose that X has the following eigenvalue decomposition

$$X = U\,\mathrm{Diag}\,(\lambda_1(X), \cdots, \lambda_n(X))\; U^\top, \tag{3}$$

where $\lambda_1(X) \geq \cdots \geq \lambda_n(X)$ are the eigenvalues of X arranged in the non-increasing order and U is a matrix of the corresponding orthonormal eigenvectors. We denote the set of all matrices U satisfying (3) as $\mathbb{O}^n(X)$. We also use $\lambda(X)$ to denote the vector whose i-th entry is $\lambda_i(X)$. Let $v_1(X) > v_2(X) > \ldots > v_r(X)$ be the distinct eigenvalues of X arranged in the decreasing order. Define

$$\alpha^l := \{i \in \{1, \ldots, n\} \mid \lambda_i(X) = v_l(X)\}, \quad l = 1, \ldots, r. \tag{4}$$

For each $i \in \{1, \ldots, n\}$, we denote $k_i(X)$ as the number of eigenvalues that equal to $\lambda_i(X)$ but are ranked before i (including i) and $o_i(X)$ as the number of eigenvalues that equal to $\lambda_i(X)$ but are ranked after i (excluding i). That is, the scalars $k_i(X)$ and $o_i(X)$ satisfy

$$\begin{aligned} &\lambda_1(X) \geq \ldots \geq \lambda_{i-k_i(X)}(X) > \lambda_{i-k_i(X)+1}(X) = \ldots = \lambda_i(X) = \ldots = \lambda_{i+o_i(X)}(X) \\ &> \lambda_{i+o_i(X)+1}(X) \geq \ldots \geq \lambda_n(X). \end{aligned} \tag{5}$$

In the subsequent discussions, when the dependence of k_i and o_i on X can be easily seen from the context, we often drop X for simplicity.

In the following, we summarize some results about the properties of the eigenvalues that are essential in our subsequent discussions. The first result is Ky Fan's inequality [17].

Lemma 1 *Let Y and Z be two matrices in $\mathbb{S}^n$. Then*

$$\langle Y, Z\rangle \leq \lambda(Y)^\top \lambda(Z)\,,$$

where the equality holds if and only if Y and Z admit a simultaneous ordered eigenvalue decomposition, i.e., there exists an orthogonal matrix $U \in \mathbb{O}^n$ such that

$$Y = U\Lambda(Y)U^\top \quad \text{and} \quad Z = U\Lambda(Z)U^\top.$$

The next lemma is about the directional differentiability of the eigenvalue function, which can be found in, for example, [18, Theorem 7] and [35, Proposition 1.4].

Lemma 2 *Let $X \in \mathbb{S}^n$ have the eigenvalue decomposition in (3). Then for any $\mathbb{S}^n \ni H \to 0$, we have*

$$\lambda_i(X + H) - \lambda_i(X) - \lambda_{k_i}(U_{\alpha^l}^\top H U_{\alpha^l}) = O(\|H\|^2), \quad i \in \alpha^l,\ l = 1, \ldots, r,$$

where for each $i \in \{1, \ldots, n\}$, k_i is defined in (5). Hence, for any given direction $H \in \mathbb{S}^n$, the eigenvalue function $\lambda_i(\cdot)$ is directionally differentiable at X with the directional derivative $\lambda_i'(X; H) = \lambda_{k_i}(U_{\alpha^l}^\top H U_{\alpha^l})$ for any $i \in \alpha^l$, $l = 1, \ldots, r$.

Let $l \in \{1, \ldots, r\}$ be fixed. Consider the following eigenvalue decomposition of the symmetric matrix $U_{\alpha^l}^\top H U_{\alpha^l} \in \mathbb{S}^{|\alpha^l|}$:

$$U_{\alpha^l}^\top H U_{\alpha^l} = R\Lambda(U_{\alpha^l}^\top H U_{\alpha^l})R^\top,$$

where $R \in \mathbb{O}^{|\alpha^l|}$. Denote the distinct eigenvalues of $U_{\alpha^l}^\top H U_{\alpha^l}$ by $\tilde{v}_1 > \tilde{v}_2 > \ldots > \tilde{v}_{\tilde{r}}$. Define

$$\tilde{\alpha}^j := \left\{\, i \in \{1, \ldots, |\alpha^l|\} \mid \lambda_i(U_{\alpha^l}^\top H U_{\alpha^l}) = \tilde{v}_j \,\right\}, \quad j = 1, \ldots, \tilde{r}.$$

For each $i \in \alpha^l$, let $\tilde{k}_i \in \{1, \ldots, |\alpha^l|\}$ and $\tilde{l} \in \{1, \ldots, \tilde{r}\}$ be such that

$$\tilde{k}_i := k_{k_i}(U_{\alpha^l}^\top H U_{\alpha^l}) \quad \text{and} \quad \tilde{k}_i \in \tilde{\alpha}^{\tilde{l}},$$

where k_i is defined by (5).

Let $\mathbb{Z}$ and $\mathbb{Z}'$ be two finite dimensional real Euclidean spaces. We say that a function $\Phi : \mathbb{Z} \to \mathbb{Z}'$ is (parabolic) second order directionally differentiable at $\mathbf{z} \in \mathbb{Z}$, if Φ is directionally differentiable at $\mathbf{z}$ and for any $\mathbf{h}, \mathbf{w} \in \mathbb{Z}$,

$$\lim_{t \downarrow 0} \frac{\Phi(\mathbf{z} + t\mathbf{h} + \frac{1}{2}t^2\mathbf{w}) - \Phi(\mathbf{z}) - t\,\Phi(\mathbf{z}; \mathbf{h})}{\frac{1}{2}t^2} \quad \text{exists.}$$

In this case, the above limit is said to be the (parabolic) second order directional derivative of Φ at $\mathbf{z}$ along the directions $\mathbf{h}$ and $\mathbf{w}$, which we denote as $\Phi''(\mathbf{z}; \mathbf{h}, \mathbf{w})$. The following proposition, which has its source from [35, Proposition 2.2], provides an explicit formula of the (parabolic) second order directional derivative of the eigenvalue function.

Lemma 3 *Let $X \in \mathbb{S}^n$ have the eigenvalue decomposition* (3)*. Then for any $H, W \in \mathbb{S}^n$,*

$$\lambda_i''(X; H, W) = \lambda_{\tilde{k}_i}\left(R_{\tilde{\alpha}^l}^\top U_{\alpha^l}^\top \left[W - 2H(X - \lambda_i I_n)^\dagger H\right] U_{\alpha^l} R_{\tilde{\alpha}^l}\right), \quad i \in \alpha^l,\ l \in \{1, \ldots, r\},$$

where $Z^\dagger \in \mathbb{R}^{p \times p}$ is the Moore-Penrose pseudo-inverse of the square matrix $Z \in \mathbb{R}^{p \times p}$.

2.2 Properties of convex piecewise affine functions

Let $\phi : \mathbb{R}^n \to (-\infty, +\infty]$ be a proper convex piecewise affine function, i.e., ϕ is a convex function whose nonempty effective domain can be represented as the union of finitely many polyhedral sets, relative to each of which $\phi(x)$ is an affine function (cf. [32, Definition 2.47]). Such a function is also called a polyhedral convex function by Rockafellar [31, Section 19]. Let ϕ be a proper convex piecewise affine function with the polyhedral effective domain

$$\operatorname{dom}\phi := \left\{ x \in \mathbb{R}^n \mid \psi(x) := \max_{1 \le i \le q} \{\langle \mathbf{b}^i, \mathbf{x}\rangle - d_i\} \le 0 \right\} \tag{6}$$

for some $\{(\mathbf{b}^i, d_i) \in \mathbb{R}^n \times \mathbb{R}\}_{i=1}^q$ with a positive integer q. It is known from [32, Theorem 2.49] that ϕ can be expressed in the form of

$$\phi(\mathbf{x}) = \underbrace{\max_{1 \le i \le p} \left\{\langle \mathbf{a}^i, \mathbf{x}\rangle - c_i\right\}}_{\text{denoted as } \phi_1(\mathbf{x})} + \underbrace{\delta_{\operatorname{dom}\phi}(\mathbf{x})}_{\text{denoted as } \phi_2(\mathbf{x})}, \quad \mathbf{x} \in \mathbb{R}^n \tag{7}$$

for some $\{(\mathbf{a}^i, c_i) \in \mathbb{R}^n \times \mathbb{R}\}_{i=1}^p$ with a positive integer p. We call $\{\mathbf{a}^i\}_{i=1}^p$ and $\{\mathbf{b}^i\}_{i=1}^q$ the bases of ϕ_1 and ϕ_2, respectively.

We denote the set of all $n \times n$ permutation matrices as $\mathbb{P}^n$. Recall that the function ϕ is called symmetric over $\mathbb{R}^n$ if for any $Q \in \mathbb{P}^n$ and any $\mathbf{x} \in \mathbb{R}^n$, it holds that $\phi(\mathbf{x}) = \phi(Q\mathbf{x})$. The following proposition characterizes the symmetric piecewise affine function.

Proposition 1 *Let $\phi : \mathbb{R}^n \to (-\infty, \infty]$ be a given proper convex piecewise affine function. Then the function ϕ is symmetric over $\mathbb{R}^n$ if and only if its decomposed components $\phi_1 : \mathbb{R}^n \to \mathbb{R}$ and $\phi_2 : \mathbb{R}^n \to (-\infty, \infty]$ in* (7) *satisfy the following conditions:*

$$\begin{cases} \phi_1(\mathbf{x}) = \max\limits_{1 \le i \le p} \left\{ \max\limits_{Q \in \mathbb{P}^n} \left\{\langle Q\mathbf{a}^i, \mathbf{x}\rangle - c_i\right\} \right\}, \\ \operatorname{dom}\phi = \left\{ \mathbf{x} \in \mathbb{R}^n \mid \max\limits_{1 \le i \le q} \left\{ \max\limits_{Q \in \mathbb{P}^n} \left\{\langle Q\mathbf{b}^i, \mathbf{x}\rangle - d_i\right\} \right\} \le 0 \right\}, \end{cases} \quad \forall\, \mathbf{x} \in \mathbb{R}^n. \tag{8}$$

Proof "$\Longleftarrow$" Suppose that ϕ_1 and ϕ_2 satisfy the conditions in (8). Consider any $\mathbf{x} \in \mathbb{R}^n$ and $Q' \in \mathbb{P}^n$. If $\mathbf{x} \notin \operatorname{dom}\phi$, then there exist $\mathbf{b}^{\hat{i}} \in \mathbb{R}^n$, $d_{\hat{i}} \in \mathbb{R}$ and $\widehat{Q} \in \mathbb{P}^n$ such that

$$\langle \widehat{Q}\mathbf{b}^{\hat{i}}, \mathbf{x}\rangle - d_{\hat{i}} > 0.$$

Since $(Q')^\top Q' = I$, we have

$$\left\langle Q'\widehat{Q}\mathbf{b}^{\hat{i}}\,,\, Q'\mathbf{x}\right\rangle - d_{\hat{i}} > 0.$$

By noting that $Q'\widehat{Q} \in \mathbb{P}^n$, we conclude that $Q'\mathbf{x} \notin \operatorname{dom}\phi$, which implies that

$$\phi(\mathbf{x}) = \phi(Q'\mathbf{x}) = +\infty.$$

Otherwise if $\mathbf{x} \in \operatorname{dom}\phi$, we deduce that

$$\begin{aligned}\max_{1\le i\le q}\left\{\max_{Q\in\mathbb{P}^n}\left\{\langle Q\mathbf{b}^i, Q'\mathbf{x}\rangle - d_i\right\}\right\} &= \max_{1\le i\le q}\left\{\max_{Q\in\mathbb{P}^n}\left\{\langle Q'^\top Q\mathbf{b}^i, \mathbf{x}\rangle - d_i\right\}\right\}\\ &= \max_{1\le i\le q}\left\{\max_{Q'^\top Q\in\mathbb{P}^n}\left\{\langle Q'^\top Q\mathbf{b}^i, \mathbf{x}\rangle - d_i\right\}\right\} \le 0,\end{aligned}$$

which implies that $Q'\mathbf{x} \in \operatorname{dom}\phi$. Moreover, we have

$$\begin{aligned}\phi(Q'\mathbf{x}) = \phi_1(Q'\mathbf{x}) &= \max_{1\le i\le p}\left\{\max_{Q\in\mathbb{P}^n}\left\{\langle Q\mathbf{a}^i, Q'\mathbf{x}\rangle - c_i\right\}\right\} = \max_{1\le i\le p}\left\{\max_{Q\in\mathbb{P}^n}\left\{\langle Q'^\top Q\mathbf{a}^i, \mathbf{x}\rangle - c_i\right\}\right\}\\ &= \max_{1\le i\le p}\left\{\max_{Q'^\top Q\in\mathbb{P}^n}\left\{\langle Q'^\top Q\mathbf{a}^i, \mathbf{x}\rangle - c_i\right\}\right\} = \phi_1(\mathbf{x}) = \phi(\mathbf{x}).\end{aligned}$$

Thus, we know that ϕ is symmetric over $\mathbb{R}^n$.

"$\Longrightarrow$" Assume that ϕ is a proper convex piecewise affine function with the decomposition in (7). Denote

$$\mathcal{Z} := \left\{\mathbf{x}\in\mathbb{R}^n \;\middle|\; \max_{1\le i\le q}\left\{\max_{Q\in\mathbb{P}^n}\left\{\langle Q\mathbf{b}^i, \mathbf{x}\rangle - d_i\right\}\right\} \le 0\right\}.$$

Obviously $\mathcal{Z} \subseteq \operatorname{dom}\phi$. Suppose that $\mathbf{x} \in \operatorname{dom}\phi$ is arbitrarily chosen . It follows from the symmetric property of ϕ that

$$\left\langle Q\mathbf{b}^i, \mathbf{x}\right\rangle - d_i = \left\langle \mathbf{b}^i, Q^\top\mathbf{x}\right\rangle - d_i \le 0, \quad \forall\, Q\in\mathbb{P}^n, \quad \forall\, i = 1,\dots,q\ ,$$

which shows that $\mathbf{x} \in \mathcal{Z}$. Therefore, we have $\operatorname{dom}\phi = \mathcal{Z}$.

Denote

$$\overline{\phi}(x) := \max_{1\le i\le p}\left\{\max_{Q\in\mathbb{P}^n}\left\{\langle Q\mathbf{a}^i, \mathbf{x}\rangle - c_i\right\}\right\}, \quad \mathbf{x}\in\operatorname{dom}\phi.$$

It is clear that $\phi(\mathbf{x}) \le \overline{\phi}(\mathbf{x})$ for any $\mathbf{x} \in \operatorname{dom}\phi$. On the other hand, for any $\mathbf{x} \in \operatorname{dom}\phi$, there exist $\overline{i} \in \{1,\dots,p\}$ and $\overline{Q} \in \mathbb{P}^n$ such that

$$\overline{\phi}(\mathbf{x}) = \left\langle \overline{Q}\mathbf{a}^{\overline{i}},\, \mathbf{x}\right\rangle - c_{\overline{i}} = \left\langle \mathbf{a}^{\overline{i}},\, \overline{Q}^\top\mathbf{x}\right\rangle - c_{\overline{i}} \le \max_{1\le i\le p}\left\{\left\langle \mathbf{a}^i,\, \overline{Q}^\top\mathbf{x}\right\rangle - c_i\right\} = \phi(\overline{Q}^\top\mathbf{x}).$$

Finally, since ϕ is symmetric over $\mathbb{R}^n$, we have $\phi(\overline{Q}^\top\mathbf{x}) = \phi(\mathbf{x})$. Thus, we know from the above inequality that $\overline{\phi}(\mathbf{x}) \le \phi(\mathbf{x})$ for any $\mathbf{x} \in \operatorname{dom}\phi$. Therefore, we know that $\phi(\mathbf{x}) = \overline{\phi}(\mathbf{x})$ for any $\mathbf{x} \in \operatorname{dom}\phi$. The proof is completed. $\square$

Remark 1 It is easy to verify from Proposition 1 that if a proper convex piecewise affine function ϕ is symmetric, then both of its components ϕ_1 and ϕ_2 in (7) are symmetric.

To proceed, we denote

$$\mathcal{D}_i := \left\{\mathbf{x}\in\operatorname{dom}\phi \mid \langle\mathbf{a}^j, \mathbf{x}\rangle - c_j \le \langle\mathbf{a}^i, \mathbf{x}\rangle - c_i,\, \forall\, j = 1,\dots,p\right\}, \quad i = 1,\dots,p.$$

It follows from [24, Proposition 3.2] that $\operatorname{dom}\phi = \bigcup_{i=1,\dots,p} \mathcal{D}_i$. For any $\overline{\mathbf{x}} \in \operatorname{dom}\phi$, we further denote the following two index sets:

$$\iota_1(\overline{\mathbf{x}}) := \{1\le i\le p \mid \overline{\mathbf{x}}\in\mathcal{D}_i\} \quad \text{and} \quad \iota_2(\overline{\mathbf{x}}) := \{1\le i\le q \mid \langle\mathbf{b}^i, \overline{\mathbf{x}}\rangle - d_i = 0\}. \tag{9}$$

It is known that the pointwise-max function ϕ_1 in (7) and ψ in (6) are directionally differentiable everywhere with the following directional derivatives (see, e.g., [4, Example 2.68])

$$\phi_1'(\overline{\mathbf{x}};\mathbf{h}) = \max_{i\in\iota_1(\overline{\mathbf{x}})}\langle\mathbf{a}^i, \mathbf{h}\rangle \quad \text{and} \quad \psi'(\overline{\mathbf{x}};\mathbf{h}) = \max_{i\in\iota_2(\overline{\mathbf{x}})}\langle\mathbf{b}^i, \mathbf{h}\rangle, \quad \mathbf{h}\in\mathbb{R}^n. \tag{10}$$

Denote ∂f as the subgradient of a convex function f. It also holds that

$$\partial\phi_1(\overline{\mathbf{x}}) = \mathrm{conv}\{\mathbf{a}^i,\ i\in\iota_1(\overline{\mathbf{x}})\} \quad \text{and} \quad \partial\phi_2(\overline{\mathbf{x}}) = \mathcal{N}_{\mathrm{dom}\,\phi}(\overline{\mathbf{x}}) = \mathrm{cone}\{\mathbf{b}^i,\ i\in\iota_2(\overline{\mathbf{x}})\}, \tag{11}$$

where 'conv $\mathcal{C}$' and 'cone $\mathcal{C}$' stand for the convex hall and conic hall of a given nonempty closed set $\mathcal{C}$, respectively (if $\iota_2(\overline{\mathbf{x}})=\emptyset$, then $\partial\phi_2(\bar{x})=\{0\}$). Therefore, for any given $\overline{\mathbf{y}}\in\partial\phi_1(\overline{\mathbf{x}})$ and $\overline{\mathbf{z}}\in\partial\phi_2(\overline{\mathbf{x}})$, we are able to define the following two index sets

$$\begin{cases} \eta_1(\overline{\mathbf{x}},\overline{\mathbf{y}}) := \big\{i\in\iota_1(\overline{\mathbf{x}}) \mid \displaystyle\sum_{i\in\iota_1(\overline{\mathbf{x}})} u_i\,\mathbf{a}^i = \overline{\mathbf{y}},\ \sum_{i\in\iota_1(\overline{\mathbf{x}})} u_i = 1,\ 0<u_i\le 1\big\}, \\ \eta_2(\overline{\mathbf{x}},\overline{\mathbf{z}}) := \big\{i\in\iota_2(\overline{\mathbf{x}}) \mid \displaystyle\sum_{i\in\iota_2(\overline{\mathbf{x}})} u_i\mathbf{b}^i = \overline{\mathbf{z}},\ u_i>0\big\}. \end{cases} \tag{12}$$

The following corollary is a direct consequence of Proposition 1.

Corollary 1 *The following two statements hold.*

(i) For any $i\in\iota_1(\overline{\mathbf{x}})$, $j\in\iota_2(\overline{\mathbf{x}})$ and $Q\in\mathbb{P}^n_{\overline{\mathbf{x}}}$ (i.e., $Q\overline{\mathbf{x}}=\overline{\mathbf{x}}$), there exist $i'\in\iota_1(\overline{\mathbf{x}})$ and $j'\in\iota_2(\overline{\mathbf{x}})$ such that $\mathbf{a}^{i'}=Q\mathbf{a}^i$ and $\mathbf{b}^{j'}=Q\mathbf{b}^j$, respectively.

(ii) For any $i\in\eta_1(\overline{\mathbf{x}},\overline{\mathbf{y}})$, $j\in\eta_2(\overline{\mathbf{x}},\overline{\mathbf{z}})$, $Q^1\in\mathbb{P}^n_{\overline{\mathbf{x}}}\cap\mathbb{P}^n_{\overline{\mathbf{y}}}$ and $Q^2\in\mathbb{P}^n_{\overline{\mathbf{x}}}\cap\mathbb{P}^n_{\overline{\mathbf{z}}}$, there exist $i'\in\eta_1(\overline{\mathbf{x}},\overline{\mathbf{y}})$ and $j'\in\eta_2(\overline{\mathbf{x}},\overline{\mathbf{z}})$ such that $\mathbf{a}^{i'}=Q^1\,\mathbf{a}^i$ and $\mathbf{b}^{j'}=Q^2\,\mathbf{b}^j$, respectively.

