SIAM J. OPTIM.
Vol. 30, No. 1, pp. 630–659

SPECTRAL OPERATORS OF MATRICES: SEMISMOOTHNESS AND CHARACTERIZATIONS OF THE GENERALIZED JACOBIAN*

CHAO DING†, DEFENG SUN‡, JIE SUN§, AND KIM-CHUAN TOH¶

Abstract. Spectral operators of matrices proposed recently in [C. Ding, D. F. Sun, J. Sun, and K. C. Toh, *Math. Program.*, 168 (2018), pp. 509–531] are a class of matrix-valued functions, which map matrices to matrices by applying a vector-to-vector function to all eigenvalues/singular values of the underlying matrices. Spectral operators play a crucial role in the study of various applications involving matrices such as matrix optimization problems that include semidefinite programming as one of most important example classes. In this paper, we will study more fundamental first- and second-order properties of spectral operators, including the Lipschitz continuity, ρ-order B(ouligand)-differentiability ($0 < \rho \leq 1$), ρ-order G-semismoothness ($0 < \rho \leq 1$), and characterization of generalized Jacobians.

Key words. spectral operators, matrix-valued functions, semismoothness, generalized Jacobian

AMS subject classifications. 90C25, 90C06, 65K05, 49J50, 49J52

DOI. 10.1137/18M1222235

1. Introduction.

Spectral operators of matrices introduced recently in [23] are a class of matrix-valued functions defined on a given real Euclidean vector space $\mathcal{X}$ of real/complex matrices over the scalar field of real numbers $\mathbb{R}$. Unlike the well-studied classical matrix functions [34, Chapter 9], [39, Chapter 6], [2, 38, 37], which are Löwner's operators generated by applying a single-variable function to each of the eigenvalues/singular values of the underlying matrices, the spectral operators introduced in [23] generate matrix-valued functions by applying a vector-to-vector function to all eigenvalues/singular values of the underlying matrices (see Definition 2.2 for details).

In addition to its intrinsic theoretical interest in linear algebra, spectral operators play a crucial role in the study of a class of optimization problems known as matrix optimization problems (MOPs), which include many important problems such as matrix norm approximation, matrix completion, rank minimization, graph theory, and machine learning [35, 82, 83, 72, 46, 8, 9, 10, 12, 88, 14, 57, 26, 42, 32, 58, 59, 93, 54]. In particular, for a given unitarily invariant proper closed convex function $f : \mathcal{X} \to (-\infty, \infty]$, the spectral operator that is closely related to MOPs is the proximal mapping [74] of f at X, which is defined by

*Received by the editors October 22, 2018; accepted for publication (in revised form) December 23, 2019; published electronically February 20, 2020.
https://doi.org/10.1137/18M1222235
Funding: The research of the first author was supported by the National Natural Science Foundation of China under projects 11671387, 11531014, and 11688101. The research of the second author was supported in part by a start-up research grant from the Hong Kong Polytechnic University.
†Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, People's Republic of China (dingchao@amss.ac.cn).
‡Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong (defeng.sun@polyu.edu.hk).
§Faculty of Science and Engineering, Curtin University, Bentley WA 6102, Australia (Jie.Sun@curtin.edu.au).
¶Department of Mathematics and Institute of Operations Research and Analytics, National University of Singapore, Singapore (mattohkc@nus.edu.sg).

$$P_f(X) := \text{argmin}_{Y\in\mathcal{X}} \left\{ f(Y) + \frac{1}{2}\|Y - X\|^2 \right\}, \quad X \in \mathcal{X}, \tag{1.1}$$

where $\mathcal{X}$ is either the real vector subspace $\mathbb{S}^m$ of $m \times m$ real symmetric or complex Hermitian matrices, or the real vector subspace $\mathbb{V}^{m\times n}$ of $m \times n$ (assume $m \leq n$) real/complex matrices. Among different MOP applications, semidefinite programming (SDP) [81] is undoubtedly one of the most influential classes of problems and its importance has been well recognized by researchers even beyond the optimization community. Recent exciting progress has been made both in the design of efficient numerical methods for solving large-scale SDPs [92, 89] and in the study of second-order variational analysis of SDP problems [25, 77, 11, 62], in which the first- and second-order properties of the special spectral operator, the projection operator over the positive semidefinite matrix cone [78, 80], have played an essential role. However, for the general MOPs arising recently from different fields, the classical theory developed for Löwner's operators has become inadequate to cope with the new theoretical developments and needs. Beyond the spectral operators of matrices arising from proximal mappings, more general spectral operators indeed have played a pivotal role in many other MOP applications [60]. Therefore, the study of the general spectral operators will provide the necessary foundations for both computational and theoretical study of the general MOPs. In particular, the first- and second-order properties of spectral operators obtained in [23] including the well-definedness, continuity, directional differentiability, and Fréchet-differentiability are of fundamental importance in the study of MOPs [22, 55, 13, 18].

In this paper, we will follow the path set in [23] to conduct extensive theoretical studies on spectral operators. A natural question one may want to ask is whether the first- and second-order properties can be transferred from a vector-valued mixed symmetric mapping g (see Definition 2.1) to the corresponding spectral operator G and vice versa. It is known that properties such as convexity, prox-regularity, proximal smoothness, and others follow the so-called transfer principle for scalar-valued spectral functions [33] (see [45, 19, 20, 21, 27, 30] for more details). However, for spectral operators, this question has not been well answered yet. More first- and second-order properties of spectral operators need to be discussed in depth. These include Lipschitz continuity, ρ-order B(ouligand)-differentiability ($0 < \rho \leq 1$), ρ-order G-semismoothness ($0 < \rho \leq 1$), and characterization of generalized Jacobians. In particular, we will study the semismoothness [61, 70] of spectral operators, which is one of the most important properties for both algorithmic design and theoretical study of the general MOPs.

Historically, the semismoothness of vector-valued functions had played a crucial role in constructing nonsmooth and smoothing Newton methods for nonlinear equations and related problems. In fact, it is shown in [70, 69, 67] that the (strong) semismoothness is the key property for the local (quadratic) superlinear convergence of the Newton method. The semismooth Newton method has became one of the most important techniques in optimization [41, 85, 92, 89, 51, 52, 91]. In particular, the several semismooth Newton based methods have been proposed for solving various large-scale optimization problems in machine learning applications such as the lasso, fused lasso, and convex clustering problems, and they have significantly outperformed a number of state-of-the-art solvers in terms of efficiency and robustness [51, 52, 91]. For MOPs, the semismoothness of the special spectral operator—the projection operator over the SDP cone—has played a key role in the development of the semismooth Newton based augmented Lagrangian method implemented in the software package

SDPNAL [92] and its enhanced version SDPNAL+ [89] for solving large-scale SDP problems. Therefore, based on this recent progress, we believe that the results on the semismoothness of spectral operators obtained in this paper will lay a foundation for the research on general MOPs. For the proximal mapping (1.1), one can obtain its semismooth property by employing the recently developed semialgebraic geometry results [3, 17]. It is shown in [4, 40] that locally Lipschitz continuous tame functions (e.g., the proximal mapping (1.1)) are semismooth. For more recent developments on semialgebraic geometry in optimization, see [1, 28, 29, 15, 50, 6, 5] and the references therein. It is worth noting that unlike our approach, by just employing its tameness, one may not able to obtain the explicit formulas of the directional derivative and, more importantly, the strong semismoothness of the proximal mapping (see section 5 for details).

Another fundamental property, which we will study, is the characterization of the Clarke generalized Jacobians [16] of the locally Lipschitz continuous spectral operators. This is an important theoretical topic in the second-order variational analysis, which is crucial for the study of many perturbation properties of MOPs such as the strong regularity [68, 77, 11] and full and tilt stability [62, 63]. In addition, for the software packages SDPNAL and SDPNAL+, due to the explicit characterization of the Clarke generalized Jacobian of the projection operator over the positive semidefinite matrix cone, it becomes possible to exploit the second-order sparsity of the SDP problems inherited from the sparse structure of the generalized Jacobian of the reformulated semismooth equations. The second-order sparsity can substantially reduce the computational cost of solving the resulting linear systems associated with the semismooth Newton directions. Indeed the efficient computation of the semismooth Newton directions is one of the biggest computational challenges in designing efficient second-order numerical methods for solving large-scale problems. To summarize, we believe that the fundamental results obtained in this paper, especially the second-order properties such as the semismoothness and the Clarke generalized Jacobian of spectral operators, are of importance in both the computational and theoretical study of general MOPs.

The remaining parts of this paper are organized as follows. In section 2, we briefly review several preliminary properties of spectral operators of matrices. We study the Lipschitz continuity and Bouligand-differentiability of spectral operators defined on a single matrix space $\mathbb{V}^{m\times n}$ in sections 3 and 4, respectively. Then, the G-semismoothness and characterization of the Clarke generalized Jacobians of spectral operators are presented in sections 5 and 6, respectively. In section 7, we extend the corresponding results to spectral operators defined on the Cartesian product of several matrix spaces and the smoothing spectral operators. We make some final remarks in section 8.

Below are some common notation and symbols to be used later in the paper:

- For any $X \in \mathbb{V}^{m\times n}$, we denote by X_{ij} the (i,j)th entry of X and x_j the jth column of X. Let $I \subseteq \{1,\ldots,m\}$ and $J \subseteq \{1,\ldots,n\}$ be two index sets. We use X_J to denote the submatrix of X obtained by removing all the columns of X not in J and X_{IJ} to denote the $|I|\times|J|$ submatrix of X obtained by removing all the rows of X not in I and all the columns of X not in J.
- For $X \in \mathbb{V}^{m\times m}$, $\operatorname{diag}(X)$ denotes the column vector consisting of all the diagonal entries of X being arranged from the first to the last. For $x \in \mathbb{R}^m$, $\operatorname{Diag}(x)$ denotes the $m \times m$ diagonal matrix whose i-th diagonal entry is x_i, $i = 1,\ldots,m$.

- We use " $\circ$ " to denote the usual Hadamard product between two matrices, i.e., for any two matrices A and B in $\mathbb{V}^{m\times n}$ the (i,j)'th entry of $Z := A\circ B \in \mathbb{V}^{m\times n}$ is $Z_{ij} = A_{ij}B_{ij}$.
- For any $X \in \mathbb{S}^m$, we use $\lambda : \mathbb{S}^m \to \mathbb{R}^m$ to denote the mapping of the ordered eigenvalues of a Hermitian matrix X satisfying $\lambda_1(X) \geq \lambda_2(X) \geq \cdots \geq \lambda_m(X)$. For any $X \in \mathbb{V}^{m\times n}$, we use $\sigma : \mathbb{V}^{m\times n} \to \mathbb{R}^m$ to denote the mapping of the ordered singular values of X satisfying $\sigma_1(X) \geq \sigma_2(X) \geq \cdots \geq \sigma_m(X) \geq 0$.
- Let $\mathbb{O}^p$ $(p = m, n)$ be the set of $p \times p$ orthogonal/unitary matrices. We denote $\mathbb{P}^p$ and $\pm\mathbb{P}^p$ to be the sets of all $p \times p$ permutation matrices and signed permutation matrices, respectively. For any $Y \in \mathbb{S}^m$ and $Z \in \mathbb{V}^{m\times n}$, we use $\mathbb{O}^m(Y)$ to denote the set of all orthogonal matrices whose columns form an orthonormal basis of eigenvectors of Y, and we use $\mathbb{O}^{m,n}(Z)$ to denote the set of all pairs of orthogonal matrices (U, V), where the columns of U and V form a compatible set of orthonormal left and right singular vectors for Z, respectively.

2. Spectral operators of matrices. The general spectral operators of matrices introduced by [23] are defined on the Cartesian product of several real or complex matrix spaces. Let us first summarize the properties of spectral operators obtained by [23], which are needed in the subsequent analysis.

For notational simplicity, we introduce the definitions and notation for the special case that $\mathcal{X} \equiv \mathbb{S}^{m_1} \times \mathbb{V}^{m_2\times n_2}$, where m_1, m_2, and n_2 are given positive integers. The corresponding generalizations can be found in [23].

Without loss of generality, we assume that $m_2 \leq n_2$. For any $X = (X_1, X_2) \in \mathcal{X}$, we have $X_1 \in \mathbb{S}^{m_1}$ and $X_2 \in \mathbb{V}^{m_2\times n_2}$. Denote $\mathcal{Y} := \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$. For any $X \in \mathcal{X}$, define $\kappa(X) \in \mathcal{Y}$ by $\kappa(X) := (\lambda(X_1), \sigma(X_2))$. Define the set $\mathcal{P}$ by

$$\mathcal{P} := \{(Q_1, Q_2) \mid Q_1 \in \mathbb{P}^{m_1} \text{ and } Q_2 \in \pm\mathbb{P}^{m_2}\}.$$

Let $g : \mathcal{Y} \to \mathcal{Y}$ be a given mapping. For any $x = (x_1, x_2) \in \mathcal{Y}$ with $x_k \in \mathbb{R}^{m_k}$ for $k = 1, 2$, we write $g(x) \in \mathcal{Y}$ in the form $g(x) = (g_1(x), g_2(x))$ with $g_k(x) \in \mathbb{R}^{m_k}$ for $k = 1, 2$.

DEFINITION 2.1. *The given mapping $g : \mathcal{Y} \to \mathcal{Y}$ is said to be mixed symmetric, with respect to $\mathcal{P}$, at $x = (x_1, x_2) \in \mathcal{Y}$ with $x_k \in \mathbb{R}^{m_k}$ for $k = 1, 2$, if*

$$g(Q_1x_1, Q_2x_2) = (Q_1g_1(x), Q_2g_2(x)) \quad \forall\ (Q_1, Q_2) \in \mathcal{P}. \tag{2.1}$$

The mapping g is said to be mixed symmetric, with respect to $\mathcal{P}$, over a set $\mathcal{D} \subseteq \mathcal{Y}$ if (2.1) *holds for every $x \in \mathcal{D}$.*

Note that the function values $g_k(x) \in \mathbb{R}^{m_k}$, $k = 1, 2$, are dependent on all x_1, x_2. When there is no danger of confusion, in later discussions we often drop the phrase "with respect to $\mathcal{P}$" from Definition 2.1. Let $\mathcal{N}$ be a given nonempty set in $\mathcal{X}$. Define $\kappa_{\mathcal{N}} := \{\kappa(X) \in \mathcal{Y} \mid X \in \mathcal{N}\}$. The following definition of the spectral operator with respect to a mixed symmetric mapping g is given by [23, Definition 1].

DEFINITION 2.2. *Suppose that $g : \mathcal{Y} \to \mathcal{Y}$ is mixed symmetric on $\kappa_{\mathcal{N}}$. The spectral operator $G : \mathcal{N} \to \mathcal{X}$ with respect to g is defined as $G(X) := (G_1(X), G_2(X))$ for $X = (X_1, X_2) \in \mathcal{N}$ such that*

$$\begin{cases} G_1(X) := P_1 \mathrm{Diag}\big(g_1(\kappa(X))\big) P_1^{\mathbb{T}}, \\ G_2(X) := U_2 \left[\mathrm{Diag}\big(g_2(\kappa(X))\big) \quad 0\right] V_2^{\mathbb{T}}, \end{cases}$$

where $P_1 \in \mathbb{O}^{m_1}(X_1)$ and $(U_2, V_2) \in \mathbb{O}^{m_2,n_2}(X_2)$.

For the well-definedness, continuity, and F(réchet)-differentiability of spectral operators, one may refer to [23] for details. It is worth mentioning that for the case that $\mathcal{X} \equiv \mathbb{S}^m$ (or $\mathbb{V}^{m\times n}$) and g has the form $g(y) = (h(y_1), \ldots, h(y_m)) \in \mathbb{R}^m$ with $y_i \in \mathbb{R}$ for some given scalar-valued function $h : \mathbb{R} \to \mathbb{R}$, the corresponding spectral operator G is just the Löwner operator coined in [80] in recognition of Löwner's original contribution on this topic in [56] (or the Löwner non-Hermitian operator [90] if $h(0) = 0$). In [90], Yang studied several important first- and second-order properties of the Löwner non-Hermitian operator, including its F-differentiability and the explicit derivative formula (the equivalent form also can be found in [64]). Furthermore, for the case that $\mathcal{X} \equiv \mathbb{S}^m$ (or $\mathbb{V}^{m\times n}$) and $g : \mathbb{R}^m \to \mathbb{R}^m$ is a given vector-valued function, the related spectral operator G covers the matrix-valued functions studied in [44, 47, 48, 49] as special cases.

Next, for further simplifying the notation, we will focus on the study of spectral operators for the case that $\mathcal{X} \equiv \mathbb{V}^{m\times n}$. The corresponding extensions for the spectral operators defined on the general Cartesian product of several matrix spaces will be presented in section 7.

Let $\mathcal{N}$ be a given nonempty open set in $\mathbb{V}^{m\times n}$. Suppose that $g : \mathbb{R}^m \to \mathbb{R}^m$ is mixed symmetric with respect to $\mathcal{P} \equiv \pm\mathbb{P}^m$ (i.e., absolutely symmetric), on an open set $\hat{\sigma}_{\mathcal{N}}$ in $\mathbb{R}^m$ containing $\sigma_{\mathcal{N}} := \{\sigma(X) \mid X \in \mathcal{N}\}$. The spectral operator $G : \mathcal{N} \to \mathbb{V}^{m\times n}$ with respect to g defined in Definition 2.2 then takes the form of

$$G(X) = U \left[\mathrm{Diag}(g(\sigma(X))) \quad 0\right] V^{\mathbb{T}}, \quad X \in \mathcal{N},$$

where $(U, V) \in \mathbb{O}^{m,n}(X)$. For a given $\overline{X} \in \mathcal{N}$, consider the singular value decomposition (SVD) of $\overline{X}$, i.e.,

$$\overline{X} = \overline{U} \left[\Sigma(\overline{X}) \quad 0\right] \overline{V}^{\mathbb{T}}, \tag{2.2}$$

where $\Sigma(\overline{X})$ is an $m \times m$ diagonal matrix whose ith diagonal entry is $\sigma_i(\overline{X})$, $\overline{U} \in \mathbb{O}^m$, and $\overline{V} = \left[\overline{V}_1 \quad \overline{V}_2\right] \in \mathbb{O}^n$ with $\overline{V}_1 \in \mathbb{V}^{n\times m}$ and $\overline{V}_2 \in \mathbb{V}^{n\times(n-m)}$.

