A STUDY ON NONSYMMETRIC MATRIX-VALUED FUNCTIONS

YANG ZHE

(Bsc., SDU)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS

NATIONAL UNIVERSITY OF SINGAPORE

2009

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Professor Sun Defeng for all of his guidance, encouragement and support. He taught me how to do research and did it with great care and patience.

I would like to thank Dr. Liu Yongjin at National University of Singapore for his patient guidance and great help.

I would also like to acknowledge National University of Singapore for providing me the financial support and the pleasant environment for my study.

Last but not least, I am also grateful to many friends at National University of Singapore for their help and support.

Yang Zhe/August 2009

Contents

Summary

The nonsymmetric matrix-valued function plays an important role in some basic issues on designing and analyzing semismooth/smoothing Newton methods for nonsymmetric matrix optimization problems, which have been recently the focus of many studies in the science and engineering community. In this thesis, we study some key properties of nonsymmetric matrix-valued functions and their smoothing counterparts. The nonsymmetric matrix-valued function is defined as follows: For any $Y \in \Re^{p\times q}$, assume that Y has the singular value decomposition

$$Y = U[\Sigma \quad 0]V^T.$$

Then, we define the nonsymmetric matrix-valued function $G : \Re^{p\times q} \to \Re^{p\times q}$ associated with the real valued function $g : \Re_+ \to \Re$ by

$$G(Y) := U[g(\Sigma) \ 0]V^T.$$

In Chapter 2, we study the well definedness of the nonsymmetric matrix-valued function. Based on the relationship between the symmetric matrix-valued function and the nonsymmetric matrix-valued function, we show that the continuity, differentiability, continuous differentiability, locally Lipschitz continuity, directional differentiability and (strongly) semismoothness are inherited by G from g. Importantly, we give the formulas for the directional derivative and the generalized Jacobian of G.

In Chapter 3, we introduce a generalized smoothing function H of the nonsmooth nonsymmetric matrix-valued function G by using the smoothing function h of the real-valued function g. We show that the smoothing function H inherits the properties of locally Lipschitz continuity, continuous differentiability, directional differentiability and (strongly) semismoothness from h.

Chapter 1

Introduction

Let $\Re^{p\times q}$ be the space of $p \times q$ real nonsymmetric matrices. We assume without loss of generality that $p \leq q$ (otherwise we can consider the transposition of the matrix). Let Y admit the following singular value decomposition:

$$Y = U[\Sigma \;\; 0]V^T = U[\Sigma \;\; 0][V_1 \;\; V_2]^T = U\Sigma V_1^T, \tag{1.1}$$

where $U \in \Re^{p\times p}$ and $V \in \Re^{q\times q}$ are orthogonal matrices, $V_1 \in \Re^{q\times p}, V_2 \in \Re^{q\times(q-p)}$ and $V = [V_1 \; V_2]$, $\Sigma = \mathrm{diag}[\sigma_1, \ldots, \sigma_p]$, and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$ are the singular values of Y. Let $g : \Re_+ \to \Re$ be a real valued function. We can then define the nonsymmetric matrix-valued function $G : \Re^{p\times q} \to \Re^{p\times q}$ associated with g by:

$$G(Y) := U[g(\Sigma) \; 0]V^T, \tag{1.2}$$

where $g(\Sigma) = \mathrm{diag}[g(\sigma_1), \ldots, g(\sigma_p)]$.

Our study of nonsymmetric matrix-valued functions is motivated by recent interest in matrix optimization problems whose variables involve nonsymmetric matrices. One particular example arising in many fields of engineering and science is the so-called nuclear norm optimization problem, which has been the focus of several recent studies. One common model is the following nuclear norm minimization problem with linear and second order cone constraints considered in [11]:

$$\min\Big\{\|X\|_* \ : \ \mathcal{A}_e(X) = b_e,\, \mathcal{A}_q(X) - b_q \in \mathcal{K}^{m_2},\, X \in \Re^{p\times q}\Big\}, \tag{1.3}$$

where $\|X\|_*$ is defined as the sum of singular values of X, the linear operators $\mathcal{A}_e : \Re^{p\times q} \to \Re^{m_1}$ and $\mathcal{A}_q : \Re^{p\times q} \to \Re^{m_2}$, the vectors $b_e \in \Re^{m_1}, b_q \in \Re^{m_2}$ are given, and $\mathcal{K}^{m_2}$ denotes the second order cone of dimension m_2; see also [2, 13] for the studies on problem (1.3) with linear equality constraints only. Another common model is the following nuclear norm regularized linear least squares problem with linear and second order cone constraints ([12]):

$$\min\Big\{\frac{1}{2}\|\mathcal{A}_u(X) - b_u\|^2 + \mu\|X\|_* : \mathcal{A}_e(X) = b_e,\, \mathcal{A}_l(X) \geq b_l,\, \mathcal{A}_q(X) - b_q \in \mathcal{K}^{m_q}\Big\}, \tag{1.4}$$

where the linear operators $\mathcal{A}_j : \Re^{p\times q} \to \Re^{m_j}, j = u, e, l, q$, the vectors $b_j \in \Re^{m_j}, j = u, e, l, q$ and $\mu > 0$ are given. For more discussions on special cases of problem (1.4), one may refer to the papers [9, 13, 22] and references therein.

For each $\tau \geq 0$, the soft thresholding operator $\mathcal{D}_\tau(\cdot)$ arising from nuclear norm optimization problems (see [9, 11, 13, 22])[1], which is defined as follows:

$$\mathcal{D}_\tau(Y) := U g_\tau(\Sigma) V^T, \quad g_\tau(\Sigma) = [\mathrm{diag}(\{\sigma_i - \tau\}_+) \ \ 0],$$

is a special case of the nonsymmetric matrix-valued functions associated with g_τ (see Example 2.3.1 for the definition of g_τ). A recent result of Jiang et al. [9] shows that the soft thresholding operator $\mathcal{D}_\tau(\cdot)$ is strongly semismooth everywhere. This property plays a key role in analyzing the quadratic convergence of generalized Newton methods for solving (1.4) with linear equalities only, see [9] for the details. Another result developed in [12] proved that a smoothing function of $\mathcal{D}_\tau(\cdot)$ based on Huber function is also strongly semismooth, which is crucial for the application

[1] Donald Goldfarb first reported the formula of the soft thresholding operator at the "Foundations of Computational Mathematics Conference'08" held at the City University of Hong Kong, Hong Kong, China, June 2008.

of the smoothing Newton methods to (1.4). These results motivate us to address the following natural questions: Does the nonsymmetric matrix-valued function G inherit properties from g in general as like in [3]? Can we extend the results in [12] to generalized smoothing functions of nonsmooth nonsymmetric matrix-valued functions? The answer to these two questions is the main purpose of the thesis.

In Chapter 2, we first discuss about the well-definedness of the nonsymmetric matrix-valued function G. We then study the continuity and differential properties of the nonsymmetric matrix-valued function G in general. In particular, we show that the properties of continuity, (locally) Lipschitz continuity, directional differentiability, differentiability, continuous differentiability, and (ρ-order) semismoothness are each inherited by G from g. These results parallel those obtained in [3] for symmetric matrix-valued functions and are useful in the design and analysis of generalized nonsmooth methods for solving nonsymmetric matrix optimization problems. Our proofs are based on a relation between the nonsymmetric matrix-valued G and a symmetric matrix-valued function defined by (2.6).

Chapter 3 is devoted to studying the smoothing functions of nonsmooth nonsymmetric matrix-valued functions. In particular, we are interested in the kind of smoothing functions: $H(\epsilon, Y) : \Re \times \Re^{p\times q} \longrightarrow \Re^{p\times q}$ such that H is continuously differentiable on $\Re \times \Re^{p\times q}$ unless $\epsilon = 0$ and $\lim\limits_{\epsilon\downarrow 0, Z\to Y} R(\epsilon, Z) = G(Y)$. We define a smoothing function H of G by

$$H(\epsilon, Y) := U\mathrm{diag}[h(\epsilon, \sigma_1(Y)), ..., h(\epsilon, \sigma_p(Y)) \;\; 0]V^T, \tag{1.5}$$

where $h : \Re \times \Re \longrightarrow \Re$ is a smoothing function of g. Our analysis shows that the properties of Lipschitz continuity, continuous differentiability, directional differentiability and (strong) semismoothness are also inherited by H from h. The property of (strong) semismoothness of the smoothing nonsmooth nonsymmetric matrix valued functions paves a way for extending the smoothing Newton methods for symmetric matrix optimization problems to nonsymmetric cases.

To make the thesis completely self-contained, we have also included two appendices. Appendix A reviews some basic properties of vector-valued functions which are continuity, (locally) Lipschitz continuity, directional differentiability, continuous differentiability and (ρ-order) semismoothness. Appendix B contains some results related to the properties of symmetric matrix-valued functions that are used to analyze the properties of nonsymmetric matrix-valued functions.

Chapter 2

Nonsymmetric matrix-valued functions

In this chapter, we first present the nonsymmetric matrix-valued function G is well-defined and then study the continuity and differential properties of the nonsymmetric matrix-valued function G in general. In particular, we show that the properties of continuity, (locally) Lipschitz continuity, directional differentiability, differentiability, continuous differentiability and (ρ-order) semismoothness are inherited by G from g.

2.1 Well-definedness

For any given real-valued function g defined on $\Re_+$ only, we first show that $g(0) = 0$ is the sufficient and necessary condition for the well-definedness of G.

Given real-valued function $\hat{g}$ defined on $\Re_+$,

$$\hat{G}(Y) = U[\hat{g}(\Sigma)\ 0]V^T = U[g(\Sigma)\ 0]V^T + U[\hat{g}(0)\ 0]V^T = U[g(\Sigma)\ 0]V^T + \hat{g}(0)UV_1^T, \tag{2.1}$$

where $g(t) := \hat{g}(t) - \hat{g}(0)$, $t \geq 0$, $g(0) = 0$.

For subsequent discussions, we need to extend the values of g to $\Re$ as follows

$$g(t) = \begin{cases} g(t) & \text{if } t \geq 0, \\ -g(-t) & \text{if } t < 0. \end{cases} \tag{2.2}$$

That is, g is odd as a function from $\Re$ to $\Re$.