Let $\mathbb{Z}$ be a finite dimensional Euclidean space and $\varpi:\mathbb{Z}\to(-\infty,\infty]$ be a proper closed convex function. The Moreau envelop and proximal mapping of ϖ are defined by

$$\chi_\varpi(\mathbf{z}) := \min_{\mathbf{w}\in\mathbb{Z}}\left\{\varpi(\mathbf{w})+\frac{1}{2}\|\mathbf{w}-\mathbf{z}\|^2\right\} \quad \text{and} \quad \mathrm{Pr}_\varpi(\mathbf{z}) := \operatorname*{argmin}_{\mathbf{w}\in\mathbb{Z}}\left\{\varpi(\mathbf{w})+\frac{1}{2}\|\mathbf{w}-\mathbf{z}\|^2\right\}, \quad \mathbf{z}\in\mathbb{Z}. \tag{13}$$

It is known that Pr_ϖ is globally Lipschitz continuous with modulus 1 [32, Proposition 12.19]. The directional derivative of the proximal mapping is closely related to the critical cone associated with the generalized equation $\overline{\mathbf{z}}\in\partial\varpi(\mathbf{z})$, which is defined as

$$\mathcal{C}(\overline{\mathbf{z}};\partial\varpi(\mathbf{z})) := \{\mathbf{d}\in\mathbb{R}^n \mid \varpi'(\mathbf{z};\mathbf{d}) = \langle\overline{\mathbf{z}}-\mathbf{z},\,\mathbf{d}\rangle\}\,.$$

The following proposition shows the directional derivative of the proximal mappings associated with $\partial\phi_1$ and $\partial\phi_2$. Necessary and sufficient conditions for them to be F(réchet)-differentiable are also provided.

Proposition 2 *Let $\phi_1:\mathbb{R}^n\to\mathbb{R}$ and $\phi_2:\mathbb{R}^n\to(-\infty,\infty]$ by given by (7). Then the following two properties hold for the corresponding proximal mappings $\mathrm{Pr}_{\phi_1}:\mathbb{R}^n\to\mathbb{R}^n$ and $\mathrm{Pr}_{\phi_2}:\mathbb{R}^n\to\mathbb{R}^n$.*

(i) Pr_{ϕ_1} and Pr_{ϕ_2} are directionally differentiable everywhere with the directional derivatives

$$\begin{cases} \mathrm{Pr}'_{\phi_1}(\mathbf{x};\mathbf{h}) = \operatorname*{argmin}_{d\in\mathbb{R}^n}\big\{\|\mathbf{d}-\mathbf{h}\|^2 \mid \mathbf{d}\in\mathcal{C}(\mathbf{x};\partial\phi_1(\mathrm{Pr}_{\phi_1}(\mathbf{x})))\big\} \\ \mathrm{Pr}'_{\phi_2}(\mathbf{x};\mathbf{h}) = \operatorname*{argmin}_{d\in\mathbb{R}^n}\big\{\|\mathbf{d}-\mathbf{h}\|^2 \mid \mathbf{d}\in\mathcal{C}(\mathbf{x};\partial\phi_2(\mathrm{Pr}_{\phi_2}(\mathbf{x})))\big\} \end{cases} \quad \mathbf{x},\mathbf{h}\in\mathbb{R}^n. \tag{14}$$

(ii) Let the index sets ι_1, ι_2, η_1 and η_2 be given by (9) and (12), respectively. Then Pr_{ϕ_1} is F-differentiable at $\mathbf{x}$ if and only if

$$\eta_1\,(\,\mathrm{Pr}_{\phi_1}(\mathbf{x}),\,\mathbf{x}-\mathrm{Pr}_{\phi_1}(\mathbf{x})\,) = \iota_1\,(\,\mathrm{Pr}_{\phi_1}(\mathbf{x})\,)\,;$$

similarly, Pr_{ϕ_2} is F-differentiable at $\mathbf{x}$ if and only if

$$\eta_2\,(\,\mathrm{Pr}_{\phi_2}(\mathbf{x}),\,\mathbf{x}-\mathrm{Pr}_{\phi_2}(\mathbf{x})\,) = \iota_2\,(\,\mathrm{Pr}_{\phi_2}(\mathbf{x})\,)\,.$$

Moreover, under the above two conditions, the derivatives $\mathrm{Pr}'_{\phi_1}(\mathbf{x})$ and $\mathrm{Pr}'_{\phi_2}(\mathbf{y})$ are given by

$$\begin{cases} \mathrm{Pr}'_{\phi_1}(\mathbf{x})\,\mathbf{h} = \operatorname*{argmin}_{d\in\mathbb{R}^n}\big\{\,\|\mathbf{d}-\mathbf{h}\|^2 \mid \langle\mathbf{d},\mathbf{a}^i-\mathbf{a}^j\rangle=0,\ i,j\in\iota_1\,(\,\mathrm{Pr}_{\phi_1}(\mathbf{x})\,)\,\big\} \\ \mathrm{Pr}'_{\phi_2}(\mathbf{x})\,\mathbf{h} = \operatorname*{argmin}_{d\in\mathbb{R}^n}\big\{\,\|\mathbf{d}-\mathbf{h}\|^2 \mid \langle\mathbf{d},\mathbf{b}^i\rangle=0,\ i\in\iota_2\,(\,\mathrm{Pr}_{\phi_2}(\mathbf{x})\,)\,\big\} \end{cases} \quad \mathbf{x},\mathbf{h}\in\mathbb{R}^n. \tag{15}$$

Proof Statement (i) follows from [3, Proposition 7.1 and Theorem 7.2]. To prove statement (ii), we first note that based on similar arguments of [16, Corollary 4.1.2], Pr_{ϕ_1} and Pr_{ϕ_2} are F(réchet)-differentiable at $\mathbf{x}$ and $\mathbf{y}$ if and only if the critical cones $\mathcal{C}(\mathbf{x};\partial\phi_1(\mathrm{Pr}_{\phi_1}(\mathbf{x})))$ and $\mathcal{C}(\mathbf{y};\partial\phi_2(\mathrm{Pr}_{\phi_2}(\mathbf{x})))$ are two linear subspaces in $\mathbb{R}^n$. The stated results then follow from [25, Proposition 3.2].

3 Variational analysis of spectral functions

In this section, we study several important variational properties of the spectral function $\theta \equiv \phi \circ \lambda$ for a symmetric piecewise affine function ϕ. According to the decomposition of $\phi = \phi_1 + \phi_2$ in (7), the function θ can be decomposed as

$$\theta(X) = \underbrace{\phi_1 \circ \lambda(X)}_{\text{denoted as } \theta_1(X)} + \underbrace{\phi_2 \circ \lambda(X)}_{\text{denoted as } \theta_2(X)}, \quad X \in \mathbb{S}^n. \tag{16}$$

The following lemma on the subdifferentials of spectral functions can be found in [20,21],

Lemma 4 *Let $\phi : \mathbb{R}^n \to (-\infty, +\infty]$ be a proper closed convex and symmetric function. Let $X \in \mathbb{S}^n$ have the eigenvalue $\lambda(X)$ in $\operatorname{dom}\phi$. Let $W \in \mathbb{S}^n$. Then $W \in \partial(\phi \circ \lambda)(X)$ if and only if $\lambda(W) \in \partial\phi(\lambda(X))$ and there exists $U \in \mathbb{O}^n(X) \cap \mathbb{O}^n(W)$. In fact, $\partial(\phi \circ \lambda)(X) = \{U \mathrm{Diag}(\mu) U^\top \mid \mu \in \partial g(\lambda(X)),\ U \in \mathbb{O}^n(X)\}$.*

In the following two subsections, we characterize the tangent sets, critical cones and the so-called "σ-term" associated with θ_1 and θ_2, respectively. Since the analysis of θ_1 and θ_2 are similar, we only give the detailed proof of the results for θ_1; the properties of θ_2 are presented without proof.

3.1 Variational properties of θ_1

We first study the variational properties of the spectral function $\theta_1 = \phi_1 \circ \lambda$.

Tangent set and its lineality space. Let $\overline{X} \in \mathbb{S}^n$ be given. Since θ_1 is Lipschitz continuous on $\mathbb{S}^n$, it follows from [4, Proposition 2.58] that the tangent cone $\mathcal{T}_{\operatorname{epi}\theta_1}(\overline{X}, \theta_1(\overline{X}))$ of the epigraph $\operatorname{epi}\theta_1$ at $(\overline{X}, \theta_1(\overline{X})) \in \operatorname{epi}\theta_1$ is given by

$$\mathcal{T}_{\operatorname{epi}\theta_1}(\overline{X}, \theta_1(\overline{X})) = \operatorname{epi}\theta_1'(\overline{X}; \bullet) = \left\{(H, h) \in \mathbb{S}^n \times \mathbb{R} \mid \theta_1'(\overline{X}; H) \le h\right\}.$$

The lineality space of $\mathcal{T}_{\operatorname{epi}\theta_1}(\overline{X}, \theta_1(\overline{X}))$, i.e., the largest linear subspace contained in $\mathcal{T}_{\operatorname{epi}\theta_1}(\overline{X}, \theta_1(\overline{X}))$, is given by

$$\begin{aligned}
\operatorname{lin}(\mathcal{T}_{\operatorname{epi}\theta_1}(\overline{X}, \theta_1(\overline{X}))) &:= \mathcal{T}_{\operatorname{epi}\theta_1}(\overline{X}, \theta_1(\overline{X})) \cap \left[-\mathcal{T}_{\operatorname{epi}\theta_1}(\overline{X}, \theta_1(\overline{X}))\right] \\
&= \left\{(H, y) \in \mathbb{S}^n \times \mathbb{R} \mid \theta_1'(\overline{X}; H) \le h \le -\theta_1'(\overline{X}; -H)\right\} \\
&= \left\{(H, y) \in \mathbb{S}^n \times \mathbb{R} \mid \theta_1'(\overline{X}; H) = h = -\theta_1'(\overline{X}; -H)\right\},
\end{aligned}$$

where the last equality follows from [31, Theorem 23.1]. We consider the following linear subspace

$$\mathcal{T}_{\theta_1}^{\mathrm{lin}}(\overline{X}) := \left\{H \in \mathbb{S}^n \mid \theta_1'(\overline{X}; H) = -\theta_1'(\overline{X}; -H)\right\}. \tag{17}$$

Let $\{\alpha^l\}_{l=1}^r$ be the index sets given by (4) with respect to $\overline{X}$. Define the index set

$$\widetilde{\mathcal{E}} := \{l \in \{1, \ldots, r\} \mid \exists\, i, j \in \alpha^l \text{ such that } (\mathbf{a}^w)_i \ne (\mathbf{a}^w)_j \text{ for some } w \in \iota_1(\lambda(\overline{X}))\},$$

where $\iota_1(\lambda(\overline{X}))$ is the set defined in (9) with respect to $\lambda(\overline{X})$. The following proposition presents the characterization of $\mathcal{T}_{\theta_1}^{\mathrm{lin}}(\overline{X})$.

Proposition 3 *For any $H \in \mathbb{S}^n$,*

$$H \in \mathcal{T}_{\theta_1}^{\mathrm{lin}}(\overline{X}) \iff \left[\langle \lambda'(\overline{X}; H), \mathbf{a}^i - \mathbf{a}^j \rangle = 0, \quad \forall\, i, j \in \iota_1(\lambda(\overline{X}))\right].$$

In fact, for each given $H \in \mathcal{T}_{\theta_1}^{\mathrm{lin}}(\overline{X})$ and $l \in \widetilde{\mathcal{E}}$, there exists a scalar $\widehat{\rho}_l$ such that for any $\overline{U} \in \mathbb{O}^n(\overline{X})$, $\overline{U}_{\alpha^l}^\top H \overline{U}_{\alpha^l} = \widehat{\rho}_l I_{|\alpha^l|}$. Moreover, we have

$$\langle \lambda'(\overline{X}; H),\, \mathbf{a}^i - \mathbf{a}^j \rangle = \sum_{l \in \widetilde{\mathcal{E}}} \left\langle \widehat{\rho}_l\, e_{|\alpha^l|}\,,\, (\mathbf{a}^i)_{\alpha^l} - (\mathbf{a}^j)_{\alpha^l} \right\rangle = 0, \quad \forall\, i, j \in \iota_1(\lambda(\overline{X})).$$

Proof Notice that for any $H \in \mathbb{S}^n$ and $U \in \mathbb{O}^n(\overline{X})$,

$$\begin{cases} \lambda'(\overline{X};H) = \left(\lambda(\overline{U}_{\alpha^1}^\top H\overline{U}_{\alpha^1}),\ldots,\lambda(\overline{U}_{\alpha^r}^\top H\overline{U}_{\alpha^r})\right)^\top, \\ \lambda'(\overline{X};-H) = \left(\lambda(-\overline{U}_{\alpha^1}^\top H\overline{U}_{\alpha^1}),\ldots,\lambda(-\overline{U}_{\alpha^r}^\top H\overline{U}_{\alpha^r})\right)^\top. \end{cases}$$

For each $1 \le l \le r$, we have $\lambda(-\overline{U}_{\alpha^l}^\top H\overline{U}_{\alpha^l}) = -\lambda^\uparrow(\overline{U}_{\alpha^l}^\top H\overline{U}_{\alpha^l})$. Moreover, we obtain from the symmetry of ϕ that for each permutation matrix $Q \in \mathbb{P}^n_{\lambda(\overline{X})}$, i.e., $Q\lambda(\overline{X}) = \lambda(\overline{X})$,

$$\phi'(\lambda(\overline{X});h) = \phi'(\lambda(\overline{X});Qh), \quad \forall\, h \in \mathbb{R}^n.$$

Therefore, $\theta_1'(\overline{X};H) = \phi_1'(\lambda(\overline{X});\lambda'(\overline{X};H)) = \sup\limits_{\mathbf{z}\in\partial\phi_1(\lambda(\overline{X}))} \langle \mathbf{z}, \lambda'(\overline{X};H)\rangle$ and

$$\begin{aligned} -\theta_1'(\overline{X};-H) &= -\phi_1'(\lambda(\overline{X});\lambda'(\overline{X};-H)) = -\phi_1'(\lambda(\overline{X});-\lambda'(\overline{X};H)) \\ &= -\sup_{\mathbf{z}\in\partial\phi_1(\lambda(\overline{X}))} \langle \mathbf{z}, -\lambda'(\overline{X};H)\rangle = \inf_{\mathbf{z}\in\partial\phi_1(\lambda(\overline{X}))} \langle \mathbf{z}, \lambda'(\overline{X};H)\rangle. \end{aligned}$$

We thus derive from (17) that $H \in \mathcal{T}^{\rm lin}_{\theta_1}(\overline{X})$ if and only if $\langle \mathbf{z}, \lambda'(\overline{X};H)\rangle$ is invariant over $\mathbf{z} \in \partial\phi_1(\lambda(\overline{X}))$. This, together with (11), implies the first claim of this proposition. The rest of the statements are consequences of Corollary 1. $\square$

Critical cone. Suppose that $\overline{Y} \in \partial\theta_1(\overline{X})$. Then the critical cone of $\partial\theta_1(\overline{X})$ at $\overline{X}+\overline{Y}$ is defined as

$$\mathcal{C}(\overline{X}+\overline{Y};\partial\theta_1(\overline{X})) := \left\{H \in \mathbb{S}^n \mid \theta_1'(\overline{X};H) = \langle \overline{Y}, H\rangle\right\}. \tag{18}$$

For each $l \in \{1,\ldots,r\}$, we further partition the index set α^l into $\{\beta^l_k\}_{k=1}^{s^l}$ such that each β^l_k contains one distinct eigenvalue of $\overline{Y}$, i.e.,

$$\begin{cases} \lambda_i(\overline{Y}) = \lambda_j(\overline{Y}) & \text{if } i,j \in \beta^l_k, \\ \lambda_i(\overline{Y}) > \lambda_j(\overline{Y}) & \text{if } i \in \beta^l_k,\ j \in \beta^l_{k'} \text{ and } k,k' \in \{1,\ldots,s^l\} \text{ with } k<k'. \end{cases} \tag{19}$$

We also denote

$$\mathcal{E}^l := \{k \in \{1,\ldots,s^l\} \mid \exists\, i,j \in \beta^l_k \text{ such that } (\mathbf{a}^w)_i \ne (\mathbf{a}^w)_j \text{ for some } w \in \eta_1(\lambda(\overline{X}),\lambda(\overline{Y}))\}, \tag{20}$$

where $\eta_1(\lambda(\overline{X}),\lambda(\overline{Y}))$ is the index set defined in (12) with respect to $\lambda(\overline{X})$ and $\lambda(\overline{Y})$. The characterization of the critical cone $\mathcal{C}(\overline{X}+\overline{Y};\partial\theta_1(\overline{X}))$ is provided in the following proposition.

Proposition 4 *Suppose that $(\overline{X},\overline{Y}) \in \operatorname{gph}\partial\theta_1$ and $\overline{U} \in \mathbb{O}^n(\overline{X}) \cap \mathbb{O}^n(\overline{Y})$. If $H \in \mathcal{C}(\overline{X}+\overline{Y};\partial\theta_1(\overline{X}))$, then the following three properties hold:*

(i) for each $l \in \{1,\ldots,r\}$, $\overline{U}_{\alpha^l}^\top H\overline{U}_{\alpha^l}$ has the following block diagonal structure:

$$\overline{U}_{\alpha^l}^\top H\overline{U}_{\alpha^l} = \operatorname{Diag}\left((\overline{U}_{\alpha^l}^\top H\overline{U}_{\alpha^l})_{\beta^l_1\beta^l_1},\ldots,(\overline{U}_{\alpha^l}^\top H\overline{U}_{\alpha^l})_{\beta^l_{s^l}\beta^l_{s^l}}\right);$$

(ii) $\langle \lambda'(\overline{X};H),\mathbf{a}^i\rangle = \langle \lambda'(\overline{X};H),\mathbf{a}^j\rangle = \max\limits_{i\in\iota_1(\lambda(\overline{X}))} \langle \lambda'(\overline{X};H),\mathbf{a}^i\rangle, \quad \forall\, i,j \in \eta_1(\lambda(\overline{X}),\lambda(\overline{Y}))$;

(iii) for each $l \in \{1,\ldots,r\}$ and $k \in \mathcal{E}^l$, there exists a scalar ρ^l_k such that $\left(\overline{U}_{\alpha^l}^\top H\overline{U}_{\alpha^l}\right)_{\beta^l_k\beta^l_k} = \rho^l_k I_{|\beta^l_k|}$.

In fact, $H \in \mathcal{C}(\overline{X}+\overline{Y};\partial\theta_1(\overline{X}))$ if and only if for any $i,j \in \eta_1(\lambda(\overline{X}),\lambda(\overline{Y}))$,

$$\left\langle \operatorname{diag}(\overline{U}^\top H\overline{U}),\, \mathbf{a}^i\right\rangle = \left\langle \operatorname{diag}(\overline{U}^\top H\overline{U}),\, \mathbf{a}^j\right\rangle = \max_{i\in\iota_1(\lambda(\overline{X}))} \left\langle \lambda'(\overline{X};H),\mathbf{a}^i\right\rangle, \tag{21}$$

where the index sets η_1 and ι_1 are defined in (12) and (9), respectively.

Proof It follows from Ky Fan's inequality in Lemma 1 that for any $H \in \mathbb{S}^n$,

$$\begin{aligned}\langle \overline{Y}, H\rangle = \langle \Lambda(\overline{Y}), \overline{U}^\top H\overline{U}\rangle &= \sum_{l=1}^{r}\langle \Lambda(\overline{Y})_{\alpha^l\alpha^l}\,,\, \overline{U}_{\alpha^l}^\top H\overline{U}_{\alpha^l}\rangle \\ &\leq \sum_{l=1}^{r}\langle \lambda(\overline{Y})_{\alpha^l}\,,\, \lambda(\overline{U}_{\alpha^l}^\top H\overline{U}_{\alpha^l})\rangle = \langle \lambda(\overline{Y})\,,\, \lambda'(\overline{X};H)\rangle \qquad (22)\\ &\leq \sup_{\mathbf{z}\in\partial\phi_1(\lambda(\overline{X}))} \big\langle \mathbf{z}, \lambda'(\overline{X};H)\big\rangle = \theta_1'(\overline{X};H). \qquad (23)\end{aligned}$$

Therefore, in order for $H \in \mathcal{C}(\overline{X}+\overline{Y};\partial\theta_1(\overline{X}))$, the equalities in (22) and (23) must hold.