We end this section by further introducing some necessary notation and results, which are used in later discussions. Let $\overline{\sigma} := \sigma(\overline{X}) \in \mathbb{R}^m$. We use $\overline{\nu}_1 > \overline{\nu}_2 > \cdots > \overline{\nu}_r > 0$ to denote the nonzero distinct singular values of $\overline{X}$. Let a_l, $l = 1, \ldots, r$, a, b, and c be the index sets defined by

$$\begin{aligned} &a_l := \{i \mid \sigma_i(\overline{X}) = \overline{\nu}_l,\ 1 \le i \le m\}, \quad l = 1, \ldots, r, \qquad a := \{i \mid \sigma_i(\overline{X}) > 0,\ 1 \le i \le m\}, \\ &b := \{i \mid \sigma_i(\overline{X}) = 0,\ 1 \le i \le m\} \quad \text{and} \quad c := \{m+1, \ldots, n\}. \end{aligned} \tag{2.3}$$

Denote $\bar{a} := \{1, \ldots, n\} \setminus a$. For each $i \in \{1, \ldots, m\}$, we also define $l_i(\overline{X})$ to be the number of singular values which are equal to $\sigma_i(\overline{X})$ but are ranked before i (including i) and $\tilde{l}_i(\overline{X})$ to be the number of singular values which are equal to $\sigma_i(\overline{X})$ but are ranked after i (excluding i), i.e., define $l_i(\overline{X})$ and $\tilde{l}_i(\overline{X})$ such that

$$\begin{aligned} &\sigma_1(\overline{X}) \ge \cdots \ge \sigma_{i-l_i(\overline{X})}(\overline{X}) > \sigma_{i-l_i(\overline{X})+1}(\overline{X}) = \cdots = \sigma_i(\overline{X}) = \cdots = \sigma_{i+\tilde{l}_i(\overline{X})}(\overline{X}) \\ &\quad > \sigma_{i+\tilde{l}_i(\overline{X})+1}(\overline{X}) \ge \cdots \ge \sigma_m(\overline{X}). \end{aligned} \tag{2.4}$$

In later discussions, when the dependence of l_i and $\tilde{l}_i$ on $\overline{X}$ is clear from the context, we often drop $\overline{X}$ from the notation for convenience. We define two linear matrix operators $S : \mathbb{V}^{p\times p} \to \mathbb{S}^p$, $T : \mathbb{V}^{p\times p} \to \mathbb{V}^{p\times p}$ by

$$S(Y) := \frac{1}{2}\left(Y + Y^{\mathbb{T}}\right), \quad T(Y) := \frac{1}{2}\left(Y - Y^{\mathbb{T}}\right), \quad Y \in \mathbb{V}^{p\times p}. \tag{2.5}$$

For any given $X \in \mathcal{N}$, let $\sigma = \sigma(X)$. For the mapping g, we define three matrices $\mathcal{E}_1^0(\sigma), \mathcal{E}_2^0(\sigma) \in \mathbb{R}^{m\times m}$ and $\mathcal{F}^0(\sigma) \in \mathbb{R}^{m\times(n-m)}$ (depending on $X \in \mathcal{N}$) by

$$(\mathcal{E}_1^0(\sigma))_{ij} := \begin{cases} (g_i(\sigma) - g_j(\sigma))/(\sigma_i - \sigma_j) & \text{if } \sigma_i \neq \sigma_j, \\ 0 & \text{otherwise,} \end{cases} \quad i, j \in \{1, \ldots, m\}, \tag{2.6}$$

$$(\mathcal{E}_2^0(\sigma))_{ij} := \begin{cases} (g_i(\sigma) + g_j(\sigma))/(\sigma_i + \sigma_j) & \text{if } \sigma_i + \sigma_j \neq 0, \\ 0 & \text{otherwise,} \end{cases} \quad i, j \in \{1, \ldots, m\}, \tag{2.7}$$

$$(\mathcal{F}^0(\sigma))_{ij} := \begin{cases} g_i(\sigma)/\sigma_i & \text{if } \sigma_i \neq 0, \\ 0 & \text{otherwise,} \end{cases} \quad i \in \{1, \ldots, m\}, \quad j \in \{1, \ldots, n-m\}. \tag{2.8}$$

When the dependence of $\mathcal{E}_1^0(\sigma)$, $\mathcal{E}_2^0(\sigma)$, and $\mathcal{F}^0(\sigma)$ on σ is clear from the context, we often drop σ from the notation. In particular, let $\overline{\mathcal{E}}_1^0$, $\overline{\mathcal{E}}_2^0 \in \mathbb{V}^{m\times m}$, and $\overline{\mathcal{F}}^0 \in \mathbb{V}^{m\times(n-m)}$ be the matrices defined by (2.6)–(2.8) with respect to $\overline{\sigma} = \sigma(\overline{X})$. Since g is absolutely symmetric at $\overline{\sigma}$, we know from [23, Proposition 1] that for all $i \in a_l$, $1 \leq l \leq r$, the function values $g_i(\overline{\sigma})$ are the same (denoted by $\bar{g}_l$). Therefore, for any $X \in \mathcal{N}$, we are able to decompose G into two parts, i.e.,

$$G_S(X) := \sum_{l=1}^{r} \bar{g}_l \mathcal{U}_l(X) \quad \text{and} \quad G_R(X) := G(X) - G_S(X), \tag{2.9}$$

where $\mathcal{U}_l(X) := \sum_{i\in a_l} u_i v_i^{\mathbb{T}}$ with $\mathbb{O}^{m,n}(X)$. It follows from [23, Lemma 1] that there exists an open neighborhood $\mathcal{B}$ of $\overline{X}$ in $\mathcal{N}$ such that G_S is twice continuously differentiable on $\mathcal{B}$, and for any $\mathbb{V}^{m\times n} \ni H \to 0$,

$$G_S(\overline{X} + H) - G_S(\overline{X}) = G_S'(\overline{X})H + O(\|H\|^2) \tag{2.10}$$

with

$$G_S'(\overline{X})H = \overline{U}\left[\overline{\mathcal{E}}_1^0 \circ S\left(\overline{U}^{\mathbb{T}} H \overline{V}_1\right) + \overline{\mathcal{E}}_2^0 \circ T\left(\overline{U}^{\mathbb{T}} H \overline{V}_1\right) \quad \overline{\mathcal{F}}^0 \circ \left(\overline{U}^{\mathbb{T}} H \overline{V}_2\right)\right] \overline{V}^{\mathbb{T}}. \tag{2.11}$$

In other words, in an open neighborhood of $\overline{X}$, G_S can be regarded as a "smooth part" of G and G_R can be regarded as the remaining "nonsmooth part" of G. As we will see in later developments, this decomposition (2.9) can simplify many of our proofs.

3. Lipschitz continuity.

In this section, we analyze the local Lipschitz continuity of the spectral operator G defined on a nonempty open set $\mathcal{N}$. Let $\overline{X} \in \mathcal{N}$ be given. Assume that g is locally Lipschitz continuous near $\overline{\sigma} = \sigma(\overline{X})$ with module $L > 0$. Therefore, there exists a positive constant $\delta_0 > 0$ such that

$$\|g(\sigma) - g(\sigma')\| \leq L\|\sigma - \sigma'\| \quad \forall\, \sigma, \sigma' \in B(\overline{\sigma}, \delta_0) := \{y \in \hat{\sigma}_{\mathcal{N}} \mid \|y - \overline{\sigma}\| \leq \delta_0\}.$$

By using the absolutely symmetric property of g on $\hat{\sigma}_{\mathcal{N}}$, we obtain the following simple observation.

PROPOSITION 3.1. *There exists a positive constant $\delta > 0$ such that for any $\sigma \in B(\overline{\sigma}, \delta)$,*

$$|g_i(\sigma) - g_j(\sigma)| \le \overline{L}|\sigma_i - \sigma_j| \quad \forall\, i, j \in \{1, \ldots, m\},\ i \neq j,\ \ \sigma_i \neq \sigma_j, \tag{3.1}$$

$$|g_i(\sigma) + g_j(\sigma)| \le \overline{L}|\sigma_i + \sigma_j| \quad \forall\, i, j \in \{1, \ldots, m\},\ \ \sigma_i + \sigma_j > 0, \tag{3.2}$$

$$|g_i(\sigma)| \le \overline{L}|\sigma_i| \quad \forall\, i \in \{1, \ldots, m\},\ \ \sigma_i > 0, \tag{3.3}$$

where $\overline{L} := \max\{(2L\delta + \tau)/\delta, \sqrt{2}L\}$ *with* $\tau := \max_{i,j}\{|g_i(\overline{\sigma}) - g_j(\overline{\sigma})|, |g_i(\overline{\sigma}) + g_j(\overline{\sigma})|\, |g_i(\overline{\sigma})|\} \ge 0$.

Proof. It is easy to check that there exists a positive constant $\delta_1 > 0$ such that for any $\sigma \in B(\overline{\sigma}, \delta_1)$,

$$|\sigma_i - \sigma_j| \ge \delta_1 > 0 \quad \forall\, i, j \in \{1, \ldots, m\},\ i \neq j,\ \ \overline{\sigma}_i \neq \overline{\sigma}_j, \tag{3.4}$$

$$|\sigma_i + \sigma_j| \ge \delta_1 > 0 \quad \forall\, i, j \in \{1, \ldots, m\},\ \ \overline{\sigma}_i + \overline{\sigma}_j > 0, \tag{3.5}$$

$$|\sigma_i| \ge \delta_1 > 0 \quad \forall\, i \in \{1, \ldots, m\},\ \ \overline{\sigma}_i > 0. \tag{3.6}$$

Let $\delta := \min\{\delta_0, \delta_1\} > 0$. Denote $L_1 := (2L\delta + \tau)/\delta$. Then, $\overline{L} := \max\{L_1, \sqrt{2}L\}$. Let σ be any fixed vector in $B(\overline{\sigma}, \delta)$.

First, we consider the case that $i, j \in \{1, \ldots, m\}$, $i \neq j$, and $\sigma_i \neq \sigma_j$. If $\overline{\sigma}_i \neq \overline{\sigma}_j$, then by the local Lipschitz continuity of g near $\overline{\sigma}$, we know from (3.4) that

$$\begin{aligned} |g_i(\sigma) - g_j(\sigma)| &= |g_i(\sigma) - g_i(\overline{\sigma}) + g_i(\overline{\sigma}) - g_j(\overline{\sigma}) + g_j(\overline{\sigma}) - g_j(\sigma)| \\ &\le 2\|g(\sigma) - g(\overline{\sigma})\| + \tau \le (2\|g(\sigma) - g(\overline{\sigma})\| + \tau)\frac{|\sigma_i - \sigma_j|}{\delta} \\ &\le \frac{2L\delta + \tau}{\delta}|\sigma_i - \sigma_j| = L_1|\sigma_i - \sigma_j|. \end{aligned} \tag{3.7}$$

If $\overline{\sigma}_i = \overline{\sigma}_j$, define $t \in \mathbb{R}^m$ by

$$t_p := \begin{cases} \sigma_p & \text{if } p \neq i, j, \\ \sigma_j & \text{if } p = i, \\ \sigma_i & \text{if } p = j, \end{cases} \qquad p = 1, \ldots, m.$$

Then, we have $\|t - \overline{\sigma}\| = \|\sigma - \overline{\sigma}\| \le \delta$. Moreover, since g is absolutely symmetric on $\hat{\sigma}_{\mathcal{N}}$, we have $g_i(t) = g_j(\sigma)$. Therefore

$$|g_i(\sigma) - g_j(\sigma)| = |g_i(\sigma) - g_i(t)| \le \|g(\sigma) - g(t)\| \le L\|\sigma - t\| = \sqrt{2}L|\sigma_i - \sigma_j|. \tag{3.8}$$

Thus, the inequality (3.1) follows from (3.7) and (3.8) immediately.

Second, consider the case $i, j \in \{1, \ldots, m\}$ and $\sigma_i + \sigma_j > 0$. If $\overline{\sigma}_i + \overline{\sigma}_j > 0$, it follows from (3.5) and the local Lipschitz continuity of g near $\overline{\sigma}$ that

$$\begin{aligned} |g_i(\sigma) + g_j(\sigma)| &= |g_i(\sigma) - g_i(\overline{\sigma}) + g_i(\overline{\sigma}) + g_j(\overline{\sigma}) - g_j(\overline{\sigma}) + g_j(\sigma)| \\ &\le 2\|g(\sigma) - g(\overline{\sigma})\| + \tau \le (2\|g(\sigma) - g(\overline{\sigma})\| + \tau)\frac{|\sigma_i + \sigma_j|}{\delta} \\ &\le \frac{2L\delta + \tau}{\delta}|\sigma_i + \sigma_j| = L_1|\sigma_i + \sigma_j|. \end{aligned} \tag{3.9}$$

If $\overline{\sigma}_i + \overline{\sigma}_j = 0$, i.e., $\overline{\sigma}_i = \overline{\sigma}_j = 0$, define the vector $\hat{t} \in \mathbb{R}^m$ by

$$\hat{t}_p := \begin{cases} \sigma_p & \text{if } p \neq i, j, \\ -\sigma_j & \text{if } p = i, \\ -\sigma_i & \text{if } p = j, \end{cases} \qquad p = 1, \ldots, m.$$

By noting that $\overline{\sigma}_i = \overline{\sigma}_j = 0$, we obtain that $\|\hat{t} - \overline{\sigma}\| = \|\sigma - \overline{\sigma}\| \leq \delta$. Again, since g is absolutely symmetric on $\hat{\sigma}_{\mathcal{N}}$, we have $g_i(\hat{t}) = -g_j(\sigma)$. Therefore,

$$|g_i(\sigma) + g_j(\sigma)| = |g_i(\sigma) - g_i(\hat{t})| \leq \|g(\sigma) - g(\hat{t})\| \leq L\|\sigma - \hat{t}\| = \sqrt{2}L|\sigma_i + \sigma_j|. \tag{3.10}$$

Thus the inequality (3.2) follows from (3.9) and (3.10).

Finally, we consider the case that $i \in \{1, \ldots, m\}$ and $\sigma_i > 0$. If $\overline{\sigma}_i > 0$, then we know from (3.6) and by the local Lipschitz continuity of g near $\overline{\sigma}$ that

$$\begin{aligned} |g_i(\sigma)| &= |g_i(\sigma) - g_i(\overline{\sigma}) + g_i(\overline{\sigma})| \leq |g_i(\sigma) - g_i(\overline{\sigma})| + |g_i(\overline{\sigma})| \\ &\leq \|g(\sigma) - g(\overline{\sigma})\| + \tau \leq (\|g(\sigma) - g(\overline{\sigma})\| + \tau)\frac{|\sigma_i|}{\delta} \leq \frac{2L\delta + \tau}{\delta}|\sigma_i| \leq L_1|\sigma_i|. \end{aligned} \tag{3.11}$$

If $\overline{\sigma}_i = 0$, define $s \in \mathbb{R}^m$ by

$$s_p := \begin{cases} \sigma_p & \text{if } p \neq i, \\ 0 & \text{if } p = i, \end{cases} \qquad p = 1, \ldots, m.$$

Then, since $\sigma_i > 0$, we know that $\|s - \overline{\sigma}\| < \|\sigma - \overline{\sigma}\| \leq \delta$. Moreover, since g is absolutely symmetric on $\hat{\sigma}_{\mathcal{N}}$, we know that $g_i(s) = 0$. Therefore, we have

$$|g_i(\sigma)| = |g_i(\sigma) - g_i(s)| \leq \|g(\sigma) - g(s)\| \leq L\|\sigma - s\| \leq L|\sigma_i|. \tag{3.12}$$

Thus, the inequality (3.1) follows from (3.11) and (3.12) immediately. This completes the proof. ∎

For any fixed $0 < \omega \leq \delta_0/\sqrt{m}$ and $y \in B(\overline{\sigma}, \delta_0/(2\sqrt{m})) := \{\|y - \overline{\sigma}\|_\infty \leq \delta_0/(2\sqrt{m})\}$, the function g is (vector-valued) integrable on $B(y, \omega/2) = \{z \in \mathbb{R}^m \mid \|y - z\|_\infty \leq \omega/2\}$ (in the sense of Lebesgue). Therefore, we know that the function

$$g(\omega, y) := \frac{1}{\omega^m}\int_{B(y,\omega/2)} g(z)dz \tag{3.13}$$

is well defined on $(0, \delta_0/\sqrt{m}\,] \times B(\overline{\sigma}, \delta_0/(2\sqrt{m}))$ and is said to be the Steklov averaged function [76] of g. For the sake of convenience, we define $g(0, y) = g(y)$. Since g is absolutely symmetric on $\hat{\sigma}_{\mathcal{N}}$, it is easy to check that for any fixed $0 < \omega \leq \delta_0/\sqrt{m}$, the function $g(\omega, \cdot)$ is also absolutely symmetric on $B(\overline{\sigma}, \delta_0/(2\sqrt{m}))$. It follows from the definition (3.13) that $g(\cdot, \cdot)$ is locally Lipschitz continuous on $(0, \delta_0/\sqrt{m}\,] \times B(\overline{\sigma}, \delta_0/(2\sqrt{m}))$ with module L. Meanwhile, by elementary calculations, we know that $g(\cdot, \cdot)$ is continuously differentiable on $(0, \delta_0/\sqrt{m}\,] \times B(\overline{\sigma}, \delta_0/(2\sqrt{m}))$ and for any fixed $\omega \in (0, \delta_0/\sqrt{m}\,]$ and $y \in B(\overline{\sigma}, \delta_0/(2\sqrt{m}))$, $\|g'_y(\omega, y)\| \leq L$. Moreover, it is well known (cf., e.g., [36, Lemma 1]) that $g(\omega, \cdot)$ converges to g uniformly on the compact set $B(\overline{\sigma}, \delta_0/(2\sqrt{m}))$ as $\omega \downarrow 0$. By using the derivative formula of spectral operators obtained in [23, Theorem 4, (38)] and Proposition 3.1, we can obtain a uniform approximation to a locally Lipschitz spectral operator through the Steklov averaged function (3.13) from [23, Theorem 4], directly. For simplicity, we omit the detailed proof here.