First we address that the nonsymmetric matrix-valued function G as in (1.2) is well defined for any given function $g : \Re_+ \to \Re$, $g(0) = 0$. For this purpose, we need to define the linear operator $\Xi : \Re^{p\times q} \to \mathcal{S}^{p+q}$ as follows:

$$\Xi(X) := \begin{bmatrix} 0 & X \\ X^T & 0 \end{bmatrix}, \quad \forall\, X \in \Re^{p\times q}. \tag{2.3}$$

Proposition 2.1.1. *Let $g : \Re_+ \to \Re$ be a real valued function, $g(0) = 0$. Assume that $Y \in \Re^{p\times q}$ has the singular value decomposition as in (1.1). Then, the corresponding nonsymmetric matrix-valued function $G(Y)$ given by (1.2) is well defined.*

Proof. First define an orthogonal matrix $Q \in \Re^{(p+q)\times(p+q)}$ by

$$Q := \frac{1}{\sqrt{2}} \begin{bmatrix} U & U & 0 \\ V_1 & -V_1 & \sqrt{2}V_2 \end{bmatrix}, \tag{2.4}$$

where U, V_1, V_2 are given as in (1.1). It follows from [7, pp. 448] that $\Xi(Y)$ has the following eigenvalue decomposition:

$$\Xi(Y) = Q \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & -\Sigma & 0 \\ 0 & 0 & 0 \end{bmatrix} Q^T. \tag{2.5}$$

Since $\Xi(Y)$ is symmetric, $F(\Xi(Y))$ (F is the symmetric matrix-valued function. See Appendix B for its definition and properties.) associated with $f = g$ is well defined (see [1]). Let us define $\Psi : \Re^{p\times q} \to \mathcal{S}^{p+q}$ by

$$\Psi(Y) := F(\Xi(Y)) = Q \begin{bmatrix} g(\Sigma) & & \\ & g(-\Sigma) & \\ & & g(0) \end{bmatrix} Q^T. \tag{2.6}$$

Then, by (2.4), (2.5), and (2.6), we obtain that

$$\begin{aligned} \Psi(Y) &= \frac{1}{2} \begin{bmatrix} U & U & 0 \\ V_1 & -V_1 & \sqrt{2}V_2 \end{bmatrix} \begin{bmatrix} g(\Sigma) & & \\ & g(-\Sigma) & \\ & & g(0) \end{bmatrix} \begin{bmatrix} U^T & V_1^T \\ U^T & -V_1^T \\ 0 & \sqrt{2}V_2^T \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} U(g(\Sigma) + g(-\Sigma))U^T & U(g(\Sigma) - g(-\Sigma))V_1^T \\ V_1(g(\Sigma) - g(-\Sigma))U^T & V_1(g(\Sigma) + g(-\Sigma))V_1^T + 2V_2 g(0) V_2^T \end{bmatrix}, \end{aligned}$$

which, together with (2.2), implies that

$$\Psi(Y) = \begin{bmatrix} 0 & Ug(\Sigma)V_1^T \\ V_1 g(\Sigma)U^T & 0 \end{bmatrix} = \begin{bmatrix} 0 & G(Y) \\ G(Y)^T & 0 \end{bmatrix}. \tag{2.7}$$

This shows that the corresponding nonsymmetric matrix-valued function $G(Y)$ is well defined. The proof is complete. $\square$

On the other hand, since UV_1 depend on the singular value decomposition of Y, from (2.1) we know that, $g(0) = 0$ is the necessary condition for the well-definedness of G .

Thus, for any real-valued function g defined on $\Re_+$ only, $g(0) = 0$ is the sufficient and necessary condition for the well-definedness of G. In the following discussion of this thesis, we assume that $g(0) = 0$.

2.2 Continuity and differential properties

In this section, we show that the properties of continuity, (locally) Lipschitz continuity, differentiability, and continuous differentiability are inherited by the nonsymmetric matrix-valued function G defined as in (1.2) from the real-valued function $g : \Re_+ \to \Re$. To this end, we review some useful perturbation results for the spectral decomposition.

Let $\mathcal{S}^n$ be the space of real symmetric matrices. For each $X \in \mathcal{S}^n$, we define the following set of orthonormal eigenvectors of X by

$$\mathcal{L}_X := \{P \in \mathcal{O} | P^T X P \in \mathcal{D}\},$$

where $\mathcal{O}$ denotes the space of $n \times n$ orthonormal matrices and $\mathcal{D}$ denotes the space of $n \times n$ real diagonal matrices with nonincreasing diagonal entries.

Lemma 2.2.1. [4, Lemma 3] *For any $X \in \mathcal{S}^n$, there exist scalars $\eta > 0$ and $\epsilon > 0$ such that*

$$\min_{P \in \mathcal{L}_X} \|P - Q\| \leq \eta \|X - Y\| \ \forall\, Y \in \mathcal{B}(X, \epsilon),\ \forall\, Q \in \mathcal{L}_Y. \tag{2.8}$$

Lemma 2.2.2. [1, p. 63] *For any $X, Y \in \mathcal{S}^n$, let $\lambda_1, \ldots, \lambda_n$ and $\mu_1, \ldots, \mu_n$ be the eigenvalues of X and Y, respectively. Then*

$$|\lambda_i - \mu_i| \leq \|X - Y\| \quad \forall\, i = 1, \ldots, n. \tag{2.9}$$

For any $Y \in \Re^{p \times q}$, assume that Y has the singular value decomposition as in (1.1), we define the following set of orthonormal eigenvectors of $\Xi(Y)$ by

$$\mathcal{O}_{\Xi(Y)} := \{Q \in \mathcal{O} \,|\, Q^T \Xi(Y) Q \in \tilde{D}\},$$

where $\tilde{D}$ denote the space of $(p+q) \times (p+q)$ real diagonal matrix $\mathrm{diag}[\lambda_1, \ldots, \lambda_{p+q}]$, where $\lambda_i = \sigma_i$, $i = 1, \ldots, p$, $\lambda_i = -\sigma_{i-p}$, $i = p+1, \ldots, 2p$, and $\lambda_i = 0$, $i = 2p+1, \ldots, p+q$.

Lemma 2.2.3. *For any $Y \in \Re^{p\times q}$, there exist scalars $\eta > 0$ and $\epsilon > 0$ such that*

$$\min_{P\in\mathcal{O}_{\Xi}(X)} \|P - Q\| \leq \eta\|\Xi(X) - \Xi(Y)\| \;\forall\; \Xi(Y) \in \mathcal{B}(\Xi(X), \epsilon),\; \forall\, Q \in \mathcal{O}_{\Xi}(Y). \quad (2.10)$$

Proof. For any $P \in \mathcal{L}_{\Xi(X)}$ and $Q \in \mathcal{L}_{\Xi(Y)}$, there exist a permutation matrix W such that $WP \in \mathcal{O}_{\Xi(X)}$ and $WP \in \mathcal{O}_{\Xi(Y)}$. Then from Lemma 2.2.1, there exist scalars $\eta > 0$ and $\epsilon > 0$ such that

$$\min_{P\in\mathcal{L}_{\Xi(X)}} \|P - Q\| = \min_{P\in\mathcal{L}_{\Xi(X)}} \|WP - WQ\| \leq \eta\|\Xi(X) - \Xi(Y)\|,$$

for any $\Xi(Y) \in \mathcal{B}(\Xi(X), \epsilon)$ and any $Q \in \mathcal{O}_{\Xi}(Y)$. Then we get (2.10).

□

Theorem 2.2.4. *Let $g : \Re_+ \rightarrow \Re$ be a real valued function. Then, the following results hold:*

(a) G is continuous at $Y \in \Re^{p\times q}$ with singular values $\sigma_1, \ldots, \sigma_p$ if and only if g is continuous at $\sigma_1, \ldots, \sigma_p$.

(b) G is continuous on $\Re^{p\times q}$ if and only if g is continuous on $\Re_+$.

Proof. (a) From (2.7), we know that G is continuous at Y if and only if Ψ is continuous at Y. We first show that if g is continuous at $\sigma_1, \ldots, \sigma_p$, Ψ is continuous at Y.

From Lemma 2.2.3, we know that there exist $\eta > 0$ and $\epsilon > 0$ such that for any $\Xi(Y + \Delta Y) \in \mathcal{B}(\Xi(Y), \epsilon)$, where $Y + \Delta Y = \bar{U}[\mathrm{diag}(\nu_1, \ldots, \nu_p) \;\; 0]\bar{V}^T$,

$$\min_{Q\in\mathcal{O}_{\Xi(Y)}} \|Q - \bar{Q}\| \leq \eta\|\Xi(\Delta Y)\|, \quad \forall\, \bar{Q} \in \mathcal{O}_{\Xi(Y+\Delta Y)}.$$

Since g defined by (2.2) is an odd function, we obtain that

$$
\begin{aligned}
&\Psi(Y) - \Psi(Y + \Delta Y) \\
&= Q\mathrm{diag}[g(\sigma_1), \ldots, g(\sigma_p), \ldots, -g(\sigma_1), \ldots, -g(\sigma_p), 0, \ldots, 0]Q^T \\
&-\bar{Q}\mathrm{diag}[g(\nu_1), \ldots, g(\nu_p), \ldots, -g(\nu_1), \ldots, -g(\nu_p), 0, \ldots, 0]\bar{Q}^T \\
&= Q\mathrm{diag}[g(\sigma_1) - g(\nu_1), \ldots, g(\sigma_p) - g(\nu_p), -g(\sigma_1) + g(\nu_1), \ldots, -g(\sigma_p) + g(\sigma_p), 0, \ldots, 0]Q^T \\
&+(Q - \bar{Q})\mathrm{diag}[g(\nu_1), \ldots, -g(\nu_p), 0, \ldots, 0]Q^T + \bar{Q}\mathrm{diag}[g(\nu_1), \ldots, -g(\nu_p), 0, \ldots, 0](Q - \bar{Q})^T \\
&\to 0 \quad \text{as} \quad \Delta Y \to 0,
\end{aligned}
$$

which shows that G is continuous at Y.