Consider the inequality in (22). For each $l \in \{1,\ldots,r\}$, we know from Ky Fan's inequality that the equality in (22) holds if and only if there exists $R^l \in \mathbb{O}^{|\alpha^l|}$ such that

$$\Lambda(\overline{Y})_{\alpha^l\alpha^l} = R^l\Lambda(\overline{Y})_{\alpha^l\alpha^l}(R^l)^\top \quad \text{and} \quad \overline{U}_{\alpha^l}^\top H\overline{U}_{\alpha^l} = R^l\Lambda(\overline{U}_{\alpha^l}^\top H\overline{U}_{\alpha^l})(R^l)^\top,$$

where R^l has the following block diagonal structure:

$$R^l = \mathrm{Diag}\left(R_1^l,\ldots,R_{s^l}^l\right) \quad \text{with} \quad R_k^l \in \mathbb{O}^{|\beta_k^l|}, \quad k=1,\ldots,s^l.$$

Here $\{\beta_k^l\}_{k=1}^{s^l}$ is defined in (19). The statement (i) thus follows. In order for the equality in (23) holds, we must have $\langle\lambda(\overline{Y}),\lambda'(\overline{X};H)\rangle = \sup\limits_{\mathbf{z}\in\partial\phi_1(\lambda(\overline{X}))}\langle \mathbf{z},\lambda'(\overline{X};H)\rangle$, which implies the statement (ii).

On the other hand, for each $l\in\{1,\ldots,r\}$, if $k\in\mathcal{E}^l$, then there exist $i,j\in\beta_k^l$ such that $(\mathbf{a}^w)_i \neq (\mathbf{a}^w)_j$ for some $w\in\eta_1(\lambda(\overline{X}),\lambda(\overline{Y}))$. Consider the $n\times n$ permutation matrix $Q^{i,j}$ satisfying

$$(Q^{i,j}\mathbf{a}^w)_z = \begin{cases}(\mathbf{a}^w)_j & \text{if } z=i,\\ (\mathbf{a}^w)_i & \text{if } z=j,\\ (\mathbf{a}^w)_z & \text{otherwise,}\end{cases} \qquad z=1,\ldots,n.$$

Since $\lambda_i(\overline{X}) = \lambda_j(\overline{X})$ and $\lambda_i(\overline{Y}) = \lambda_j(\overline{Y})$, it is clear that $Q^{i,j}\lambda(\overline{X}) = \lambda(\overline{X})$ and $Q^{i,j}\lambda(\overline{Y}) = \lambda(\overline{Y})$. It then follows from Corollary 1 that there exists $w'\in\eta_1(\lambda(\overline{X}),\lambda(\overline{Y}))$ such that $\mathbf{a}^{w'} = Q^{i,j}\mathbf{a}^w$. Therefore, we derive from (ii) that

$$\langle\lambda'(\overline{X};H),\mathbf{a}^w-\mathbf{a}^{w'}\rangle = (\lambda_i'(\overline{X};H)-\lambda_j'(\overline{X};H))((\mathbf{a}^w)_i-(\mathbf{a}^w)_j) = 0,$$

which implies that

$$\lambda_i'(\overline{X};H) = \lambda_j'(\overline{X};H).$$

For any $i'\in\beta_k^l$ with $i'\neq i$ and $i'\neq j$, if $(\mathbf{a}^w)_{i'}\neq(\mathbf{a}^w)_i$, by replacing i by i' and j by i in the above argument, we obtain that

$$\lambda_{i'}'(\overline{X};H) = \lambda_i'(\overline{X};H) = \lambda_j'(\overline{X};H);$$

otherwise if $(\mathbf{a}^w)_{i'} = (\mathbf{a}^w)_i$, then by replacing i by i' in the above argument, we can also obtain the above equality. Consequently, we know that for any $k\in\mathcal{E}^l$, there exists some $\rho_k^l\in\mathbb{R}$ such that for any $i\in\beta_k^l$,

$$\lambda_i'(\overline{X};H) = \rho_k^l,$$

which, together with Lemma 2, shows the property (ii).

To establish the last statement of this proposition, we observe that for each $l\in\{1,\ldots,r\}$, if $k\notin\mathcal{E}^l$, then for any $w\in\eta_1(\lambda(\overline{X}),\lambda(\overline{Y}))$, there exists a scalar $\tilde{\rho}_k^l$ such that

$$(\mathbf{a}^w)_i = (\mathbf{a}^w)_j = \tilde{\rho}_k^l, \quad \forall\, i,j\in\beta_k^l,$$

which yields

$$\langle \lambda'(\overline{X}, H), \mathbf{a}^w \rangle = \sum_{l=1}^{r} \Big(\sum_{k\in\mathcal{E}^l} \big\langle \lambda'(\overline{X}, H)_{\beta_k^l}, (\mathbf{a}^w)_{\beta_k^l} \big\rangle + \sum_{k\notin\mathcal{E}^l} \big\langle \lambda'(\overline{X}, H)_{\beta_k^l}, (\mathbf{a}^w)_{\beta_k^l} \big\rangle \Big)$$

$$= \sum_{l=1}^{r} \Big(\sum_{k\in\mathcal{E}^l} \big\langle \rho_k^l e_{|\beta_k^l|}, (\mathbf{a}^w)_{\beta_k^l} \big\rangle + \sum_{k\notin\mathcal{E}^l} \big\langle \lambda((\overline{U}_{\alpha^l}^{\top} H \overline{U}_{\alpha^l})_{\beta_k^l\beta_k^l}),\, \tilde{\rho}_k^l e_{|\beta_k^l|} \big\rangle \Big)$$

$$= \sum_{l=1}^{r} \Big(\sum_{k\in\mathcal{E}^l} \big\langle \rho_k^l e_{|\beta_k^l|}, (\mathbf{a}^w)_{\beta_k^l} \big\rangle + \sum_{k\notin\mathcal{E}^l} \tilde{\rho}_k^l \operatorname{trace}\big((\overline{U}_{\alpha^l}^{\top} H \overline{U}_{\alpha^l})_{\beta_k^l\beta_k^l}\big) \Big)$$

$$= \sum_{l=1}^{r} \Big(\sum_{k\in\mathcal{E}^l} \big\langle \operatorname{diag}\big((\overline{U}_{\alpha^l}^{\top} H \overline{U}_{\alpha^l})_{\beta_k^l\beta_k^l}\big),\, (\mathbf{a}^w)_{\beta_k^l} \big\rangle + \sum_{k\notin\mathcal{E}^l} \big\langle \operatorname{diag}\big((\overline{U}_{\alpha^l}^{\top} H \overline{U}_{\alpha^l})_{\beta_k^l\beta_k^l}\big),\, \tilde{\rho}_k^l\, e_{|\beta_k^l|} \big\rangle \Big)$$

$$= \Big\langle \operatorname{diag}(\overline{U}^{\top} H \overline{U}), \mathbf{a}^w \Big\rangle.$$

Conversely, suppose that $H \in \mathbb{S}^n$ satisfies (21). We have

$$\langle \overline{Y}, H \rangle = \langle \Lambda(\overline{Y}), \overline{U}^{\top} H \overline{U} \rangle = \langle \lambda(\overline{Y}), \operatorname{diag}(\overline{U}^{\top} H \overline{U}) \rangle = \sum_{i\in\eta_1(\lambda(\overline{X}),\lambda(\overline{Y}))} u_i \langle \mathbf{a}^i, \operatorname{diag}(\overline{U}^{\top} H \overline{U}) \rangle = \theta_1'(\overline{X}; H),$$

which shows that $H \in \mathcal{C}(\overline{X} + \overline{Y}; \partial\theta_1(\overline{X}))$. The proof of this proposition is thus completed. □

Based on the above proposition, we can further characterize the affine hull of $\mathcal{C}(\overline{X}+\overline{Y}; \partial\theta_1(\overline{X}))$, which we denoted as aff $(\mathcal{C}(A; \partial\theta_1(\overline{X})))$. The proof can be directly obtained from Proposition 4. For simplicity, we omit the details here.

Proposition 5 *Suppose that $(\overline{X}, \overline{Y}) \in \operatorname{gph} \partial\theta_1$ and $\overline{U} \in \mathbb{O}^n(\overline{X}) \cap \mathbb{O}^n(\overline{Y})$. Then $H \in \operatorname{aff}(\mathcal{C}(A; \partial\theta_1(\overline{X})))$ if and only if it satisfies the the properties (i) and (iii) in Proposition 4, and*

$$\langle \operatorname{diag}(\overline{U}^{\top} H \overline{U}), \mathbf{a}^i \rangle = \langle \operatorname{diag}(\overline{U}^{\top} H \overline{U}), \mathbf{a}^j \rangle, \quad \forall\, i, j \in \eta_1(\lambda(\overline{X}), \lambda(\overline{Y})). \tag{24}$$

<u>σ-term</u>. Suppose that $\overline{Y} \in \partial\theta_1(\overline{X})$. Let $H \in \mathcal{C}(\overline{X}+\overline{Y}, \partial\theta_1(\overline{X}))$ be arbitrarily given. Since ϕ_1 is Lipschitz continuous, we know from [2, Lemma 3.1] that θ_1 is (parabolic) second-order directionally differentiable with the second order directional derivative

$$\digamma_{\overline{X},H}(W) := \theta_1''(\overline{X}; H, W) = \phi_1''(\lambda(\overline{X}); \lambda'(\overline{X}; H), \lambda''(\overline{X}; H, W)), \quad W \in \mathbb{S}^n. \tag{25}$$

Moreover, it is easy to see that $\digamma_{\overline{X},H} : \mathbb{S}^n \to \mathbb{R}$ is convex. We define the "σ-term" associated with the spectral function $\theta_1 = \phi_1 \circ \lambda$ at $\overline{Y} \in \partial\theta_1(\overline{X})$ as the conjugate function (cf. [31] for the definition) of $\digamma_{\overline{X},H}$ at $\overline{Y}$, that is, we consider the function

$$\digamma^*_{\overline{X},H}(\overline{Y}) = \sup_{W\in\mathbb{S}^n} \Big\{ \langle W, \overline{Y} \rangle - \digamma_{\overline{X},H}(W) \Big\}.$$

The proposition below characterizes the property of $\digamma^*_{\overline{X},H}$.

Proposition 6 *Suppose that $(\overline{X}, \overline{Y}) \in \operatorname{gph} \partial\theta_1$ and $\overline{U} \in \mathbb{O}^n(\overline{X}) \cap \mathbb{O}^n(\overline{Y})$. Denote $\bar{v}_1 > \bar{v}_2 > \cdots > \bar{v}_r$ as the distinct eigenvalues of $\overline{X}$. Let $H \in \mathcal{C}(\overline{X} + \overline{Y}, \partial\theta_1(\overline{X}))$ be given. Then*

$$\digamma^*_{\overline{X},H}(\overline{Y}) = 2 \sum_{l=1}^{r} \langle \Lambda(\overline{Y})_{\alpha^l\alpha^l}\,,\, \overline{U}_{\alpha^l}^{\top} H (\overline{X} - \bar{v}_l I)^{\dagger} H \overline{U}_{\alpha^l} \rangle, \tag{26}$$

where for each $l \in \{1, \ldots, r\}$, $(\overline{X} - \bar{v}_l I)^{\dagger}$ is the Moore-Penrose pseudo-inverse of $\overline{X} - \bar{v}_l I$.

Proof For any $W \in \mathbb{S}^n$, we have

$$\begin{aligned}
\langle W, \overline{Y} \rangle &= \langle W, \overline{U} \Lambda(\overline{Y}) \overline{U}^\top \rangle = \langle \overline{U}^\top W \overline{U}, \Lambda(\overline{Y}) \rangle \\
&= \sum_{l=1}^{r} \langle \Lambda(\overline{Y})_{\alpha^l \alpha^l}, \overline{U}_{\alpha^l}^\top W \overline{U}_{\alpha^l} \rangle = \sum_{l=1}^{r} \langle \Lambda(\overline{Y})_{\alpha^l \alpha^l}, \overline{U}_{\alpha^l}^\top (W - 2H(\overline{X} - \bar{v}_l I)^\dagger H) \overline{U}_{\alpha^l} \rangle \\
&\quad + 2 \sum_{l=1}^{r} \langle \Lambda(\overline{Y})_{\alpha^l \alpha^l}, \overline{U}_{\alpha^l}^\top H (\overline{X} - \bar{v}_l I)^\dagger H \overline{U}_{\alpha^l} \rangle.
\end{aligned}$$

It follows from [4, Example 2.68] that

$$\phi_1''(\lambda(\overline{X}); \lambda'(\overline{X}; H), \lambda''(\overline{X}; H, W)) = \max_{i \in \xi_1(\lambda(\overline{X}), \lambda'(\overline{X}; H))} \langle \lambda''(\overline{X}; H, W), \mathbf{a}^i \rangle,$$

where $\xi_1(\lambda(\overline{X}), \lambda'(\overline{X}; H)) \subseteq \iota_1(\lambda(\overline{X}))$ is defined by

$$\xi_1(\lambda(\overline{X}), \lambda'(\overline{X}; H)) := \left\{ i \in \iota_1(\lambda(\overline{X})) \mid \langle \lambda'(\overline{X}; H), \mathbf{a}^i \rangle = \max_{j \in \iota_1(\lambda(\overline{X}))} \langle \lambda'(\overline{X}; H), \mathbf{a}^j \rangle \right\}.$$

We then have

$$\begin{aligned}
\digamma^*_{\overline{X}, H}(\overline{Y}) &= \sup_{W \in \mathbb{S}^n} \left\{ \langle W, \overline{Y} \rangle - \phi_1''(\lambda(\overline{X}); \lambda'(\overline{X}; H), \lambda''(\overline{X}; H, W)) \right\} \\
&= 2 \sum_{l=1}^{r} \langle \Lambda(\overline{Y})_{\alpha^l \alpha^l}, \overline{U}_{\alpha^l}^\top H (\overline{X} - \bar{v}_l I)^\dagger H \overline{U}_{\alpha^l} \rangle \\
&+ \sup_{W \in \mathbb{S}^n} \Big\{ \underbrace{\sum_{l=1}^{r} \langle \Lambda(\overline{Y})_{\alpha^l \alpha^l}, \overline{U}_{\alpha^l}^\top (W - 2H(\overline{X} - \bar{v}_l I)^\dagger H) \overline{U}_{\alpha^l} \rangle - \max_{i \in \xi_1(\lambda(\overline{X}), \lambda'(\overline{X}; H))} \langle \lambda''(\overline{X}; H, W), \mathbf{a}^i \rangle}_{\text{denoted as } \Xi(W)} \Big\}.
\end{aligned}$$

Therefore, in order to prove this proposition, it suffices to show that $\sup_{W \in \mathbb{S}^n} \Xi(W) = 0$. In fact, since $H \in \mathcal{C}(\overline{X} + \overline{Y}, \partial\theta_1(\overline{X}))$, we know from Proposition 4(i) that for each $l \in \{1, \ldots, r\}$, there exists $R^l \in \mathbb{O}^{|\alpha^l|}$ such that

$$\Lambda(\overline{Y})_{\alpha^l \alpha^l} = R^l \Lambda(\overline{Y})_{\alpha^l \alpha^l} (R^l)^\top \quad \text{and} \quad \overline{U}_{\alpha^l}^\top H \overline{U}_{\alpha^l} = R^l \Lambda(\overline{U}_{\alpha^l}^\top H \overline{U}_{\alpha^l}) (R^l)^\top.$$

Therefore, for any $W \in \mathbb{S}^n$ and $l \in \{1, \ldots, r\}$,

$$\begin{aligned}
&\left\langle \Lambda(\overline{Y})_{\alpha^l \alpha^l}, \overline{U}_{\alpha^l}^\top (W - 2H(\overline{X} - \bar{v}_l I)^\dagger H) \overline{U}_{\alpha^l} \right\rangle \\
&= \left\langle \Lambda(\overline{Y})_{\alpha^l \alpha^l}, (R^l)^\top \overline{U}_{\alpha^l}^\top (W - 2H(\overline{X} - \bar{v}_l I)^\dagger H) \overline{U}_{\alpha^l} R^l \right\rangle \\
&= \sum_{\tilde{l}=1}^{\tilde{r}} \left\langle (\Lambda(\overline{Y})_{\alpha^l \alpha^l})_{\tilde{\alpha}^{\tilde{l}} \tilde{\alpha}^{\tilde{l}}}, (R^l_{\tilde{\alpha}^{\tilde{l}}})^\top \overline{U}_{\alpha^l}^\top \left[W - 2H(\overline{X} - \bar{v}_l I)^\dagger H \right] \overline{U}_{\alpha^l} R^l_{\tilde{\alpha}^{\tilde{l}}} \right\rangle \\
&\leq \sum_{\tilde{l}=1}^{\tilde{r}} \sum_{i \in \tilde{\alpha}^{\tilde{l}}} \lambda_i(\overline{Y}) \lambda_i \left(R_{\tilde{\alpha}^{\tilde{l}}}^\top \overline{U}_{\alpha^l}^\top \left[W - 2H(\overline{X} - \bar{v}_l I)^\dagger H \right] \overline{U}_{\alpha^l} R_{\tilde{\alpha}^{\tilde{l}}} \right) = \sum_{i \in \alpha^l} \lambda_i(\overline{Y}) \lambda_i''(\overline{X}; H, W),
\end{aligned}$$

where the first inequality follows from Ky Fan's inequality in Lemma 1 and the last equality is due to Lemma 3. Since $\lambda(\overline{Y}) \in \partial\phi_1(\lambda(\overline{X}))$, we know from (11) that there exists $\{u_i \in [0,1]\}_{i \in \iota_1(\lambda(\overline{X}))}$ with $\sum_{i \in \iota_1(\lambda(\overline{X}))} u_i = 1$ such that $\lambda(\overline{Y}) = \sum_{i \in \iota_1(\lambda(\overline{X}))} u_i \, \mathbf{a}^i$. It then follows from Proposition 4 (ii) that if $H \in \mathcal{C}(\overline{X} + \overline{Y}, \partial\theta_1(\overline{X}))$, then

$$\eta_1(\lambda(\overline{X}), \lambda(\overline{Y})) \subseteq \xi_1(\lambda(\overline{X}), \lambda'(\overline{X}; H)),$$

where the index set $\eta_1(\lambda(\overline{X}),\lambda(\overline{Y}))$ is defined in (12). We then derive

$$
\begin{aligned}
\varXi(W) &\le \sum_{l=1}^{r}\sum_{i\in\alpha^l}\lambda_i(\overline{Y})\lambda_i''(\overline{X};H,W) - \max_{i\in\xi_1(\lambda(\overline{X}),\lambda'(\overline{X};H))}\langle\lambda''(\overline{X};H,W),\mathbf{a}^i\rangle \\
&= \left\langle \sum_{i\in\iota_1(\lambda(\overline{X}))} u_i\,\mathbf{a}^i,\lambda''(\overline{X};H,W)\right\rangle - \max_{i\in\xi_1(\lambda(\overline{X}),\lambda'(\overline{X};H))}\langle\lambda''(\overline{X};H,W),\mathbf{a}^i\rangle \\
&\le \sum_{i\in\eta_1(\lambda(\overline{X}),\lambda(\overline{Y}))} u_i\langle\lambda''(\overline{X};H,W),\mathbf{a}^i\rangle - \max_{i\in\xi_1(\lambda(\overline{X}),\lambda'(\overline{X};H))}\langle\lambda''(\overline{X};H,W),\mathbf{a}^i\rangle \\
&\le \sum_{i\in\eta_1(\lambda(\overline{X}),\lambda(\overline{Y}))} u_i \max_{i\in\eta_1(\lambda(\overline{X}),\lambda(\overline{Y}))}\langle\lambda''(\overline{X};H,W),\mathbf{a}^i\rangle - \max_{i\in\xi_1(\lambda(\overline{X}),\lambda'(\overline{X};H))}\langle\lambda''(\overline{X};H,W),\mathbf{a}^i\rangle \\
&= \max_{i\in\eta_1(\lambda(\overline{X}),\lambda(\overline{Y}))}\langle\lambda''(\overline{X};H,W),\mathbf{a}^i\rangle - \max_{i\in\xi_1(\lambda(\overline{X}),\lambda'(\overline{X};H))}\langle\lambda''(\overline{X};H,W),\mathbf{a}^i\rangle \le 0.
\end{aligned}
$$

On the other hand, it is easy to see that $\varXi(\widehat{W})=0$ if

$$
\overline{U}_{\alpha^l}^{\top}\,\widehat{W}\,\overline{U}_{\alpha^l} = 2\,\overline{U}_{\alpha^l}^{\top}H(\overline{X}-\bar{v}_l I)^{\dagger}H\overline{U}_{\alpha^l}, \quad l=1,\ldots,r.
$$

Therefore, we know that $\sup\limits_{W\in\mathbb{S}^n}\varXi(W)=0$. This completes the proof of the proposition. □

Remark 2 In fact, for any given $\overline{X}\in\mathbb{S}^n$ and any $\overline{Y},H\in\mathbb{S}^n$ (not necessary in $\mathcal{C}(\overline{X}+\overline{Y},\partial\theta_1(\overline{X}))$), we can define the function $\varUpsilon^1_{\overline{X}}:\mathbb{S}^n\times\mathbb{S}^n\to\mathbb{R}$ as the right side of (26), i.e.,

$$
\varUpsilon^1_{\overline{X}}\big(\overline{Y},H\big) := 2\sum_{l=1}^{r}\langle\varLambda(\overline{Y})_{\alpha^l\alpha^l},\overline{U}_{\alpha^l}^{\top}H(\overline{X}-\bar{v}_l I)^{\dagger}H\overline{U}_{\alpha^l}\rangle, \tag{27}
$$

where $\overline{U}\in\mathbb{O}^n(\overline{X})\cap\mathbb{O}^n(\overline{Y})$. Notice that if $\overline{Y}\in\partial\theta_1(\overline{X})$, it holds that

$$
\varUpsilon^1_{\overline{X}}\big(\overline{Y},H\big) = -2\sum_{1\le l<l'\le r}\sum_{i\in\alpha^l}\sum_{j\in\alpha^{l'}}\frac{\lambda_i(\overline{Y})-\lambda_j(\overline{Y})}{\lambda_i(\overline{X})-\lambda_j(\overline{X})}(\overline{U}_{\alpha^l}^{\top}H\overline{U}_{\alpha^{l'}})^2_{ij}. \tag{28}
$$

Since for any $i\in\alpha^l$ and $j\in\alpha^{l'}$ with $1\le l<l'\le r$, $\dfrac{\lambda_i(\overline{Y})-\lambda_j(\overline{Y})}{\lambda_i(\overline{X})-\lambda_j(\overline{X})}\ge 0$, we conclude that

$$
\varUpsilon^1_{\overline{X}}\big(\overline{Y},H\big)\le 0, \quad \forall\, H\in\mathbb{S}^n.
$$

3.2 Variational properties of θ_2

In this subsection, we present analogue results with respect to the function θ_2. Recall the definition of the convex piecewise affine function ψ in (6). For notational simplicity, we denote $\zeta:=\psi\circ\lambda$ as the spectral function associated with ψ. Thus, the function θ_2 can be viewed as the the indictor function of the closed convex set $\mathcal{K}$ that is defined in the following way

$$
\mathcal{K} := \{X\in\mathbb{S}^n \mid \lambda(X)\in\operatorname{dom}\phi\} = \{X\in\mathbb{S}^n \mid \zeta(X)\le 0\}\,. \tag{29}
$$

Let $\overline{X}\in\mathcal{K}$ be given. Denote $\mathcal{N}_{\mathcal{K}}(\overline{X})$ as the normal cone of $\mathcal{K}$ at $\overline{X}\in\mathbb{S}^n$ in the sense of convex analysis [31]. In the rest of the paper, we assume the following Slater condition for the closed convex set $\mathcal{K}$.