PROPOSITION 3.2. *Suppose that g is locally Lipschitz continuous near $\overline{\sigma}$ with module L. Let $g(\cdot, \cdot)$ be the corresponding Steklov averaged function defined in* (3.13). *Then, for any given $\omega \in (0, \delta_0/\sqrt{m}\,]$, the spectral operator $G(\omega, \cdot)$ with respect to $g(\omega, \cdot)$ is continuously differentiable on $B(\overline{X}, \delta_0/(2\sqrt{m})) := \{X \in \mathcal{X} \mid \|\sigma(X) - \overline{\sigma}\|_\infty \leq \delta_0/(2\sqrt{m})\}$, and there exists a positive constant $\delta_2 > 0$ such that*

$$\|G'(\omega, X)\| \le \overline{L} \quad \forall\, 0 < \omega \le \min\{\delta_0/\sqrt{m}, \delta_2\} \text{ and } X \in B(\overline{X}, \delta_0/(2\sqrt{m})), \tag{3.14}$$

where $\overline{L} > 0$ *is the positive constant defined in Proposition* 3.1. *Moreover,* $G(\omega, \cdot)$ *converges to* G *uniformly in the compact set* $B(\overline{X}, \delta_0/(2\sqrt{m}))$ *as* $\omega \downarrow 0$.

Proposition 3.2 allows us to derive the following result on the local Lipschitz continuity of spectral operators.

THEOREM 3.3. *Suppose that* $\overline{X}$ *has the SVD* (2.2). *If* g *is locally Lipschitz continuous near* $\overline{\sigma} = \sigma(\overline{X})$ *with modulus* $L > 0$, *then the spectral operator* G *is locally Lipschitz continuous near* $\overline{X}$ *with modulus* $\overline{L} > 0$, *where* $\overline{L} > 0$ *is the positive constant defined in Proposition* 3.1. *Conversely, if* G *is locally Lipschitz continuous near* $\overline{X}$ *with modulus* $\overline{L} > 0$, *then* g *is locally Lipschitz continuous near* $\overline{\sigma}$ *with the same modulus.*

Proof. Suppose that g is locally Lipschitz continuous near $\overline{\sigma} = \sigma(\overline{X})$ with modulus $L > 0$, i.e., there exists a positive constant $\delta_0 > 0$ such that

$$\|g(\sigma) - g(\sigma')\| \le L\|\sigma - \sigma'\| \quad \forall\, \sigma, \sigma' \in B(\overline{\sigma}, \delta_0).$$

By Proposition 3.2, for any $\omega \in (0, \delta_0/\sqrt{m}\,]$, the spectral operator $G(\omega, \cdot)$ defined with respect to the Steklov averaged function $g(\omega, \cdot)$ is continuously differentiable. Since $G(\omega, \cdot)$ converges to G uniformly in the compact set $B(\overline{X}, \delta_0/(2\sqrt{m}))$ as $\omega \downarrow 0$, we know that for any $\varepsilon > 0$, there exists a constant $\delta_2 > 0$ such that for any $0 < \omega \le \delta_2$,

$$\|G(\omega, X) - G(X)\| \le \varepsilon \quad \forall\, X \in B(\overline{X}, \delta_0/(2\sqrt{m})).$$

Fix any $X, X' \in B(\overline{X}, \delta_0/(2\sqrt{m}))$ with $X \ne X'$. By Proposition 3.2, we know that there exists $\delta_1 > 0$ such that (3.14) holds. Let $\bar{\delta} := \min\{\delta_1, \delta_2, \delta_0/\sqrt{m}\}$. Then, by the mean value theorem, we know that

$$\begin{aligned}\|G(X) - G(X')\| &= \|G(X) - G(\omega, X) + G(\omega, X) - G(\omega, X') + G(\omega, X') - G(X')\| \\ &\le \left\| \int_0^1 G'(\omega, X + t(X - X'))dt \right\| + 2\varepsilon \le \overline{L} \| X - X' \| + 2\varepsilon \quad \forall\, 0 < \omega < \bar{\delta},\end{aligned}$$

where $\overline{L} > 0$ is the positive constant defined in Proposition 3.1. Since $X, X' \in B(\overline{X}, \delta_0/(2\sqrt{m}))$ and $\varepsilon > 0$ are arbitrary, by letting $\varepsilon \downarrow 0$, we obtain that

$$\|G(X) - G(X')\| \le \overline{L}\|X - X'\| \quad \forall\, X, X' \in B(\overline{X}, \delta_0/(2\sqrt{m})).$$

Thus G is locally Lipschitz continuous near $\overline{X}$.

Conversely, suppose that G is locally Lipschitz continuous near $\overline{X}$ with modulus $\overline{L} > 0$, i.e., there exists an open neighborhood $\mathcal{B}$ of $\overline{X}$ in $\mathcal{N}$ such that for any $X, X' \in \mathcal{B}$,

$$\|G(X) - G(X')\| \le \overline{L}\|X - X'\|.$$

Let $(\overline{U}, \overline{V}) \in \mathbb{O}^{m \times n}(\overline{X})$ be fixed. For any $y \in \hat{\sigma}_{\mathcal{N}}$, we define $Y := \overline{U}\,[\mathrm{Diag}(y) \quad 0]\,\overline{V}^{\mathbb{T}}$. Then, we know from [23, Proposition 3] that $G(Y) = \overline{U}\,[\mathrm{Diag}(g(y)) \quad 0]\,\overline{V}^{\mathbb{T}}$. Therefore, we obtain that there exists an open neighborhood $\mathcal{B}_{\overline{\sigma}}$ of $\overline{\sigma}$ in $\hat{\sigma}_{\mathcal{N}}$ such that

$$\|g(y) - g(y')\| = \|G(Y) - G(Y')\| \le \overline{L}\|Y - Y'\| = \overline{L}\|y - y'\| \quad \forall\, y, y' \in \mathcal{B}_{\overline{\sigma}}.$$

This completes the proof. $\square$

It seems that the smoothing function approach employed above may not be extended to the Hölder continuity of the general spectral operator, since we may not be able to find a uniform approximation to a Hölder continuous function through some smoothing functions (see [86, p. 144] and [7, Theorem 2] for counterexamples). However, if $\mathcal{X} \equiv \mathbb{S}^m$ (or $\mathbb{V}^{m\times n}$), then the Hölder continuity with exponent $0 < q < 1$ of the Löwner operator with respect to the scalar-valued function is studied in [87, Theorem 1.1].

4. Bouligand-differentiability. In this section, we shall study the ρ-order Bouligand-differentiability of spectral operators with $0 < \rho \leq 1$, which is a slightly stronger property than the directional differentiability studied in [23, Theorem 3].

Let $\mathcal{Z}$ be a finite dimensional real Euclidean space equipped with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$. Let $\mathcal{O}$ be an open set in $\mathcal{Z}$ and let $\mathcal{Z}'$ be another finite dimensional real Euclidean space. The function $F : \mathcal{O} \subseteq \mathcal{Z} \to \mathcal{Z}'$ is said to be *B(ouligand)-differentiable* [73] (see also [65, 31, 66] for more details) at $z \in \mathcal{O}$ if for any $h \in \mathcal{Z}$ with $h \to 0$,

$$F(z+h) - F(z) - F'(z;h) = o(\|h\|).$$

It is well known (cf. [75]) that if F is locally Lipschitz continuous, then F is B-differentiable at $z \in \mathcal{O}$ if and only if F is directionally differentiable at z. If the spectral operator G is directionally differentiable, then the corresponding directional derivative formula is presented in [23, Theorem 3, (21)]. More precisely, since g is absolutely symmetric on the nonempty open set $\hat{\sigma}_{\mathcal{N}}$, it is easy to see that the directional derivative $\phi := g'(\overline{\sigma}; \cdot) : \mathbb{R}^m \to \mathbb{R}^m$ satisfies

$$g'(\overline{\sigma}; Qh) = Qg'(\overline{\sigma}; h) \quad \forall\, Q \in \pm\mathbb{P}^m_{\overline{\sigma}} \quad \text{and} \quad \forall\, h \in \mathbb{R}^m\,, \tag{4.1}$$

where $\pm\mathbb{P}^m_{\overline{\sigma}}$ is the subset defined with respect to $\overline{\sigma}$ by $\pm\mathbb{P}^m_{\overline{\sigma}} := \{Q \in \pm\mathbb{P}^m \mid \overline{\sigma} = Q\overline{\sigma}\}$. Thus, we know that the function ϕ is a mixed symmetric mapping, with respect to $\mathbb{P}^{|a_1|}\times\cdots\times\mathbb{P}^{|a_r|}\times\pm\mathbb{P}^{|b|}$, over $\mathcal{V} := \mathbb{R}^{|a_1|}\times\cdots\times\mathbb{R}^{|a_r|}\times\mathbb{R}^{|b|}$. Let $\Psi := G'(\overline{X}; \cdot) : \mathbb{V}^{m\times n} \to \mathbb{V}^{m\times n}$ be the directional derivative of G at $\overline{X}$. Let $\mathcal{W} := \mathbb{S}^{|a_1|}\times\cdots\times\mathbb{S}^{|a_r|}\times\mathbb{V}^{|b|\times(n-|a|)}$. We know from [23, (21) Theorem 3] that for any $H \in \mathbb{V}^{m\times n}$,

$$\begin{aligned}\Psi(H) = G'(\overline{X}; H) &= \overline{U}\left[\overline{\mathcal{E}}^0_1 \circ S\left(\overline{U}^{\mathbb{T}} H \overline{V}_1\right) + \overline{\mathcal{E}}^0_2 \circ T\left(\overline{U}^{\mathbb{T}} H \overline{V}_1\right) \quad \overline{\mathcal{F}}^0 \circ \overline{U}^{\mathbb{T}} H \overline{V}_2\right]\overline{V}^{\mathbb{T}} \\ &\quad + \overline{U}\widehat{\Phi}(D(H))\overline{V}^{\mathbb{T}} \\ &= G'_S(\overline{X}; H) + \overline{U}\widehat{\Phi}(D(H))\overline{V}^{\mathbb{T}},\end{aligned} \tag{4.2}$$

where $D(H) = (S(\widetilde{H}_{a_1a_1}), \ldots, S(\widetilde{H}_{a_ra_r}), \widetilde{H}_{b\bar{a}}) \in \mathcal{W}$, $\widetilde{H} = \overline{U}^{\mathbb{T}} H \overline{V}$, $\Phi : \mathcal{W} \to \mathcal{W}$ is the spectral operator defined with respect to the mixed symmetric mapping $\phi = g'(\overline{\sigma}; \cdot)$, and $\widehat{\Phi} : \mathcal{W} \to \mathbb{V}^{m\times n}$ is defined by

$$\widehat{\Phi}(W) := \begin{bmatrix} \operatorname{Diag}(\Phi_1(W), \ldots, \Phi_r(W)) & 0 \\ 0 & \Phi_{r+1}(W) \end{bmatrix} \quad \forall\, W \in \mathcal{W}. \tag{4.3}$$

A stronger notion than B-differentiability is ρ-order B-differentiability with $\rho > 0$. The function $F : \mathcal{O} \subseteq \mathcal{Z} \to \mathcal{Z}'$ is said to be *ρ-order B-differentiable* at $z \in \mathcal{O}$ if for any $h \in \mathcal{Z}$ with $h \to 0$,

$$F(z+h) - F(z) - F'(z;h) = O(\|h\|^{1+\rho}).$$

Let $\overline{X} \in \mathbb{V}^{m\times n}$ be given. We have the following results on the ρ-order B-differentiability of spectral operators.

THEOREM 4.1. *Suppose that $\overline{X} \in \mathcal{N}$ has the SVD (2.2). Let $0 < \rho \le 1$ be given.*
(i) *If g is locally Lipschitz continuous near $\sigma(\overline{X})$ and ρ-order B-differentiable at $\sigma(\overline{X})$, then G is ρ-order B-differentiable at $\overline{X}$.*
(ii) *If G is ρ-order B-differentiable at $\overline{X}$, then g is ρ-order B-differentiable at $\sigma(\overline{X})$.*

Proof. Without loss of generality, we only prove the results for the case that $\rho = 1$.

(i) For any $H \in \mathbb{V}^{m\times n}$, denote $X = \overline{X} + H$. Let $U \in \mathbb{O}^m$ and $V \in \mathbb{O}^n$ be such that

$$X = U[\Sigma(X) \quad 0]V^{\mathbb{T}}. \tag{4.4}$$

Denote $\sigma = \sigma(X)$. Let $G_S(X)$ and $G_R(X)$ be defined by (2.9). Therefore, by (2.10), we know that for any $H \to 0$,

$$G_S(X) - G_S(\overline{X}) = G_S'(\overline{X})H + O(\|H\|^2), \tag{4.5}$$

where $G_S'(\overline{X})H$ is given by (2.11). For $H \in \mathbb{V}^{m\times n}$ sufficiently small, we have $\mathcal{U}_l(X) = \sum_{i\in a_l} u_i v_i^{\mathbb{T}}$, $l = 1, \ldots, r$. Therefore, we know that

$$G_R(X) = G(X) - G_S(X) = \sum_{l=1}^{r+1} \Delta_l(H), \tag{4.6}$$

where

$$\Delta_l(H) = \sum_{i\in a_l} (g_i(\sigma) - g_i(\overline{\sigma}))u_i v_i^{\mathbb{T}}, \quad l = 1, \ldots, r, \quad \text{and} \quad \Delta_{r+1}(H) = \sum_{i\in b} g_i(\sigma)u_i v_i^{\mathbb{T}}.$$

(a) We first consider the case that $\overline{X} = [\Sigma(\overline{X}) \quad 0]$. Then, we know from the directional differentiability of single values (cf., e.g., [43, Theorem 7], [84, Proposition 1.4], and [49, section 5.1]) that for any H sufficiently small,

$$\sigma = \overline{\sigma} + \sigma'(\overline{X}; H) + O(\|H\|^2), \tag{4.7}$$

where $\sigma'(\overline{X}; H) = (\lambda(S(H_{a_1a_1})), \ldots, \lambda(S(H_{a_ra_r})), \sigma([H_{bb} \quad H_{bc}])) \in \mathbb{R}^m$. Denote $h := \sigma'(\overline{X}; H)$. Since g is locally Lipschitz continuous near $\overline{\sigma}$ and 1-order B-differentiable at $\overline{\sigma}$, we know that for any H sufficiently small,

$$\begin{aligned} g(\sigma) - g(\overline{\sigma}) &= g(\overline{\sigma} + h + O(\|H\|^2)) - g(\overline{\sigma}) = g(\overline{\sigma} + h) - g(\overline{\sigma}) + O(\|H\|^2) \\ &= g'(\overline{\sigma}; h) + O(\|H\|^2). \end{aligned}$$

Let $\phi = g'(\overline{\sigma}; \cdot)$. Since $u_i v_i^{\mathbb{T}}$, $i = 1, \ldots, m$, are uniformly bounded, we obtain that for H sufficiently small,

$$\begin{aligned} \Delta_l(H) &= U_{a_l}\mathrm{Diag}(\phi_l(h))V_{a_l}^{\mathbb{T}} + O(\|H\|^2), \quad l = 1, \ldots, r, \\ \Delta_{r+1}(H) &= U_b\mathrm{Diag}(\phi_{r+1}(h))V_b^{\mathbb{T}} + O(\|H\|^2). \end{aligned}$$

Again, we know from [24, Proposition 7] that there exist $Q_l \in \mathbb{O}^{|a_l|}$, $M \in \mathbb{O}^{|b|}$, and $N = [N_1 \quad N_2] \in \mathbb{O}^{n-|a|}$ with $N_1 \in \mathbb{V}^{(n-|a|)\times|b|}$ and $N_2 \in \mathbb{V}^{(n-|a|)\times(n-m)}$ (depending on H) such that

$$U_{a_l} = \begin{bmatrix} O(\|H\|) \\ Q_l + O(\|H\|) \\ O(\|H\|) \end{bmatrix}, \quad V_{a_l} = \begin{bmatrix} O(\|H\|) \\ Q_l + O(\|H\|) \\ O(\|H\|) \end{bmatrix}, \ l = 1, \ldots, r,$$

$$U_b = \begin{bmatrix} O(\|H\|) \\ M + O(\|H\|) \end{bmatrix}, \quad [V_b \quad V_c] = \begin{bmatrix} O(\|H\|) \\ N + O(\|H\|) \end{bmatrix}.$$

Since g is locally Lipschitz continuous near $\overline{\sigma}$ and directionally differentiable at $\overline{\sigma}$, we know from [73, Theorem A.2] or [70, Lemma 2.2] that the directional derivative ϕ is globally Lipschitz continuous on $\mathbb{R}^m$. Thus, for H sufficiently small, we have $\|\phi(h)\| = O(\|H\|)$. Therefore, we obtain that

$$\Delta_l(H) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & Q_l \mathrm{Diag}(\phi_l(h)) Q_l^{\mathbb{T}} & 0 \\ 0 & 0 & 0 \end{bmatrix} + O(\|H\|^2), \quad l = 1, \dots, r, \tag{4.8}$$

$$\Delta_{r+1}(H) = \begin{bmatrix} 0 & 0 \\ 0 & M \mathrm{Diag}(\phi_{r+1}(h)) N_1^{\mathbb{T}} \end{bmatrix} + O(\|H\|^2). \tag{4.9}$$

Again, it follows from [24, Proposition 7] that

$$S(H_{a_l a_l}) = Q_l (\Sigma(X)_{a_l a_l} - \overline{\nu}_l I_{|a_l|}) Q_l^{\mathbb{T}} + O(\|H\|^2), \quad l = 1, \dots, r, \tag{4.10}$$

$$[H_{bb} \quad H_{bc}] = M(\Sigma(X)_{bb} - \overline{\nu}_{r+1} I_{|b|}) N_1^{\mathbb{T}} + O(\|H\|^2). \tag{4.11}$$

Since g is locally Lipschitz continuous near $\overline{\sigma} = \sigma(\overline{X})$, we know from Theorem 3.3 that the spectral operator G is locally Lipschitz continuous near $\overline{X}$. Therefore, we know from [23, Theorem 3 and Remark 1] that G is directionally differentiable at $\overline{X}$. Thus, from [73, Theorem A.2] or [70, Lemma 2.2], we know that $G'(\overline{X}, \cdot)$ is globally Lipschitz continuous on $\mathbb{V}^{m\times n}$. Moreover, from the definition of directional derivative and the absolute symmetry of g on the nonempty open set $\hat{\sigma}_{\mathcal{N}}$, it is easy to see that the directional derivative $\phi := g'(\overline{\sigma}; \cdot)$ is actually a mixed symmetric mapping over the space $\mathcal{V} := \mathbb{R}^{|a_1|} \times \cdots \times \mathbb{R}^{|a_r|} \times \mathbb{R}^{|b|}$. Let $\mathcal{W} := \mathbb{S}^{|a_1|} \times \cdots \times \mathbb{S}^{|a_r|} \times \mathbb{V}^{|b| \times (n-|a|)}$. Thus, the corresponding spectral operator Φ defined with respect to ϕ is globally Lipschitz continuous on the space $\mathcal{W}$. Hence, we know from (4.6) that for H sufficiently small,

$$G_R(X) = \widehat{\Phi}(D(H)) + O(\|H\|^2), \tag{4.12}$$

where $D(H) = (S(H_{a_1 a_1}), \dots, S(H_{a_r a_r}), H_{b\bar{a}}) \in \mathcal{W}$ and $\widehat{\Phi}$ is defined by (4.3)

(b) Next, consider the general case that $\overline{X} \in \mathbb{V}^{m\times n}$. For any $H \in \mathbb{V}^{m\times n}$, we rewrite (4.4) by using the SVD of $\overline{X}$ as follows: $\widetilde{X} := [\Sigma(\overline{X}) \quad 0] + \overline{U}^{\mathbb{T}} H \overline{V} = \overline{U}^{\mathbb{T}} U [\Sigma(X) \quad 0] V^{\mathbb{T}} \overline{V}$. Then, since $\overline{U}$ and $\overline{V}$ are unitary matrices, we know from (4.12) that

$$G_R(X) = \overline{U} G_R(\widetilde{X}) \overline{V}^{\mathbb{T}} = \overline{U} \widehat{\Phi}(D(H)) \overline{V}^{\mathbb{T}} + O(\|H\|^2), \tag{4.13}$$

where $D(H) = (S(\widetilde{H}_{a_1 a_1}), \dots, S(\widetilde{H}_{a_r a_r}), \widetilde{H}_{b\bar{a}})$ and $\widetilde{H} = \overline{U}^{\mathbb{T}} H \overline{V}$. Thus, by combining (4.2), (4.5), and (4.13) and noting that $G(\overline{X}) = G_S(\overline{X})$, we obtain that for any $H \in \mathbb{V}^{m\times n}$ sufficiently close to 0,

$$\begin{aligned} & G(X) - G(\overline{X}) - G'(\overline{X}; H) \\ & \quad = G_R(X) + G_S(X) - G_S(\overline{X}) - G'(\overline{X}; H) \\ & \quad = G_R(X) - \overline{U} \widehat{\Phi}(D(H)) \overline{V}^{\mathbb{T}} + O(\|H\|^2) = O(\|H\|^2), \end{aligned}$$

where the directional derivative $G'(\overline{X}; H)$ of G at $\overline{X}$ along H is given by (4.2). This implies that G is 1-order B-differentiable at $\overline{X}$.