Suppose instead G is continuous at Y. Fix any orthogonal matrices U and V such that $Y = U[\Sigma\ 0]V^T$, where $\Sigma = \mathrm{diag}[\sigma_1, \ldots, \sigma_p]$. Then for any $i \in \{1, \ldots, p\}$,

$$Z = U[\mathrm{diag}[\sigma_1, \ldots, \sigma_{i-1}, \mu_i, \sigma_{i+1}, \ldots, \sigma_p] \quad 0]V^T \to Y \quad \text{as} \quad \mu_i \to \sigma_i,$$

and hence $G(Z) \to G(Y)$. By the definition of G, we know that $g(\mu_i) \to g(\sigma_i)$, that is, g is continuous at σ_i.

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

Now assume that the function $g : \Re \to \Re$ defined by (2.2) is differentiable at $\sigma_1, \ldots, \sigma_p$, we denote by Ω the $(p+q) \times (p+q)$ symmetric matrix whose (i,j)th entry is given by

$$
(\Omega)_{ij} = \begin{cases}
\dfrac{g(\lambda_i) - g(\lambda_j)}{\lambda_i - \lambda_j} & \text{if } \lambda_i \neq \lambda_j, \\
g'(\lambda_i) & \text{if } \lambda_i = \lambda_j, \text{ and } i \in \{1, \ldots, 2p\},\ j \in \{1, \ldots, p+q\}, \\
g'(0) & \text{if } \lambda_i = \lambda_j = 0, \text{ and } i \in \{2p+1, \ldots, p+q\},\ j \in \{1, \ldots, 2p\}, \\
0 & \text{if } i, j \in \{2p+1, \ldots, p+q\}.
\end{cases}
$$

Lemma 2.2.5. *Ψ is differentiable at Y if and only if g is differentiable at $\sigma_1, \ldots, \sigma_p$. Furthermore, if Ψ is differentiable at Y, we have*

$$\Psi'(Y)H = Q(\Omega \circ (Q^T \Xi(H) Q))Q^T \quad \forall H \in \Re^{p \times q}. \tag{2.11}$$

Proof. Suppose first that g is differentiable at $\sigma_1, \ldots, \sigma_p$. Then, it is also differentiable at $-\sigma_1, \ldots, -\sigma_p$, that is, g is differentiable at $\lambda_1, \ldots, \lambda_{2p}$.

By Lemma 2.2.3, we know that there exist scalars $\eta > 0$ and $\epsilon > 0$ such that

$$\min_{Q \in \mathcal{O}_{\Xi(Y)}} \|Q - \bar{Q}\| \leq \eta \|\Xi(Y) - \Xi(\bar{Y})\|, \ \forall \ \bar{Y} \in \mathcal{B}(Y, \epsilon), \ \forall \ \bar{Q} \in \mathcal{O}_{\Xi(\bar{Y})}.$$

We show below that for any $H \in \Re^{p \times q}$ with $\|H\| \leq \epsilon$, there exists $Q \in \mathcal{O}_{\Xi(Y)}$ such that

$$\Psi(Y + H) - \Psi(Y) - Q(\Omega \circ (Q^T \Xi(H) Q))Q^T = o(\|H\|). \tag{2.12}$$

This together with the independence of the third term on Q (see [1]) would show that Ψ is differentiable at Y and $\Psi'(Y)$ is given by (2.11).

Let $\nu_1, \ldots, \nu_{p+q}$ be the eigenvalues of $\Xi(Y + H)$ and $\tau_1, \ldots, \tau_p$ be the singular value of $Y + H$. Fix any $\bar{Q} \in \mathcal{O}_{\Xi(Y+H)}$, then $\nu_i = \tau_i$ $(i = 1, \ldots, p)$, $\nu_i = -\tau_{i-p}$ $(i = p + 1, \ldots, 2p)$ and $\nu_i = 0$ $(i = 2p + 1, \ldots, p + q)$. By Lemma 2.2.3, we know that there exists $Q \in \mathcal{O}_{\Xi(Y)}$ satisfying

$$\|Q - \bar{Q}\| \leq \eta \|\Xi(H)\|. \tag{2.13}$$

For simplicity, let r denote the left-hand side of (2.12), i.e.,

$$r := \Psi(Y + H) - \Psi(Y) - Q(\Omega \circ (Q^T \Xi(H) Q))Q^T,$$

and denote $\bar{r} := Q^T r Q$ and $\bar{h} := Q^T \Xi(H) Q$. Then we have

$$\bar{r} = o^T b o - a - \Omega \circ \bar{h}, \tag{2.14}$$

where for simplicity we denote $a := \operatorname{diag}[g(\lambda_1), \ldots, g(\lambda_{p+q})]$, $b := \operatorname{diag}[g(\nu_1), \ldots, g(\nu_{p+q})]$, and $o := \bar{Q}^T Q$. Note that

$$o = \bar{Q}^T Q = (\bar{Q} - Q)^T Q + I,$$

which, together with (2.13), implies that

$$o_{ij} = O(\|\Xi(H)\|) \ \ \forall i \neq j. \tag{2.15}$$

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

$$1 = o_{ii}^2 + \sum_{k\neq i} o_{ki}^2 = o_{ii}^2 + O(\|\Xi(H)\|^2), \quad i = 1, \ldots, p+q, \tag{2.16}$$

$$0 = o_{ii}o_{ij} + o_{ji}o_{jj} + \sum_{k\neq i,j} o_{ki}o_{kj} = o_{ii}o_{ij} + o_{ji}o_{jj} + O(\|\Xi(H)\|^2) \quad \forall i \neq j. \tag{2.17}$$

On the other hand, since

$$\mathrm{diag}[\lambda_1, \ldots, \lambda_{p+q}] = Q^T \Xi(Y) Q = o^T \mathrm{diag}[\nu_1, \ldots, \nu_{p+q}] o - \bar{h},$$

we have

$$\sum_{k=1}^{p+q} o_{ki}o_{kj}\nu_k - \bar{h}_{ij} = \begin{cases} \lambda_i & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases} \qquad i, j = 1, \ldots, p+q. \tag{2.18}$$

We now show that $\bar{r} = o(\|\Xi(H)\|) = o(\|H\|)$, which, by $\|r\| = \|\bar{r}\|$, would prove (2.12). For any $i \in \{1, \ldots, 2p\}$, from (2.14), (2.18) and the fact that $g(\nu_k) = g(0) = 0$ when $k \geq 2p+1$, we have that

$$\begin{aligned}
\bar{r}_{ii} &= \sum_{k=1}^{2p} o_{ki}^2 g(\nu_k) - g(\lambda_i) - g'(\lambda_i)\bar{h}_{ii} \\
&= \sum_{k=1}^{2p} o_{ki}^2 g(\nu_k) - g(\lambda_i) - g'(\lambda_i)(-\lambda_i + \sum_{k=1}^{2p} o_{ki}^2 \nu_k) \\
&= o_{ii}^2 g(\nu_i) - g(\lambda_i) - g'(\lambda_i)(-\lambda_i + o_{ii}^2\nu_i) + O(\|\Xi(H)\|^2) \\
&= (1 + O(\|\Xi(H)\|^2))g(\nu_i) - g(\lambda_i) - g'(\lambda_i)(-\lambda_i + (1 + O(\|\Xi(H)\|^2))\nu_i) \\
&\quad + O(\|\Xi(H)\|^2) \\
&= g(\nu_i) - g(\lambda_i) - g'(\lambda_i)(\nu_i - \lambda_i) + O(\|\Xi(H)\|^2),
\end{aligned}$$

where the third and fifth equalities use (2.15), (2.16), and the local boundedness of g. Since g is differentiable at $\lambda_1, \ldots, \lambda_{2p}$ ($\lambda_i = \sigma_i$, $i = 1, \ldots, p$ and $\lambda_i = -\sigma_i$, $i = p+1, \ldots, 2p$), by Lemma 2.2.2, we know that the right hand side is $o(\|\Xi(H)\|)$.

For $i \in \{2p+1, \dots, p+q\}$, since $k \neq i$, we have

$$\begin{aligned} \bar{r}_{ii} &= \sum_{k=1}^{2p} o_{ki}^2 g(\nu_k) - g(\lambda_i) - 0 \cdot \bar{h}_{ii} \\ &= -g(\lambda_i) + O(\|H\|^2). \end{aligned}$$

Since $\lambda_i = 0$, it hold that $\bar{r}_{ii} = o(\|H\|)$.

For any $i, j \in \{1, \dots, p+q\}$ with $i \neq j$, from (2.14), (2.18) and $g(\nu_k) = g(0) = 0$ when $k \geq 2p+1$, we obtain that

$$\begin{aligned} \bar{r}_{ij} &= \sum_{k=1}^{p+q} o_{ki} o_{kj} g(\nu_k) - \Omega_{ij} \bar{h}_{ij} \\ &= \sum_{k=1}^{p+q} o_{ki} o_{kj} g(\nu_k) - \Omega_{ij} \sum_{k=1}^{p+q} o_{ki} o_{kj} \nu_k \\ &= o_{ii} o_{ij} g(\nu_i) + o_{ji} o_{jj} g(\nu_j) - \Omega_{ij} (o_{ii} o_{ij} \nu_i + o_{ji} o_{jj} \nu_j) + O(\|\Xi(H)\|^2) \\ &= (o_{ii} o_{ij} + o_{ji} o_{jj}) g(\nu_i) + o_{ji} o_{jj} (g(\nu_j) - g(\nu_i)) \\ &\quad - \Omega_{ij} ((o_{ii} o_{ij} + o_{ji} o_{jj}) \nu_i + o_{ji} o_{jj} (\nu_j - \nu_i)) + O(\|\Xi(H)\|^2) \\ &= o_{ji} o_{jj} (g(\nu_j) - g(\nu_i) - \Omega_{ij} (\nu_j - \nu_i)) + O(\|\Xi(H)\|^2), \end{aligned}$$

where the third and fifth equalities use (2.15), (2.17) and the local boundedness of g. We consider the following six cases to prove $r = o(\|H\|)$.