Assumption 1 *There exists $\widetilde{X}\in\mathcal{K}$ such that $\zeta(\widetilde{X})<0$.*

It is worth mentioning that the above assumption automatically holds for many interesting matrix optimization problems, such as the negative semidefinite programming (where $\mathcal{K}$ is the negative semidefinite matrix cone).

Recall the index sets $\{\alpha^l\}_{l=1}^r$ given by (4) with respect to $\overline{X}$. Three variational properties with respect to θ_2 are in order.

Tangent cone and its lineality space. Let $\overline{X} \in \mathcal{K}$ be such that $\zeta(\overline{X}) = 0$. Since $\zeta : \mathbb{S}^n \to \mathbb{R}$ is a closed convex function, it follows from [4, Proposition 2.61] that the tangent cone $\mathcal{T}_{\mathcal{K}}(\overline{X})$ of the closed convex set $\mathcal{K}$ is given by

$$\mathcal{T}_{\mathcal{K}}(\overline{X}) := \left\{ H \in \mathbb{S}^n \mid \zeta'(\overline{X}; H) \le 0 \right\}.$$

Let $\iota_2(\lambda(\overline{X}))$ be the index set defined by (9) with respect to $\lambda(\overline{X})$, i.e.,

$$\iota_2(\lambda(\overline{X})) = \{1 \le i \le q \mid \langle \mathbf{b}^i, \lambda(\overline{X}) \rangle - d_i = 0\}.$$

It then follows from [4, Example 2.68] that for any $H \in \mathbb{S}^n$,

$$\zeta'(\overline{X}; H) = \psi'(\lambda(\overline{X}); \lambda'(\overline{X}; H)) = \max_{i \in \iota_2(\lambda(\overline{X}))} \langle \mathbf{b}^i, \lambda'(\overline{X}; H) \rangle.$$

Thus, the tangent cone $\mathcal{T}_{\mathcal{K}}(\overline{X})$ of the convex set $\mathcal{K}$ can be re-written as

$$\mathcal{T}_{\mathcal{K}}(\overline{X}) = \left\{ H \in \mathbb{S}^n \mid \langle \mathbf{b}^i, \lambda'(\overline{X}; H) \rangle \le 0, \ \forall\, i \in \iota_2(\lambda(\overline{X})) \right\}. \tag{30}$$

Moreover, the corresponding lineality space $\mathrm{lin}(\mathcal{T}_{\mathcal{K}}(\overline{X}))$ of $\mathcal{T}_{\mathcal{K}}(\overline{X})$ is given by

$$\begin{aligned} \mathrm{lin}(\mathcal{T}_{\mathcal{K}}(\overline{X})) &:= \mathcal{T}_{\mathcal{K}}(\overline{X}) \cap (-\mathcal{T}_{\mathcal{K}}(\overline{X})) = \left\{ H \in \mathbb{S}^n \mid \zeta'(\overline{X}; H) \le 0 \le -\zeta'(\overline{X}; -H) \right\} \\ &= \left\{ H \in \mathbb{S}^n \mid \zeta'(\overline{X}; H) = -\zeta'(\overline{X}; -H) = 0 \right\}, \end{aligned} \tag{31}$$

where the last equality follows from [31, Theorem 23.1]. Define the index set

$$\widetilde{\mathcal{F}} := \{ l \in \{1, \ldots, r\} \mid \exists\, i, j \in \alpha^l \text{ such that } (\mathbf{a}^w)_i \ne (\mathbf{a}^w)_j \text{ for some } w \in \iota_2(\lambda(\overline{X}))\}.$$

By employing similar arguments in the proof of Proposition 3, we obtain the following characterization of $\mathrm{lin}(\mathcal{T}_{\mathcal{K}}(\overline{X}))$ based on Corollary 1.

Proposition 7 *For any $H \in \mathbb{S}^n$,*

$$H \in \mathrm{lin}(\mathcal{T}_{\mathcal{K}}(\overline{X})) \iff \left[\, \langle \lambda'(\overline{X}; H), \mathbf{b}^i \rangle = 0, \quad \forall\, i \in \iota_2(\lambda(\overline{X})) \,\right].$$

In fact, for each given $H \in \mathrm{lin}(\mathcal{T}_{\mathcal{K}}(\overline{X}))$ and $l \in \{1, \ldots, r\}$, there exists a scalar $\widetilde{\rho}_l$ such that for any $\overline{V} \in \mathbb{O}^n(\overline{X})$,

$$\overline{V}_{\alpha^l}^\top H \overline{V}_{\alpha^l} = \widetilde{\rho}_l\, I_{|\alpha^l|}.$$

Critical cone. Let $\overline{X} \in \mathcal{K}$ and $\overline{Z} \in \mathcal{N}_{\mathcal{K}}(\overline{X})$. The critical cone of $\mathcal{N}_{\mathcal{K}}(\overline{X})$ at $\overline{X} + \overline{Z}$ is defined by

$$\mathcal{C}(\overline{X} + \overline{Z}; \mathcal{N}_{\mathcal{K}}(\overline{X})) := \mathcal{T}_{\mathcal{K}}(\overline{X}) \cap \overline{Z}^{\perp} = \left\{ H \in \mathbb{S}^n \mid \zeta'(\overline{X}; H) \le 0, \ \langle \overline{Z}, H \rangle = 0 \right\}. \tag{32}$$

For each $l \in \{1, \ldots, r\}$, similar to the definition of the index sets β_k^l in (19), we use the notation $\{\gamma_k^l\}_{k=1}^{t_l}$ to further partition the set α^l based on the eigenvalue of $\overline{Z}$ as

$$\begin{cases} \lambda_i(\overline{Z}) = \lambda_j(\overline{Z}) & \text{if } i, j \in \gamma_k^l \text{ and } k \in \{1, \ldots, t_l\}, \\ \lambda_i(\overline{Z}) > \lambda_j(\overline{Z}) & \text{if } i \in \gamma_k^l,\ j \in \gamma_{k'}^l \text{ and } k, k' \in \{1, \ldots, t_l\} \text{ with } k < k'. \end{cases} \tag{33}$$

Recall the index set $\eta_2(\lambda(\overline{X}), \lambda(\overline{Z}))$ defined in (12) with respect to $\lambda(\overline{X})$ and $\lambda(\overline{Z})$. For each $l \in \{1, \ldots, r\}$, define the index set

$$\mathcal{F}_l := \{ k \in \{1, \ldots, t_l\} \mid \exists\, i, j \in \gamma_k^l \text{ such that } (\mathbf{b}^w)_i \ne (\mathbf{b}^w)_j \text{ for some } w \in \eta_2(\lambda(\overline{X}), \lambda(\overline{Z}))\}. \tag{34}$$

The following result on the characterization of $\mathcal{C}(\overline{X} + \overline{Z}; \mathcal{N}_{\mathcal{K}}(\overline{X}))$ can be obtained similarly as Proposition 4 for θ_1. For brevity, we omit the proof here.

Proposition 8 *Suppose that $(\overline{X},\overline{Z})\in \operatorname{gph}\mathcal{N}_{\mathcal{K}}$ and $\overline{V}\in \mathbb{O}^n(\overline{X})\cap\mathbb{O}^n(\overline{Z})$. If $H\in \mathcal{C}(\overline{X}+\overline{Z};\mathcal{N}_{\mathcal{K}}(\overline{X}))$, then the following three conditions hold:*

(i) for each $l\in\{1,\ldots,r\}$, $\overline{V}_{\alpha^l}^{\top}H\overline{V}_{\alpha^l}$ has the following block diagonal structure, i.e.,

$$\overline{V}_{\alpha^l}^{\top}H\overline{V}_{\alpha^l}=\operatorname{Diag}\left((\overline{V}_{\alpha^l}^{\top}H\overline{V}_{\alpha^l})_{\gamma_1^l\gamma_1^l},\cdots,(\overline{V}_{\alpha^l}^{\top}H\overline{V}_{\alpha^l})_{\gamma_{t_l}^l\gamma_{t_l}^l}\right);$$

(ii) $\langle\lambda'(\overline{X};H),\mathbf{b}^i\rangle=\max\limits_{i\in\iota_2(\lambda(\overline{X}))}\langle\lambda'(\overline{X};H),\mathbf{b}^i\rangle=0,\quad\forall\, i\in\eta_2(\lambda(\overline{X}),\lambda(\overline{Y}))$;

(iii) for each $l\in\{1,\ldots,r\}$ and $k\in\mathcal{F}_l$, there exists a scalar $\rho_k^l\in\mathbb{R}$ such that $(\overline{V}_{\alpha^l}^{\top}H\overline{V}_{\alpha^l})_{\gamma_k^l\gamma_k^l}=\rho_k^l\, I_{|\gamma_k^l|}$.

In fact, $H\in\mathcal{C}(\overline{X}+\overline{Z};\mathcal{N}_{\mathcal{K}}(\overline{X}))$ if and only if for any $i\in\eta_2(\lambda(\overline{X}),\lambda(\overline{Z}))$,

$$\langle\operatorname{diag}(\overline{V}^{\top}H\overline{V}),\mathbf{b}^i\rangle=\max_{i\in\iota_2(\lambda(\overline{X}))}\langle\lambda'(\overline{X};H),\mathbf{b}^i\rangle=0,$$

where the index set η_2 and ι_2 are defined in (12) *and* (9)*.*

The results below on the characterization of the affine hull of the critical cone $\mathcal{C}(\overline{X}+\overline{Z};\mathcal{N}_{\mathcal{K}}(\overline{X}))$ follows from Proposition 8.

Proposition 9 *Suppose that $(\overline{X},\overline{Z})\in\operatorname{gph}\mathcal{N}_{\mathcal{K}}$. Let $\overline{V}\in\mathbb{O}^n(\overline{X})\cap\mathbb{O}^n(\overline{Z})$. Then $H\in\operatorname{aff}(\mathcal{C}(\overline{X}+\overline{Z};\mathcal{N}_{\mathcal{K}}(\overline{X})))$ if and only if it satisfies the properties (i) and (iii) in Proposition 8, and for any $i\in\eta_2(\lambda(\overline{X}),\lambda(\overline{Z}))$,*

$$\langle\operatorname{diag}(\overline{V}^{\top}H\overline{V}),\mathbf{b}^i\rangle=0,$$

where the index set $\eta_2(\lambda(\overline{X}),\lambda(\overline{Z}))\subseteq\iota_2(\lambda(\overline{X}))$ is defined in (12) *with respect to $\lambda(\overline{X})$ and $\lambda(\overline{Z})$.*

<u>**σ-term**</u>. Suppose that $(\overline{X},\overline{Z})\in\operatorname{gph}\mathcal{N}_{\mathcal{K}}$. Let $H\in\mathcal{C}(\overline{X}+\overline{Z},\mathcal{N}_{\mathcal{K}}(\overline{X}))$ be arbitrarily given. Since $\zeta:\mathbb{S}^n\to\mathbb{R}$ is Lipschitz continuous, we know from [2, Lemma 3.1] that ζ is (parabolic) second-order directionally differentiable and for any $W\in\mathbb{S}^n$,

$$\zeta''(\overline{X};H,W)=\psi''(\lambda(\overline{X});\lambda'(\overline{X};H),\lambda''(\overline{X};H,W)).$$

Since $\mathcal{K}$ is $\mathcal{C}^2$-cone reducible (see [4, Definition 3.135] for the definition) and Assumption 1 holds, the second order tangent set of $\mathcal{K}$ at $\overline{X}$ along H is given by

$$\mathcal{T}_{\mathcal{K}}^2(\overline{X},H)=\{W\in\mathbb{S}^n\mid\zeta''(\overline{X};H,W)\le 0\}.$$

As in the conventional conic programming, the "σ-term" associated with $\mathcal{K}$ is defined as the support function of its second order tangent set, whose explicit expression is given in the following proposition. The proof can be obtained in a similar fashion as that of Proposition 6.

Proposition 10 *Suppose that $(\overline{X},\overline{Z})\in\operatorname{gph}\mathcal{N}_{\mathcal{K}}$ and $\overline{V}\in\mathbb{O}^n(\overline{X})\cap\mathbb{O}^n(\overline{Z})$. Let $H\in\mathcal{C}(\overline{X}+\overline{Z};\mathcal{N}_{\mathcal{K}}(\overline{X}))$ be given. Then the support function of $\mathcal{T}_{\mathcal{K}}^2(\overline{X},H)$ at $\overline{Z}$ takes the following form*

$$\delta^*_{\mathcal{T}_{\mathcal{K}}^2(\overline{X},H)}(\overline{Z})=2\sum_{l=1}^{r}\langle\Lambda(\overline{Z})_{\alpha^l\alpha^l},\overline{V}_{\alpha^l}^{\top}H(\overline{X}-\bar{v}_lI)^{\dagger}H\overline{V}_{\alpha^l}\rangle.\tag{35}$$

Remark 3 Similarly as that for θ_1, for any given $\overline{X}\in\mathbb{S}^n$, define the function $\Upsilon_{\overline{X}}^2:\mathcal{N}_{\mathcal{K}}(\overline{X})\times\mathbb{S}^n\to\mathbb{R}$ as the value of the right side of (35), i.e.,

$$\Upsilon_{\overline{X}}^2\left(\overline{Z},H\right):=2\sum_{l=1}^{r}\langle\Lambda(\overline{Z})_{\alpha^l\alpha^l},\overline{V}_{\alpha^l}^{\top}H(\overline{X}-\bar{v}_lI)^{\dagger}H\overline{V}_{\alpha^l}\rangle,\quad\overline{Z}\in\mathcal{N}_{\mathcal{K}}(\overline{X})\quad\text{and}\quad H\in\mathbb{S}^n,\tag{36}$$

where $\overline{V}\in\mathbb{O}^n(\overline{X})$. If $\overline{Z}\in\mathcal{N}_{\mathcal{K}}(\overline{X})$, then for $\overline{V}\in\mathbb{O}^n(\overline{X})\cap\mathbb{O}^n(\overline{Z})$,

$$\Upsilon_{\overline{X}}^2\left(\overline{Z},H\right)=-2\sum_{1\le l<l'\le r}\sum_{i\in\alpha^l}\sum_{j\in\alpha^{l'}}\frac{\lambda_i(\overline{Z})-\lambda_j(\overline{Z})}{\lambda_i(\overline{X})-\lambda_j(\overline{X})}(\overline{V}_{\alpha^l}^{\top}H\overline{V}_{\alpha^{l'}})_{ij}^2.\tag{37}$$

Moreover, since for any $i\in\alpha^l$ and $j\in\alpha^{l'}$ with $1\le l<l'\le r$, $\dfrac{\lambda_i(\overline{Z})-\lambda_j(\overline{Z})}{\lambda_i(\overline{X})-\lambda_j(\overline{X})}\ge 0$, we know that

$$\Upsilon_{\overline{X}}^2\left(\overline{Z},H\right)\le 0,\quad\forall\, H\in\mathbb{S}^n.$$

4 Characterization of the strong regularity

This section is devoted to the characterization of the strong regularity of the solution to the KKT optimality condition for problem (1). Based on the decomposition of θ in (16), we can rewrite problem (1) as follows:

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{X}}{\text{minimize}} \quad & f(\mathbf{x}) + \theta_1(g(\mathbf{x})) \\ \text{subject to} \quad & h(\mathbf{x}) = 0, \\ & g(\mathbf{x}) \in \mathcal{K}, \end{aligned}$$

where the closed convex set $\mathcal{K}$ is given by (29). In fact, all the subsequent analysis does not require the function g in the objective and constraint to be the same. In order to make the discussions more general, we allow two different continuously differentiable functions g_1 and g_2 in this problem, i.e., we consider the problem

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{X}}{\text{minimize}} \quad & f(\mathbf{x}) + \theta_1(g_1(\mathbf{x})) \\ \text{subject to} \quad & h(\mathbf{x}) = 0, \\ & g_2(\mathbf{x}) \in \mathcal{K}. \end{aligned} \tag{38}$$

The Lagrangian function $\mathcal{L} : \mathbb{X} \times \mathbb{S}^n \times \mathbb{Y} \times \mathbb{S}^n \to \mathbb{R}$ of the above problem can be written as

$$\mathcal{L}(\mathbf{x}, \mathbf{y}, Y, Z) := f(\mathbf{x}) + \langle Y, g_1(\mathbf{x})\rangle + \langle \mathbf{y}, h(\mathbf{x})\rangle + \langle Z, g_2(\mathbf{x})\rangle, \quad (\mathbf{x}, \mathbf{y}, Y, Z) \in \mathbb{X} \times \mathbb{S}^n \times \mathbb{Y} \times \mathbb{S}^n,$$

yielding the following KKT optimality condition of (38):

$$\begin{cases} \mathcal{L}'_{\mathbf{x}}(\mathbf{x}, \mathbf{y}, Y, Z) = 0, \quad h(\mathbf{x}) = 0, \\ Y \in \partial\theta_1(g_1(\mathbf{x})), \quad Z \in \mathcal{N}_{\mathcal{K}}(g_2(\mathbf{x})), \end{cases} \tag{39}$$

where $\mathcal{L}'_{\mathbf{x}}(\mathbf{x}, \mathbf{y}, Y, Z)$ is the partial derivative of $\mathcal{L}$ with respect to $\mathbf{x}$. For any $(\mathbf{x}, \mathbf{y}, Y, Z) \in \mathbb{X}\times\mathbb{Y}\times\mathbb{S}^n\times\mathbb{S}^n$ satisfying (39), we call $\mathbf{x}$ a stationary point, $(\mathbf{y}, Y, Z)$ the corresponding multiplier and $(\mathbf{x}, \mathbf{y}, Y, Z)$ a KKT point of (38), respectively. We also use $\mathcal{M}(\mathbf{x})$ to denote the set of multipliers $(\mathbf{y}, Y, Z)$ for any stationary point $\mathbf{x}$ such that $(\mathbf{x}, \mathbf{y}, Y, Z)$ is a KKT point.

The following concept of nondegeneracy for the nonsmooth matrix optimization (38) is adopted from Robinson [30], which is an analogue of the LICQ for the conventional nonlinear programming.