(ii) Suppose that G is 1-order B-differentiable at $\overline{X}$. Let $(\overline{U},\overline{V}) \in \mathbb{O}^{m\times n}(\overline{X})$ be fixed. For any $h \in \mathbb{R}^m$, let $H = \overline{U}[\mathrm{Diag}(h) \quad 0]\overline{V}^{\mathbb{T}} \in \mathbb{V}^{m\times n}$. We know from [23, Proposition 3] that for all h sufficiently close to 0, $G(\overline{X}+H) = \overline{U}\mathrm{Diag}(g(\overline{\sigma}+h))\overline{V}_1^{\mathbb{T}}$. Therefore, we know from the assumption that

$$\mathrm{Diag}(g(\overline{\sigma}+h)-g(\overline{\sigma})) = \overline{U}^{\mathbb{T}}\left(G(\overline{X}+H)-G(\overline{X})\right)\overline{V}_1 = \overline{U}^{\mathbb{T}}G'(\overline{X};H)\overline{V}_1 + O(\|H\|^2).$$

This shows that g is 1-order B-differentiable at $\overline{\sigma}$. The proof is completed. □

5. G-semismoothness. Let $\mathcal{Z}$ and $\mathcal{Z}'$ be two finite dimensional real Euclidean spaces and $\mathcal{O}$ be an open set in $\mathcal{Z}$. Suppose that $F:\mathcal{O}\subseteq\mathcal{Z}\to\mathcal{Z}'$ is a locally Lipschitz continuous function on $\mathcal{O}$. Then, according to Rademacher's theorem, F is almost everywhere differentiable (in the sense of Fréchet) in $\mathcal{O}$. Let $\mathcal{D}_F$ be the set of points in $\mathcal{O}$ where F is differentiable. Let $F'(z)$ be the derivative of F at $z\in\mathcal{D}_F$. Then the *B(ouligand)-subdifferential* of F at $z\in\mathcal{O}$ is denoted by [69],

$$\partial_B F(z) := \left\{ \lim_{\mathcal{D}_F \ni z^k \to z} F'(z^k) \right\},$$

and the *Clarke generalized Jacobian* of F at $z\in\mathcal{O}$ [16] takes the form

$$\partial F(z) = \mathrm{conv}\{\partial_B F(z)\},$$

where "conv" stands for the convex hull in the usual sense of convex analysis [74]. The function F is said to be G-semismooth at a point $z\in\mathcal{O}$ if for any $y\to z$ and $V\in\partial F(y)$,

$$F(y)-F(z)-V(y-z) = o(\|y-z\|).$$

A stronger notion than G-semismoothness is ρ-order G-semismoothness with $\rho>0$. The function F is said to be ρ-order G-semismooth at z if for any $y\to z$ and $V\in\partial F(y)$,

$$F(y)-F(z)-V(y-z) = O(\|y-z\|^{1+\rho}).$$

In particular, the function F is said to be strongly G-semismooth at z if F is 1-order G-semismooth at z. Furthermore, the function F is said to be (ρ-order, strongly) semismooth at $z\in\mathcal{O}$ if (i) the directional derivative of F at z along any direction $d\in\mathcal{Z}$, denoted by $F'(z;d)$, exists; and (ii) F is (ρ-order, strongly) G-semismooth.

The following result taken from [78, Theorem 3.7] provides a convenient tool for proving the G-semismoothness of Lipschitz functions.

LEMMA 5.1. *Let $F:\mathcal{O}\subseteq\mathcal{Z}\to\mathcal{Z}'$ be a locally Lipschitz continuous function on the open set $\mathcal{O}$, and let $\rho>0$ be a constant. F is ρ-order G-semismooth (G-semismooth) at z if and only if for any $\mathcal{D}_F\ni y\to z$,*

$$F(y)-F(z)-F'(y)(y-z) = O(\|y-z\|^{1+\rho}) \quad \left(=o(\|y-z\|)\right). \tag{5.1}$$

Let $\overline{X}\in\mathcal{N}$ be given. Assume that g is locally Lipschitz continuous near $\overline{\sigma}=\sigma(\overline{X})$. Then from Theorem 3.3 we know that the corresponding spectral operator G is locally Lipschitz continuous near $\overline{X}$. The following theorem is on the G-semismoothness of the spectral operator G.

THEOREM 5.2. *Suppose that $\overline{X}\in\mathcal{N}$ has the SVD (2.2). Let $0<\rho\le 1$ be given. G is ρ-order G-semismooth at $\overline{X}$ if and only if g is ρ-order G-semismooth at $\overline{\sigma}$.*

Proof. Without loss of generality, we only prove the result for the case that $\rho = 1$.

" $\Longleftarrow$ " For any $H \in \mathbb{V}^{m\times n}$, denote $X = \overline{X} + H$. Let $U \in \mathbb{O}^m$ and $V \in \mathbb{O}^n$ be such that

$$X = U[\Sigma(X) \quad 0]V^{\mathbb{T}}. \tag{5.2}$$

Denote $\sigma = \sigma(X)$. Recall the mappings G_S and G_R defined in (2.9). We know from [24, Proposition 8] that there exists an open neighborhood $\mathcal{B} \subseteq \mathcal{N}$ of $\overline{X}$ such that G_S twice continuously differentiable on $\mathcal{B}$ and

$$\begin{aligned} G_S(X) - G_S(\overline{X}) &= \sum_{l=1}^{r} \bar{g}_l\, \mathcal{U}_l'(X)\, H + O(\|H\|^2) \\ &= \sum_{l=1}^{r} \bar{g}_l \left\{ U[\Gamma_l(X) \circ S(U^{\mathbb{T}}HV_1) + \Xi_l(X) \circ T(U^{\mathbb{T}}HV_1)]V_1^{\mathbb{T}} \right. \\ &\qquad \left. + U(\Upsilon_l(X) \circ U^{\mathbb{T}}HV_2)V_2^{\mathbb{T}} \right\} + O(\|H\|^2), \end{aligned} \tag{5.3}$$

where for each $l \in \{1, \ldots, r\}$, $\Gamma_l(X)$, $\Xi_l(X)$ and $\Upsilon_l(X)$ are given by [24, (40)–(42)], respectively. By taking a smaller $\mathcal{B}$ if necessary, we may assume that for any $X \in \mathcal{B}$ and $l, l' \in \{1, \ldots, r\}$,

$$\sigma_i(X) > 0, \quad \sigma_i(X) \neq \sigma_j(X) \quad \forall\, i \in a_l,\ j \in a_{l'} \text{ and } l \neq l'. \tag{5.4}$$

Since g is locally Lipschitz continuous near $\overline{\sigma}$, we know that for any H sufficiently small,

$$\bar{g}_l = g_i(\sigma) + O(\|H\|) \quad \forall\, i \in a_l, \quad l = 1, \ldots, r. \tag{5.5}$$

By noting that $U \in \mathbb{O}^m$ and $V \in \mathbb{O}^n$ are uniformly bounded, we know from (5.3) and (5.5) that for any $X \in \mathcal{B}$ (shrinking $\mathcal{B}$ if necessary),

$$\begin{aligned} G_S(X) - G_S(\overline{X}) = U &\left[\mathcal{E}_1^0 \circ S\left(U^{\mathbb{T}}HV_1\right) + \mathcal{E}_2^0 \circ T\left(U^{\mathbb{T}}HV_1\right) \quad \mathcal{F}^0 \circ U^{\mathbb{T}}HV_2\right] V^{\mathbb{T}} \\ &+ O(\|H\|^2), \end{aligned} \tag{5.6}$$

where $\mathcal{E}_1^0$, $\mathcal{E}_2^0$, and $\mathcal{F}^0$ are the corresponding real matrices defined in (2.6)–(2.8) (depending on X), respectively.

Let $X \in \mathcal{D}_G \cap \mathcal{B}$, where $\mathcal{D}_G$ is the set of points in $\mathbb{V}^{m\times n}$ for which G is (F-)differentiable. Define the corresponding index sets in $\{1, \ldots, m\}$ for X by $a' := \{i \mid \sigma_i(X) > 0\}$ and $b' := \{i \mid \sigma_i(X) = 0\}$. By (5.4), we have

$$a' \supseteq a \quad \text{and} \quad b' \subseteq b. \tag{5.7}$$

We know from [23, Theorem 4] that

$$\begin{aligned} G'(X)H = U &\left[\mathcal{E}_1 \circ S\left(U^{\mathbb{T}}HV_1\right) + \mathcal{E}_2 \circ T\left(U^{\mathbb{T}}HV_1\right) + \mathrm{Diag}\left(\mathcal{C}\mathrm{diag}\left(S(U^{\mathbb{T}}HV_1)\right)\right),\right. \\ &\left. \mathcal{F} \circ U^{\mathbb{T}}HV_2\right] V^{\mathbb{T}}, \end{aligned} \tag{5.8}$$

where η, $\mathcal{E}_1$, $\mathcal{E}_2$, $\mathcal{F}$, and $\mathcal{C}$ are defined by [23, (33)–(36)] with respect to σ, respectively. Denote $\Delta(H) := G'(X)H - (G_S(X) - G_S(\overline{X}))$. Moreover, since there exists an integer $j \in \{0, \ldots, |b|\}$ such that $|a'| = |a| + j$, we can define two index sets $b_1 :=$

$\{|a|+1,\ldots,|a|+j\}$ and $b_2 := \{|a|+j+1,\ldots,|a|+|b|\}$ such that $a' = a \cup b_1$ and $b' = b_2$. From (5.6) and (5.8), we obtain that

$$\Delta(H) = U\widehat{R}(H)V^{\mathbb{T}} + O(\|H\|^2), \tag{5.9}$$

where $\widehat{R}(H) \in \mathbb{V}^{m\times n}$ is defined by

$$\widehat{R}(H) := \begin{bmatrix} \mathrm{Diag}\,(R_1(H),\ldots,R_r(H)) & 0 \\ 0 & R_{r+1}(H) \end{bmatrix},$$

(5.10)
$$R_l(H) = (\mathcal{E}_1)_{a_l a_l} \circ S\left(U_{a_l}^{\mathbb{T}} H V_{a_l}\right) + \mathrm{Diag}\left((\mathcal{C}\mathrm{diag}\left(S(U^{\mathbb{T}}HV_1)\right))_{a_l a_l}\right),\ l = 1,\ldots,r,$$

(5.11)
$$R_{r+1}(H) = \begin{bmatrix} (\mathcal{E}_1)_{b_1b_1} \circ S(U_{b_1}^{\mathbb{T}} H V_{b_1}) + \mathrm{Diag}\left((\mathcal{C}\mathrm{diag}(S(U^{\mathbb{T}}HV_1)))_{b_1b_1}\right) & 0 & 0 \\ 0 & \gamma U_{b_2}^{\mathbb{T}} H V_{b_2} & \gamma U_{b_2}^{\mathbb{T}} H V_2 \end{bmatrix},$$

and $\gamma := (g'(\sigma))_{ii}$ for any $i \in b_2$. By (2.2), we obtain from (5.2) that

$$\left[\Sigma(\overline{X}) \quad 0\right] + \overline{U}^{\mathbb{T}} H \overline{V} = \overline{U}^{\mathbb{T}} U\left[\Sigma(X) \quad 0\right] V^{\mathbb{T}}\overline{V}.$$

Let $\widehat{H} := \overline{U}^{\mathbb{T}} H\overline{V}$, $\widehat{U} := \overline{U}^{\mathbb{T}} U$, and $\widehat{V} := \overline{V}^{\mathbb{T}} V$. Then, $U^{\mathbb{T}}HV = \widehat{U}^{\mathbb{T}}\overline{U}^{\mathbb{T}}H\overline{V}\widehat{V} = \widehat{U}^{\mathbb{T}}\widehat{H}\widehat{V}$. We know from [24, Proposition 7, (31)] that there exist $Q_l \in \mathbb{O}^{|a_l|}$, $l = 1,\ldots,r$, and $M \in \mathbb{O}^{|b|}$, $N \in \mathbb{O}^{n-|a|}$ such that

$$U_{a_l}^{\mathbb{T}} H V_{a_l} = \widehat{U}_{a_l}^{\mathbb{T}} \widehat{H} \widehat{V}_{a_l} = Q_l^{\mathbb{T}} \widehat{H}_{a_l a_l} Q_l + O(\|H\|^2), \quad l = 1,\ldots,r,$$
$$\left[U_b^{\mathbb{T}} H V_b \quad U_b^{\mathbb{T}} H V_2\right] = \left[\widehat{U}_b^{\mathbb{T}} \widehat{H} \widehat{V}_b \quad \widehat{U}_b^{\mathbb{T}} \widehat{H} \widehat{V}_2\right] = M^{\mathbb{T}}\left[\widehat{H}_{bb} \quad \widehat{H}_{bc}\right] N + O(\|H\|^2).$$

Moreover, from [24, Proposition 7, (32)–(33)], we obtain that

$$\begin{aligned} S(U_{a_l}^{\mathbb{T}} H V_{a_l}) &= Q_l^{\mathbb{T}} S(\widehat{H}_{a_l a_l}) Q_l + O(\|H\|^2) \\ &= \Sigma(X)_{a_l a_l} - \Sigma(\overline{X})_{a_l a_l} + O(\|H\|^2), \quad l = 1,\ldots,r, \\ \left[U_b^{\mathbb{T}} H V_b \quad U_b^{\mathbb{T}} H V_2\right] &= M^{\mathbb{T}}\left[\widehat{H}_{bb} \quad \widehat{H}_{bc}\right] N = \left[\Sigma(X)_{bb} - \Sigma(\overline{X})_{bb} \quad 0\right] + O(\|H\|^2). \end{aligned}$$

Denote $h = \sigma'(X;H) \in \mathbb{R}^m$. Since the singular value functions are strongly semismooth [79], we know that

$$\begin{aligned} S(U_{a_l}^{\mathbb{T}} H V_{a_l}) &= \mathrm{Diag}(h_{a_l}) + O(\|H\|^2), \quad l = 1,\ldots,r, \\ S(U_{b_1}^{\mathbb{T}} H V_{b_1}) &= \mathrm{Diag}(h_{b_1}) + O(\|H\|^2), \\ \left[U_{b_2}^{\mathbb{T}} H V_{b_2} \quad U_{b_2}^{\mathbb{T}} H V_2\right] &= \left[\mathrm{Diag}(h_{b_2}) \quad 0\right] + O(\|H\|^2). \end{aligned}$$

Therefore, since $\mathcal{C} = g'(\sigma) - \mathrm{Diag}(\eta)$, by (5.10) and (5.11), we obtain from (5.9) that

$$\Delta(H) = U\left[\mathrm{Diag}\,(g'(\sigma)h) \quad 0\right] V^{\mathbb{T}} + O(\|H\|^2). \tag{5.12}$$

On the other hand, for X sufficiently close to $\overline{X}$, we have $\mathcal{U}_l(X) = \sum_{i\in a_l} u_i v_i^{\mathbb{T}}$, $l = 1,\ldots,r$. Therefore,

$$G_R(X) = G(X) - G_S(X) = \sum_{l=1}^{r}\sum_{i\in a_l}[g_i(\sigma) - g_i(\overline{\sigma})]u_i v_i^{\mathbb{T}} + \sum_{i\in b} g_i(\sigma) u_i v_i^{\mathbb{T}}. \tag{5.13}$$

Note that by definition, $G_R(\overline{X}) = 0$. We know from [23, Theorem 4] that G is differentiable at X if and only if g is differentiable at σ. Since g is 1-order G-semismooth at $\overline{\sigma}$ and $\sigma(\cdot)$ is strongly semismooth, we obtain that for any $X \in \mathcal{D}_G \cap \mathcal{B}$ (shrinking $\mathcal{B}$ if necessary),

$$\begin{aligned} g(\sigma) - g(\overline{\sigma}) &= g'(\sigma)(\sigma - \overline{\sigma}) + O(\|H\|^2) = g'(\sigma)(h + O(\|H\|^2)) + O(\|H\|^2) \\ &= g'(\sigma)h + O(\|H\|^2). \end{aligned}$$