Case 1: $\lambda_i = \lambda_j$ and $i \in \{1, \dots, 2p\}$, $j \in \{1, \dots, p+q\}$. The preceding relation together with (2.15), (2.16) and $|\nu_i - \lambda_i| \leq \|\Xi(H)\|$, $|\nu_j - \lambda_j| \leq \|\Xi(H)\|$ and the continuity of g at λ_i yields

$$\bar{r}_{ij} = o(\|\Xi(H)\|).$$

Case 2: $\lambda_i = \lambda_j$, $i \in \{2p+1, \dots, p+q\}$ and $j \in \{1, \dots, 2p\}$. We know that $\nu_i = 0$, so $\bar{r}_{ij} = o_{ji} o_{jj} (g(\nu_j) - g'(0) \nu_j)$. Together with (2.15), (2.16), $|\nu_j - 0| \leq \|\Xi(H)\|$, and the continuity of g at 0, we have $\bar{r}_{ij} = o(\|\Xi(H)\|)$.

Case 3: $i, j \in \{2p+1, \ldots, p+q\}$. In this case, we have $\nu_i = \nu_j = 0$ and hence $\bar{r}_{ij} = o(\|\Xi(H)\|)$.

Case 4: $\lambda_i \neq \lambda_j$ and $i, j \in \{1, \ldots, 2p\}$. Then, we know that $\Omega_{ij} = (g(\lambda_i) - g(\lambda_j))/(\lambda_i - \lambda_j)$ in this case. The preceding relation yields

$$\begin{aligned}\bar{r}_{ij} &= o_{ji}o_{jj}(g(\nu_j) - g(\nu_i) - \frac{g(\lambda_i) - g(\lambda_j)}{\lambda_i - \lambda_j}(\nu_j - \nu_i)) + O(\|\Xi(H)\|^2) \\ &= o_{ji}o_{jj}(g(\nu_j) - g(\nu_i) - (g(\lambda_j) - g(\lambda_i))(1 + \frac{\nu_j - \nu_i - \lambda_j + \lambda_i}{\lambda_j - \lambda_i})) + O(\|\Xi(H)\|^2).\end{aligned}$$

This together with (2.15), (2.16) and $|\nu_i - \lambda_i| \leq \|\Xi(H)\|$, $|\nu_j - \lambda_j| \leq \|\Xi(H)\|$ and the continuity of g at λ_i and λ_j yields $\bar{r}_{ij} = o(\|\Xi(H)\|)$.

Case 5: $\lambda_i \neq \lambda_j$, $i \in \{1, \ldots, 2p\}$ and $j \in \{2p+1, \ldots, p+q\}$. Then, we know that $\Omega_{ij} = g(\lambda_i)/\lambda_i$ in this case. The preceding relation yields

$$\begin{aligned}\bar{r}_{ij} &= o_{ji}o_{jj}(-g(\nu_i) + \frac{g(\lambda_i)}{\lambda_i}\nu_i) + O(\|\Xi(H)\|^2) \\ &= o_{ji}o_{jj}(-g(\nu_i) + g(\lambda_i)(1 + \frac{\nu_i - \lambda_i}{\lambda_i})) + O(\|\Xi(H)\|^2).\end{aligned}$$

This together with (2.15), (2.16) and $|\nu_i - \lambda_i| \leq \|\Xi(H)\|$, and the continuity of g at λ_i yields $\bar{r}_{ij} = o(\|\Xi(H)\|)$.

Case 6: $\lambda_i \neq \lambda_j$, $i \in \{2p+1, \ldots, p+q\}$ and $j \in \{1, \ldots, 2p\}$. The analysis is the same as Case 5.

Consequently, we can draw the conclusion that $r = o(\|\Xi(H)\|) = o(\|H\|)$. This shows that Ψ is differentiable at Y and $\Psi'(Y)$ is given by (2.11). $\square$

Remark 2.2.1. If $\sigma_p = 0$, then g is differentiable at 0. From [3, Proposition 4.3], F is differentiable at $\Xi(Y)$. Then, by the chain rule of composite function, we know that Ψ is differentiable at Y and

$$\Psi'(Y)(H) = F'(\Xi(Y))\Xi(H). \tag{2.19}$$

Although when $i, j \in \{2p+1, \ldots, p+q\}$, $\Omega_{ij} = g'(0)$ may not be 0, $(\Xi(H))_{ij} = 0$. So (2.19) coincides with (2.11).

In what follows, we want to give the formula of the differential of G. Since $\lambda_i = \sigma_i$ for $i = 1, \ldots, p$, $\lambda_i = -\sigma_{i-p}$ for $i = p+1, \ldots, 2p$, and $\lambda_i = 0$ for $i = 2p+1, \ldots, p+q$, we define three index sets: $\alpha = \{1, \ldots, p\}$, $\beta = \{p+1, \ldots, 2p\}$ and $\gamma = \{2p+1, \ldots, p+q\}$ and divide Ω into 9 parts,

$$\Omega = \begin{bmatrix} \Omega_{\alpha\alpha} & \Omega_{\alpha\beta} & \Omega_{\alpha\gamma} \\ \Omega_{\beta\alpha} & \Omega_{\beta\beta} & \Omega_{\beta\gamma} \\ \Omega_{\gamma\alpha} & \Omega_{\gamma\beta} & \Omega_{\gamma\gamma} \end{bmatrix}, \tag{2.20}$$

where

$$\Omega_{\alpha\alpha} \in \Re^{p\times p} \text{ and } (\Omega_{\alpha\alpha})_{ij} = \begin{cases} \dfrac{g(\sigma_i) - g(\sigma_j)}{\sigma_i - \sigma_j} & \text{if } \sigma_i \neq \sigma_j, \\ g'(\sigma_i) & \text{if } \sigma_i = \sigma_j, \end{cases}$$

$$\Omega_{\alpha\beta} \in \Re^{p\times p} \text{ and } (\Omega_{\alpha\beta})_{ij} = \begin{cases} \dfrac{g(\sigma_i) + g(\sigma_j)}{\sigma_i + \sigma_j} & \text{if } \sigma_i \neq -\sigma_j \neq 0, \\ g'(0) & \text{if } \sigma_i = -\sigma_j = 0, \end{cases}$$

$$\Omega_{\alpha\gamma} \in \Re^{p\times(q-p)} \text{ and } (\Omega_{\alpha\gamma})_{ij} = \begin{cases} \dfrac{g(\sigma_i)}{\sigma_i} & \text{if } \sigma_i \neq 0, \\ g'(0) & \text{if } \sigma_i = 0, \end{cases}$$

$$\Omega_{\beta\alpha} \in \Re^{p\times p} \text{ and } (\Omega_{\beta\alpha})_{ij} = \begin{cases} \dfrac{g(\sigma_i) + g(\sigma_j)}{\sigma_i + \sigma_j} & \text{if } -\sigma_i \neq \sigma_j \neq 0, \\ g'(0) & \text{if } -\sigma_i = \sigma_j = 0, \end{cases}$$

$$\Omega_{\beta\beta} \in \Re^{p\times p} \text{ and } \Omega_{\beta\beta} = \begin{cases} \dfrac{g(\sigma_i) - g(\sigma_j)}{\sigma_i - \sigma_j} & \text{if } -\sigma_i \neq -\sigma_j, \\ g'(\sigma_i) & \text{if } -\sigma_i = -\sigma_j, \end{cases}$$

$$\Omega_{\beta\gamma} \in \Re^{p\times(q-p)} \text{ and } (\Omega_{\beta\gamma}) = \begin{cases} \dfrac{g(\sigma_i)}{\sigma_i} & \text{if } -\sigma_i \neq 0, \\ g'(0) & \text{if } -\sigma_i = 0, \end{cases}$$

$$\Omega_{\gamma\alpha} \in \Re^{(p-q)\times p} \text{ and } (\Omega_{\gamma\alpha})_{ij} = \begin{cases} \dfrac{g(\sigma_j)}{\sigma_j} & \text{if } \sigma_j \neq 0, \\ g'(0) & \text{if } \sigma_j = 0, \end{cases}$$

$$\Omega_{\gamma\beta} \in \Re^{(p-q)\times p} \text{ and } (\Omega_{\gamma\beta})_{ij} = \begin{cases} \dfrac{g(\sigma_j)}{\sigma_j} & \text{if } -\sigma_j \neq 0, \\ g'(0) & \text{if } -\sigma_j = 0, \end{cases}$$

$$\Omega_{\gamma\gamma} \in \Re^{(q-p)\times(q-p)} \text{ and } (\Omega_{\gamma\gamma})_{ij} = 0.$$

It should be noted that we have:

$$\Omega_{\beta\alpha} := \Omega_{\alpha\beta}^T, \quad \Omega_{\gamma\alpha} := \Omega_{\alpha\gamma}^T, \quad \Omega_{\gamma\beta} := \Omega_{\gamma\beta}^T.$$

Theorem 2.2.6. *For any $Y \in \Re^{p\times q}$, assume that Y adopts the singular value decomposition as in (1.1). Then, G is differentiable at Y with singular values $\sigma_1, \ldots, \sigma_p$ if and only if g is differentiable at $\sigma_1, \ldots, \sigma_p$. Moreover, $G'(Y)$ is given by*

$$G'(Y)\Delta Y = \frac{1}{2}U[\Omega_{\alpha\alpha}\circ(A^T+A)+\Omega_{\alpha\beta}\circ(A-A^T)]V_1^T+U(\Omega_{\alpha\gamma}\circ B)V_2^T \quad \forall\, \Delta Y \in \Re^{p\times q}. \tag{2.21}$$

where $A := U^T\Delta Y V_1 \in \Re^{p\times p}$, $B := U^T\Delta Y V_2 \in \Re^{p\times(q-p)}$.

Proof. From Lemma 2.11, we know that Ψ is differentiable at Y and $\Psi'(Y)$ is given by (2.11). By (2.7), the differentiability of Ψ at Y means the differentiability of G at Y.