Definition 1 The constraint nondegeneracy of problem (38) is defined as

$$\begin{bmatrix} h'(\overline{\mathbf{x}}) \\ g_1'(\overline{\mathbf{x}}) \\ g_2'(\overline{\mathbf{x}}) \end{bmatrix} \mathbb{X} + \begin{bmatrix} \{0\} \\ \mathcal{T}^{\text{lin}}_{\theta_1}(g_1(\overline{\mathbf{x}})) \\ \text{lin}\,(\mathcal{T}_{\mathcal{K}}(g_2(\overline{\mathbf{x}}))) \end{bmatrix} = \begin{bmatrix} \mathbb{Y} \\ \mathbb{S}^n \\ \mathbb{S}^n \end{bmatrix}, \tag{40}$$

where the affine spaces $\mathcal{T}^{\text{lin}}_{\theta_1}(g_1(\overline{\mathbf{x}}))$ and $\text{lin}\,(\mathcal{T}_{\mathcal{K}}(g_2(\overline{\mathbf{x}})))$ are defined in (17) and (31).

Let $\overline{\mathbf{x}} \in \mathbb{X}$ be a stationary point of problem (38) and $(\overline{\mathbf{y}}, \overline{Y}, \overline{Z}) \in \mathcal{M}(\overline{\mathbf{x}})$. Since $\mathcal{M}(\overline{\mathbf{x}})$ is nonempty, the critical cone of (38) can be defined as

$$\mathcal{C}(\overline{\mathbf{x}}) := \left\{ \mathbf{d} \in \mathbb{X} \;\middle|\; \begin{array}{l} h'(\overline{\mathbf{x}})\mathbf{d} = 0, \;\; g_1'(\overline{\mathbf{x}})\mathbf{d} \in \mathcal{C}(g_1(\overline{\mathbf{x}}) + \overline{Y}; \partial\theta_1(g_1(\overline{\mathbf{x}}))), \\ g_2'(\overline{\mathbf{x}})\mathbf{d} \in \mathcal{C}(g_2(\overline{\mathbf{x}}) + \overline{Z}; \mathcal{N}_{\mathcal{K}}(g_2(\overline{\mathbf{x}}))) \end{array} \right\}. \tag{41}$$

where $\mathcal{C}(g_1(\overline{\mathbf{x}}) + \overline{Y}; \partial\theta_1(g_1(\overline{\mathbf{x}})))$ and $\mathcal{C}(g_2(\overline{\mathbf{x}}) + \overline{Z}; \mathcal{N}_{\mathcal{K}}(g_2(\overline{\mathbf{x}})))$ are the critical cones defined in (18) and (32), respectively.

For notationally simplicity, we define the outer approximation set to $\mathcal{C}(\overline{\mathbf{x}})$ with respect to $(\overline{\mathbf{y}}, \overline{Y}, \overline{Z}) \in \mathcal{M}(\overline{\mathbf{x}})$ as

$$\text{app}(\overline{\mathbf{y}}, \overline{Y}, \overline{Z}) := \left\{ \mathbf{d} \in \mathbb{X} \;\middle|\; \begin{array}{l} h'(\overline{\mathbf{x}})\mathbf{d} = 0, \;\; g_1'(\overline{\mathbf{x}})\mathbf{d} \in \text{aff}\left(\mathcal{C}(g_1(\overline{\mathbf{x}}) + \overline{Y}; \partial\theta_1(g(\overline{\mathbf{x}})))\right), \\ g_2'(\overline{\mathbf{x}})\mathbf{d} \in \text{aff}\left(\mathcal{C}(g_2(\overline{\mathbf{x}}) + \overline{Z}; \mathcal{N}_{\mathcal{K}}(g_2(\overline{\mathbf{x}})))\right) \end{array} \right\}. \tag{42}$$

The following definition of the strong second order sufficient condition of problem (38) generalizes the concept from the conventional nonlinear programming introduced by Robinson [29] to the nonsmooth matrix optimization.

Definition 2 Let $\overline{\mathbf{x}} \in \mathbb{X}$ be a stationary point of the problem (38). We say the strong second order sufficient condition holds at $\overline{\mathbf{x}}$ if

$$\sup_{(\overline{\mathbf{y}},\overline{Y},\overline{Z})\in\mathcal{M}(\overline{\mathbf{x}})} \left\{\langle \mathbf{d}, \mathcal{L}''_{\mathbf{xx}}(\overline{\mathbf{x}},\overline{\mathbf{y}},\overline{Y},\overline{Z})\mathbf{d}\rangle - \varUpsilon^1_{g_1(\overline{\mathbf{x}})}\left(\overline{Y}, g_1'(\overline{\mathbf{x}})\mathbf{d}\right) - \varUpsilon^2_{g_2(\overline{\mathbf{x}})}\left(\overline{Z}, g_2'(\overline{\mathbf{x}})\mathbf{d}\right)\right\} > 0, \quad \forall\, \mathbf{d} \in \bigcap_{(\overline{\mathbf{y}},\overline{Y},\overline{Z})\in\mathcal{M}(\overline{\mathbf{x}})} \mathrm{app}(\overline{\mathbf{y}},\overline{Y},\overline{Z}) \setminus \{0\}. \tag{43}$$

Let $\overline{\mathbf{x}}$ be a local optimal solution to (38) with $\mathcal{M}(\overline{\mathbf{x}}) \neq \emptyset$. Then there exists $(\overline{\mathbf{y}},\overline{Y},\overline{Z}) \in \mathbb{Y}\times\mathbb{S}^n\times\mathbb{S}^n$ such that the KKT condition (39) holds, i.e., $(\overline{\mathbf{x}},\overline{\mathbf{y}},\overline{Y},\overline{Z})$ is a solution of the following generalized equation:

$$0 \in \begin{bmatrix} \mathcal{L}'_{\mathbf{x}}(\mathbf{x},\mathbf{y},Y,Z) \\ h(\mathbf{x}) \\ -g_1(\mathbf{x}) \\ -g_2(\mathbf{x}) \end{bmatrix} + \begin{bmatrix} \{0\} \\ \{0\} \\ \partial\theta_1^*(Y) \\ \partial\delta^*_{\mathcal{K}}(Z) \end{bmatrix}, \tag{44}$$

where $\delta^*_{\mathcal{K}}$ is the support function of the nonempty closed convex set $\mathcal{K}$. The following concept of strong regularity for a solution of the generalized equation (44) is adapted from Robinson [29].

Definition 3 Let $\mathbb{T} \equiv \mathbb{X}\times\mathbb{Y}\times\mathbb{S}^n\times\mathbb{S}^n$. We say that $(\overline{\mathbf{x}},\overline{\mathbf{y}},\overline{Y},\overline{Z}) \in \mathbb{T}$ is a strongly regular solution of the generalized equation (44) if there exist neighborhoods $\mathcal{U}$ of the origin 0 and $\mathcal{V}$ of $(\overline{\mathbf{x}},\overline{\mathbf{y}},\overline{Y},\overline{Z})$ such that for every $\delta \in \mathcal{U}$, the following generalized equation

$$\delta \in \begin{bmatrix} \mathcal{L}'_{\mathbf{x}}(\mathbf{x},\mathbf{y},Y,Z) \\ h(\mathbf{x}) \\ -g_1(\mathbf{x}) \\ -g_2(\mathbf{x}) \end{bmatrix} + \begin{bmatrix} \{0\} \\ \{0\} \\ \partial\theta_1^*(Y) \\ \partial\delta^*_{\mathcal{K}}(Z) \end{bmatrix} \tag{45}$$

has a unique solution in $\mathcal{V}$, denoted by $S_{\mathcal{V}}(\delta)$, and the mapping $S_{\mathcal{V}} : \mathcal{U} \to \mathcal{V}$ is Lipschitz continuous.

In fact, the solution of the generalized equation (44) can be viewed as the solution of the following nonsmooth equation

$$F(\mathbf{x},\mathbf{y},Y,Z) := \begin{bmatrix} \mathcal{L}'_{\mathbf{x}}(\mathbf{x},\mathbf{y},Y,Z) \\ h(\mathbf{x}) \\ g_1(\mathbf{x}) - \mathrm{Pr}_{\theta_1}(g_1(\mathbf{x}) + Y) \\ g_2(\mathbf{x}) - \varPi_{\mathcal{K}}(g_2(\mathbf{x}) + Z) \end{bmatrix} = 0, \tag{46}$$

where $\mathrm{Pr}_{\theta_1} : \mathbb{S}^n \to \mathbb{S}^n$ is the proximal mapping of θ_1 and $\varPi_{\mathcal{K}} : \mathbb{S}^n \to \mathbb{S}^n$ is the metric projection onto $\mathcal{K}$. The function F is said to be locally Lipschitz homeomorphism near S if there exists an open neighborhood $\mathcal{U}$ such that the restricted mapping $S|_{\mathcal{U}} : \mathcal{U} \to S(\mathcal{U})$ is Lipschitz continuous and bijective, and its inverse is also Lipschitz continuous. The following result on the relationship between the strong regularity of (44) and the locally Lipschitz homeomorphism of F in (46) can be obtained directly from their definitions.

Lemma 5 *Suppose that $F(\overline{\mathbf{x}},\overline{\mathbf{y}},\overline{Y},\overline{Z}) = 0$. Then F is locally Lipschitz homeomorphism near $(\overline{\mathbf{x}},\overline{\mathbf{y}},\overline{Y},\overline{Z})$ if and only if $(\overline{\mathbf{x}},\overline{\mathbf{y}},\overline{Y},\overline{Z})$ is a strongly regular solution of the generalized equation* (44).

Let $\overline{S} = (\overline{\mathbf{x}},\overline{\mathbf{y}},\overline{Y},\overline{Z}) \in \mathbb{T}$ be such that $F(\overline{\mathbf{x}},\overline{\mathbf{y}},\overline{Y},\overline{Z}) = 0$. Denote $\overline{X} := g(\overline{\mathbf{x}})$, $A := \overline{X} + \overline{Y}$ and $B := \overline{X} + \overline{Z}$. By [8, Lemma 1], we know that $\boldsymbol{W} \in \partial F(\overline{\mathbf{x}},\overline{\mathbf{y}},\overline{Y},\overline{Z})$ (respectively, $\boldsymbol{W} \in \partial_B F(\overline{\mathbf{x}},\overline{\mathbf{y}},\overline{Y},\overline{Z})$) if and only if there exist $\mathcal{S}^1 \in \partial\mathrm{Pr}_{\theta_1}(A)$ (respectively, $\mathcal{S}^1 \in \partial_B\mathrm{Pr}_{\theta_1}(A)$) and $\mathcal{S}^2 \in \partial\varPi_{\mathcal{K}}(B)$ (respectively, $\mathcal{S}^2 \in \partial_B\varPi_{\mathcal{K}}(B)$) such that for any $(\triangle\mathbf{x},\triangle\mathbf{y},\triangle Y,\triangle Z) \in \mathbb{T}$,

$$\boldsymbol{W}(\triangle\mathbf{x},\triangle\mathbf{y},\triangle Y,\triangle Z) = \begin{bmatrix} \mathcal{L}''_{\mathbf{xx}}(\overline{\mathbf{x}},\overline{\mathbf{y}},\overline{Y},\overline{Z})\triangle\mathbf{x} + h'(\overline{\mathbf{x}})^*\triangle\mathbf{y} + g_1'(\overline{\mathbf{x}})^*\triangle Y + g_2'(\overline{\mathbf{x}})^*\triangle Z \\ h'(\overline{\mathbf{x}})\triangle\mathbf{x} \\ g_1'(\overline{\mathbf{x}})\triangle\mathbf{x} - \mathcal{S}^1(g_1'(\overline{\mathbf{x}})\triangle\mathbf{x} + \triangle Y) \\ g_2'(\overline{\mathbf{x}})\triangle\mathbf{x} - \mathcal{S}^2(g_2'(\overline{\mathbf{x}})\triangle\mathbf{x} + \triangle Z) \end{bmatrix}. \tag{47}$$

Next, we shall provide the explicit formula of the generalized Jacobian $\partial F(\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{Y}, \overline{Z})$. We first consider the characterization of $\mathcal{S}^1 \in \partial \mathrm{Pr}_{\theta_1}(A)$. Recall that the proximal mapping $\mathrm{Pr}_{\theta_1} : \mathbb{S}^n \to \mathbb{S}^n$ is the spectral operator with respect to the proximal mapping $\mathrm{Pr}_{\phi_1} : \mathbb{R}^n \to \mathbb{R}^n$, i.e.,

$$\mathrm{Pr}_{\theta_1}(A) = \overline{X} = \overline{U}\mathrm{Diag}(\mathrm{Pr}_{\phi_1}(\lambda(A)))\overline{U}^\top,$$

where $\overline{U} \in \mathbb{O}^n(\overline{X})$. To proceed, we denote $\mathbb{W}^1 := \prod_{l=1}^r \mathbb{S}^{|\beta_1^l|} \times \ldots \times \mathbb{S}^{|\beta_{s^l}^l|}$ and

$$\mathcal{A}_{ij} := \begin{cases} (\lambda_i(\overline{X}) - \lambda_j(\overline{X}))/(\lambda_i(A) - \lambda_j(A)) & \text{if } \lambda_i(A) \neq \lambda_j(A), \\ 0 & \text{otherwise,} \end{cases} \quad i, j \in \{1, \ldots, n\}. \tag{48}$$

We also define the spectral operator with respect to the directional derivative $\mathrm{Pr}'_{\phi}(\lambda(\overline{X}); \bullet)$

$$\Sigma := (\Sigma_1^1, \ldots, \Sigma_{s^1}^1, \ldots, \Sigma_1^r, \ldots, \Sigma_{s^r}^r) : \mathbb{W}^1 \to \mathbb{W}^1,$$

and

$$D^1(H) = \left(\overline{U}_{\beta_1^1}^\top H\overline{U}_{\beta_1^1}, \ldots, \overline{U}_{\beta_{s^1}^1}^\top H\overline{U}_{\beta_{s^1}^1}, \ldots, \overline{U}_{\beta_1^r}^\top H\overline{U}_{\beta_1^r}, \ldots, \overline{U}_{\beta_{s^r}^r}^\top H\overline{U}_{\beta_{s^r}^r}\right) \in \mathbb{W}^1. \tag{49}$$

Since Pr_{ϕ_1} is globally Lipchitz continuous and directionally differentiable at $\lambda(A)$, we know from [13, Remark 1 and Theorem 6] that Pr_{θ_1} is directionally differentiable at A and the directional derivative $\mathrm{Pr}'_{\theta_1}(A; H)$ at A along $H \in \mathbb{S}^n$ is given by

$$\mathrm{Pr}'_{\theta_1}(A; H) = \overline{U}\mathrm{Pr}_{\phi_1}^{[1]}(A; H)\overline{U}^\top,$$

where $\mathrm{Pr}_{\phi_1}^{[1]}(A; H)$ is the first divided directional difference of Pr_{ϕ_1} at A along H with the expression

$$\mathrm{Pr}_{\phi_1}^{[1]}(A; H) := \mathcal{A} \circ \overline{U}^\top H\overline{U} + \mathrm{Diag}\left(\Sigma_1^1(D^1(H)), \ldots, \Sigma_{s^r}^r(D^1(H))\right) \in \mathbb{S}^n,$$

It follows from Proposition 2 that $\Sigma : \mathbb{W}^1 \to \mathbb{W}^1$ is actually the metric projection operator over the following nonempty closed convex set $\Delta^1 \subseteq \mathbb{W}^1$,

$$\Delta^1 := \left\{ W \in \mathbb{W} \mid \langle \mu(W), \mathbf{a}^i \rangle = \langle \mu(W), \mathbf{a}^j \rangle = \max_{i \in \iota_1(\lambda(\overline{X}))} \langle \mu(W), \mathbf{a}^i \rangle, \ \forall\, i, j \in \eta_1(\lambda(\overline{X}), \lambda(\overline{Y})) \right\}, \tag{50}$$

where the index set η_1 is defined in (12), and for any $W = (W_1^1, \ldots, W_{s^1}^1, \ldots, W_1^r, \ldots, W_{s^r}^r) \in \mathbb{W}^1$,

$$\mu(W) := (\lambda(W_1^1), \ldots, \lambda(W_{s^1}^1), \ldots, \lambda(W_1^r), \ldots, \lambda(W_{s^r}^r)).$$

Finally, since Pr_{ϕ_1} is piecewise affine, we know from [3, Section 7.3] that

$$\mathrm{Pr}_{\phi_1}(\mathbf{x} + \mathbf{h}) - \mathrm{Pr}_{\phi_1}(\mathbf{x}) = \mathrm{Pr}'_{\phi_1}(\mathbf{x}; \mathbf{h}), \quad \forall\, \mathbf{x}, \mathbf{h} \in \mathbb{R}^n.$$

Thus, it follows from [14, Theorem 7.8] that

$$\partial \mathrm{Pr}_{\theta_1}(A) = \partial \Psi(0),$$

where $\Psi := \mathrm{Pr}'_{\theta_1}(A; \bullet) : \mathbb{S}^n \to \mathbb{S}^n$ is the directional derivative of Pr_{θ_1} at A. Therefore, we obtain the following result on the characterization of $\partial \mathrm{Pr}_{\theta_1}(A)$.

Lemma 6 *It holds that $\mathcal{S} \in \partial \mathrm{Pr}_{\theta_1}(A)$ if and only if there exists $\mathcal{U} := (\mathcal{U}_1^1, \ldots, \mathcal{U}_{s^1}^1, \ldots, \mathcal{U}_1^r, \ldots, \mathcal{U}_{s^r}^r) \in \partial \Pi_{\Delta^1}(0)$ such that for any $H \in \mathbb{S}^n$,*

$$\mathcal{S}(H) = \overline{U}\left[\mathcal{A} \circ \left(\overline{U}^\top H\overline{U}\right)\right]\overline{U}^\top + \overline{U}\,\mathrm{Diag}\left(\mathcal{U}_1^1(D^1(H)), \ldots, \mathcal{U}_{s^r}^r(D^1(H))\right)\overline{U}^\top,$$

where $D^1(H) \in \mathbb{W}^1$ is defined by (49)*, Δ^1 is the nonempty convex set defined by* (50) *and Π_{Δ^1} denotes the metric projection onto Δ^1.*

Similarly, for $B = \overline{X} + \overline{Z}$ with $(\overline{X}, \overline{Z}) \in \mathrm{gph}\,\mathcal{N}_{\mathcal{K}}$, we have the following characterization of $\partial \Pi_{\mathcal{K}}(B)$.

Lemma 7 *It holds that $\mathcal{S} \in \partial \Pi_{\mathcal{K}}(B)$ if and only if there exists $\mathcal{V} := (\mathcal{V}_1^1, \ldots, \mathcal{V}_{t^1}^1, \ldots, \mathcal{V}_1^r, \ldots, \mathcal{V}_{t^r}^r) \in \partial \Pi_{\Delta^2}(0)$ such that for any $H \in \mathbb{S}^n$,*

$$\mathcal{S}(H) = \overline{V} \left[\mathcal{B} \circ \left(\overline{V}^\top H \overline{V} \right) \right] \overline{V}^\top + \overline{V} \operatorname{Diag} \left(\mathcal{V}_1^1(D^2(H)), \ldots, \mathcal{V}_{t^r}^r(D^2(H)) \right) \overline{V}^\top,$$

with $\mathbb{W}^2 := \prod_{l=1}^r \mathbb{S}^{|\gamma_1^l|} \times \ldots \times \mathbb{S}^{|\gamma_{t^l}^l|}$,

$$\mathcal{B}_{ij} := \begin{cases} (\lambda_i(\overline{X}) - \lambda_j(\overline{X}))/(\lambda_i(B) - \lambda_j(B)) & \text{if } \lambda_i(B) \neq \lambda_j(B), \\ 0 & \text{otherwise,} \end{cases} \quad i, j \in \{1, \ldots, n\},$$

$$D^2(H) = \left(\overline{V}_{\gamma_1^1}^\top H \overline{V}_{\gamma_1^1}, \ldots, \overline{V}_{\gamma_{t^1}^1}^\top H \overline{V}_{\gamma_{t^1}^1}, \ldots, \overline{V}_{\gamma_1^r}^\top H \overline{V}_{\gamma_1^r}, \ldots, \overline{V}_{\gamma_{t^r}^r}^\top H \overline{V}_{\gamma_{t^r}^r} \right) \in \mathbb{W}^2,$$

and $\Delta^2 \subseteq \mathbb{W}^2$ is the nonempty convex set defined by

$$\Delta^2 := \left\{ W \in \mathbb{W}^2 \mid 0 = \langle \nu(W), \mathbf{b}^i \rangle = \max_{i \in \iota_2(\lambda(\overline{X}))} \langle \nu(W), \mathbf{b}^i \rangle, \ \forall\, i \in \eta_2(\lambda(\overline{X}), \lambda(\overline{Z})) \right\},$$

where the index set $\eta_2(\lambda(\overline{X}), \lambda(\overline{Y}))$ is defined in (12) with respect to $\lambda(\overline{X})$ and $\lambda(\overline{Z})$, and for any $W = (W_1^1, \ldots, W_{t^1}^1, \ldots, W_1^r, \ldots, W_{t^r}^r) \in \mathbb{W}^2$,

$$\nu(W) := (\lambda(W_1^1), \ldots, \lambda(W_{t^1}^1), \ldots, \lambda(W_1^r), \ldots, \lambda(W_{t^r}^r)).$$

By comparing the characterizations of Clarke's generalized Jacobian of the proximal mapping Pr_{θ_1} in Lemma 6 with $\operatorname{aff}\left(\mathcal{C}(\overline{X} + \overline{Y}; \partial\theta_1(\overline{X}))\right)$ in Proposition 5, we derive the following lemma.