Then, since $U \in \mathbb{O}^m$ and $U \in \mathbb{O}^n$ are uniformly bounded, we obtain from (5.13) that

$$G_R(X) = U\,[\mathrm{Diag}\,(g'(\sigma)h) \quad 0]\,V^{\mathbb{T}} + O(\|H\|^2).$$

Thus, from (5.12), we obtain that $\Delta(H) = G_R(X) + O(\|H\|^2)$. That is, for any $X \in \mathcal{D}_G$ converging to $\overline{X}$,

$$\begin{aligned} G(X) - G(\overline{X}) - G'(X)H &= G_R(X) + G_S(X) - G_S(\overline{X}) - G'(X)H \\ &= G_R(X) - \Delta(H) = O(\|H\|^2). \end{aligned}$$

"$\Longrightarrow$" Suppose that G is 1-order G-semismooth at $\overline{X}$. Let $(\overline{U}, \overline{V}) \in \mathbb{O}^{m\times n}(\overline{X})$ be fixed. Assume that $\sigma = \overline{\sigma} + h \in \mathcal{D}_g$ and $h \in \mathbb{R}^m$ is sufficiently small. Let $X = \overline{U}\,[\mathrm{Diag}(\sigma) \quad 0]\,\overline{V}^{\mathbb{T}}$ and $H = \overline{U}\,[\mathrm{Diag}(h) \quad 0]\,\overline{V}^{\mathbb{T}}$. Then, $X \in \mathcal{D}_G$ and converges to $\overline{X}$ if h goes to zero. We know from [23, Proposition 3] that for all h sufficiently close to 0, $G(X) = \overline{U}\mathrm{Diag}(g(\sigma))\overline{V}_1^{\mathbb{T}}$. Therefore, for any h sufficiently close to 0,

$$\mathrm{Diag}(g(\overline{\sigma} + h) - g(\overline{\sigma})) = \overline{U}^{\mathbb{T}}\left(G(X) - G(\overline{X})\right)\overline{V}_1 = \overline{U}^{\mathbb{T}}G'(X)H\overline{V}_1 + O(\|H\|^2).$$

Hence, since obviously $\mathrm{Diag}(g'(\sigma)h) = \overline{U}^{\mathbb{T}}G'(X)H\overline{V}_1$, we know that for h sufficiently small, $g(\overline{\sigma} + h) - g(\overline{\sigma}) = g'(\sigma)h + O(\|h\|^2)$. Thus, g is 1-order G-semismooth at $\overline{\sigma}$. ☐

It is worth mentioning that for MOPs, we are able to obtain the semismoothness of the proximal point mapping P_f defined by (1.1) by employing the corresponding results on tame functions. We first recall the following concept on the *o(rder)-minimal structure* (cf. [17, Definition 1.4]).

DEFINITION 5.3. *An o-minimal structure of $\mathbb{R}^n$ is a sequence $\mathcal{M} = \{\mathcal{M}_i\}_{i=1}^{\infty}$ such that for each $i \geq 1$, $\mathcal{M}_i$ is a collection of subsets of $\mathbb{R}^i$ satisfying the following axioms:*

(i) *For every i, $\mathcal{M}_i$ is closed under Boolean operators (finite unions, intersections, and complement).*
(ii) *If $A \in \mathcal{M}_i$ and $B \in \mathcal{M}_{i'}$, then $A \times B$ belongs to $\mathcal{M}_{i+i'}$.*
(iii) *$\mathcal{M}_i$ contains all the subsets of the form $\{x \in \mathbb{R}^i \mid p(x) = 0\}$, where $p : \mathbb{R}^i \to \mathbb{R}$ is a polynomial function.*
(iv) *Let $\Pi : \mathbb{R}^{i+1} \to \mathbb{R}^i$ be the projection on the first i coordinates. If $A \in \mathcal{M}_{i+1}$, then $\Pi(A) \in \mathcal{M}_i$.*
(v) *The elements of $\mathcal{M}_1$ are exactly the finite union of points and intervals.*

The elements of o-minimal structure are called definable sets. A map $F : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is called definable if its graph is a definable subset of $\mathbb{R}^{n+m}$.

A set of $\mathbb{R}^n$ is called *tame* with respect to an o-minimal structure if its intersection with the interval $[-r, r]^n$ for every $r > 0$ is definable in this structure, i.e., the element of this structure. A mapping is tame if its graph is tame. One most frequently used o-minimal structure is the class of semialgebraic subsets of $\mathbb{R}^n$. A set in $\mathbb{R}^n$ is *semialgebraic* if it is a finite union of sets of the form

$$\{x \in \mathbb{R}^n \,|\, p_i(x) > 0,\ q_j(x) = 0, \quad i = 1, \ldots, a,\ j = 1, \ldots, b\}\ ,$$

where $p_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, a$, and $q_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \ldots, b$, are polynomials. A mapping is semialgebraic if its graph is semialgebraic.

For tame functions, we have the following proposition of the semismoothness [4, 40].

PROPOSITION 5.4. *Let $\xi : \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz continuous mapping.*
(i) *If ξ is tame, then ξ is semismooth.*
(ii) *If ξ is semialgebraic, then ξ is γ-order semismooth with some $\gamma > 0$.*

Let $\mathcal{Z}$ be a finite dimensional Euclidean space. If the closed proper convex function $f : \mathcal{Z} \to (-\infty, \infty]$ is semialgebraic, then the Moreau–Yosida regularization $\psi_f(x) := \min_{z\in\mathcal{Z}} \left\{ f(z) + \frac{1}{2}\|z - x\|^2 \right\}$, $x \in \mathcal{Z}$, of f is semialgebraic. Moreover, since the graph of the corresponding proximal point mapping P_f is of the form

$$\operatorname{gph} P_f = \left\{ (x, z) \in \mathcal{Z} \times \mathcal{Z} \mid f(z) + \frac{1}{2}\|z - x\|^2 = \psi_f(x) \right\},$$

we know that P_f is also semialgebraic (cf. [40]). Since P_f is globally Lipschitz continuous, according to Proposition 5.4(ii), it yields that P_f is γ-order semismooth with some $\gamma > 0$. On the other hand, most unitarily invariant closed proper convex functions $f : \mathcal{X} \to (-\infty, \infty]$ in MOPs are semialgebraic. For example, it is easy to verify that the indicator function $\delta_{\mathbb{S}^n_+}(\cdot)$ of the positive semidefinite matrix cone and the matrix Ky Fan k-norm $\|\cdot\|_{(k)}$ (the sum of k-largest singular values of matrices) are all semialgebraic. Therefore, we know that the corresponding proximal point mapping P_f defined by (1.1) for MOPs are γ-order semismooth with some $\gamma > 0$. However, since γ is not known explicitly, by this approach, we may not be able to show the strong semismoothness of the spectral operator $G = P_f$ even if the corresponding symmetric mapping g is strongly semismooth. However, the order ρ is very important for both algorithmic design and theoretical study of large-scale MOPs. For instance, it is well known (cf., e.g., [70, Theorem 3.2]) that the semismooth Newton method has a $\rho+1$-order local convergence rate for ρ-order semismooth functions. In contrast, the tame function approach does not quantify the order ρ even if one knows the order of semismoothness for the mixed symmetric function. Moreover, the explicit formulas of the derivatives obtained via the spectral operator framework are vital for applications.

6. Characterization of Clarke's generalized Jacobian. Let $\overline{X} \in \mathcal{N}$ be given. In this section, we assume that g is locally Lipschitz continuous near $\overline{\sigma} = \sigma(\overline{X})$ and directionally differentiable at $\overline{\sigma}$. Therefore, from Theorem 3.3 and [23, Theorem 3 and Remark 1], we know that the corresponding spectral operator G is locally Lipschitz continuous near $\overline{X}$ and directionally differentiable at $\overline{X}$. Furthermore, we define the function $d : \mathbb{R}^m \to \mathbb{R}^m$ by

$$d(h) := g(\overline{\sigma} + h) - g(\overline{\sigma}) - g'(\overline{\sigma}; h), \quad h \in \mathbb{R}^m. \tag{6.1}$$

Consequently, we know that the function d is also a mixed symmetric mapping, with respect to $\mathbb{P}^{|a_1|} \times \cdots \times \mathbb{P}^{|a_r|} \times \pm\mathbb{P}^{|b|}$, over $\mathcal{V} = \mathbb{R}^{|a_1|} \times \cdots \times \mathbb{R}^{|a_r|} \times \mathbb{R}^{|b|}$. Again, since g is locally Lipschitz continuous near $\overline{\sigma}$ and directionally differentiable at $\overline{\sigma}$, we know from [75] that g is B-differentiable at $\overline{\sigma}$. Thus, d is differentiable at zero with the derivative $d'(0) = 0$. Furthermore, if we assume that the function d is also strictly differentiable at zero, then we have

$$\lim_{\substack{w, w' \to 0 \\ w \neq w'}} \frac{d(w) - d(w')}{\|w - w'\|} = 0. \tag{6.2}$$

By using the mixed symmetric property of d, one can easily obtain the following results. We omit the details of the proof here.

LEMMA 6.1. *Let $d : \mathbb{R}^m \to \mathbb{R}^m$ be the function given by* (6.1). *Suppose that d is strictly differentiable at zero. Let $\{w^k\}$ be a given sequence in $\mathbb{R}^m$ converging to zero. Then, if there exist $i, j \in a_l$ for some $l \in \{1, \ldots, r\}$ or $i, j \in b$ such that $w_i^k \neq w_j^k$ for all k sufficiently large, then*

$$\lim_{k\to\infty} \frac{d_i(w^k) - d_j(w^k)}{w_i^k - w_j^k} = 0; \tag{6.3}$$

if there exist $i, j \in b$ such that $w_i^k + w_j^k \neq 0$ for all k sufficiently large, then

$$\lim_{k\to\infty} \frac{d_i(w^k) + d_j(w^k)}{w_i^k + w_j^k} = 0; \tag{6.4}$$

and if there exists $i \in b$ such that $w_i^k \neq 0$ for all k sufficiently large, then

$$\lim_{k\to\infty} \frac{d_i(w^k)}{w_i^k} = 0. \tag{6.5}$$

Again, since the spectral operator G is locally Lipschitz continuous near $\overline{X}$, we know that $\Psi = G'(\overline{X}; \cdot)$ is globally Lipschitz continuous (cf. [73, Theorem A.2] or [70, Lemma 2.2]). Therefore, $\partial_B \Psi(0)$ and $\partial \Psi(0)$ are well defined. Furthermore, we have the following characterization of the B-subdifferential and Clarke's subdifferential of the spectral operator G at $\overline{X}$ in terms of those of Ψ at 0, whose detailed proof can be found in the appendix. It is worth mentioning that these characterizations play essential roles for the study of the inverse function theorem for semismooth functions (cf., e.g., [68, Theorem 6]) and provide necessary theoretical foundation for designing semismooth Newton algorithms for solving MOPs (cf., e.g., [92, 89, 53]).

THEOREM 6.2. *Suppose that the given $\overline{X} \in \mathcal{N}$ has the decomposition* (2.2). *Suppose that there exists an open neighborhood $\mathcal{B} \subseteq \mathbb{R}^m$ of $\overline{\sigma}$ in $\hat{\sigma}_{\mathcal{N}}$ such that g is differentiable at $\sigma \in \mathcal{B}$ if and only if $g'(\overline{\sigma}; \cdot)$ is differentiable at $\sigma - \overline{\sigma}$. Assume further that the function $d : \mathbb{R}^m \to \mathbb{R}^m$ defined by* (6.1) *is strictly differentiable at zero. Then, we have*

$$\partial_B G(\overline{X}) = \partial_B \Psi(0) \quad \text{and} \quad \partial G(\overline{X}) = \partial \Psi(0).$$

7. Extensions. In this section, we consider the extensions of the related results obtained in previous sections for the case that $\mathcal{X} \equiv \mathbb{V}^{m\times n}$ to the general spectral operators defined on the Cartesian product of several real or complex matrices.

Let s be a positive integer and $0 \leq s_0 \leq s$ be a nonnegative integer. For given positive integers $m_1, \ldots, m_s$ and $n_{s_0+1}, \ldots, n_s$, define the real vector space $\mathcal{X}$ by

$$\mathcal{X} := \mathbb{S}^{m_1} \times \cdots \times \mathbb{S}^{m_{s_0}} \times \mathbb{V}^{m_{s_0+1}\times n_{s_0+1}} \times \cdots \times \mathbb{V}^{m_s\times n_s}. \tag{7.1}$$

Without loss of generality, we assume that $m_k \leq n_k$, $k = s_0 + 1, \ldots, s$. For any $X = (X_1, \ldots, X_s) \in \mathcal{X}$, we have for $1 \leq k \leq s_0$, $X_k \in \mathbb{S}^{m_k}$ and $s_0 + 1 \leq k \leq s$, $X_k \in \mathbb{V}^{m_k\times n_k}$. Denote

$$\mathcal{Y} := \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_{s_0}} \times \mathbb{R}^{m_{s_0}} \times \cdots \times \mathbb{R}^{m_s}. \tag{7.2}$$

For any $X \in \mathcal{X}$, define $\kappa(X) \in \mathcal{Y}$ by $\kappa(X) := (\lambda(X_1), \ldots, \lambda(X_{s_0}), \sigma(X_{s_0+1}), \ldots, \sigma(X_s))$. Define the set $\mathcal{P}$ by

$$\mathcal{P} := \{(Q_1, \ldots, Q_s) \mid Q_k \in \mathbb{P}^{m_k},\ 1 \leq k \leq s_0,\ \text{and } Q_k \in \pm\mathbb{P}^{m_k},\ s_0 + 1 \leq k \leq s\}.$$

Let $g : \mathcal{Y} \to \mathcal{Y}$ be a given mapping. For any $x = (x_1, \ldots, x_s) \in \mathcal{Y}$ with $x_k \in \mathbb{R}^{m_k}$, we write $g(x) \in \mathcal{Y}$ in the form $g(x) = (g_1(x), \ldots, g_s(x))$ with $g_k(x) \in \mathbb{R}^{m_k}$ for $1 \le k \le s$. The following definition of the mixed symmetric property and the general spectral operator is taken from [23, Definition 1].

Definition 7.1. *The given mapping $g : \mathcal{Y} \to \mathcal{Y}$ is said to be mixed symmetric, with respect to $\mathcal{P}$, at $x = (x_1, \ldots, x_s) \in \mathcal{Y}$ with $x_k \in \mathbb{R}^{m_k}$, if*

$$g(Q_1 x_1, \ldots, Q_s x_s) = (Q_1 g_1(x), \ldots, Q_s g_s(x)) \quad \forall\ (Q_1, \ldots, Q_s) \in \mathcal{P}. \tag{7.3}$$

The mapping g is said to be mixed symmetric, with respect to $\mathcal{P}$, over a set $\mathcal{D} \subseteq \mathcal{Y}$ if (7.3) *holds for every $x \in \mathcal{D}$.*

Let $\mathcal{N}$ be a given nonempty set in $\mathcal{X}$. Define $\kappa_{\mathcal{N}} := \{\kappa(X) \in \mathcal{Y} \mid X \in \mathcal{N}\}$. The following definition of the spectral operator with respect to a mixed symmetric mapping g is given by [23, Definition 2].

Definition 7.2. *Suppose that $g : \mathcal{Y} \to \mathcal{Y}$ is mixed symmetric on $\kappa_{\mathcal{N}}$. The spectral operator $G : \mathcal{N} \to \mathcal{X}$ with respect to g is defined as $G(X) := (G_1(X), \ldots, G_s(X))$ for $X = (X_1, \ldots, X_s) \in \mathcal{N}$ such that*

$$G_k(X) := \begin{cases} P_k \mathrm{Diag}\big(g_k(\kappa(X))\big) P_k^{\mathbb{T}} & \text{if } 1 \le k \le s_0, \\ U_k \left[\mathrm{Diag}\big(g_k(\kappa(X))\big) \quad 0\right] V_k^{\mathbb{T}} & \text{if } s_0 + 1 \le k \le s, \end{cases}$$

where $P_k \in \mathbb{O}^{m_k}(X_k)$, $1 \le k \le s_0$, $(U_k, V_k) \in \mathbb{O}^{m_k, n_k}(X_k)$, $s_0 + 1 \le k \le s$.

7.1. The spectral operators defined on the general matrix spaces. In fact, the corresponding properties of the general spectral operators defined on the vector space $\mathcal{X}$ given by (7.1), including locally Lipschitzian continuity, ρ-order B-differentiability, ρ-order G-semismoothness, and the characterization of the Clarke generalized Jacobian, can be studied in the same fashion as those in sections 4–6. For simplicity, we omit the proofs here. For readers who are interested in seeking the details, we refer to [22].

Let $\mathcal{X}$ and $\mathcal{Y}$ be the vector spaces defined by (7.1) and (7.2), respectively. Suppose that $\mathcal{N}$ is a given nonempty open set in $\mathcal{X}$. Let $G : \mathcal{X} \to \mathcal{X}$ be the spectral operator defined in Definition 7.2 with respect to $g : \mathcal{Y} \to \mathcal{Y}$, which is mixed symmetric on an open set $\hat{\kappa}_{\mathcal{N}}$ in $\mathcal{Y}$ containing $\kappa_{\mathcal{N}} := \{\kappa(X) \mid X \in \mathcal{N}\}$. For the given $\overline{X} = (\overline{X}_1, \ldots, \overline{X}_{s_0}, \overline{X}_{s_0+1}, \ldots, \overline{X}_s) \in \mathcal{X}$, recall that $\kappa(\overline{X}) = (\lambda(\overline{X}_1), \ldots, \lambda(\overline{X}_{s_0}), \sigma(\overline{X}_{s_0+1}), \ldots, \sigma(\overline{X}_s)) \in \mathcal{Y}$. We first consider the locally Lipschitzian continuity of spectral operators of matrices.

Theorem 7.3. *Let $\overline{X} \in \mathcal{N}$ be given. The spectral operator G is locally Lipschitz continuous near $\overline{X}$ if and only if the corresponding mixed symmetric function g is locally Lipschitz continuous near $\kappa(\overline{X})$.*

For the ρ-order B(ouligand)-differentiability ($0 < \rho \le 1$) of the general spectral operators, we have the following theorem.