Next we show below $G'(Y)$ is given by (2.21). Let Q is given as in (2.4). By a direct calculation, we obtain that

$$\begin{aligned} Q^T(\Xi(\Delta Y))Q &= \frac{1}{2}\begin{bmatrix} U^T & V_1^T \\ U^T & -V_1^T \\ 0 & \sqrt{2}V_2^T \end{bmatrix}\begin{bmatrix} 0 & \Delta Y \\ \Delta Y^T & 0 \end{bmatrix}\begin{bmatrix} U & U & 0 \\ V_1 & -V_1 & \sqrt{2}V_2 \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} A+A^T & A^T-A & \sqrt{2}B \\ A-A^T & -A^T-A & \sqrt{2}B \\ \sqrt{2}B^T & \sqrt{2}B^T & 0 \end{bmatrix}. \end{aligned} \tag{2.22}$$

Denote $A := U^T\Delta Y V_1$ and $B := U^T \Delta Y V_2$.

Let us denote

$$M := (\Omega \circ Q^T(\Xi(\Delta Y))Q)Q^T = \frac{1}{2\sqrt{2}}\begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \\ M_{31} & M_{32} \end{bmatrix}.$$

Then, by simple calculations, we get

$$\begin{aligned} M_{11} &= [\Omega_{\alpha\alpha} \circ (A+A^T) + \Omega_{\alpha\beta} \circ (-A+A^T)]U^T, \\ M_{12} &= [\Omega_{\alpha\alpha} \circ (A+A^T) - \Omega_{\alpha\beta} \circ (-A+A^T)]V_1^T + 2(\Omega_{\alpha\gamma} \circ B)V_2^T, \\ M_{21} &= [\Omega_{\beta\alpha} \circ (A-A^T) + \Omega_{\beta\beta} \circ (-A-A^T)]U^T, \\ M_{22} &= [\Omega_{\beta\alpha} \circ (A-A^T) - \Omega_{\beta\beta} \circ (-A-A^T)]V_1^T + 2(\Omega_{\beta\gamma} \circ B)V_2^T, \\ M_{31} &= \sqrt{2}(\Omega_{\gamma\alpha} \circ B^T + \Omega_{\gamma\beta} \circ B^T)U^T, \\ M_{32} &= \sqrt{2}(\Omega_{\gamma\alpha} \circ B^T - \Omega_{\gamma\beta} \circ B^T)V_2^T. \end{aligned}$$

Consequently,

$$
\begin{aligned}
& Q(\Omega \circ Q^T(\Xi(\Delta Y))Q)Q^T \\
= \; & \frac{1}{4}\begin{bmatrix} U & U & 0 \\ V_1 & -V_1 & \sqrt{2}V_2, \end{bmatrix}\begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \\ M_{31} & M_{32} \end{bmatrix} \\
= \; & \frac{1}{4}\begin{bmatrix} U(M_{11}+M_{21}) & U(M_{12}+M_{22}) \\ V_1(M_{11}-M_{21})+\sqrt{2}V_2M_{31} & V_1(M_{12}-M_{22})+\sqrt{2}V_2M_{32} \end{bmatrix}. \quad (2.23)
\end{aligned}
$$

Note that

$$\Omega_{\alpha\alpha} = \Omega_{\gamma\gamma}, \quad \Omega_{\alpha\gamma} = \Omega_{\beta\gamma},$$

we obtain from (2.23) that

$$\Psi'(Y)(\Delta Y) = \frac{1}{2}\begin{bmatrix} 0 & UM_{12} \\ (UM_{12})^T & 0 \end{bmatrix},$$

which, combining with

$$\Psi'(Y)(\Delta Y) = \begin{bmatrix} 0 & G'(Y)\Delta Y \\ (G'(Y)\Delta Y)^T & 0 \end{bmatrix},$$

yields (2.21).

On the other hand, suppose that G is differentiable at Y. Suppose for the purpose of a contradiction that $g : \Re \to \Re$ is not differentiable at σ_i for some $i \in \{1, \ldots, p\}$. Then either g is not directionally differentiable at σ_i, or if it is, the right and the left derivatives at σ_i are unequal. In either case, this means there exists two sequences of nonzero scalars t^υ and τ^υ, $\upsilon = 1, 2, \ldots$, converging to zero, such that the limits

$$\lim_{\upsilon\to\infty} \frac{g(\sigma_i + t^\upsilon) - g(\sigma_i)}{t^\upsilon}, \qquad \lim_{\upsilon\to\infty} \frac{g(\sigma_i + \tau^\upsilon) - g(\sigma_i)}{\tau^\upsilon}$$

exist and either are unequal or are both equal to ∞ or are both equal to $-\infty$. Consider any $U \in \Re^{p\times p}$ and $V \in \Re^{q\times q}$ satisfying $Y = U[\Sigma\ 0]V^T$. Let $\Delta Y = U[\mathrm{diag}[0,\ldots,1,\ldots,0]\ 0]V^T$ with 1 being in the ith diagonal, we obtain that $Y + t\Delta Y = U[\mathrm{diag}[\sigma_1,\ldots,\sigma_i+t,\ldots,\sigma_p]\ 0]V^T$ for all $t \in \Re$ and hence

$$\lim_{\upsilon\to\infty} \frac{G(Y+t^{\upsilon}\Delta Y)-G(Y)}{t^{\upsilon}} = U[\mathrm{diag}[0,\ldots,\lim_{\upsilon\to\infty}\frac{g(\sigma_i+t^{\upsilon})-g(\sigma_i)}{t^{\upsilon}},\ldots,0]\ 0]V^T,$$

$$\lim_{\upsilon\to\infty} \frac{G(Y+\tau^{\upsilon}\Delta Y)-G(Y)}{\tau^{\upsilon}} = U[\mathrm{diag}[0,\ldots,\lim_{\upsilon\to\infty}\frac{g(\sigma_i+\tau^{\upsilon})-g(\sigma_i)}{\tau^{\upsilon}},\ldots,0]\ 0]V^T.$$

It follows that these two limits either are unequal or both nonfinite, which implies that G is not differentiable at Y. This contradicts to the fact that G is differentiable at Y. Therefore, g is differentiable at $\sigma_1,\ldots,\sigma_p$. □

Theorem 2.2.7. *The nonsymmetric matrix-valued function G is continuously differentiable if and only if g is continuously differentiable.*

Proof. By similar proof as in [4, Lemma 4] we know that Ψ is continuously differentiable at Y. This, together with (2.7), implies that G is continuously differentiable.

To see "only if" direction, suppose G is continuously differentiable. Then it follows from (2.21) and the definition of $\Omega_{\alpha\alpha}$ that $g'(\lambda)$ is well defined for all $\lambda \in \Re$. Moreover, $G'([\mathrm{diag}(\lambda,0,\ldots,0)\ ])$ is continuous we get $g'(\lambda)$ is continuous. This shows that g is continuously differentiable. □

2.3 Semismoothness and the generalized Jacobian

In this section, we show that G inherits the locally Lipschitz continuity, directional continuity and (strongly) semismoothness from g. First we introduce some notations.

For any $X \in \mathcal{S}^n$, $\lambda_1(X), \ldots, \lambda_n(X)$ be the eigenvalues of X and $e_1(X), \ldots, e_n(X)$ be a set of corresponding orthonormal eigenvectors. Assume that F is defined as in (B.2), then

$$F(X) = \sum_{i=1}^{n} f(\lambda_i(X))e_i(X)e_i(X)^T.$$

Let $\mu_1, \ldots, \mu_t$ be the distinct values of $\lambda_1(X), \ldots, \lambda_n(X)$ and $r_1, \ldots, r_t$ the multiplicities, i.e., $\mu_j = \lambda_{s_j+1}(X) = \ldots = \lambda_{s_j+r_j}(X),\ j = 1, \ldots, t$, where

$$s_1 := 0,\ s_2 := r_1,\ \ldots,\ s_t := r_1 + \ldots + r_{t-1}.$$

We denote by $E_j(X)$ the $n \times r_j$ matrix whose columns are formed by the eigenvectors $e_{s_j+1}(X), \ldots, e_{s_j+r_j}(X),\ j = 1, \ldots, t$, and define $P_j(X) := E_j(X)E_j(X)^T$. Then we have

$$X = \sum_{j=1}^{t} \mu_j P_j \quad \text{and} \quad F(X) = \sum_{j=1}^{t} f(\mu_j)P_j.$$

We need the following lemmas in our sequent analysis, for the details, see [18] and the references therein.

Lemma 2.3.1. *For any $j \in \{1, \ldots, t\}$, the mapping $X \mapsto P_j(X)$ is analytic in a neighborhood of X and*

$$P_j'(X)H = \sum_{k \neq j; k=1}^{t} \frac{1}{\mu_j - \mu_k}(P_jHP_k + P_kHP_j). \tag{2.24}$$

Lemma 2.3.2. [10, Theorem 7] *The directional derivatives $\lambda'_{s_j+i}(X, H),\ i = 1, \ldots, r_j$ exist and coincide with the corresponding eigenvalues of the matrix $E_j^T H E_j$ arranged in decreasing order.*

Lemma 2.3.3. [20, Theorem 4.7] *The eigenvalue function $\lambda_i : \mathcal{S}^n \to \Re$, $i = 1, \ldots, n$, are strongly semismooth at every $X \in \mathcal{S}^n$.*

Let $\phi_j(\cdot) := f'(\mu_j, \cdot),\quad j = 1, \ldots, t$ and $\Phi_j : \mathcal{S}^{r_j} \mapsto \mathcal{S}^{r_j}$ be the corresponding matrix functions.

Let $\mu_1, \ldots, \mu_m$ be the distinct values of $\sigma_1, \ldots, \sigma_p$, $\mu_{m+1}, \ldots, \mu_{2m}$ be the distinct value of $-\sigma_1, \ldots, -\sigma_p$ and $\mu_{2m+1} = 0$ be the value of $\lambda_i(\Xi(Y))$ with $i \geq 2p+1$.