Lemma 8 *Suppose that $\overline{Y} \in \partial\theta_1(\overline{X})$ and $\mathcal{S} \in \partial \mathrm{Pr}_{\theta_1}(\overline{X} + \overline{Y})$. Then*

$$\mathcal{S}(H) \in \operatorname{aff}\left(\mathcal{C}(\overline{X} + \overline{Y}; \partial\theta_1(\overline{X}))\right), \quad \forall\, H \in \mathbb{S}^n. \tag{51}$$

In addition, if $\mathcal{S}(H) = 0$ for some $H \in \mathbb{S}^n$, then the following two conditions hold:

(i) the matrix $\overline{U}^\top H \overline{U} \in \mathbb{S}^n$ has the following block diagonal structure:

$$\overline{U}^\top H \overline{U} = \operatorname{Diag}\left(\overline{U}_{\alpha^1}^\top H \overline{U}_{\alpha^1}, \cdots, \overline{U}_{\alpha^r}^\top H \overline{U}_{\alpha^r} \right);$$

(ii) for $l = 1, \ldots, r$, if $k \in \mathcal{E}^l$, then there exists $\{\kappa_{ij} \in \mathbb{R}\}_{i,j \in \iota_1(\lambda(\overline{X}))}$ such that

$$\operatorname{tr}\left(\overline{U}_{\beta_k^l}^\top H \overline{U}_{\beta_k^l} \right) = \sum_{i,j \in \iota_1(\lambda(\overline{X}))} \kappa_{ij} \left\langle e_{|\beta_k^l|}, (\mathbf{a}^i - \mathbf{a}^j)_{\beta_k^l} \right\rangle;$$

otherwise if $k \notin \mathcal{E}^l$, then $\overline{U}_{\beta_k^l}^\top H \overline{U}_{\beta_k^l} = 0$, where the index set $\mathcal{E}^l$ is defined by (20).

Proof To prove the inclusion (51) in this lemma, it suffices to check the three conditions in Proposition 5 hold. For any given $\mathcal{S} \in \partial \mathrm{Pr}_{\theta_1}(\overline{X} + \overline{Y})$ and $H \in \mathbb{S}^n$, we obtain from Lemma 6 that for each $l \in \{1, \ldots, r\}$, there exists $\mathcal{U} = (\mathcal{U}_1^1, \ldots, \mathcal{U}_{s^r}^r) \in \partial \Pi_{\Delta^1}(0)$ such that

$$\overline{U}_{\alpha^l}^\top \mathcal{S}(H) \overline{U}_{\alpha^l} = \mathcal{A}_{\alpha^l \alpha^l} \circ \left(\overline{U}_{\alpha^l}^\top H \overline{U}_{\alpha^l} \right) + \operatorname{Diag}\left(\mathcal{U}_1^l(D^1(H)), \ldots, \mathcal{U}_{s^l}^l(D^1(H)) \right).$$

For each $l \in \{1, \ldots, r\}$, since $\lambda_i(\overline{X}) = \lambda_j(\overline{X})$ for any $i, j \in \alpha_l$, it follows from (48) that $\mathcal{A}_{\alpha_l \alpha_l} = 0$, which implies that the condition (i) in Proposition 5 holds.

Let $\mathcal{D}_{\Pi_{\Delta^1}} \subseteq \mathbb{W}^1$ be the set of all points at which Π_{Δ^1} is differentiable. We define

$$\daleth := \left\{ W \in \mathbb{W}^1 \mid \text{for each } l \in \{1, \ldots, r\} \text{ and } k \in \{1, \ldots, s^l\}, \text{ the eigenvalues of } W_k^l \text{ are distinct} \right\}.$$

Since $\mathbb{W}^1 \setminus \daleth$ has measure zero (in the sense of Lebesgue), we know from [38, Theorem 4] that

$$\partial \Pi_{\Delta^1}(0) = \operatorname{conv}\left\{ \lim_{W \to 0} \Pi'_{\Delta^1}(W) \mid W \in \mathcal{D}_{\Pi_{\Delta^1}} \cap \daleth \right\}. \tag{52}$$

Then for any $\Theta \in \left\{ \lim_{W \to 0} \Pi'_{\Delta^1}(W) \mid W \in \mathcal{D}_{\Pi_{\Delta^1}} \cap \urcorner \right\}$, there exits a sequence

$$\left\{ W^q := \left((W_1^1)^q, \ldots, (W_1^{s^1})^q, \ldots, (W_r^1)^q, \ldots, (W_r^{s^r})^q \right) \right\} \subseteq \mathcal{D}_{\Pi_{\Delta^1}} \cap \urcorner$$

converging to $0 \in \mathbb{W}$ such that for any $H \in \mathbb{S}^n$,

$$\Theta(D^1(H)) = \lim_{q \to \infty} \Pi'_{\Delta^1}(W^q)(D^1(H)).$$

Consider any fixed $l = 1, \ldots, r$ and $k = 1, \ldots, s^l$. For each q, assume that $(W_k^l)^q$ admits the eigenvalue decomposition

$$(W_k^l)^q = (R_k^l)^q \Lambda((W_k^l)^q)((R_k^l)^q)^\top \quad \text{with} \quad (R_k^l)^q \in \mathbb{O}^{|\beta_k^l|}.$$

Notice that $\Pi_{\Delta^1} : \mathbb{W}^1 \to \mathbb{W}^1$ is the spectral operator with respect to the symmetric mapping $\pi := \mathrm{Pr}'_{\phi_1}(\lambda(\overline{X}); \bullet)$ defined by (14). Thus, we know from [13, Theorem 7] that Π_{Δ^1} is differentiable at W^q if and only if π is differentiable at

$$\mu^q := \mu(W^q) = (\lambda((W_1^1)^q), \ldots, \lambda((W_{s^r}^r)^q)).$$

For each $l \in \{1, \ldots, r\}$ and $k \in \{1, \ldots, s^l\}$, denote

$$\begin{cases} (\mathcal{A}_k^l(\mu^q))_{ij} = \begin{cases} \dfrac{(\pi_k^l(\mu^q))_i - (\pi_k^l(\mu^q))_j}{\lambda_i((W_k^l)^q) - \lambda_j((W_k^l)^q)} & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} \quad \forall\, i, j \in \{\, 1, \ldots, |\beta_k^l| \,\}, \\ h^q = \left(\mathrm{diag}\big(((R_1^1)^q)^\top \widetilde{H}_{\beta_1^1 \beta_1^1} (R_1^1)^q \big), \ldots, \mathrm{diag}\big(((R_{s^r}^r)^q)^\top \widetilde{H}_{\beta_{s^r}^r \beta_{s^r}^r} (R_{s^r}^r)^q \big) \right), \\ \widetilde{H} = \overline{U}^\top H \overline{U}, \\ (\Omega_k^l)^q = (R_k^l)^q \left[\mathcal{A}_k^l(\mu^q) \circ \left[((R_k^l)^q)^\top \widetilde{H}_{\beta_k^l \beta_k^l} (R_k^l)^q \right] + \mathrm{Diag}\left((\pi'(\mu^q) h^q)_{\beta_k^l} \right) \right] ((R_k^l)^q)^\top. \end{cases} \tag{53}$$

It has been shown in [13, Theorem 7] that for each q, the derivative of Π_{Δ^1} at W^q is given by

$$\Pi'_{\Delta^1}(W^q)(D^1(H)) = \left((\Omega_1^1)^q, \ldots, (\Omega_{s^1}^1)^q, \ldots, (\Omega_1^r)^q, \ldots, (\Omega_{s^r}^r)^q \right) \in \mathbb{W}^1.$$

For each $l \in \{1, \ldots, r\}$, recall the index set $\mathcal{E}^l$ defined by (20). We know from Proposition 2 that for each $l \in \{1, \ldots, r\}$, $k \in \mathcal{E}^l$ and q,

$$\mathcal{A}_k^l(\mu^q) = 0. \tag{54}$$

Denote $X^q := \overline{X} + W^q$. Based on [25, Theorem 2.1], for any q sufficient large, it holds that $\eta_1(\lambda(\overline{X}), \lambda(\overline{Y})) \subseteq \iota_1(\lambda(X^q))$. Again using Proposition 2, we obtain that for each $l \in \{1, \ldots, r\}$, $k \in \mathcal{E}^l$ and q, there exists $\rho_{k^l}^l \in \mathbb{R}$ such that

$$(\pi'(\mu^q) h^q)_{\beta_k^l} = \rho_k^l e_{|\beta_k^l|},$$

and for any $i, j \in \eta_1(\lambda(\overline{X}), \lambda(\overline{Y}))$,

$$\begin{aligned} \langle\, \mathrm{diag}(\Omega^q), \mathbf{a}^i - \mathbf{a}^j \,\rangle &= \sum_{l=1}^r \sum_{k \in \mathcal{E}^l} \left\langle \mathrm{diag}((\Omega_k^l)^q),\, (\mathbf{a}^i)_{\beta_k^l} - (\mathbf{a}^j)_{\beta_k^l} \right\rangle \\ &= \sum_{l=1}^r \sum_{k \in \mathcal{E}^l} \left\langle \rho_k^l e_{|\beta_k^l|},\, (\mathbf{a}^i)_{\beta_k^l} - (\mathbf{a}^j)_{\beta_k^l} \right\rangle = 0, \end{aligned}$$

where $\mathrm{diag}(\Omega^q) = \left(\mathrm{diag}((\Omega_1^1)^q), \ldots, \mathrm{diag}((\Omega_{s^1}^1)^q), \ldots, \mathrm{diag}((\Omega_1^r)^q), \ldots, \mathrm{diag}((\Omega_{s^r}^r)^q) \right)$. Thus, we know that $\mathcal{S}(H)$ satisfies condition (ii) and (24) of Proposition 5. The above arguments show that (51) holds.

In the following, we assume $\mathcal{S}(H) = 0$ for some $H \in \mathbb{S}^n$ and prove the rest of the lemma. Since $\mathcal{S}(H) = 0$, it is clear that $\overline{U}^\top \mathcal{S}(H) \overline{U} = 0$. It then follows from Lemma 6 that for any $1 \le l < l' \le r$,

$$0 = \overline{U}_{\alpha^l}^\top \mathcal{S}(H) \overline{U}_{\alpha^{l'}} = \left(\overline{U}_{\alpha^{l'}}^\top \mathcal{S}(H) \overline{U}_{\alpha^l} \right)^\top = \mathcal{A}_{\alpha^l \alpha^{l'}} \circ \left(\overline{U}_{\alpha^l}^\top H \overline{U}_{\alpha^{l'}} \right).$$

By the definition of $\mathcal{A}$ in (48), we have $\mathcal{A}_{\alpha^l\alpha^{l'}} = (\mathcal{A}_{\alpha^{l'}\alpha^l})^\top \neq 0$ for all $1 \leq l < l' \leq r$, which implies that

$$\overline{U}_{\alpha^l}^\top H \overline{U}_{\alpha^{l'}} = \left(\overline{U}_{\alpha^l}^\top H \overline{U}_{\alpha^{l'}}\right)^\top = 0, \quad 1 \leq l < l' \leq r.$$

This proves the statement (i).

We further obtain from Lemma 6 that

$$0 = \overline{U}^\top \mathcal{S}(H) \overline{U} = \mathrm{Diag}\left(\mathcal{U}_1^1(D^1(H)), \ldots, \mathcal{U}_{s^r}^r(D^1(H))\right). \tag{55}$$

By Carathéodory's theorem, we obtain from (52) that there exist

$$\mathcal{U}^{\{i\}} \in \left\{\lim_{W \to 0} \Pi'_{\Delta^1}(W) \mid W \in \mathcal{D}_{\Pi_{\Delta^1}} \cap \daleth\right\}, \quad i = 1, \ldots, w$$

for some positive integer w such that $\mathcal{U} = (\mathcal{U}_1^1, \ldots, \mathcal{U}_{s^1}^1, \ldots, \mathcal{U}_1^r, \ldots, \mathcal{U}_{s^r}^r) \in \partial \Pi_{\Delta^1}(0)$ can be written as

$$\mathcal{U} = \sum_{i=1}^{w} \varsigma_i \mathcal{U}^{\{i\}} \quad \text{for some } \varsigma_i \geq 0,\ i = 1, \ldots, w, \text{ and } \sum_{i=1}^{w} \varsigma_i = 1.$$

For each $\mathcal{U}^{\{i\}}$, there exits a sequence $\{W^q\} \subseteq \mathcal{D}_{\Pi_{\Delta^1}} \cap \daleth$ converging to $0 \in \mathbb{W}$ such that

$$\mathcal{U}^{\{i\}}(D^1(H)) = \lim_{q \to \infty} \Pi'_{\Delta^1}(W^q)(D^1(H)).$$

Following the same notation in (53) with respect to the newly defined sequence $\{W^q\}$, we know from Proposition 2 that for each $l \in \{1, \ldots, r\}$ and $k \in \mathcal{E}^l$ and all q,

$$(\mathcal{A}_k^l(\mu^q))_{ij} = \begin{cases} 0 & \text{if } k \in \mathcal{E}^l \\ 1 & \text{if } k \notin \mathcal{E}^l \end{cases} \quad \forall\, i, j \in \{1, \ldots, |\beta_k^l|\} \text{ with } i \neq j.$$

For each q, denote $X^q := \overline{X} + W^q$. Then, by [25, Theorem 2.1], we know that for all q sufficient large, $\eta_1(\lambda(\overline{X}), \lambda(\overline{Y})) \subseteq \iota_1(\lambda(X^q))$. Therefore, if $k \in \mathcal{E}_l$ for some $l \in \{1, \ldots, r\}$, then there exists a scalar $(\rho_{k^l}^l)^q$ such that

$$(\pi'(\mu^q)h^q)_{\beta_k^l} = \begin{cases} (\rho_k^l)^q\, e_{|\beta_k^l|} & \text{if } k \in \mathcal{E}^l, \\ (h^q)_{\beta_k^l} & \text{if } k \notin \mathcal{E}^l, \end{cases}$$

yielding

$$(\Omega_k^l)^q = \begin{cases} \mathrm{Diag}\left((\pi'(\mu^q)h^q)_{\beta_k^l}\right) = (\rho_k^l)^q\, I_{|\beta_k^l|} & \text{if } k \in \mathcal{E}^l, \\ \widetilde{H}_{\beta_k^l\beta_k^l} & \text{if } k \notin \mathcal{E}^l. \end{cases}$$

This further implies the existence of a scalar $\hat{\rho}_k^l$ such that

$$\left(\mathcal{U}^{\{i\}}\right)_k^l (D^1(H)) = \lim_{q \to \infty} (\Omega_k^l)^q = \begin{cases} \hat{\rho}_{k^l}^l I_{|\beta_k^l|} & \text{if } k \in \mathcal{E}^l, \\ \widetilde{H}_{\beta_k^l\beta_k^l} & \text{if } k \notin \mathcal{E}^l. \end{cases}$$

Taking into account the equality in (55), we derive

$$\mathrm{tr}\left(\widetilde{H}_{\beta_k^l\beta_k^l}\right) = \sum_{i,j \in \iota_1(\lambda(\overline{X}))} \kappa_{ij} \left\langle e_{|\beta_k^l|}, (\mathbf{a}^i - \mathbf{a}^j)_{\beta_k^l} \right\rangle \quad \text{for some scalars } \{\kappa_{ij}\}_{i,j \in \iota_1(\lambda(\overline{X}))}.$$

If $k \notin \mathcal{E}^l$, then $\widetilde{H}_{\beta_k^l\beta_k^l} = 0$. This completes the proof of this lemma. $\square$

By comparing the characterization of Clarke's generalized Jacobian of the proximal mapping $\Pi_\mathcal{K}$ in Lemma 7 with aff $(\mathcal{C}(\overline{X} + \overline{Z}; \mathcal{N}_\mathcal{K}(\overline{X})))$ in Proposition 9, we can obtain the following results with respect to $\partial\theta_2$. Its proof can be obtained similarly as that of Lemma 8. We omit the details here for brevity.

Lemma 9 *Suppose that* $\overline{Z} \in \mathcal{N}_{\mathcal{K}}(\overline{X})(= \partial\theta_2(\overline{X}))$ *and* $\mathcal{S} \in \partial\Pi_{\mathcal{K}}(\overline{X}+\overline{Z})$. *Then*

$$\mathcal{S}(H) \in \mathrm{aff}\,(\mathcal{C}(\overline{X}+\overline{Z};\mathcal{N}_{\mathcal{K}}(\overline{X}))), \quad \forall\, H \in \mathbb{S}^n.$$

In addition, if $\mathcal{S}(H)=0$ *for some* $H \in \mathbb{S}^n$, *then the following two conditions hold:*

(i) $\overline{V}^\top H \overline{V} \in \mathbb{S}^n$ *has the following block diagonal structure:*

$$\overline{V}^\top H\overline{V} = \mathrm{Diag}\left(\overline{V}_{\alpha^1}^\top H\overline{V}_{\alpha^1}, \cdots, \overline{V}_{\alpha^r}^\top H\overline{V}_{\alpha^r}\right); \tag{56}$$

(ii) for $l=1,\ldots,r$, *let* $\mathcal{F}^l$ *be the index set defined by (34). If* $k \in \mathcal{F}^l$, *then there exist* $\{\kappa_{ij} \in \mathbb{R}\}_{i,j\in\iota_2(\lambda(\overline{X}))}$ *such that*

$$\mathrm{tr}\left(\overline{V}_{\gamma_k^l}^\top H\overline{V}_{\gamma_k^l}\right) = \sum_{i,j\in\iota_2(\lambda(\overline{X}))} \kappa_{ij}\left\langle e_{|\gamma_k^l|}\,,\,(\mathbf{b}^i-\mathbf{b}^j)_{\gamma_k^l}\right\rangle;$$

otherwise if $k \notin \mathcal{F}^l$, *then* $\overline{V}_{\gamma_k^l}^\top H\overline{V}_{\gamma_k^l} = 0$.

Finally, we establish a connection between the function $\Upsilon^1_{\overline{X}}$ defined in Remark 2 and the Clarke generalized Jacobian of Pr_{θ_1} given by Proposition 6. This result plays a key role in our subsequent analysis.