Theorem 7.4. *Let $\overline{X} \in \mathcal{N}$ and $0 < \rho \le 1$ be given. Then, we have the following results:*

(i) *If g is locally Lipschitz continuous near $\kappa(\overline{X})$ and ρ-order B-differentiable at $\kappa(\overline{X})$, then G is ρ-order B-differentiable at $\overline{X}$.*
(ii) *If G is ρ-order B-differentiable at $\overline{X}$, then g is ρ-order B-differentiable at $\kappa(\overline{X})$.*

Suppose that g is locally Lipschitz continuous near $\kappa(\overline{X})$. Then we know from Theorem 7.3 that the corresponding spectral operator G is also locally Lipschitz continuous near $\overline{X}$. We have the following theorem on the G-semismoothness of spectral operators.

THEOREM 7.5. *Let $\overline{X} \in \mathcal{N}$ be given. Suppose that $0 < \rho \leq 1$. Then, the spectral operator G is ρ-order G-semismooth at $\overline{X}$ if and only if g is ρ-order G-semismooth at $\kappa(\overline{X})$.*

Finally, we assume that g is locally Lipschitz continuous near $\overline{\kappa} = \kappa(\overline{X})$ and directionally differentiable at $\overline{\kappa}$. From Theorem 7.3 and [23, Theorem 6 and Remark 1], the spectral operator G is also locally Lipschitz continuous near $\overline{X}$ and directionally differentiable at $\overline{X}$. Then, we have the following results on the characterization of the Clarke generalized Jacobian of G.

THEOREM 7.6. *Let $\overline{X} \in \mathcal{N}$ be given. Suppose that there exists an open neighborhood $\mathcal{B} \subseteq \mathcal{Y}$ of $\overline{\kappa}$ in $\hat{\kappa}_{\mathcal{N}}$ such that g is differentiable at $\kappa \in \mathcal{B}$ if and only if $\phi = g'(\overline{\kappa}; \cdot)$ is differentiable at $\kappa - \overline{\kappa}$. Assume that the function $d : \mathcal{Y} \to \mathcal{Y}$ defined by*

$$d(h) = g(\overline{\kappa} + h) - g(\overline{\kappa}) - g'(\overline{\kappa}; h), \quad h \in \mathcal{Y},$$

is strictly differentiable at zero. Then, we have

$$\partial_B G(\overline{X}) = \partial_B \Psi(0) \quad \text{and} \quad \partial G(\overline{X}) = \partial \Psi(0),$$

where $\Psi := G'(\overline{X}; \cdot) : \mathcal{X} \to \mathcal{X}$ is the directional derivative of G at $\overline{X}$.

7.2. The smoothing spectral operators. In this subsection, we consider the smoothing spectral operators of matrices. The corresponding properties obtained here are important and useful for designing globally convergent smoothing Newton methods for solving MOPs, which can often be solved via nonsmooth equations involving the nonsmooth spectral operators. Note that semismooth Newton methods usually only converge locally. For globalized nonsmooth Newton methods, one needs smoothing functions as demonstrated in [71]. For simplicity, we mainly focus on the case $\mathcal{X} \equiv \mathbb{R} \times \mathbb{V}^{m\times n}$. The corresponding results can be obtained as special cases for the spectral operators defined on the general matrix space $\mathcal{X}$ given by (7.1).

Let $\mathcal{N}$ be a given nonempty open set in $\mathbb{V}^{m\times n}$. Suppose that $g : \mathbb{R}^m \to \mathbb{R}^m$ is mixed symmetric with respect to $\mathcal{P} \equiv \pm\mathbb{P}^m$ on an open set $\hat{\sigma}_{\mathcal{N}}$ in $\mathbb{R}^m$ containing $\sigma_{\mathcal{N}} = \{\sigma(X) \mid X \in \mathcal{N}\}$. Let $\overline{X} \in \mathcal{N}$ be given. Assume that g is Lipschitz continuous near $\overline{\sigma} = \sigma(\overline{X})$. Suppose there exists a mapping $\theta : \mathbb{R}_{++} \times \hat{\sigma}_{\mathcal{N}} \to \mathbb{R}^m$ such that for any $x \in \hat{\sigma}_{\mathcal{N}}$ and $(\omega, z) \in \mathbb{R}_{++} \times \hat{\sigma}_{\mathcal{N}}$ close to $(0, x)$, θ is continuously differentiable around (ω, z) unless $\omega = 0$ and $\theta(\omega, z) \to g(x)$ as $(\omega, z) \to (0, x)$. For convenience, for any $x \in \hat{\sigma}_{\mathcal{N}}$, we always define $\theta(0, x) = g(x)$ and $\theta(\omega, x) = \theta(-\omega, x)$ for any $\omega < 0$. Furthermore, we assume that for any fixed ω close to 0, $\theta(\omega, \cdot)$ is also mixed symmetric on $\hat{\sigma}_{\mathcal{N}}$. Then, the mapping θ is said to be a smoothing approximation of g on $\hat{\sigma}_{\mathcal{N}}$. For a given mixed symmetric mapping g, there are many ways to construct such a smoothing approximation. For example, as mentioned in section 3, the Steklov averaged function defined by (3.13) is a smoothing approximation of the mixed symmetric mapping g.

Define $\pi : \mathbb{R} \times \hat{\sigma}_{\mathcal{N}} \to \mathbb{R} \times \mathbb{R}^m$ by $\pi(\omega, x) = (\omega, \theta(\omega, x))$, $(\omega, x) \in \mathbb{R} \times \hat{\sigma}_{\mathcal{N}}$. Then, it is easy to verify that π is mixed symmetric (Definition 2.1) over $\mathbb{R} \times \mathbb{R}^m$ with respect to $\pm\mathbb{P}^1 \times \pm\mathbb{P}^m$. Note that $\mathbb{R} \equiv \mathbb{V}^{1\times 1}$. The spectral operator $\Pi : \mathbb{V}^{1\times 1} \times \mathbb{V}^{m\times n} \to \mathbb{V}^{1\times 1} \times \mathbb{V}^{m\times n}$ defined with respect to π takes the form

$$\Pi(\omega, X) = (\omega, \Theta(\omega, X)), \quad (\omega, X) \in \mathbb{V}^{1\times 1} \times \mathcal{N},$$

where $\Theta(\omega, X) := U \left[\mathrm{Diag}\big(\theta(\omega, \sigma(X))\big) \quad 0\right] V^{\mathbb{T}}$ and $(U, V) \in \mathbb{O}^{m,n}(X)$. We call $\Theta : \mathbb{V}^{1\times 1} \times \mathcal{N} \to \mathbb{V}^{m\times n}$ the smoothing spectral operator of G with respect to θ. It follows from [23, Theorem 1] that Θ is well defined. Moreover, since θ is continuously differentiable at any $(\omega, z) \in \mathbb{R} \times \hat{\sigma}_{\mathcal{N}}$ with ω close to 0, we know from [23, Theorem 7] that Θ is also continuously differentiable at any $(\omega, X) \in \mathbb{R}\times\mathcal{N}$, and the corresponding derivative formula can be found in [23, Theorem 7]. For the case $\omega = 0$, the continuity and Hadamard directional differentiability of Θ follow directly from [23, Theorem 6]. Next, we study the locally Lipschitz continuity, ρ-order B-differentiable $(0 < \rho \leq 1)$, ρ-order G-semismooth $(0 < \rho \leq 1)$, and the characterization of the Clarke generalized Jacobian of Θ at $(0, \overline{X})$. The first property we consider is the local Lipschitzian continuity of Θ near $(0, \overline{X})$.

THEOREM 7.7. *Let $\overline{X} \in \mathcal{N}$ be given. Suppose that the smoothing approximation θ of g is locally Lipschitz continuous near $(0, \overline{\sigma})$. Then, the smoothing spectral operator Θ with respect to θ is locally Lipschitz continuous near $(0, \overline{X})$.*

The following theorem is on the ρ-order B-differentiability $(0 < \rho \leq 1)$ of the smoothing spectral operator Θ at $(0, \overline{X})$.

THEOREM 7.8. *Let $\overline{X} \in \mathcal{N}$ and $0 < \rho \leq 1$ be given. If the smoothing approximation θ of g is locally Lipschitz continuous near $(0, \overline{\sigma})$ and ρ-order B-differentiable at $(0, \overline{\sigma})$, then the smoothing spectral operator Θ is ρ-order B-differentiable at $(0, \overline{X})$.*

Suppose that the smoothing approximation θ of g is locally Lipschitz continuous near $(0, \sigma(\overline{X}))$. Then, by Theorem 7.7, the smoothing spectral operator Θ is also locally Lipschitz continuous near $\overline{X}$. Moreover, we have the following results on the G-semismoothness of the smoothing spectral operator Θ at $(0, \overline{X})$.

THEOREM 7.9. *Let $\overline{X} \in \mathcal{N}$ be given. Suppose that the smoothing approximation θ of g is ρ-order G-semismooth $(0 < \rho \leq 1)$ at $(0, \sigma(\overline{X}))$. Then, the corresponding smoothing spectral operator Θ is ρ-order G-semismooth at $(0, \overline{X})$.*

Finally, suppose that the smoothing approximation θ of g is locally Lipschitz continuous near $(0, \overline{\sigma})$ and directionally differentiable at $(0, \overline{\sigma})$. It then follows from Theorem 7.7 and [23, Theorem 3] that the smoothing spectral operator Θ is also locally Lipschitz continuous near $(0, \overline{X})$ and directionally differentiable at $(0, \overline{X})$. Furthermore, we have the following results on the characterization of the Clarke generalized Jacobian of Θ at $(0, \overline{X})$.

THEOREM 7.10. *Let $\overline{X} \in \mathcal{N}$ be given. Suppose that there exists an open neighborhood $\mathcal{B} \subseteq \mathbb{R} \times \hat{\sigma}_{\mathcal{N}}$ of $(0, \overline{\sigma})$ such that θ is differentiable at $(\tau, \sigma) \in \mathcal{B}$ if and only if $\theta'((0, \overline{\sigma}); (\cdot, \cdot))$ is differentiable at $(\tau, \sigma - \overline{\sigma})$. Assume that the function $d : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$ defined by*

$$d(\tau, h) := \theta(\tau, \overline{\sigma} + h) - \theta(0, \overline{\sigma}) - \theta'((0, \overline{\sigma}); \tau, h), \quad (\tau, h) \in \mathbb{R} \times \mathbb{R}^m,$$

is strictly differentiable at zero. Then, we have

$$\partial_B \Theta(0, \overline{X}) = \partial_B \Psi(0, 0) \quad \text{and} \quad \partial\Theta(0, \overline{X}) = \partial\Psi(0, 0),$$

where $\Psi := \Theta'((0, \overline{X}); (\cdot, \cdot))$ is the directional derivative of Θ at $(0, \overline{X})$.

8. Conclusions. We conduct extensive studies on spectral operators initiated in [23]. Several fundamental first- and second-order properties of spectral operators, including locally Lipschitz continuity, ρ-order B(ouligand)-differentiability $(0 < \rho \leq 1)$, ρ-order G-semismooth $(0 < \rho \leq 1)$, and the characterization of Clarke's generalized

Jacobian, are systematically studied. These results, together with the results obtained in [23], provide the necessary theoretical foundations for both the computational and theoretical aspects of many applications. In particular, based on the recent exciting progress made in solving large-scale SDP problems, we believe that the properties of the spectral operators studied here, such as the semismoothness and the characterization of Clarke's generalized Jacobian, constitute the backbone for future developments on both designing some efficient numerical methods for solving large-scale MOPs and conducting second-order variational analysis of the general MOPs. The work done on spectral operators of matrices is by no means complete. Due to the rapid advances in the applications of matrix optimization in different fields, spectral operators of matrices will become even more important and many other properties of spectral operators are waiting to be explored.

Appendix A.

Proof of Theorem 6.2. We only need to prove the result for the B-subdifferentials. Let $\mathcal{V}$ be any element of $\partial_B G(\overline{X})$. Then, there exists a sequence $\{X^k\}$ in $\mathcal{D}_G$ converging to $\overline{X}$ such that $\mathcal{V} = \lim_{k\to\infty} G'(X^k)$. Now we present two preparatory steps before proving that $\mathcal{V} \in \partial_B \Psi(0)$.

(a) For each X^k, let $U^k \in \mathbb{O}^m$ and $V^k \in \mathbb{O}^n$ be the matrices such that

$$X^k = U^k[\Sigma(X^k) \quad 0](V^k)^{\mathbb{T}}.$$

For each X^k, denote $\sigma^k = \sigma(X^k)$. Then, we know from [23, Theorem 4] that for each k, $\sigma^k \in \mathcal{D}_g$. For k sufficiently large, we know from [23, Lemma 1] that for each k, G_S is twice continuously differentiable at $\overline{X}$. Thus, $\lim_{k\to\infty} G'_S(X^k) = G'_S(\overline{X})$. Hence, we have for any $H \in \mathbb{V}^{m\times n}$,

(A.1)
$$\begin{aligned}\lim_{k\to\infty} G'_S(X^k)H &= G'_S(\overline{X})H \\ &= \overline{U}\left[\overline{\mathcal{E}}_1^0 \circ S\left(\overline{U}^{\mathbb{T}} H \overline{V}_1\right) + \overline{\mathcal{E}}_2^0 \circ T\left(\overline{U}^{\mathbb{T}} H \overline{V}_1\right) \quad \overline{\mathcal{F}}^0 \circ \overline{U}^{\mathbb{T}} H \overline{V}_2\right]\overline{V}^{\mathbb{T}}.\end{aligned}$$

Moreover, we know that the mapping $G_R = G - G_S$ is also differentiable at each X^k for k sufficiently large. Therefore, we have

$$\mathcal{V} = \lim_{k\to\infty} G'(X^k) = G'_S(\overline{X}) + \lim_{k\to\infty} G'_R(X^k). \tag{A.2}$$

From the continuity of the singular value function $\sigma(\cdot)$, by taking a subsequence if necessary, we assume that for each X^k and $l, l' \in \{1,\ldots,r\}$, $\sigma_i(X^k) > 0$, $\sigma_i(X^k) \neq \sigma_j(X^k)$ for any $i \in a_l$, $j \in a_{l'}$, and $l \neq l'$. Since $\{U^k\}$ and $\{V^k\}$ are uniformly bounded, by taking subsequences if necessary, we may also assume that $\{U^k\}$ and $\{V^k\}$ converge and denote the limits by $U^\infty \in \mathbb{O}^m$ and $V^\infty \in \mathbb{O}^n$, respectively. It is clear that $(U^\infty, V^\infty) \in \mathbb{O}^{m,n}(\overline{X})$. Therefore, we know from [24, Proposition 5] that there exist $Q_l \in \mathbb{O}^{|a_l|}$, $l = 1,\ldots,r$, $Q' \in \mathbb{O}^{|b|}$, and $Q'' \in \mathbb{O}^{n-|a|}$ such that $U^\infty = \overline{U}M$ and $V^\infty = \overline{V}N$, where $M = \mathrm{Diag}(Q_1,\ldots,Q_r,Q') \in \mathbb{O}^m$ and $N = \mathrm{Diag}(Q_1,\ldots,Q_r,Q'') \in \mathbb{O}^n$. Let $H \in \mathbb{V}^{m\times n}$ be arbitrarily given. For each k, denote $\widetilde{H}^k := (U^k)^{\mathbb{T}} H V^k$. Since $\{(U^k, V^k)\} \in \mathbb{O}^{m,n}(X^k)$ converges to $(U^\infty, V^\infty) \in \mathbb{O}^{m,n}(\overline{X})$, we know that $\lim_{k\to\infty} \widetilde{H}^k = (U^\infty)^{\mathbb{T}} H V^\infty$. For notational simplicity, we denote $\widetilde{H} := \overline{U}^{\mathbb{T}} H \overline{V}$ and $\widehat{H} := (U^\infty)^{\mathbb{T}} H V^\infty$.