Lemma 2.3.4. *If g is locally Lipschitz continuous at $\sigma_1, \ldots, \sigma_p$, then Ψ is locally Lipschitz continuous at Y.*

Proof. Since g is locally Lipschitz continuous at $\sigma_1, \ldots, \sigma_p$, it is also locally Lipschitz continuous at $-\sigma_1, \ldots, -\sigma_p$. If $\sigma_1 \geq \ldots \geq \sigma_p > 0$. Then, from

$$\Psi(Y) = \sum_{i=1}^{2m} g(\mu_i) P_i(\Xi(Y)) + \sum_{i=1}^{2m} \sum_{k=s_i+1}^{s_i+r_i} [g(\lambda_k(\Xi(Y))) - g(\mu_i)] e_k(\Xi(Y)) e_k(\Xi(Y))^T,$$

we obtain that

$$\|\Psi(\bar{Y}) - \Psi(Y)\| \leq \sum_{i=1}^{2m} |g(\mu_i)| \|P_i(\Xi(\bar{Y})) - P_i(\Xi(Y))\|$$
$$+ \sum_{i=1}^{2m} \sum_{k=s_i+1}^{s_i+r_i} |g(\lambda_k(\Xi(\bar{Y})) - g(\mu_i)| \|e_k(\Xi(\bar{Y})) e_k(\Xi(\bar{Y}))^T\|.$$

Since $\|e_k(\Xi(\bar{Y})) e_k(\Xi(\bar{Y}))^T\|$ are uniformly bounded, the conclusion then follows from the locally Lipschitz continuity of the eigenvalue function $\lambda_k(\cdot)$ and of $P_k(\cdot)$.

If $\sigma_p = 0$. Then, from [3, Proposition 4.6], we know that F is locally Lipschitz continuous at $\Xi(Y)$, i.e., there exists $L > 0$ such that

$$\|F(\Xi(Y) + \Xi(H)) - F(\Xi(Y))\| \leq L \|\Xi(H)\|.$$

By the definition of Ψ, we have

$$\|\Psi(Y + H) - \Psi(Y)\| \leq \hat{L} \|H\|,$$

which means Ψ is locally Lipschitz continuous at Y. □

Theorem 2.3.5. *The following results hold:*

(a) G is locally Lipschitz continuous at $Y \in \Re^{p \times q}$ if and only if g is locally Lipschitz continuous at $\sigma_1, \ldots, \sigma_p$.

(b) G is locally Lipschitz continuous on $\Re^{p\times q}$ if and only if g is locally Lipschitz continuous on $\Re_+$.

Proof. (a) As shown in Lemma 2.3.4, Ψ is locally Lipschitz continuous at Y. From (2.7), we know that G is locally Lipschitz continuous at Y.

Suppose instead that G is locally Lipschitz continuous at Y and Y adopts the singular decomposition (1.1). Then, there exist $\delta > 0$ and $\kappa > 0$ such that

$$\|G(X) - G(Z)\| \leq \kappa\|X - Z\|, \quad \forall X, Z \text{ such that } \|X - Y\| \leq \delta,\ \|Z - Y\| \leq \delta,$$

Choose ν, τ such that $|\nu - \sigma_i| \leq \delta$, $|\tau - \sigma_i| \leq \delta$. Let $X = U[\mathrm{diag}(\sigma_1, \ldots, \nu, \ldots, \sigma_p)\ 0]V^T$ and $Z = U[\mathrm{diag}(\sigma_1, \ldots, \tau, \ldots, \sigma_p)\ 0]V^T$. Then, we know that $\|X - Y\| \leq \delta$ and $\|Z - Y\| \leq \delta$ and hence $|g(\nu) - g(\tau)| = \|G(X) - G(Z)\| \leq \kappa\|X - Z\| = \kappa|\nu - \tau|$. So, g is locally Lipschitz continuous at σ_i, $i = 1, \ldots, p$.

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

From Lemma 2.3.4, we know that Ψ is also locally Lipschitz continuous if $g : \Re \to \Re$ is locally Lipschitz continuous. Hence, $\partial_B \Psi(Y)$ is well defined for any $Y \in \Re^{p\times q}$. Now we study the structure of this generalized Jacobian. Here we denote by Γ the $(p+q) \times (p+q)$ symmetric matrix whose (i,j)th entry is

$$(\Gamma)_{ij} = \begin{cases} \dfrac{g(\lambda_i) - g(\lambda_j)}{\lambda_i - \lambda_j} & \text{if } \lambda_i \neq \lambda_j, \\ \in \partial g(\lambda_i) & \text{if } \lambda_i = \lambda_j, \text{ and } i \in \{1, \ldots, 2p\},\ j \in \{1, \ldots, p+q\}, \\ \in \partial g(0) & \text{if } \lambda_i = \lambda_j = 0, \text{ and } i \in \{2p+1, \ldots, p+q\},\ j \in \{1, \ldots, 2p\}, \\ 0 & \text{if } i, j \in \{2p+1, \ldots, p+q\}. \end{cases}$$

Lemma 2.3.6. *If $g : \Re \to \Re$ is locally Lipschitz continuous at $\sigma_1, \ldots, \sigma_p$, the generalized Jacobian of Ψ at Y is well defined and nonempty. For any $V \in \partial_B \Psi(Y)$, one has*

$$VH = Q(\Gamma \circ (Q^T \Xi(H) Q))Q^T \quad \forall H \in \Re^{p\times q}, \tag{2.25}$$

for some $Q \in \mathcal{O}_{\Xi(Y)}$.

Proof. Fix any $V \in \partial_B \Psi(Y)$. According to the definition of $\partial_B \Psi(Y)$, there exists a sequence $\{Y_k\} \subseteq \Re^{p\times q}$ converging to Y such that Ψ is differentiable at Y_k for all k and $V = \lim_{k\to\infty} \Psi'(Y_k)$. Let σ_i, and σ_i^k be the singular value of Y and Y^k respectively. Let λ_i, and λ_i^k $(i = 1,\ldots,p+q)$ be the eigenvalue of $\Xi(Y)$ and $\Xi(Y_k)$ respectively. Then $\lambda_i = \sigma_i$ $(i = 1,\ldots,p)$, $\lambda_i = -\sigma_{i-p}$ $(i = p+1,\ldots,2p)$, and $\lambda_i = 0$ $(i = 2p+1,\ldots,p+q)$; $\lambda_i^k = \sigma_i^k$ $(i = 1,\ldots,p)$, $\lambda_i^k = -\sigma_{i-p}^k$ $(i = p+1,\ldots,2p)$, and $\lambda_i^k = 0$ $(i = 2p+1,\ldots,p+q)$. Choose any $Q_k \in \mathcal{O}_{\Xi(Y_k)}$. By Lemma 2.2.3, there exist $\eta > 0$ and $\bar{Q}_k \in \mathcal{O}_{\Xi(Y)}$ satisfying

$$\|Q_k - \bar{Q}_k\| \leq \eta \|\Xi(Y) - \Xi(Y_k)\|$$

for all k sufficiently large. By passing to a subsequence if necessary, we assume that this holds for all k and that $\{Q_k\}$ converges. By Lemma 2.2.2, we have $\lambda_i^k \to \lambda_i$ for $i = 1,\ldots,p+q$. Denote $\lambda^k = (\lambda_1^k,\ldots,\lambda_{p+q}^k)^T$. Then, from Theorem 2.2.6, we get that

$$\Psi'(Y_k)H = Q_k((Q_k^T \Xi(H) Q_k) \circ \Gamma^k) Q_K^T \quad \forall H \in \Re^{p\times q}, \tag{2.26}$$

where

$$\Gamma_{ij}^k = \begin{cases} \dfrac{g(\lambda_i^k) - g(\lambda_j^k)}{\lambda_i^k - \lambda_j^k} & \text{if } \lambda_i^k \neq \lambda_j^k, \\ g'(\lambda_i^k) & \text{if } \lambda_i^k = \lambda_j^k, \text{ and } i \in \{1,\ldots,2p\},\ j \in \{1,\ldots,p+q\}, \\ g'(0) & \text{if } \lambda_i^k = \lambda_j^k = 0, \text{ and } i \in \{2p+1,\ldots,p+q\},\ j \in \{1,\ldots,2p\}, \\ 0 & i,j = 2p+1,\ldots,p+q. \end{cases} \tag{2.27}$$

Since g is locally Lipschitz continuous, then $\{\Gamma_{ij}^k\}$ is bounded for all i,j. By passing to a subsequence if necessary, we can assume that $\{\Gamma_{ij}^k\}$ converges to some $\Gamma_{ij} \in \Re$ for all i,j.

Case 1. For each i, $i = 1, \ldots, 2p$, we have

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

Case 2. For each $i \neq j$ such that $\lambda_i \neq \lambda_j$, we have $\lambda_i^k \neq \lambda_j^k$ for all k sufficiently large and hence

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

Case 3. For each $i \neq j$ such that $\lambda_i = \lambda_j$ and $i = 1, \ldots, 2p$, $j = 1, \ldots, p+q$. If $\lambda_i^k = \lambda_j^k$ for k along some subsequence, then

$$\Gamma_{ij}^k = g'(\lambda_i^k) \to \Gamma_{ii} \in \partial_B g(\lambda_i) \subseteq \partial g(\lambda_i).$$

If $\lambda_i^k \neq \lambda_j^k$ for k along some subsequences, then a mean-value theorem of Lebourg yields

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

for some $\hat{\lambda}_{ij}^k$ in the interval between λ_i^k and λ_j^k. Since ∂g is upper semicontinuous, this together with $\hat{\lambda}_{ij}^k \to \lambda_i = \lambda_j$ implies the limit of $\{\Gamma_{ij}^k\}$ belongs to $\partial g(\lambda_i)$.

Case 4. For each $i \neq j$ with $i \in \{2p+1, \ldots, p+q\}$ and $j \in \{1, \ldots, 2p\}$, $\lambda_i = \lambda_j = 0$, the argument is similar to that in Case 3.