Lemma 10 *Suppose that* $\overline{Y} \in \partial\theta_1(\overline{X})$. *Then, for any* $\mathcal{S} \in \partial\mathrm{Pr}_{\theta_1}(\overline{X}+\overline{Y})$ *and* $(\triangle X, \triangle Y) \in \mathbb{S}^n\times\mathbb{S}^n$ *such that* $\triangle X = \mathcal{S}(\triangle X + \triangle Y)$, *it holds that*

$$\langle \triangle X, \triangle Y\rangle \geq -\Upsilon^1_{\overline{X}}\left(\overline{Y}, \triangle X\right).$$

Proof Denote $H := \triangle X + \triangle Y$. For any given $\mathcal{S} \in \partial\mathrm{Pr}_{\theta_1}(\overline{X}+\overline{Y})$, it is known from Lemma 8 that

$$\overline{U}^\top \triangle X\overline{U} = \mathcal{A}\circ\overline{U}^\top H\overline{U} + \mathrm{Diag}\left(\mathcal{U}_1^1(D^1(\overline{U}^\top H\overline{U})),\ldots,\mathcal{U}^r_{s^r}(D^1(\overline{U}^\top H\overline{U}))\right),$$

where $\overline{U} \in \mathbb{O}^n(\overline{X})\cap\mathbb{O}^n(\overline{Y})$. This further yields that

$$\begin{cases} \overline{U}_{\alpha^l}^\top \triangle X\overline{U}_{\alpha^{l'}} = \mathcal{A}_{\alpha^l\alpha^{l'}}\circ\overline{U}_{\alpha^l}^\top(\triangle X+\triangle Y)\overline{U}_{\alpha^{l'}} & \forall\, l,l' \in \{1,\ldots,r\} \text{ with } l\neq l', \\ \overline{U}_{\alpha^l}^\top \triangle X\overline{U}_{\alpha^l} = \mathrm{Diag}\left(\mathcal{U}_1^l(D^1(\overline{U}^\top H\overline{U})),\ldots,\mathcal{U}^l_{s^l}(D^1(\overline{U}^\top H\overline{U}))\right) & \forall\, l \in \{1,\ldots,r\}. \end{cases} \tag{57}$$

Therefore, we have

$$\begin{aligned}
&\langle \triangle X, \triangle Y\rangle = \left\langle \overline{U}^\top\triangle X\overline{U}, \overline{U}^\top\triangle Y\overline{U}\right\rangle \\
&= \sum_{l=1}^r \langle \overline{U}_{\alpha^l}^\top\triangle X\overline{U}_{\alpha^l}, \overline{U}_{\alpha^l}^\top\triangle Y\overline{U}_{\alpha^l}\rangle + 2\sum_{1\leq l<l'\leq r}\langle \overline{U}_{\alpha^l}^\top\triangle X\overline{U}_{\alpha^{l'}}, \overline{U}_{\alpha^l}^\top\triangle Y\overline{U}_{\alpha^{l'}}\rangle \\
&= \sum_{l=1}^r\sum_{k=1}^{s^l} \langle \overline{U}_{\beta_k^l}^\top\triangle X\overline{U}_{\beta_k^l}, \overline{U}_{\beta_k^l}^\top\triangle Y\overline{U}_{\beta_k^l}\rangle + 2\sum_{1\leq l<l'\leq r}\langle \overline{U}_{\alpha^l}^\top\triangle X\overline{U}_{\alpha^{l'}}, \overline{U}_{\alpha^l}^\top\triangle Y\overline{U}_{\alpha^{l'}}\rangle \\
&= \sum_{l=1}^r\sum_{k=1}^{s^l} \langle \mathcal{U}_k^l(D^1(\overline{U}^\top H\overline{U})), \overline{U}_{\beta_k^l}^\top\triangle Y\overline{U}_{\beta_k^l}\rangle + 2\sum_{1\leq l<l'\leq r}\langle \overline{U}_{\alpha^l}^\top\triangle X\overline{U}_{\alpha^{l'}}, \overline{U}_{\alpha^l}^\top\triangle Y\overline{U}_{\alpha^{l'}}\rangle \\
&= \sum_{l=1}^r\sum_{k=1}^{s^l} \langle \mathcal{U}_k^l(D^1(\overline{U}^\top H\overline{U})), \overline{U}_{\beta_k^l}^\top H\overline{U}_{\beta_k^l} - \overline{U}_{\beta_k^l}^\top\triangle X\overline{U}_{\beta_k^l}\rangle + 2\sum_{1\leq l<l'\leq r}\langle \overline{U}_{\alpha^l}^\top\triangle X\overline{U}_{\alpha^{l'}}, \overline{U}_{\alpha^l}^\top\triangle Y\overline{U}_{\alpha^{l'}}\rangle \\
&= \sum_{l=1}^r\sum_{k=1}^{s^l} \langle \mathcal{U}_k^l(D^1(\overline{U}^\top H\overline{U})), \overline{U}_{\beta_k^l}^\top H\overline{U}_{\beta_k^l} - \mathcal{U}_k^l(D(\overline{U}^\top H\overline{U}))\rangle + 2\sum_{1\leq l<l'\leq r}\langle \overline{U}_{\alpha^l}^\top\triangle X\overline{U}_{\alpha^{l'}}, \overline{U}_{\alpha^l}^\top\triangle Y\overline{U}_{\alpha^{l'}}\rangle \\
&= \left\langle \mathcal{U}(D^1(\overline{U}^\top H\overline{U})), D^1(\overline{U}^\top H\overline{U}) - \mathcal{U}(D^1(\overline{U}^\top H\overline{U}))\right\rangle + 2\sum_{1\leq l<l'\leq r}\langle \overline{U}_{\alpha^l}^\top\triangle X\overline{U}_{\alpha^{l'}}, \overline{U}_{\alpha^l}^\top\triangle Y\overline{U}_{\alpha^{l'}}\rangle.
\end{aligned}$$

Since $\mathcal{U} \in \partial \Pi_{\Delta^1}(0)$ and Δ^1 is a nonempty closed convex set defined by (50), we know from [22, Proposition 1 (c)] that

$$\left\langle \mathcal{U}(D^1(\overline{U}^\top H\overline{U})), D^1(\overline{U}^\top H\overline{U}) - \mathcal{U}(D^1(\overline{U}^\top H\overline{U}))\right\rangle \geq 0.$$

Therefore,

$$\langle \triangle X, \triangle Y\rangle \geq 2 \sum_{1\leq l<l'\leq r} \langle \overline{U}_{\alpha^l}^\top \triangle X \overline{U}_{\alpha^{l'}}, \overline{U}_{\alpha^l}^\top \triangle Y \overline{U}_{\alpha^{l'}}\rangle. \tag{58}$$

On the other hand, one can easily verify from Remark 2 and (57) that

$$-\Upsilon^1_{\overline{X}}\left(\overline{Y}, \triangle X\right) = 2 \sum_{1\leq l<l'\leq r} \langle \overline{U}_{\alpha^l}^\top \triangle X \overline{U}_{\alpha^{l'}}, \overline{U}_{\alpha^l}^\top \triangle Y \overline{U}_{\alpha^{l'}}\rangle.$$

Thus, together with (58), establishes the desired result in this lemma. □

We also have the following parallel results with respect to the function θ_2.

Lemma 11 *Suppose that $\overline{Z} \in \partial\theta_2(\overline{X})$. Then for any $\mathcal{S} \in \partial\Pi_{\mathcal{K}}(B)$ and $(\triangle X, \triangle Z) \in \mathbb{S}^n \times \mathbb{S}^n$ such that $\triangle X = \mathcal{S}(\triangle X + \triangle Z)$, it holds that*

$$\langle \triangle X, \triangle Z\rangle \geq -\Upsilon^2_{\overline{X}}\left(\overline{Z}, \triangle X\right). \tag{59}$$

The following theorem, which is the main result of our paper, establishes the relationship between the strong second order sufficient condition (43) and constraint nondegeneracy (40) for problem (38), the non-singularity of Clarke's Jacobian of the mapping F and the strong regularity of a solution to the generalized equation (44).

Theorem 1 *Let $\overline{\mathbf{x}} \in \mathbb{X}$ be a feasible solution to problem* (38) *with $\mathcal{M}(\overline{\mathbf{x}}) \neq \emptyset$. Suppose that $(\overline{\mathbf{y}}, \overline{Y}, \overline{Z}) \in \mathcal{M}(\overline{\mathbf{x}})$. Consider the following three statements:*

(i) the strong second order sufficient condition (43) *and constraint nondegeneracy* (40) *hold at $\overline{\mathbf{x}}$ for problem* (38)*;*

(ii) every element in $\partial F(\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{Y}, \overline{Z})$ is nonsingular;

(iii) $(\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{Y}, \overline{Z})$ is a strongly regular solution of the generalized equation (44)*.*

It holds that (i) $\Longrightarrow$ (ii) $\Longrightarrow$ (iii).

Proof "(i) $\Longrightarrow$ (ii)" Since the constraint nondegeneracy (40) holds at $\overline{\mathbf{x}}$, we know that $\mathcal{M}(\overline{\mathbf{x}}) = \{(\overline{\mathbf{y}}, \overline{Y}, \overline{Z})\}$. Then the strong second order sufficient condition in (43) reduces to

$$\langle \mathbf{d}, \mathcal{L}''_{\mathbf{xx}}(\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{Y}, \overline{Z})\mathbf{d}\rangle - \Upsilon^1_{g_1(\overline{\mathbf{x}})}\left(\overline{Y}, g_1'(\overline{\mathbf{x}})\mathbf{d}\right) - \Upsilon^2_{g_2(\overline{\mathbf{x}})}\left(\overline{Z}, g_2'(\overline{\mathbf{x}})\mathbf{d}\right) > 0, \quad \forall\, \mathbf{d} \in \operatorname{app}(\overline{\mathbf{y}}, \overline{Y}, \overline{Z}) \setminus \{0\}. \tag{60}$$

Let $\boldsymbol{W}$ be an arbitrary element in $\partial F(\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{Y}, \overline{Z})$. We shall show that $\boldsymbol{W}$ is nonsingular. Suppose that $(\triangle\mathbf{x}, \triangle\mathbf{y}, \triangle Y, \triangle Z) \in \mathbb{X} \times \mathbb{Y} \times \mathbb{S}^n \times \mathbb{S}^n$ satisfies $\boldsymbol{W}(\triangle\mathbf{x}, \triangle\mathbf{y}, \triangle Y, \triangle Z) = 0$. By (47), we know that there exists $\mathcal{S}^1 \in \partial\mathrm{Pr}_{\theta_1}(g_1(\overline{\mathbf{x}}) + \overline{Y})$ and $\mathcal{S}^2 \in \partial\Pi_{\mathcal{K}}(g_2(\overline{\mathbf{x}}) + \overline{Z})$ such that

$$\boldsymbol{W}\left(\triangle\mathbf{x}, \triangle\mathbf{y}, \triangle Y, \triangle Z\right) = \begin{bmatrix} \mathcal{L}''_{\mathbf{xx}}(\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{Y}, \overline{Z})\triangle\mathbf{x} + h'(\overline{\mathbf{x}})^*\triangle\mathbf{y} + g_1'(\overline{\mathbf{x}})^*\triangle Y + g_2'(\overline{\mathbf{x}})^*\triangle Z \\ h'(\overline{\mathbf{x}})\triangle\mathbf{x} \\ g_1'(\overline{\mathbf{x}})\triangle\mathbf{x} - \mathcal{S}^1(g_1'(\overline{\mathbf{x}})\triangle\mathbf{x} + \triangle Y) \\ g_2'(\overline{\mathbf{x}})\triangle\mathbf{x} - \mathcal{S}^2(g_2'(\overline{\mathbf{x}})\triangle\mathbf{x} + \triangle Z) \end{bmatrix} = 0.$$

It then follows from Lemma 8 and Lemma 9 that

$$\begin{cases} g_1'(\overline{\mathbf{x}})\triangle\mathbf{x} = \mathcal{S}^1(g_1'(\overline{\mathbf{x}})\triangle\mathbf{x} + \triangle Y) \in \operatorname{aff}\left(\mathcal{C}(g_1(\overline{\mathbf{x}}) + \overline{Y}; \partial\theta_1(g_1(\overline{\mathbf{x}})))\right), \\ g_2'(\overline{\mathbf{x}})\triangle\mathbf{x} = \mathcal{S}^2(g_2'(\overline{\mathbf{x}})\triangle\mathbf{x} + \triangle Z) \in \operatorname{aff}\left(\mathcal{C}(g_2(\overline{\mathbf{x}}) + \overline{Z}; \mathcal{N}_{\mathcal{K}}(g_2(\overline{\mathbf{x}})))\right). \end{cases}$$

We thus obtain from (42) that $\triangle\mathbf{x} \in \text{app}(\overline{\mathbf{y}}, \overline{Y}, \overline{Z})$. In addition, we derive from Propositions 10 and 11 that

$$\begin{aligned}
0 &= \Big\langle \triangle\mathbf{x},\, \mathcal{L}''_{\mathbf{xx}}(\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{Y}, \overline{Z})\triangle\mathbf{x} + h'(\overline{\mathbf{x}})^*\triangle\mathbf{y} + g_1'(\overline{\mathbf{x}})^*\triangle Y + g_2'(\overline{\mathbf{x}})^*\triangle Z \Big\rangle \\
&= \Big\langle \triangle\mathbf{x},\, \mathcal{L}''_{\mathbf{xx}}(\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{Y}, \overline{Z})\triangle\mathbf{x}\rangle + \langle g_1'(\overline{\mathbf{x}})\triangle\mathbf{x}, \triangle Y\rangle + \langle g_2'(\overline{\mathbf{x}})\triangle\mathbf{x}, \triangle Z \Big\rangle \\
&\geq \Big\langle \triangle\mathbf{x}, \mathcal{L}''_{\mathbf{xx}}(\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{Y}, \overline{Z})\triangle\mathbf{x} \Big\rangle - \Upsilon^1_{g_1(\overline{\mathbf{x}})}\left(\overline{Y}, g_1'(\overline{\mathbf{x}})\triangle\mathbf{x}\right) - \Upsilon^2_{g_2(\overline{\mathbf{x}})}\left(\overline{Z}, g_2'(\overline{\mathbf{x}})\triangle\mathbf{x}\right).
\end{aligned} \tag{61}$$

Since $\triangle\mathbf{x} \in \text{app}(\overline{\mathbf{y}}, \overline{Y}, \overline{Z})$, we conclude from (60) and (61) that $\triangle\mathbf{x} = 0$ and

$$\begin{bmatrix} h'(\overline{\mathbf{x}})^*\triangle\mathbf{y} + g_1'(\overline{\mathbf{x}})^*\triangle Y + g_2'(\overline{\mathbf{x}})^*\triangle Z \\ \mathcal{S}^1(\triangle Y) \\ \mathcal{S}^2(\triangle Z) \end{bmatrix} = 0. \tag{62}$$

By the assumed constraint nondegeneracy condition (40), we know that there exist $\mathbf{h} \in \mathbb{X}$, $H^1 \in \mathcal{T}^{\text{lin}}_{\theta_1}(g_1(\overline{\mathbf{x}}))$ and $H^2 \in \text{lin}\,(\mathcal{T}_{\mathcal{K}}(g_2(\overline{\mathbf{x}})))$ such that

$$h'(\overline{\mathbf{x}})\mathbf{h} = \triangle\mathbf{y}, \quad g_1'(\overline{\mathbf{x}})\mathbf{h} + H^1 = \triangle Y \quad \text{and} \quad g_2'(\overline{\mathbf{x}})\mathbf{h} + H^2 = \triangle Z.$$

It then follows from the first equation of (62) that

$$\begin{aligned}
&\langle \triangle\mathbf{y}, \triangle\mathbf{y}\rangle + \langle \triangle Y, \triangle Y\rangle + \langle \triangle Z, \triangle Z\rangle \\
&= \langle h'(\overline{\mathbf{x}})\mathbf{h},\, \triangle\mathbf{y}\rangle + \langle g_1'(\overline{\mathbf{x}})\mathbf{h} + H^1,\, \triangle Y\rangle + \langle g_2'(\overline{\mathbf{x}})\mathbf{h} + H^2,\, \triangle Z\rangle \\
&= \langle \mathbf{h}, h'(\overline{\mathbf{x}})^*\triangle\mathbf{y} + g_1'(\overline{\mathbf{x}})^*\triangle Y + g_2'(\overline{\mathbf{x}})^*\triangle Z\rangle + \langle H^1, \triangle Y\rangle + \langle H^2, \triangle Z\rangle \\
&= \langle H^1, \triangle Y\rangle + \langle H^2, \triangle Z\rangle = \Big\langle \overline{U}^\top H^1 \overline{U},\, \overline{U}^\top \triangle Y \overline{U} \Big\rangle + \Big\langle \overline{V}^\top H^2 \overline{V},\, \overline{V}^\top \triangle Z \overline{V} \Big\rangle.
\end{aligned}$$

Combining Proposition 3, Proposition 7, Lemma 8 and Lemma 9, we derive

$$\begin{aligned}
\Big\langle \overline{U}^\top H^1 \overline{U}, \overline{U}^\top \triangle Y \overline{U} \Big\rangle &= \sum_{1\leq l\leq r} \sum_{k\in\mathcal{E}^l} \Big\langle \overline{U}^\top_{\beta^l_k} H^1 \overline{U}_{\beta^l_k},\, \overline{U}^\top_{\beta^l_k} \triangle Y \overline{U}_{\beta^l_k} \Big\rangle \\
&= \sum_{1\leq l\leq r} \sum_{k\in\mathcal{E}^l} \Big\langle \widehat{\rho}_l\, e_{|\beta^l_k|},\, \text{diag}\left((R^l_k)^\top \overline{U}^\top_{\beta^l_k} \triangle Y\, \overline{U}_{\beta^l_k} R^l_k\right) \Big\rangle = 0
\end{aligned}$$

and

$$\begin{aligned}
\Big\langle \overline{V}^\top H^2 \overline{V},\, \overline{V}^\top \triangle Z \overline{V} \Big\rangle &= \sum_{1\leq l\leq r} \sum_{k\in\mathcal{F}^l} \Big\langle \overline{V}^\top_{\gamma^l_k} H^2 \overline{V}_{\gamma^l_k},\, \overline{V}^\top_{\gamma^l_k} \triangle Z \overline{V}_{\gamma^l_k} \Big\rangle \\
&= \sum_{1\leq l\leq r} \sum_{k\in\mathcal{F}^l} \Big\langle \widetilde{\rho}_l\, e_{|\gamma^l_k|},\, \text{diag}\left((R^l_k)^\top \overline{V}^\top_{\gamma^l_k} \triangle Z\, \overline{V}_{\gamma^l_k} R^l_k\right) \Big\rangle = 0.
\end{aligned}$$

Therefore, we get

$$\langle \triangle\mathbf{y}, \triangle\mathbf{y}\rangle + \langle \triangle Y, \triangle Y\rangle + \langle \triangle Z, \triangle Z\rangle = 0,$$

which implies that $\boldsymbol{W} \in \partial F(\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{Y}, \overline{Z})$ is nonsingular.

"(ii) $\Longrightarrow$ (iii)" By Clarke's inverse function theorem [9,10], we know that F is a locally Lipschitz homeomorphism near $(\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{Y}, \overline{Z})$. Thus, we know from Lemma 5 that $(\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{Y}, \overline{Z})$ is a strongly regular solution of the generalized equation (44). □

It has been proved in [34] that for the nonlinear semidefinite programming, the three statements in Theorem 1 are actually equivalent. Regrettably, we fail to establish the implication (iii) $\Longrightarrow$ (i) for the general CMatOPs. Nevertheless, for a class of special nonsmooth matrix optimization problem involving the largest eigenvalue, we can show the equivalence of these three conditions via the relationship between the so-called uniform second order growth condition and the strong second order sufficient condition. This is the subject of the next section.

5 Applications to CMatOPs involving the largest eigenvalue

In this section, we apply our obtained results to a class of CMatOPs involving the largest eigenvalue of a symmetric matrix. Relying on its special structure, we improve the results in Theorem 1 by showing that the three statements therein are actually equivalent.

Specifically, we consider the following problem

$$
\begin{aligned}
&\underset{\mathbf{x}\in\mathbb{X}}{\text{minimize}} && f(\mathbf{x})+\lambda_1(g(\mathbf{x}))\\
&\text{subject to} && h(\mathbf{x})=0,
\end{aligned}
\tag{63}
$$

where $\lambda_1(g(\mathbf{x}))$ denotes the largest eigenvalue of a symmetric matrix $g(\mathbf{x})$. This corresponds to a special case of problem (1) where

$$
\phi(\mathbf{x})=\max_{1\le i\le p}\left\{\langle \mathbf{e}^i,\mathbf{x}\rangle\right\},\quad \mathbf{x}\in\mathbb{R}^n
$$

with $\mathbf{e}^i$ being the unit vector whose i-th component is 1 and others are zero. Based on the formulas derived in Section 3, we get the following results.

Tangent cone and its lineality space. Recall the definitions of $\{\alpha^l\}$ in (4) and $\iota_1(\lambda(\overline{X}))$ in (9). We have

$$
\iota_1(\lambda(\overline{X}))=\left\{1\le i\le n \mid \lambda_i(\overline{X})=\bar{v}_1\right\}=\alpha^1.
$$

It follows from Proposition 3 that

$$
H\in\mathcal{T}^{\mathrm{lin}}_{\theta_1}(\overline{X}) \iff \left[\,\lambda_i'(\overline{X};H)=\lambda_j'(\overline{X};H),\quad \forall\, i,j\in\alpha^1\,\right].
$$

Based on Lemma 2, the above right-side is further equivalent to the existence of a scalar $\widehat{\rho}$ such that

$$
\overline{U}_{\alpha^1}^{\top}H\overline{U}_{\alpha^1}=\widehat{\rho}\, I_{|\alpha^1|}\ \text{ for some } \overline{U}\in\mathbb{O}^n(\overline{X}).
$$

The value of $\widehat{\rho}$ is independent of the selected orthogonal matrix $\overline{U}$ in $\mathbb{O}^n(\overline{X})$ ([13, Proposition 2]).