For k sufficiently large, we know from [24, Proposition 8] and [23, Theorem 4, (38)] that for any $H \in \mathbb{V}^{m\times n}$, $G_R'(X^k)H = U^k\Delta^k(V^k)^{\mathbb{T}}$ with

$$\Delta^k := \begin{bmatrix} \mathrm{Diag}\left(\Delta_1^k,\dots,\Delta_r^k\right) & 0 \\ 0 & \Delta_{r+1}^k \end{bmatrix} \in \mathbb{V}^{m\times n},$$

where for each k, $\Delta_l^k = (\mathcal{E}_1(\sigma^k))_{a_l a_l} \circ S(\widetilde{H}_{a_l a_l}^k) + \mathrm{Diag}((\mathcal{C}(\sigma)\mathrm{diag}(S(\widetilde{H}^k)))_{a_l})$, $l = 1,\dots,r$,

$$\Delta_{r+1}^k = \Big[(\mathcal{E}_1(\sigma^k))_{bb} \circ S(\widetilde{H}_{bb}^k) + \mathrm{Diag}((\mathcal{C}(\sigma^k)\mathrm{diag}(S(\widetilde{H}^k)))_b) \\ +(\mathcal{E}_2(\sigma^k))_{bb} \circ T(\widetilde{H}_{bb}^k) \quad (\mathcal{F}(\sigma^k))_{bc} \circ \widetilde{H}_{bc}^k\Big]$$

and $\mathcal{E}_1(\sigma^k)$, $\mathcal{E}_2(\sigma^k)$, $\mathcal{F}(\sigma^k)$, and $\mathcal{C}(\sigma^k)$ are defined for σ^k by [23, (34)–(36)], respectively. Again, since $\{U^k\}$ and $\{V^k\}$ are uniformly bounded, we know that

$$\text{(A.3)} \qquad \lim_{k\to\infty} G_R'(X^k)H = U^\infty\left(\lim_{k\to\infty}\Delta^k\right)(V^\infty)^{\mathbb{T}} = \overline{U}M\left(\lim_{k\to\infty}\Delta^k\right)N^{\mathbb{T}}\overline{V}^{\mathbb{T}}.$$

(b) For each k, denote $w^k := \sigma^k - \overline{\sigma} \in \mathbb{R}^m$. Moreover, for each k, we can define $W_l^k := Q_l\mathrm{Diag}(w_{a_l}^k)Q_l^{\mathbb{T}} \in \mathbb{S}^{|a_l|}$, $l = 1,\dots,r$, and $W_{r+1}^k := Q'[\mathrm{Diag}(w_b^k) \quad 0]Q''^{\mathbb{T}} \in \mathbb{V}^{|b|\times(n-|a|)}$. Therefore, it is clear that for each k, $W^k := (W_1^k,\dots,W_l^k,W_{r+1}^k) \in \mathcal{W}$ and $\kappa(W^k) = w^k$, where $\mathcal{W} = \mathbb{S}^{|a_1|}\times\cdots\times\mathbb{S}^{|a_r|}\times\mathbb{V}^{|b|\times(n-|a|)}$. Moreover, since $\lim_{k\to\infty}\sigma^k = \overline{\sigma}$, we know that $\lim_{k\to\infty}W^k = 0$ in $\mathcal{W}$. From the assumption, we know that $\phi = g'(\overline{\sigma};\cdot)$ and $d(\cdot)$ are differentiable at each w^k and $\phi'(w^k) = g'(\sigma^k) - d'(w^k)$ for all w^k. Since d is strictly differentiable at zero, it can be checked easily that $\lim_{k\to\infty} d'(w^k) = d'(0) = 0$. By taking a subsequence if necessary, we may assume that $\lim_{k\to\infty} g'(\sigma^k)$ exists. Therefore, we have

$$\text{(A.4)} \qquad \lim_{k\to\infty}\phi'(w^k) = \lim_{k\to\infty} g'(\sigma^k).$$

Since Φ is the spectral operator with respect to the mixed symmetric mapping ϕ, from [23, Theorem 7] we know that Φ is differentiable at $W \in \mathcal{W}$ if and only if ϕ is differentiable at $\kappa(W)$. Recall that $\widehat{\Phi} : \mathcal{W} \to \mathbb{V}^{m\times n}$ is defined by (4.3). Then, for k sufficiently large, $\widehat{\Phi}$ is differentiable at W^k. Moreover, for each k, we define the matrix $C^k \in \mathbb{V}^{m\times n}$ by

$$C^k = \overline{U}\begin{bmatrix} \mathrm{Diag}\left(W_1^k,\dots,W_r^k\right) & 0 \\ 0 & W_{r+1}^k \end{bmatrix}\overline{V}^{\mathbb{T}}.$$

Then, we know that for k sufficiently large, Ψ is differentiable at C^k and $\lim_{k\to\infty}C^k = 0$ in $\mathbb{V}^{m\times n}$. Thus, we know from (4.2) that for each k,

$$\Psi'(C^k)H = G_S'(\overline{X})H + \overline{U}\left[\widehat{\Phi}'(W^k)D(H)\right]\overline{V}^{\mathbb{T}} \quad \forall\, H \in \mathbb{V}^{m\times n},$$

where $D(H) = (S(\widetilde{H}_{a_1a_1}),\dots,S(\widetilde{H}_{a_ra_r}),\widetilde{H}_{b\bar{a}})$ with $\widetilde{H} = \overline{U}^{\mathbb{T}}H\overline{V}$ and $\widehat{\Phi}'(W^k)D(H)$ can be derived from [23, Theorem 7]. By comparing with (A.2) and (A.3), we know that $\mathcal{V} \in \partial_B\Psi(0)$ if we can show that

$$\text{(A.5)} \qquad \lim_{k\to\infty}\Delta^k = \lim_{k\to\infty} M^{\mathbb{T}}\widehat{\Phi}'(W^k)D(H)N.$$

To show that (A.5) holds, for any $(i,j) \in \{1,\ldots,m\} \times \{1,\ldots,n\}$, consider the following cases.

Case 1: $i = j$. It is easy to check that for each k,

$$(\Delta^k)_{ii} = (g'(\sigma^k)h^k)_i \quad \text{and} \quad \Big(M^{\mathbb{T}}\widehat{\Phi}'(W^k)D(H)N\Big)_{ii} = (\phi'(w^k)\widehat{h})_i,$$

where $h^k = (\mathrm{diag}(S(\widetilde{H}^k_{aa})), \mathrm{diag}(\widetilde{H}^k_{bb}))$ and $\widehat{h} = (\mathrm{diag}(S(\widehat{H}_{aa})), \mathrm{diag}(\widehat{H}_{bb}))$. Therefore, we know from (A.4) that

$$\lim_{k\to\infty}(\Delta^k)_{ii} = \lim_{k\to\infty}(g'(\sigma^k)h^k)_i = \lim_{k\to\infty}(\phi'(w^k)\widehat{h})_i = \lim_{k\to\infty}\Big(M^{\mathbb{T}}\widehat{\Phi}'(W^k)D(H)N\Big)_{ii}.$$

Case 2: $i,j \in a_l$ for some $l \in \{1,\ldots,r\}$, $i \neq j$, and $\sigma_i^k \neq \sigma_j^k$ for k sufficiently large. We obtain that for k sufficiently large,

$$(\Delta^k)_{ij} = \frac{g_i(\sigma^k) - g_j(\sigma^k)}{\sigma_i^k - \sigma_j^k}(S(\widetilde{H}^k_{a_la_l}))_{ij},$$
$$\Big(M^{\mathbb{T}}\widehat{\Phi}'(W^k)D(H)N\Big)_{ij} = \frac{\phi_i(w^k) - \phi_j(w^k)}{w_i^k - w_j^k}(S(\widehat{H}_{a_la_l}))_{ij}.$$

Since $\overline{\sigma}_i = \overline{\sigma}_j$ and $g_i(\overline{\sigma}) = g_j(\overline{\sigma})$, we know that for k sufficiently large,

$$\frac{g_i(\sigma^k)-g_j(\sigma^k)}{\sigma_i^k - \sigma_j^k} = \frac{g_i(\overline{\sigma}+w^k)-g_j(\overline{\sigma}+w^k)}{w_i^k - w_j^k} = \frac{g_i(\overline{\sigma}+w^k)-g_i(\overline{\sigma}) + g_j(\overline{\sigma}) - g_j(\overline{\sigma}+w^k)}{w_i^k - w_j^k}$$
$$= \frac{d_i(w^k) - d_j(w^k)}{w_i^k - w_j^k} + \frac{\phi_i(w^k) - \phi_j(w^k)}{w_i^k - w_j^k}. \tag{A.6}$$

Therefore, we know from (6.3) that

$$\lim_{k\to\infty}\frac{g_i(\sigma^k) - g_j(\sigma^k)}{\sigma_i^k - \sigma_j^k}(S(\widetilde{H}^k_{a_la_l}))_{ij} = \lim_{k\to\infty}\frac{\phi_i(w^k) - \phi_j(w^k)}{w_i^k - w_j^k}(S(\widehat{H}_{a_la_l}))_{ij},$$

which implies $\lim_{k\to\infty}(\Delta^k)_{ij} = \lim_{k\to\infty}(M^{\mathbb{T}}\widehat{\Phi}'(W^k)D(H)N)_{ij}$.

Case 3: $i,j \in a_l$ for some $l \in \{1,\ldots,r\}$, $i \neq j$, and $\sigma_i^k = \sigma_j^k$ for k sufficiently large. We have for k sufficiently large,

$$(\Delta^k)_{ij} = \big((g'(\sigma^k))_{ii} - (g'(\sigma^k))_{ij}\big)\,(S(\widetilde{H}^k_{a_la_l}))_{ij},$$
$$\Big(M^{\mathbb{T}}\widehat{\Phi}'(W^k)D(H)N\Big)_{ij} = \big((\phi'(w^k))_{ii} - (\phi'(w^k))_{ij}\big)\,(S(\widehat{H}_{a_la_l}))_{ij}.$$

Therefore, we obtain from (A.4) that

$$\lim_{k\to\infty}\big((g'(\sigma^k))_{ii}-(g'(\sigma^k))_{ij}\big)\,(S(\widetilde{H}^k_{a_la_l}))_{ij} = \lim_{k\to\infty}\big((\phi'(w^k))_{ii}-(\phi'(w^k))_{ij}\big)\,(S(\widehat{H}_{a_la_l}))_{ij}.$$

Thus, we have $\lim_{k\to\infty}(\Delta^k)_{ij} = \lim_{k\to\infty}(M^{\mathbb{T}}\widehat{\Phi}'(W^k)D(H)N)_{ij}$.

Case 4: $i,j \in b$, $i \neq j$, and $\sigma_i^k = \sigma_j^k > 0$ for k sufficiently large. We have for k large,

$$(\Delta^k)_{ij} = \left((g'(\sigma^k))_{ii} - (g'(\sigma^k))_{ij}\right)(S(\widetilde{H}^k_{bb}))_{ij} + \frac{g_i(\sigma^k)+g_j(\sigma^k)}{\sigma^k_i+\sigma^k_j}(T(\widetilde{H}^k_{bb}))_{ij},$$

$$\left(M^{\mathbb{T}}\widehat{\Phi}'(W^k)D(H)N\right)_{ij} = \left((\phi'(w^k))_{ii} - (\phi'(w^k))_{ij}\right)(S(\widehat{H}_{bb}))_{ij} + \frac{\phi_i(w^k)+\phi_j(w^k)}{w^k_i+w^k_j}(T(\widehat{H}_{bb}))_{ij}.$$

Since $\overline{\sigma}_i = \overline{\sigma}_j = 0$ and $g_i(\overline{\sigma}) = g_j(\overline{\sigma}) = 0$, we get

$$\text{(A.7)} \qquad \frac{g_i(\sigma^k)+g_j(\sigma^k)}{\sigma^k_i+\sigma^k_j} = \frac{d_i(w^k)+d_j(w^k)}{w^k_i+w^k_j} + \frac{\phi_i(w^k)+\phi_j(w^k)}{w^k_i+w^k_j}.$$

Therefore, we know from (6.4) and (A.4) that $\lim_{k\to\infty}(\Delta^k)_{ij} = \lim_{k\to\infty}(M^{\mathbb{T}}\widehat{\Phi}'(W^k) D(H)N)_{ij}$.

Case 5: $i,j \in b$, $i \neq j$, and $\sigma^k_i \neq \sigma^k_j$ for k sufficiently large. For large k, we have

$$(\Delta^k)_{ij} = \frac{g_i(\sigma^k)-g_j(\sigma^k)}{\sigma^k_i-\sigma^k_j}(S(\widetilde{H}^k_{bb}))_{ij} + \frac{g_i(\sigma^k)+g_j(\sigma^k)}{\sigma^k_i+\sigma^k_j}(T(\widetilde{H}^k_{bb}))_{ij},$$

$$\left(M^{\mathbb{T}}\widehat{\Phi}'(W^k)D(H)N\right)_{ij} = \frac{\phi_i(w^k)-\phi_j(w^k)}{w^k_i-w^k_j}(S(\widehat{H}_{bb}))_{ij} + \frac{\phi_i(w^k)+\phi_j(w^k)}{w^k_i+w^k_j}(T(\widehat{H}_{bb}))_{ij}.$$

Thus, by (A.6) and (A.7), we know from (6.3) and (6.4) that $\lim_{k\to\infty}(\Delta^k)_{ij} = \lim_{k\to\infty}(M^{\mathbb{T}}\widehat{\Phi}'(W^k)D(H)N)_{ij}$.

Case 6: $i,j \in b$, $i \neq j$, and $\sigma^k_i = \sigma^k_j = 0$ for k sufficiently large. We know that for k sufficiently large,

$$(\Delta^k)_{ij} = \left((g'(\sigma^k))_{ii} - (g'(\sigma^k))_{ij}\right)(S(\widetilde{H}^k_{bb}))_{ij} + (g'(\sigma^k))_{ii}(T(\widetilde{H}^k_{bb}))_{ij},$$

$$\left(M^{\mathbb{T}}\widehat{\Phi}'(W^k)D(H)N\right)_{ij} = \left((\phi'(w^k))_{ii} - (\phi'(w^k))_{ij}\right)(S(\widehat{H}_{bb}))_{ij} + (\phi'(w^k))_{ii}(T(\widehat{H}_{bb}))_{ij}.$$

Again, we obtain from (A.4) that $\lim_{k\to\infty}(\Delta^k)_{ij} = \lim_{k\to\infty}(M^{\mathbb{T}}\widehat{\Phi}'(W^k)D(H)N)_{ij}$.

Case 7: $i \in b$, $j \in c$, and $\sigma^k_i > 0$ for k sufficiently large. We have for k sufficiently large,

$$(\Delta^k)_{ij} = \frac{g_i(\sigma^k)}{\sigma^k_i}(\widetilde{H}^k_{bc})_{ij}, \quad \left(M^{\mathbb{T}}\widehat{\Phi}'(W^k)D(H)N\right)_{ij} = \frac{\phi_i(w^k)}{w^k_i}(\widehat{H}_{bc})_{ij}.$$

Since $\overline{\sigma}_i = 0$ and $g_i(\overline{\sigma}) = 0$, we get

$$\frac{g_i(\sigma^k)}{\sigma^k_i} = \frac{g_i(\overline{\sigma}+w^k)-g_i(\overline{\sigma})}{w^k_i} = \frac{d_i(w^k)}{w^k_i} + \frac{\phi_i(w^k)}{w^k_i}.$$

Therefore, by (6.5), we obtain that $\lim_{k\to\infty}(\Delta^k)_{ij} = \lim_{k\to\infty}(M^{\mathbb{T}}\widehat{\Phi}'(W^k)D(H)N)_{ij}$.

Case 8: $i \in b$, $j \in c$, and $\sigma^k_i = 0$ for k sufficiently large. We have for k sufficiently large,

$$(\Delta^k)_{ij} = (g'(\sigma^k))_{ii}(\widetilde{H}^k_{bc})_{ij}, \quad \left(M^{\mathbb{T}}\widehat{\Phi}'(W^k)D(H)N\right)_{ij} = (\phi'(w^k))_{ii}(\widehat{H}_{bc})_{ij}.$$

Therefore, by (A.4), we obtain that $\lim_{k\to\infty}(\Delta^k)_{ij} = \lim_{k\to\infty}(M^{\mathbb{T}}\widehat{\Phi}'(W^k)D(H)N)_{ij}$.

Thus, we know that (A.5) holds. Therefore, by (A.2) and (A.3), we obtain that $\mathcal{V} \in \partial_B \Psi(0)$.

Conversely, suppose that $\mathcal{V} \in \partial_B \Psi(0)$ is arbitrarily chosen. Then, from the definition of $\partial_B \Psi(0)$, we know that there exists a sequence $\{C^k\} \subseteq \mathbb{V}^{m\times n}$ converging to zero such that Ψ is differentiable at each C^k and $\mathcal{V} = \lim_{k\to\infty} \Psi'(C^k)$. For each k, we know from (4.2) that Ψ is differentiable at C^k if and only if the spectral operator $\Phi : \mathcal{W} \to \mathcal{W}$ is differentiable at $W^k := D(C^k) = (S(\widetilde{C}^k_{a_1a_1}), \ldots, S(\widetilde{C}^k_{a_ra_r}), \widetilde{C}^k_{b\bar{a}}) \in \mathcal{W}$, where for each k, $\widetilde{C}^k = \overline{U}^{\mathbb{T}} C^k \overline{V}$. Moreover, for each k, we have the following decompositions:

$$S(\widetilde{C}^k_{a_la_l}) = Q^k_l \Lambda(S(\widetilde{C}^k_{a_la_l}))(Q^k_l)^{\mathbb{T}},\ l = 1, \ldots, r, \quad \widetilde{C}^k_{b\bar{a}} = Q'^k \begin{bmatrix} \Sigma(\widetilde{C}^k_{b\bar{a}}) & 0 \end{bmatrix} (Q''^k)^{\mathbb{T}},$$

where $Q^k_l \in \mathbb{O}^{|a_l|}$, $Q'^k \in \mathbb{O}^{|b|}$, and $Q''^k \in \mathbb{O}^{n-|a|}$. For each k, let

$$w^k := \Big(\lambda(S(\widetilde{C}^k_{a_1a_1})), \ldots, \lambda(S(\widetilde{C}^k_{a_ra_r})), \sigma(\widetilde{C}^k_{b\bar{a}})\Big) \in \mathbb{R}^m,$$
$$M^k := \mathrm{Diag}\Big(Q^k_1, \ldots, Q^k_r, Q'^k\Big) \in \mathbb{O}^m, \quad N^k := \mathrm{Diag}\Big(Q^k_1, \ldots, Q^k_r, Q''^k\Big) \in \mathbb{O}^n.$$

Since $\{M^k\}$ and $\{N^k\}$ are uniformly bounded, by taking subsequences if necessary, we know that there exist $Q_l \in \mathbb{O}^{|a_l|}$, $Q' \in \mathbb{O}^{|b|}$, and $Q'' \in \mathbb{O}^{n-|b|}$ such that

$$\lim_{k\to\infty} M^k = M := \mathrm{Diag}\Big(Q_1, \ldots, Q_r, Q'\Big), \quad \lim_{k\to\infty} N^k = N := \mathrm{Diag}\Big(Q_1, \ldots, Q_r, Q''\Big).$$

For each k, by [23, Theorem 7], we know that for any $H \in \mathbb{V}^{m\times n}$,

$$\text{(A.8)} \quad \begin{aligned} \Psi'(C^k)H = {} & \overline{U}\left[\overline{\mathcal{E}}^0_1 \circ S\left(\overline{U}^{\mathbb{T}} H \overline{V}_1\right) + \overline{\mathcal{E}}^0_2 \circ T\left(\overline{U}^{\mathbb{T}} H \overline{V}_1\right) \quad \overline{\mathcal{F}}^0 \circ \overline{U}^{\mathbb{T}} H \overline{V}_2\right] \overline{V}^{\mathbb{T}} \\ & + \overline{U}\left[\widehat{\Phi}'(W^k) D(H)\right] \overline{V}^{\mathbb{T}}, \end{aligned}$$

where $D(H) = (S(\widetilde{H}_{a_1a_1}), \ldots, S(\widetilde{H}_{a_ra_r}), \widetilde{H}_{b\bar{a}})$ with $\widetilde{H} = \overline{U}^{\mathbb{T}} H \overline{V}$. Let $R^k := \Phi'_k(W^k)$ $D(H)$, $k = 1, \ldots, r+1$.