Case 5. For each $i, j = 2p+1, \ldots, p+q$, then $\lambda_i = \lambda_j = \lambda_i^k = \lambda_j^k = 0$. Then

$$\Gamma_{ij}^k = 0 = \Gamma_{ij}.$$

Thus, taking limits on both sides of (2.26) and using the above results, we obtain (2.25) for some $Q \in \mathcal{O}_{\Xi(Y)}$ and $\Gamma \in \mathcal{S}^{p+q}$, which are the limit of Q_k and $\Gamma(\lambda^k)$, respectively. $\square$

Next we give a formula for the generalized Jacobian of G. Since $\lambda_i = \sigma_i$ for $i = 1, \ldots, p$, $\lambda_i = -\sigma_{i-p}$ for $i = p+1, \ldots, 2p$, and $\lambda_i = 0$ for $i = 2p+1, \ldots, p+q$, we define three index sets: $\alpha = \{1, \ldots, p\}$, $\beta = \{p+1, \ldots, 2p\}$ and $\gamma = \{2p + 1, \ldots, p+q\}$ and divide Γ into 9 parts,

$$\Gamma = \begin{bmatrix} \Gamma_{\alpha\alpha} & \Gamma_{\alpha\beta} & \Gamma_{\alpha\gamma} \\ \Gamma_{\beta\alpha} & \Gamma_{\beta\beta} & \Gamma_{\beta\gamma} \\ \Gamma_{\gamma\alpha} & \Gamma_{\gamma\beta} & \Gamma_{\gamma\gamma} \end{bmatrix}, \tag{2.28}$$

Proposition 2.3.7. *Assume that g is locally Lipschitz continuous, then, for any $Y \in \Re^{p\times q}$, the generalized Jacobian $\partial_B G(Y)$ is well defined and nonempty. Moreover, for any $W \in \partial_B G(Y)$ and any $H \in \Re^{p\times q}$, we have*

$$WH = \frac{1}{2}U[\Gamma_{\alpha\alpha} \circ (A^T + A) + \Gamma_{\alpha\beta} \circ (A - A^T)]V_1^T + 2U(\Gamma_{\alpha\gamma} \circ B)V_2^T, \tag{2.29}$$

for some U, V such that $Y = U[\Sigma \ 0]V^T$, and $A = U^T H V_1 \in \Re^{p\times p}$ and $B = U^T H V_2 \in \Re^{p\times(q-p)}$.

Proof. Fix any $W \in \partial_B G(Y)$ for any $Y \in \Re^{p\times q}$. By the definition of B-subdifferential, we know that there exists $\{Y^k\} \in \Re^{p\times q}$ such that G is differentiable at Y^k and for any $H \in \Re^{p\times q}$,

$$WH = \lim_{Y^k \to Y} G'(Y^k)H. \tag{2.30}$$

Since G is differentiable at Y^k, combining with (2.7), we obtain that Ψ is differentiable at Y^k. Moreover,

$$\lim_{Y^k\to Y} \Psi'(Y^k)H = \begin{bmatrix} 0 & \lim\limits_{Y^k\to Y} G'(Y^k)H \\ (\lim\limits_{Y^k\to Y} G'(Y^k)H)^T & 0 \end{bmatrix} = \begin{bmatrix} 0 & WH \\ (WH)^T & 0 \end{bmatrix}.$$

Since $\lim_{Y^k\to Y} \Psi(Y^k) \in \partial_B \Psi(Y)$, from Lemma 2.3.6,

$$\lim_{Y^k\to Y} \Psi(Y^k)H = Q(\Gamma \circ (Q^T \Xi(H) Q))Q^T.$$

It follows from the same calculation as in (2.2.6) we get that WH is given by (2.29). $\square$

In the next two theorems, we show that G inherits the directional differentiability and semismoothness from g.

Assume that $\sigma_1 \geq \ldots \geq \sigma_p > 0$ and $g : \Re \to \Re$ is directionally differentiable at $\sigma_1, \ldots, \sigma_p$. For any $H \in \Re^{p\times q}$, we denote by Λ the $(p+q)\times(p+q)$ symmetric matrix whose (i,j) entry is

$$\Lambda_{ij} := \begin{cases} \dfrac{g(\lambda_i) - g(\lambda_j)}{\lambda_i - \lambda_j}(Q^T\Xi(H)Q)_{ij} & \text{if } \lambda_i \neq \lambda_j, \\ g'(\lambda_i; (Q^T\Xi(H)Q)_{ij}) & \text{if } \lambda_i = \lambda_j \text{ and } i,j = 1,\ldots,2p, \\ 0 & \text{if } i,j = 2p+1,\ldots,p+q. \end{cases} \tag{2.31}$$

Lemma 2.3.8. *If g is directionally differentiable at $\sigma_1, \ldots, \sigma_p$ and $\sigma_p > 0$. Then, Ψ is directionally differentiable at Y. Moreover, for any $H \in \Re^{p\times q}$, one has*

$$\Psi'(Y; H) = Q\Lambda Q^T. \tag{2.32}$$

Proof. Let $\mu_1, \ldots, \mu_m$ be the distinct values of $\sigma_1, \ldots, \sigma_p$, $\mu_{m+1}, \ldots, \mu_{2m}$ be the distinct value of $-\sigma_1, \ldots, -\sigma_p$ and $\mu_{2m+1} = 0$ be the value of $\lambda_i(\Xi(Y))$ with $i \geq 2p+1$. Using the above notations, we have

$$\Psi(Y) = \sum_{j=1}^{2m} g(\lambda_j)P_j(\Xi(Y)). \tag{2.33}$$

Consider the decomposition of Ψ at $\bar{Y} = Y + tH$. Since

$$g(\mu_{2m+1}(\Xi(\bar{Y}))) = g(\mu_{2m+1}) = g(0) = 0$$

and

$$\Psi(\bar{Y}) = \sum_{j=1}^{2m} g(\mu_j)P_j(\Xi(\bar{Y})) + \sum_{j=1}^{2m}\sum_{k=s_j+1}^{s_j+r_j}[g(\mu_k(\Xi(\bar{Y}))) - g(\mu_j)]e_k(\Xi(\bar{Y}))e_k(\Xi(\bar{Y}))^T,$$

we have

$$\begin{aligned} \Psi(\bar{Y}) - \Psi(Y) &= \sum_{j=1}^{2m} g(\mu_j)[P_j(\Xi(\bar{Y})) - P_j] \\ &+ \sum_{j=1}^{2m}\sum_{k=s_1+1}^{s_1+r_1}[g(\lambda_k(\Xi(\bar{Y}))) - g(\mu_j)]e_k(\Xi(\bar{Y}))e_k(\Xi(\bar{Y}))^T. \end{aligned} \tag{2.34}$$

First, we have that

$$\lim_{t\downarrow 0} t^{-1}\sum_{j=1}^{2m} g(\mu_j)[P_j(\Xi(\bar{Y}))-P_j]=\sum_{j=1}^{2m} g(\mu_j)P_j'(\Xi(Y))\Xi(H),$$

and by Lemma 2.3.1, we further have that

$$\begin{aligned}\sum_{j=1}^{2m} g(\mu_j)DP_j(\Xi(Y))\Xi(H) &= \sum_{j=1}^{2m}\sum_{k\neq j;k=1}^{2m+1}\frac{g(\mu_j)}{\mu_j-\mu_k}(P_j\Xi(H)P_k+P_k\Xi(H)P_j)\\ &= \sum_{1\le j<k\le 2m}\frac{g(\mu_j)-g(\mu_k)}{\mu_j-\mu_k}(P_j\Xi(H)P_k+P_k\Xi(H)P_j)\\ &+ \sum_{j=1}^{2m}\frac{g(\mu_j)}{\mu_j}(P_j\Xi(H)P_{2m+1}+P_{2m+1}\Xi(H)P_j). \quad (2.35)\end{aligned}$$

Next, for $t>0$ and $j=1$, let

$$\Delta_1(t):=t^{-1}\sum_{k=s_1+1}^{s_1+r_1}[g(\lambda_k(\Xi(\bar{Y})))-g(\mu_1)]e_k(\Xi(\bar{Y}))e_k(\Xi(\bar{Y}))^T.$$

Note that

$$\lim_{t\downarrow 0} t^{-1}[g(\lambda_k(\Xi(\bar{Y})))-g(\mu_1)]=g'(\mu_1,\lambda_k'(\Xi(Y),\Xi(H))),$$

by [18], we know that any accumulation point of $E_1(\Xi(\bar{Y}))$ is a matrix $\tilde{E}_1(\Xi(Y))$ whose columns $\tilde{e}_1(\Xi(Y)),\ldots,\tilde{e}_{r_1}(\Xi(Y))$ satisfy the following two conditions

(a) $\tilde{e}_i^T\Xi(H)\tilde{e}_i=0$ for $i\neq j\in\{1,\ldots,r_1\}$.

(b) $\tilde{e}_1^T\Xi(H)\tilde{e}_1,\ldots,\tilde{e}_{r_1}^T\Xi(H)\tilde{e}_{r_1}$ form the eigenvalues of the $r_1\times r_1$ matrix $\tilde{E}_1^T\Xi(H)\tilde{E}_1$ arranged in the decreasing order.