Critical cone. Given $(\overline{X},\overline{Y})\in\operatorname{gph}\partial\lambda_1$ and let $\overline{U}\in\mathbb{O}^n(\overline{X})\cap\mathbb{O}^n(\overline{Y})$. It follows from Lemma 4 and [40, Lemma 2.2] (see also [26, Lemma 3.1]) that

$$
\begin{cases}
0\le\lambda_i(\overline{Y})\le 1,\quad \forall\, i\in\alpha^1 \quad\text{and}\quad \displaystyle\sum_{i\in\alpha^1}\lambda_i(\overline{Y})=1,\\
\lambda_i(\overline{Y})=0,\quad \forall\, i\in\alpha^l,\quad l=2,\ldots,r.
\end{cases}
\tag{64}
$$

Denote

$$
\mu:=\{i\in\alpha^1\mid\lambda_i(\overline{Y})>0\}\quad\text{and}\quad \nu:=\{i\in\alpha^1\mid\lambda_i(\overline{Y})=0\}. \tag{65}
$$

For each $l\in\{1,\ldots,r\}$, we further partition the index set α^l by $\{\beta^l_k\}_{k=1}^{s^l}$ as in (19) based on $\lambda(\overline{Y})$. We then obtain from (64) that

$$
\iota_1(\lambda(\overline{X}))=\alpha^1=\mu\cup\nu,\quad \mu=\bigcup_{k=1}^{s^1-1}\beta^1_k,\quad \nu=\beta^1_{s^1}\quad\text{and}\quad \alpha^l=\beta^l_1\quad\text{for}\quad l=2,\ldots,r.
$$

Recall the index set $\eta_1(\lambda(\overline{X}),\lambda(\overline{Y}))\subseteq\iota_1(\lambda(\overline{X}))$ defined in (12). Obviously $\eta_1(\lambda(\overline{X}),\lambda(\overline{Y}))=\mu$. It follows from Proposition 4 that

$$
H\in\mathcal{C}(\overline{X}+\overline{Y};\partial\lambda_1(\overline{X})) \iff \left[\left(\operatorname{diag}(\overline{U}^{\top}H\overline{U})\right)_i=\lambda_1(\overline{U}_{\alpha^1}^{\top}H\overline{U}_{\alpha^1}),\quad\forall\, i\in\mu\right].
$$

One can also derive from Proposition 5 that

$$
H\in\operatorname{aff}\left(\mathcal{C}(A;\partial\lambda_1(\overline{X}))\right) \iff \overline{U}_{\alpha^1}^{\top}H\overline{U}_{\alpha^1}=\begin{bmatrix}\rho\, I_{|\mu|} & 0\\ 0 & \overline{U}_{\nu}^{\top}H\overline{U}_{\nu}\end{bmatrix}\quad\text{for some } \rho\in\mathbb{R}. \tag{66}
$$

$\underline{\boldsymbol{\sigma}\textbf{-term}}$. Also given $(\overline{X},\overline{Y}) \in \operatorname{gph}\partial\lambda_1$ and let $\overline{U} \in \mathbb{O}^n(\overline{X}) \cap \mathbb{O}^n(\overline{Y})$. Denote $\omega := \bigcup_{l=2}^{r} \alpha^l$. By noting that $\lambda_i(\overline{Y}) = 0$ for any $i \in \nu \cup \omega$, we derive from (28) and (64) that

$$\Upsilon^1_{\overline{X}}(\overline{Y}, H) = -2\sum_{i\in\mu}\sum_{j\in\omega} \frac{\lambda_i(\overline{Y})}{\lambda_i(\overline{X}) - \lambda_j(\overline{X})} (\overline{U}_\mu^\top H \overline{U}_\omega)^2_{ij} \le 0, \quad H \in \mathbb{S}^n. \tag{67}$$

In the rest of this section, we show that the strong regularity of the generalized equation for the KKT system at a local optimal optimal of problem (63) implies the strong second order sufficient condition and the constraint nondegeneracy at the same point, i.e., the three statements in Theorem 1 are equivalent.

Given a feasible point $\overline{\mathbf{x}}$ of problem (63), we say Robinson's constraint qualification (CQ) [28] at $\overline{\mathbf{x}}$ holds if

$$h'(\overline{\mathbf{x}})\mathbb{X} = \mathbb{Y}. \tag{68}$$

It has been proved in [11, Proposition 3.3] that the function $\lambda_1(\,\bullet\,)$ is $\mathcal{C}^2$-cone reducible at any point so that the set $\operatorname{epi}\lambda_1$ is second order regular [4, Proposition 3.136] (see [4, Definitions 3.85 & 3.135] for the definitions of $\mathcal{C}^2$-cone reducibility and second order regularity). One can then obtain the following second order necessary and sufficient conditions of (63) by adapting the proof of [4, Theorems 3.45 & 3.86]. For brevity, we omit the detailed proof here.

Proposition 11 *Suppose that $\overline{\mathbf{x}} \in \mathbb{X}$ is a local optimal solution of* (63) *and the Robinson's CQ* (68) *holds at $\overline{\mathbf{x}}$. Then the following second order necessary condition holds at $\overline{\mathbf{x}}$:*

$$\sup_{(\overline{\mathbf{y}},\overline{Y})\in\mathcal{M}(\overline{\mathbf{x}})} \left\{ \langle \mathbf{d}, \mathcal{L}''_{\mathbf{xx}}(\overline{\mathbf{x}},\overline{\mathbf{y}},\overline{Y})\mathbf{d}\rangle - \Upsilon^1_{g(\overline{\mathbf{x}})}\left(\overline{Y}, g'(\overline{\mathbf{x}})\mathbf{d}\right)\right\} \ge 0, \quad \forall\, \mathbf{d} \in \mathcal{C}(\overline{\mathbf{x}}),$$

where $\Upsilon^1_{g(\overline{\mathbf{x}})}$ is given by (67)*. Conversely, let $\overline{\mathbf{x}}$ be a feasible point of* (63) *and assume Robinson's CQ* (68) *holds at $\overline{\mathbf{x}}$. Then the following condition*

$$\sup_{(\overline{\mathbf{y}},\overline{Y})\in\mathcal{M}(\overline{\mathbf{x}})} \left\{ \langle \mathbf{d}, \mathcal{L}''_{\mathbf{xx}}(\overline{\mathbf{x}},\overline{\mathbf{y}},\overline{Y})\mathbf{d}\rangle - \Upsilon^1_{g(\overline{\mathbf{x}})}\left(\overline{Y}, g'(\overline{\mathbf{x}})\mathbf{d}\right)\right\} > 0, \quad \forall\, \mathbf{d} \in \mathcal{C}(\overline{\mathbf{x}}) \setminus \{0\}$$

is necessary and sufficient for the existence of a positive scalar ρ and a neighborhood $\mathcal{N}$ of $\overline{\mathbf{x}}$ such that

$$f(\mathbf{x}) + \lambda_1(g(\mathbf{x})) \ge f(\overline{\mathbf{x}}) + \lambda_1(g(\overline{\mathbf{x}})) + \rho\|\mathbf{x} - \overline{\mathbf{x}}\|^2 \quad \forall\, \mathbf{x} \in \mathcal{N} \text{ such that } h(\mathbf{x}) = 0. \tag{69}$$

The inequality (69) is usually called the quadratic growth condition at $\bar{x}$ of problem (63). In the conventional nonlinear programming, there is a stronger concept termed uniform quadratic growth condition [4, Definition 5.16]. Let $\mathbb{U}$ be a Banach space and consider functions $f : \mathbb{X} \times \mathbb{U} \to \mathbb{R}$, $g : \mathbb{X} \times \mathbb{U} \to \mathbb{S}^n$ and $h : \mathbb{X} \times \mathbb{U} \to \mathbb{Y}$. We say that $(f(\mathbf{x}, \mathbf{u}), g(\mathbf{x}, \mathbf{u}), h(\mathbf{x}, \mathbf{u}))$ is a C^2-smooth parameterization of (63) if $f(\,\bullet\,,\,\bullet\,)$, $g(\,\bullet\,,\,\bullet\,)$ and $h(\,\bullet\,,\,\bullet\,)$ are twice continuously differentiable and there exits $\overline{\mathbf{u}} \in \mathbb{U}$ such that $f(\,\bullet\,, \overline{\mathbf{u}}) \equiv f(\,\bullet\,)$, $g(\,\bullet\,, \overline{\mathbf{u}}) \equiv g(\,\bullet\,)$ and $h(\,\bullet\,, \overline{\mathbf{u}}) \equiv h(\,\bullet\,)$. Let $\overline{\mathbf{x}} \in \mathbb{X}$ be a stationary point of problem (63). We say that the uniform second order growth condition holds at $\overline{\mathbf{x}}$ with respect to a C^2-smooth parameterization $(f(\mathbf{x}, \mathbf{u}), g(\mathbf{x}, \mathbf{u}), h(\mathbf{x}, \mathbf{u}))$ if there exist $\rho > 0$ and neighborhoods $\mathcal{V}$ of $\overline{\mathbf{x}}$ and $\mathcal{U}$ of $\overline{\mathbf{u}}$ such that for any $\mathbf{u} \in \mathcal{U}$ and any stationary point $\mathbf{x}(\mathbf{u}) \in \mathcal{V}$ of the corresponding parameterized problem, the following holds:

$$f(\mathbf{x}, \mathbf{u}) + \lambda_1(g(\mathbf{x}, \mathbf{u})) \ge f(\mathbf{x}(\mathbf{u}), \mathbf{u}) + \lambda_1(g(\mathbf{x}(\mathbf{u}), \mathbf{u})) + \rho\|\mathbf{x} - \mathbf{x}(\mathbf{u})\|^2, \quad \forall\, \mathbf{x} \in \mathcal{V} \text{ such that } h(\mathbf{x}, \mathbf{u}) = 0.$$

We say that the uniform second order growth condition holds at $\overline{\mathbf{x}}$ if the above inequality holds for every C^2-smooth parameterization of (63).

The following proposition shows that the uniform second order growth condition of problem (63) at a stationary solution implies the strong second order sufficient condition at that point. Its proof is similar to [34, Lemma 4.1] for the nonlinear semidefinite programming problem.

Proposition 12 *Let $\overline{\mathbf{x}} \in \mathbb{X}$ be a stationary point of problem* (63)*. Suppose that Robinson's CQ* (68) *holds at $\overline{\mathbf{x}}$. If the uniform second order growth condition holds at $\overline{\mathbf{x}}$, then the strong second order sufficient condition* (43) *holds at $\overline{\mathbf{x}}$.*

Proof Denote $\overline{X} := g(\overline{\mathbf{x}})$. Let $(\overline{\mathbf{y}}, \overline{Y}) \in \mathcal{M}(\overline{\mathbf{x}})$ and $\overline{U} \in \mathbb{O}^n(\overline{X}) \cap \mathbb{O}^n(\overline{Y})$. Recall the index sets μ and ν defined by (65). Given a positive scalar τ, we consider the following problem:

$$\begin{array}{ll} \underset{\mathbf{x}\in\mathbb{X}}{\text{minimize}} & f(\mathbf{x}) + \lambda_1\left(g(\mathbf{x}) - \tau\overline{U}_\nu\overline{U}_\nu^\top\right) \\ \text{subject to} & h(\mathbf{x}) = 0. \end{array}$$

Let $\mathcal{M}_\tau(\overline{\mathbf{x}})$ be the set of all $(\mathbf{y}, Y) \in \mathbb{Y} \times \mathbb{S}^n$ such that $(\overline{\mathbf{x}}, \mathbf{y}, Y)$ satisfies the KKT optimality condition of the above problem. Similarly, let $\mathcal{C}_\tau(\overline{\mathbf{x}})$ be the the critical cone defined in (41) with respect to the above problem. For all sufficiently small τ, we have

$$\iota_1\left(\lambda\left(\overline{X} - \tau\overline{U}_\nu\overline{U}_\nu^\top\right)\right) \equiv \eta_1(\lambda(\overline{X}), \lambda(\overline{Y})) = \mu,$$

where the set ι_1 and η_1 are defined in (9) and (12). It then follows from (64) and the above equality that for such sufficiently small τ, $(\overline{\mathbf{y}}, \overline{Y}) \in \mathcal{M}_\tau(\overline{\mathbf{x}}) \subseteq \mathcal{M}(\overline{\mathbf{x}})$. In addition, we know from Propositions 4 and 5 that

$$\mathcal{C}_\tau(\overline{\mathbf{x}}) \supseteq \operatorname{app}(\overline{\mathbf{y}}, \overline{Y}),$$

where $\operatorname{app}(\overline{\mathbf{y}}, \overline{Y})$ is defined in (42). Therefore, Proposition 11 implies that for all sufficiently small τ,

$$\sup_{(\mathbf{y},Y)\in\mathcal{M}_\tau(\overline{\mathbf{x}})} \left\{\langle \mathbf{d}, \mathcal{L}''_{\mathbf{xx}}(\overline{\mathbf{x}}, \mathbf{y}, Y)\mathbf{d}\rangle - \Upsilon^1_{\left(\overline{X}-\tau\overline{U}_\nu\overline{U}_\nu^\top\right)}(Y, g_1'(\overline{\mathbf{x}})\mathbf{d})\right\} > 0, \quad \forall\, \mathbf{d} \in \mathcal{C}_\tau(\overline{\mathbf{x}}) \setminus \{0\}.$$

Since $\lambda_i(\overline{Y}) = 0$ for all $i \in \nu$, we obtain from (66) and (67) that

$$\Upsilon^1_{\left(\overline{X}-\tau\overline{U}_\nu\overline{U}_\nu^\top\right)}\left(\overline{Y}, g'(\overline{\mathbf{x}})\mathbf{d}\right) = \Upsilon^1_{\overline{X}}\left(\overline{Y}, g'(\overline{\mathbf{x}})\mathbf{d}\right).$$

Combining the above derivations together, we know that the strong second order sufficient condition (43) holds at $\bar{\mathbf{x}}$. $\square$

By considering the epigraphical formulation of (63) and following the proof of [4, Theorem 5.20], one can show that the strong regularity of a KKT solution $(\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{Y})$ implies the uniform second order growth condition at $\overline{\mathbf{x}}$. We again omit the proof here.

Now we are ready to present the main result of this section pertaining to the necessary and sufficient conditions for the strong regularity of (63) for problem (63), which can be directly obtained by combining Theorem 1 and Proposition 12.

Theorem 2 *Let $\overline{\mathbf{x}} \in \mathbb{X}$ be a locally optimal solution of problem* (63)*. Suppose that Robinson's CQ* (68) *holds at $\overline{\mathbf{x}}$. Let $(\overline{\mathbf{y}}, \overline{Y}) \in \mathcal{M}(\overline{\mathbf{x}})$. Then the following statements are equivalent:*

(i) the strong second order sufficient condition (43) *and constraint nondegeneracy* (40) *hold at $\overline{\mathbf{x}}$;*

(ii) every element in $\partial F(\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{Y})$ is nonsingular;

(iii) $(\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{Y})$ is a strongly regular solution of the generalized equation for (63)*.*

(iv) The uniform second order growth condition and constraint nondegeneracy (40) *hold at $\overline{\mathbf{x}}$.*

6 Conclusion

In this paper, we conduct an extensive study on the characterization of strong regularity of the KKT solutions for a class of nonsmooth composite matrix optimization problems (CMatOPs). Due to its non-polyhedrality, the classical perturbation analysis developed for the nonlinear programming has become inadequate for CMatOPs. We have systemically analyzed second-order variational properties of spectral functions associated with piecewise affine symmetric functions, including the characterizations of their induced tangent sets, lineality spaces, critical cones and the "σ-term". These variational results provide the necessary tools for the characterization of the strong regularity for the general CMatOPs. The work done on CMatOPs in this paper is by no means complete. Due to the rapid advances in matrix optimization applications in emerging fields, we believe that the fundamental perturbation analysis of CMatOPs will become even more important and many other variational properties are waiting to be explored.

References

1. Alizadeh, F.: Interior point methods in semidefinite programming with applications to combinatorial optimization. SIAM J. Optim. **5** 13–51 (1995)
2. Ben-Tal, A., Zowe, J.: Necessary and sufficient optimality conditions for a class of nonsmooth minimization problems. Math. Program. **24**, 70–91 (1982)
3. Bonnans, J.F., Cominetti, R., Shapiro, A.: Sensitivity analysis of optimization problems under second order regular constraints. Math. Oper. Res. **23**, 806–831 (1998)
4. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer, New York (2000)
5. Boyd, S., El Ghaoui, L., Feron, E., Balakrishnan, V.: Linear matrix inequalities in system and control theory. Vol. 15 Studies in Applied Mathematics, SIAM, Philadelphia (1994)
6. Boyd, S., Diaconis, P., Parrilo, P.A., Xiao, L.: Fastest mixing Markov chain on graphs with symmetries. SIAM J. Optim. **20**, 792–819 (2009)
7. Candés, E.J., Tao, T.: The power of convex relaxation: near-optimal matrix completion. IEEE Trans. Inf. Theory **56**, 2053–2080 (2009)
8. Chan, Z.X., Sun, D.F.: Constraint nondegeneracy, strong regularity, and nonsingularity in semidefinite programming. SIAM J. Optim. **19**, 370–396 (2008)
9. Clarke, F.H.: On the inverse function theorem. Pacific J. Math **64**, 97–102 (1976)
10. Clarke, F.H.: Optimization and Nonsmooth Analysis. John Wiley & Sons, New York (1983)
11. Cui, Y., Ding, C., Zhao, X.Y.: Quadratic growth conditions for convex matrix optimization problems associated with spectral functions. SIAM J. Optim. **27**, 2332–2355 (2017)
12. Ding, C., Sun, D.F., Toh, K.-C.: An introduction to a class of matrix cone programming, Math. Program. **144**, 141–179 (2014)
13. Ding, C., Sun, D.F., Sun, J., Toh, K.-C.: Spectral operators of matrices. Math. Program. **168**, 509–531 (2018)
14. Ding, C., Sun, D.F., Sun, J., Toh, K.-C.: Spectral operators of matrices: semismoothness and characterizations of the generalized Jacobian. arXiv:1810.09856, 1–26 (2018)
15. Dontchev, A.L., Rockafellar, R.T.: Characterizations of strong regularity for variational inequalities over polyhedral convex sets. SIAM J. Optim. **6**, 1087–1105 (1996)
16. Facchinei, F., Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer, New York (2003)
17. Fan, K.: On a theorem of Weyl concerning eigenvalues of affine transformations. P. Natl. Acad. Sci. USA. **35**, 652–655 (1949)
18. Lancaster, P.: On eigenvalues of matrices dependent on a parameter. Numer. Math. **6**, 377–387 (1964)
19. Lewis, A.S.: The convex analysis of unitarily invariant matrix functions. J. Convex Anal. **2**, 173–183 (1995)
20. Lewis, A.S.: Convex analysis on the Hermitian matrices. SIAM J. Optim. **6**, 164–177 (1996)
21. Lewis, A.S.: Derivatives of spectral functions. Math. Oper. Res. **21**, 576–588 (1996)
22. Meng, F., Sun, D., Zhao, G.: Semismoothness of solutions to generalized equations and the Moreau-Yosida regularization. Math. Program. **104**, 561–581 (2005)
23. Mordukhovich, B.S., Nghia T.T.A., Rockafellar, R.T.: Full stability in finite-dimensional optimization. Math. Oper. Res. **40**, 226–252 (2014)
24. Mordukhovich, B.S., Sarabi, M.E.: Generalized differentiation of piecewise affine functions in second-order variational analysis. Nonlinear Anal-Theor. **132**, 240–273 (2016)
25. Mordukhovich, B.S., Sarabi, M.E.: Critical multipliers in variational systems via second-order generalized differentiation. Math. Program. **169**, 605–648 (2018)
26. Overton, M.L., Womersley, R.S.: Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices. Math. Program. **62**, 321–357 (1993).
27. Recht, B., Fazel, M., Parrilo, P.A.: Guaranteed minimum rank solutions to affine matrix equations via nuclear norm minimization. SIAM Rev. **52**, 471–501 (2010)
28. Robinson, S.M.: First order conditions for general nonlinear optimization. SIAM J. Appl. Math. **30**, 597–607 (1976)
29. Robinson, S.M.: Strongly regular generalized equations. Math. Oper. Res. **5**, 43–62 (1980)
30. Robinson, S.M.: Local structure of feasible sets in nonlinear programming, Part II: Nondegeneracy, Math. Program. Stud. **22**, 217–230 (1984).
31. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
32. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, New York (1998)
33. Sturm, J.F.: Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. Optim. Method. Softw. **11** 625–653 (1999)
34. Sun, D.F.: The strong second order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. Math. Oper. Res. **31**, 761–776 (2006)
35. Torki, M.: Second-order directional derivatives of all eigenvalues of a symmetric matrix. Nonlinear Anal-Theor. **46**, 1133–1150 (2001)
36. Toh, K.-C., Todd, M.J., Tütüncü, R.H.: SDPT3 – a Matlab software package for semidefinite programming. Optim. Method. Softw. **11** 545–581 (1999)
37. Toh, K.-C., Trefethen, L. N.: The Chebyshev polynomials of a matrix. SIAM J. Matrix Anal. A. **20**, 400–419 (1998)
38. Warga, J.: Fat homeomorphisms and unbounded derivate containers. J. Math. Anal. Appl. **81**, 545–560 (1981)
39. Watson, G.A.: On matrix approximation problems with Ky Fan k norms. Numer. Algorithms **5**, 263–272 (1993)
40. Wu, B., Ding, C., Sun, D., Toh, K.-C.: On the Moreau-Yosida regularization of the vector k-norm related functions. SIAM J. Optim. **24**, 766–794 (2014).
41. Yang, L., Sun, D.F., Toh, K.-C.: SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints. Mathe. Program. Comput. **7** 331–366 (2015)
42. Zhao, X.Y., Sun, D.F., Toh, K.-C.: A Newton-CG augmented Lagrangian method for semidefinite programming. SIAM J. Optim. **20**, 1737–1765 (2010)