For each k, define $\sigma^k := \overline{\sigma} + w^k \in \mathbb{R}^m$. Since $\lim_{k\to\infty} w^k = 0$ and for each k, $w^k_i \geq 0$ for all $i \in b$, we have $\sigma^k \geq 0$ for k sufficiently large. Therefore, for k sufficiently large, we are able to define

$$X^k := \overline{U} M [\mathrm{Diag}(\sigma^k) \quad 0] N^{\mathbb{T}} \overline{V}^{\mathbb{T}} \in \mathbb{V}^{m\times n}.$$

For simplicity, denote $U = \overline{U}M \in \mathbb{O}^m$ and $V = \overline{V}N \in \mathbb{O}^n$. It is clear that the sequence $\{X^k\}$ converges to $\overline{X}$. From the assumption, we know that g is differentiable at each σ^k and d is differentiable at each w^k with $g'(\sigma^k) = \phi'(w^k) + d'(w^k)$ for all σ^k. Therefore, by [23, Theorem 4], we know that G is differentiable at each X^k. By taking subsequences if necessary, we may assume that $\lim_{k\to\infty} \phi'(w^k)$ exists. Thus, since d is strictly differentiable at zero, we know that (A.4) holds. Since the derivative formula (2.11) is independent of $(U, V) \in \mathbb{O}^{m,n}(\overline{X})$, we know from [23, (38) in Theorem 4] that for any $H \in \mathbb{V}^{m\times n}$,

$$\text{(A.9)} \quad \begin{aligned} G'(X^k)H = {} & \overline{U}\left[\overline{\mathcal{E}}^0_1 \circ S(\overline{U}^{\mathbb{T}} H \overline{V}_1) + \overline{\mathcal{E}}^0_2 \circ T(\overline{U}^{\mathbb{T}} H \overline{V}_1) \quad \overline{\mathcal{F}}^0 \circ \overline{U}^{\mathbb{T}} H \overline{V}_2\right] \overline{V}^{\mathbb{T}} \\ & + \overline{U}\begin{bmatrix} \mathrm{Diag}\left(Q_1 \Omega^k_1 Q_1^{\mathbb{T}}, \ldots, Q_r \Omega^k_r Q_r^{\mathbb{T}}\right) & 0 \\ 0 & Q' \Omega^k_{r+1} Q''^{\mathbb{T}} \end{bmatrix} \overline{V}^{\mathbb{T}}, \end{aligned}$$

where for each k, $\Omega_l^k = (\mathcal{E}_l(\sigma^k))_{a_l a_l} \circ S(\widehat{H}_{a_l a_l}) + \mathrm{Diag}((\mathcal{C}(\sigma^k)\mathrm{diag}(S(\widehat{H})))_{a_l})$, $l = 1, \ldots, r$, and

$$\Omega_{r+1}^k = \Big[(\mathcal{E}_1(\sigma^k))_{bb} \circ S(\widehat{H}_{bb}) + \mathrm{Diag}((\mathcal{C}(\sigma^k)\mathrm{diag}(S(\widehat{H})))_b) \\ +(\mathcal{E}_2(\sigma^k))_{bb} \circ T(\widehat{H}_{bb}) \quad (\mathcal{F}_2(\sigma^k))_{bc} \circ \widehat{H}_{bc}\Big],$$

$\mathcal{E}_1(\sigma^k)$, $\mathcal{E}_2(\sigma^k)$, and $\mathcal{F}(\sigma^k)$ are defined by [23, (34)–(36)], respectively, and $\widehat{H} := M^{\mathbb{T}}\overline{U}^{\mathbb{T}} H \overline{V} N = M^{\mathbb{T}} \widetilde{H} N$. Therefore, by comparing (A.8) and (A.9), we know that the inclusion $\mathcal{V} \in \partial_B G(\overline{X})$ follows if we can show that

$$\lim_{k\to\infty} \left(R_1^k, \ldots, R_r^k, R_{r+1}^k\right) = \lim_{k\to\infty} \left(Q_1 \Omega_1^k Q_1^{\mathbb{T}}, \ldots, Q_r \Omega_r^k Q_r^{\mathbb{T}}, Q' \Omega_{r+1}^k Q''^{\mathbb{T}}\right). \tag{A.10}$$

Similar to the proofs for Cases 1–8 in the first part, by using (A.4) and (6.3)–(6.5) in Lemma 6.1, we can show that (A.10) holds. For simplicity, we omit the details here. Therefore, we obtain that $\partial_B G(\overline{X}) = \partial_B \Psi(0)$. This completes the proof. ∎

Acknowledgments. The authors would like to thank the two anonymous referees and the associate editor for their many valuable comments and constructive suggestions, which have substantially helped improve the quality and the presentation of this paper.

REFERENCES

[1] H. Attouch, J. Bolte, and B. F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods*, Math. Program., 137 (2013), pp. 91–129.

[2] R. Bhatia, *Matrix Analysis*, Springer, New York, 1997.

[3] J. Bochnak, M. Coste, and M.-F. Roy, *Real Algebraic Geometry*, Ergeb. Math. Grenzgeb. (3) 36, Springer, Berlin, 1998.

[4] J. Bolte, A. Daniilidis, and A. S. Lewis, *Tame functions are semismooth*, Math. Program., 117 (2009), pp. 5–19.

[5] J. Bolte, A. Hochart, and E. Pauwels, *Qualification conditions in semialgebraic programming*, SIAM J. Optim., 28 (2018), pp. 1867–1891.

[6] J. Bolte and E. Pauwels, *Majorization-minimization procedures and convergence of SQP methods for semi-algebraic and tame programs*, Math. Oper. Res., 41 (2016), pp. 442–465.

[7] J. B. Brown and G. Kozlowski, *Smooth Interpolation, Hölder Continuity, and the Takagi—van der Waerden Function*, Amer. Math. Monthly, 110 (2003), pp. 142–147.

[8] E. J. Candès and B. Recht, *Exact matrix completion via convex optimization*, Found. Comput. Math., 9 (2008), pp. 717–772.

[9] E. J. Candès and T. Tao, *The power of convex relaxation: near-optimal matrix completion*, IEEE Trans. Inform. Theory, 56 (2009), pp. 2053–2080.

[10] E. J. Candès, X. Li, Y. Ma, and J. Wright, *Robust principal component analysis?*, J. ACM, 58 (2011), 11.

[11] Z. X. Chan and D. F. Sun, *Constraint nondegeneracy, strong regularity, and nonsingularity in semidefinite programming*, SIAM J. Optim., 19 (2008), pp. 370–396.

[12] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. Willsky, *Rank-sparsity incoherence for matrix decomposition*, SIAM J. Optim., 21 (2011), pp. 572–596.

[13] C. H. Chen, Y. J. Liu, D. F. Sun, and K. C. Toh, *A semismooth Newton-CG dual proximal point algorithm for matrix spectral norm approximation problems*, Math. Program., 155 (2016), pp. 435–470.

[14] M. Chu, R. Funderlic, and R. Plemmons, *Structured low rank approximation*, Linear Algebra Appl., 366 (2003), pp. 157–172.

[15] C. B. Chua and L. T. K. Hien, *A superlinearly convergent smoothing Newton continuation algorithm for variational inequalities over definable sets*, SIAM J. Optim., 25 (2015), pp. 1034–1063.

[16] F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, 1983.

[17] M. Coste, *An Introduction to o-minimal Geometry*, RAAG Notes, Institut de Recherche Mathématiques de Rennes, 1999.

[18] Y. Cui, C. Ding, and X.-Y. Zhao, *Quadratic growth conditions for convex matrix optimization problems associated with spectral functions*, SIAM J. Optim., 27 (2017), pp. 2332–2355.

[19] A. Daniilidis, A. S. Lewis, J. Malick, and H. Sendov, *Prox-regularity of spectral functions and spectral sets*, J. Convex Anal., 15 (2008), pp. 547–560.

[20] A. Daniilidis, J. Malick, and H. Sendov, *Spectral (isotropic) manifolds and their dimension*, J. Anal. Math., 128 (2016), pp. 369–397.

[21] A. Daniilidis, D. Drusvyatskiy, and A. S. Lewis, *Orthogonal invariance and identifiability*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 580–598.

[22] C. Ding, *An Introduction to a Class of Matrix Optimization Problems*, Ph.D. thesis, National University of Singapore, 2012, http://www.mypolyuweb.hk/~dfsun/DingChao_Thesis_final.pdf.

[23] C. Ding, D. F. Sun, J. Sun, and K. C. Toh, *Spectral operators of matrices*, Math. Program., 168 (2018), pp. 509–531.

[24] C. Ding, D. F. Sun, and K. C. Toh, *An introduction to a class of matrix cone programming*, Math. Program., 144 (2014), pp. 141–179.

[25] C. Ding, D. F. Sun, and J. J. Ye, *First order optimality conditions for mathematical programs with semidefinite cone complementarity constraints*, Math. Program., 147 (2014), pp. 539–579.

[26] V. Dobrynin, *On the rank of a matrix associated with a graph*, Discrete Math., 276 (2004), pp. 169–175.

[27] D. Drusvyatskiy, H. L. Lee, G. Ottaviani, and R. R. Thomas, *The Euclidean distance degree of orthogonally invariant matrix varieties*, Israel J. Math., 221 (2017), pp. 291–316.

[28] D. Drusvyatskiy and A. D. Ioffe, *Quadratic growth and critical point stability of semi-algebraic functions*, Math. Program., 153 (2015), pp. 635–653.

[29] D. Drusvyatskiy, A. D. Ioffe, and A. S. Lewis, *Curves of descent*, SIAM J. Control Optim., 53 (2015), pp. 114–138.

[30] D. Drusvyatskiy and Paquette, *Variational analysis of spectral functions simplified*, J. Convex Anal., 25 (2018), pp. 119–134.

[31] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer, New York, 2003.

[32] M. Fazel, T. K. Pong, D. Sun, and P. Tseng, *Hankel matrix rank minimization with applications in system identification and realization*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 946–977.

[33] S. Friedland, *Convex spectral functions*, Linear Multilinear Algebra, 9 (1981), pp. 299–316.

[34] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed., Johns Hopkins University Press, Baltimore, MD, 2012.

[35] A. Greenbaum and L. N. Trefethen, *GMRES/CR and Arnoldi/Lanczos as matrix approximation problems*, SIAM J. Sci. Comput., 15 (1994), pp. 359–368.

[36] A. M. Gupal, *A method for the minimization of almost-differentiable functions*, Kibernetika, 1 (1977), pp. 114–116.

[37] J. B. Hawkins and A. Ben-Israel, *On generalized matrix functions*, Linear Multilinear Algebra, 1 (1973), pp. 163–171.

[38] N. J. Higham, *Functions of Matrices*, SIAM, Philadelphia, 2008.

[39] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1994.

[40] A. Ioffe, *An invitation to tame optimization*, SIAM J. Optim., 19 (2009), pp. 1894–1917.

[41] A. F. Izmailov, A. S. Kurennoy, and M. V. Solodov, *The Josephy–Newton method for semismooth generalized equations and semismooth SQP for optimization*, Set-Valued Var. Anal., 21 (2013), pp. 17–45.

[42] A. Kotlov, L. Lovász, and S. Vempala, *The Colin de Verdière number and sphere representations of a graph*, Combinatorica, 17 (1997), pp. 483–521.

[43] P. Lancaster, *On eigenvalues of matrices dependent on a parameter*, Numer. Math., 6 (1964), pp. 377–387.

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

[45] A. S. Lewis, *Convex analysis on the Hermitian matrices*, SIAM J. Optim., 6 (1996), pp. 164–177.

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

[47] A. S. Lewis and H. S. Sendov, *Twice differentiable spectral functions*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 368–386.
[48] A. S Lewis and H. S. Sendov, *Nonsmooth analysis of singular values. Part* I: *Theory*, Set-Valued Anal., 13 (2005), pp. 213–241.
[49] A. S. Lewis and H. S. Sendov, *Nonsmooth analysis of singular values. Part* II: *Applications*, Set-Valued Anal., 13 (2005), pp. 243–264.
[50] G. Li, B. S. Mordukhovich, and T. S. Phạm, *New fractional error bounds for polynomial systems with applications to Hölderian stability in optimization and spectral theory of tensors*, Math. Program., 153 (2015), pp. 333–362.
[51] X. Li, D. F. Sun, and K. C. Toh, *A highly efficient semismooth Newton augmented Lagrangian method for solving Lasso problems*, SIAM J. Optim., 28 (2018), pp. 433–458.
[52] X. Li, D. F. Sun, and K. C. Toh, *On efficiently solving the subproblems of a level-set method for fused lasso problems*, SIAM J. Optim., 28 (2018), pp. 1842–1866.
[53] X. Li, D. F. Sun, and K. C. Toh, *QSDPNAL: A two-phase proximal augmented Lagrangian method for convex quadratic semidefinite programming*, Math. Program. Comput., 10 (2018), pp. 703–743.
[54] T. X. Liu, Z. S. Lu, X. J. Chen, and Y. H. Dai, *An exact penalty method for semidefinite-box constrained low-rank matrix optimization problems*, IMA J. Numer. Anal., 40 (2020), pp. 563–586.
[55] Y. J. Liu, D. F. Sun, and K. C. Toh, *An implementable proximal point algorithmic framework for nuclear norm minimization*, Math. Program., 133 (2012), pp. 399–436.
[56] K. Löwner, *Über monotone matrixfunktionen*, Math. Z., 38 (1934), pp. 177–216.
[57] L. Lovász, *On the Shannon capacity of a graph*, IEEE Trans. Inform. Theory, 25 (1979), pp. 1–7.
[58] Z. S. Lu, Y. Zhang, and X. R. Li, *Penalty decomposition methods for rank minimization*, Optim. Methods Softw., 30 (2015), pp. 531–558.
[59] Z. S. Lu, Y. Zhang, and J. Lu, ℓ_p *regularized low-rank approximation via iterative reweighed singular value minimization*, Comput. Optim. Appl., 68 (2017), pp. 619–642.
[60] W. M. Miao, D. F. Sun, and S. H. Pan, *A rank-corrected procedure for matrix completion with fixed basis coefficients*, Math. Program., 159 (2016), pp. 289–338.
[61] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optim., 15 (1977), pp. 959–972.
[62] B. S. Mordukhovich, T. T. A. Nghia, and R. T. Rockafellar, *Full stability in finite-dimensional optimization*, Math. Oper. Res., 40 (2015), pp. 226–252.
[63] B. S. Mordukhovich and R. T. Rockafellar, *Second-order subdifferential calculus with applications to tilt stability in optimization*, SIAM J. Optim., 22 (2012), pp. 953–986.
[64] V. Noferini, *A formula for the Fréchet derivative of a generalized matrix function*, SIAM J. Matrix Anal. Appl., 38 (2017), pp. 434–457.
[65] J. S. Pang, *Newton's method for B-differentiable equations*, Math. Oper. Res., 15 (1990), pp. 311–341.
[66] J. S. Pang, *A B-differentiable equation-based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems*, Math. Program., 51 (1991), pp. 101–131.
[67] J. S. Pang and L. Q. Qi, *Nonsmooth equations: Motivation and algorithms*, SIAM J. Optim., 3 (1993), pp. 443–465.
[68] J. S. Pang, D. F. Sun, and J. Sun, *Semismooth homeomorphisms and strong stability of semidefinite and Lorentz complementarity problems*, Math. Oper. Res., 28 (2002), pp. 39–63.
[69] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res., 18 (1993), pp. 227–244.
[70] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Program., 58 (1993), pp. 353–367.
[71] L. Qi, D. F. Sun, and G. Zhou, *A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities*, Math. Program., 87 (2000), pp. 1–35.
[72] B. Recht, M. Fazel, and P. A. Parrilo, *Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization*, SIAM Rev., 52 (2010), pp. 471–501.
[73] S. M. Robinson, *Local structure of feasible sets in nonlinear programming, Part* III: *Stability and sensitivity*, Math. Program. Study, 30 (1987), pp. 45–66.
[74] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[75] A. Shapiro, *On concepts of directional differentiability*, J. Optim. Theory Appl., 66 (1990), pp. 477–487.

[76] V. A. Steklov, *On the asymptotic representation of certain functions defined by a linear differential equation of the second order, and their application to the problem of expanding an arbitrary function into a series of these functions*, Kharkov, 1957 (in Russian).

[77] D. F. Sun, *The strong second order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications*, Math. Oper. Res., 31 (2006), pp. 761–776.

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

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

[80] D. F. Sun and J. Sun, *Löwner's operator and spectral functions in Euclidean Jordan algebras*, Math. Oper. Res., 33 (2008), pp. 421–445.

[81] M. J. Todd, *Semidefinite optimization*, Acta Numer., 10 (2001), pp. 515–560.

[82] K. C. Toh, *GMRES vs. ideal GMRES*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 30–36.

[83] K. C. Toh and L. N. Trefethen, *The Chebyshev polynomials of a matrix*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 400–419.

[84] M. Torki, *Second-order directional derivatives of all eigenvalues of a symmetric matrix*, Nonlinear Anal., 46 (2001), pp. 1133–1150.

[85] M. Ulbrich, S. Ulbrich, and D. Bratzke, *A multigrid semismooth Newton method for semilinear contact problems*, J. Comput. Math., 35 (2017), pp. 484–526.

[86] H. Whitney, *On totally differentiable and smooth functions*, Pacific J. Math., 1 (1951), pp. 143–159.

[87] T. P. Wihler, *On the Hölder continuity of matrix functions for normal matrices*, J. Inequal. Pure Appl. Math., 10 (2009), 91.

[88] J. Wright, Y. Ma, A. Ganesh, and S. Rao, *Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization*, in Proceedings of Advances in Neural Information Processing Systems, Y. Bengio, D. Schuurmans, J. Lafferty, and C. Williams, eds., 2009.

[89] L. Q. Yang, D. F. Sun, and K. C. Toh, *SDPNAL+: A majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints*, Math. Program. Comput., 7 (2015), pp. 331–366.

[90] Z. Yang, *A Study on Nonsymmetric Matrix-Valued Functions*, Master's thesis, National University of Singapore, 2009, https://www.polyu.edu.hk/ama/profile/dfsun/Main_YZ.pdf.

[91] Y. Yuan, D. F. Sun, and K. C. Toh, *An efficient semismooth Newton based algorithm for convex clustering*, in Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.

[92] X. Y. Zhao, D. F. Sun, and K. C. Toh, *A Newton-CG augmented Lagrangian method for semidefinite programming*, SIAM J. Optim., 20 (2010), pp. 1737–1765.

[93] Z. R. Zhou and A. M. C. So, *A unified approach to error bounds for structured convex optimization problems*, Math. Program., 165 (2017), pp. 689–728.