Then, by Lemma 2.3.2, we get

$$g'(\mu_1,\lambda_k'(\Xi(Y),\Xi(H)))=g'(\mu_1,\tilde{e}_k^T\Xi(H)\tilde{e}_k).$$

Moreover, since the eigenvalues of $\tilde{E}_1^T\Xi(H)\tilde{E}_1$ coincide with the corresponding eigenvalues of $E_1^T\Xi(H)E_1$, it follows that

$$\lim_{t\downarrow 0}\Delta_1(t)=\sum_{k=s_1+1}^{s_1+r_1} g'(\lambda_1,\tilde{e}_k^T\Xi(H)\tilde{e}_k)\tilde{e}_k\tilde{e}_k^T=E_1[\Phi_1(E_1^T\Xi(H)E_1)]E_1^T.$$

The same calculations can be performed for every $j \in \{1, \ldots, 2m\}$. Together with (2.35), we have

$$\Psi'(Y, H) = \sum_{1 \leq j < k \leq 2m} \frac{g(\mu_j) - g(\mu_k)}{\mu_j - \mu_k} (P_j \Xi(H) P_k + P_k \Xi(H) P_j)$$

$$+ \sum_{j=1}^{2m} \frac{g(\mu_j)}{\mu_j} (P_j \Xi(H) P_{2m+1} + P_{2m+1} \Xi(H) P_j) + \sum_{j=1}^{2m} E_j [\Phi_j (E_j^T \Xi(H) E_j)] E_j^T, \quad (2.36)$$

which is the same as (2.32). □

Remark 2.3.1. If $\sigma_1 \geq \ldots \geq \sigma_p = 0$. Then, g is also directionally differentiable at 0 since g is directionally differentiable at $\sigma_1, \ldots, \sigma_p$. By [3, Proposition 4.2], we know that F is directionally differentiable at $\Xi(Y)$. Since

$$\begin{aligned} \Psi'(Y; H) &= \lim_{t \downarrow 0} \frac{F(\Xi(Y + tH)) - F(\Xi(Y))}{t} \\ &= \lim_{t \downarrow 0} \frac{F(\Xi(Y) + t\Xi(H)) - F(\Xi(Y))}{t} \\ &= F'(\Xi(Y); \Xi(H)), \end{aligned} \quad (2.37)$$

we obtain that Ψ is directionally differentiable and $\Psi'(Y; H) = Q \Sigma Q^T$, where

$$\Lambda_{ij} = \begin{cases} \dfrac{g(\lambda_i) - g(\lambda_j)}{\lambda_i - \lambda_j} (\Xi(H))_{ij} & \text{if } \lambda_i \neq \lambda_j, \\ g'(\lambda_i; (\Xi(H))_{ij} & \text{if } \lambda_i = \lambda_j. \end{cases}$$

Since for $i, j \in \{2p+1, \ldots, p+q\}$, $\lambda_i = \lambda_j = 0$ and $\Xi(H)_{ij} = 0$, we have $g'(\lambda_i; \Xi(H))_{ij} = 0$. Thus, we get $\Psi'(Y; H) = Q \Lambda Q^T$, where

$$\Lambda_{ij} = \begin{cases} \dfrac{g(\lambda_i) - g(\lambda_j)}{\lambda_i - \lambda_j} (Q^T \Xi(H) Q)_{ij} & \text{if } \lambda_i \neq \lambda_j, \\ g'(\lambda_i; (Q^T \Xi(H) Q)_{ij}) & \text{if } \lambda_i = \lambda_j \text{ and } i \in \{1, \ldots, 2p\},\ j \in \{1, \ldots, p+q\} \\ g'(\lambda_i; (Q^T \Xi(H) Q)_{ij}) & \text{if } \lambda_i = \lambda_j \text{ and } i \in \{1, \ldots, p+q\}\ j \in \{1, \ldots, 2p\} \\ 0 & \text{if } i, j \in \{2p+1, \ldots, p+q\}. \end{cases}$$

Theorem 2.3.9. *Let Y have the singular value decomposition as in (1.1). Then, G is directionally differentiable at $Y \in \Re^{p\times q}$ if and only if g is directionally differentiable at $\sigma_1, \ldots, \sigma_p$. Moreover, for any nonzero $\Delta Y \in \Re^{p\times q}$,*

$$G'(Y;H) = U(\Lambda_{\alpha\alpha} - \Lambda_{\alpha\beta})V_1^T + \sqrt{2}U\Lambda_{\alpha\gamma}V_2^T, \tag{2.38}$$

where $A = U^T H V_1$ and $B = U^T H V_2$.

Proof. Suppose first that g is directionally differentiable at $\sigma_1, \ldots, \sigma_p$. Then, from Lemma 2.3.8 and the above arguments, we conclude that G is directionally differentiable at Y.

Next we calculate the directional derivative of G.

$$\begin{aligned}
\Psi'(Y,H) &= \begin{bmatrix} 0 & \lim\limits_{t\downarrow 0} \dfrac{G(Y+tH)-G(Y)}{t} \\ \left(\lim\limits_{t\downarrow 0} \dfrac{G(Y+tH)-G(Y)}{t}\right)^T & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 & G'(Y;H) \\ (G'(Y;H))^T & 0 \end{bmatrix}.
\end{aligned} \tag{2.39}$$

From (2.22), we know that

$$Q^T \Xi(H) Q = \begin{bmatrix} A + A^T & A^T - A & \sqrt{2}B \\ A - A^T & -A^T - A & \sqrt{2}B \\ \sqrt{2}B^T & \sqrt{2}B^T & 0 \end{bmatrix}.$$

Divide Λ into 9 parts as follows:

$$\Lambda = \begin{bmatrix} \Lambda_{\alpha\alpha} & \Lambda_{\alpha\beta} & \Lambda_{\alpha\gamma} \\ \Lambda_{\beta\alpha} & \Lambda_{\beta\beta} & \Lambda_{\beta\gamma} \\ \Lambda_{\gamma\alpha} & \Lambda_{\gamma\beta} & \Lambda_{\gamma\gamma} \end{bmatrix}.$$

By the definition of Λ, we have

$$(\Lambda_{\alpha\alpha})_{ij} = \begin{cases} \dfrac{g(\sigma_i) - g(\sigma_j)}{\sigma_i - \sigma_j}(a_{ij} + a_{ji}) & \text{if } \sigma_i \neq \sigma_j, \\ g'(\sigma_i; a_{ij} + a_{ji}) & \text{if } \sigma_i = \sigma_j, \end{cases}$$

$$(\Lambda_{\alpha\beta})_{ij} = \begin{cases} \dfrac{g(\sigma_i)+g(\sigma_j)}{\sigma_i+\sigma_j}(a_{ji}-a_{ij}) \text{ if } \sigma_i \neq 0, \\ g'(0; a_{ji}-a_{ij}) \text{ if } \sigma_i = -\sigma_j = 0, \end{cases}$$

$$(\Lambda_{\alpha\gamma})_{ij} = \begin{cases} \sqrt{2}\dfrac{g(\sigma_i)}{\sigma_i} b_{ij} \text{ if } \sigma_i \neq 0, \\ g'(0, \sqrt{2}b_{ij}) \text{ if } \sigma_i = 0, \end{cases}$$

$$(\Lambda_{\beta\alpha})_{ij} = \begin{cases} \dfrac{g(\sigma_i)+g(\sigma_j)}{\sigma_i+\sigma_j}(a_{ij}-a_{ji}) \text{ if } \sigma_i \neq 0, \\ g'(0; a_{ij}-a_{ji}) \text{ if } -\sigma_i = \sigma_j = 0, \end{cases}$$

$$(\Lambda_{\beta\beta})_{ij} = \begin{cases} -\dfrac{g(\sigma_i)-g(\sigma_j)}{\sigma_i-\sigma_j}(a_{ij}+a_{ji}) & \text{if } \sigma_i \neq \sigma_j, \\ g'(-\sigma_i; -(a_{ij}+a_{ji})) = -g'(\sigma_i; a_{ij}+a_{ji}) & \text{if } \sigma_i = \sigma_j, \end{cases}$$

$$(\Lambda_{\beta\gamma})_{ij} = \begin{cases} \sqrt{2}\dfrac{g(\sigma_i)}{\sigma_i} b_{ij} \text{ if } \sigma_i \neq 0, \\ g'(0; \sqrt{2}b_{ij}) \text{ if } \sigma_i = 0, \end{cases} \qquad (\Lambda_{\gamma\alpha})_{ij} = \begin{cases} \sqrt{2}\dfrac{g(\sigma_j)}{\sigma_j} b_{ji}, \text{ if } \sigma_j \neq 0, \\ g'(0; \sqrt{2}b_{ji}) \text{ if } \sigma_j = 0, \end{cases}$$

$$(\Lambda_{\gamma\beta})_{ij} = \begin{cases} \sqrt{2}\dfrac{g(\sigma_j)}{\sigma_j} b_{ji} \text{ if } \sigma_j \neq 0, \\ g'(0; \sqrt{2}b_{ji}) \text{ if } \sigma_j = 0, \end{cases} \qquad (\Lambda_{\gamma\gamma})_{ij} = 0.$$

By calculation, we know that

$$\begin{aligned} \Psi'(Y; H) &= \frac{1}{2}\begin{bmatrix} U & U & 0 \\ V_1 & -V_1 & \sqrt{2}V_2 \end{bmatrix} \Lambda \begin{bmatrix} U^T & V_1^T \\ U^T & -V_1^T \\ 0 & \sqrt{2}V_2^T \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} M_1 & M_2 \\ M_3 & M_4 \end{bmatrix}, \end{aligned}$$

where

$$
\begin{aligned}
M_1 &= U(\Lambda_{\alpha\alpha} + \Lambda_{\beta\alpha}\Lambda_{\alpha\beta} + \Lambda_{\beta\beta})U^T, \\
M_2 &= U(\Lambda_{\alpha\alpha} + \Lambda_{\beta\alpha} - \Lambda_{\alpha\beta} - \Lambda_{\beta\beta})V_1^T + \sqrt{2}U(\Lambda_{\alpha\gamma} + \Lambda_{\beta\gamma})V_2^T, \\
M_3 &= V_1(\Lambda_{\alpha\alpha} - \Lambda_{\beta\alpha} + \Lambda_{\alpha\beta} - \Lambda_{\beta\beta})U^T + \sqrt{2}V_2(\Lambda_{\gamma\alpha} - \Lambda_{\gamma\beta})U^T, \\
M_4 &= V_1(\Lambda_{\alpha\alpha} - \Lambda_{\beta\alpha} - \Lambda_{\alpha\beta} + \Lambda_{\beta\beta})V_1^T + \sqrt{2}V_2(\Lambda_{\gamma\alpha} - \Lambda_{\gamma\beta})V_1^T + \sqrt{2}V_1(\Lambda_{\alpha\gamma} - \Lambda_{\beta\gamma})V_2^T.
\end{aligned}
$$

By the definition of Λ, we know that

$$
\Lambda_{\alpha\alpha} = -\Lambda_{\beta\beta},\ \ \Lambda_{\alpha\beta} = -\Lambda_{\beta\alpha},\ \ \Lambda_{\alpha\gamma} = \Lambda_{\beta\gamma},\ \ \Lambda_{\gamma\alpha} = \Lambda_{\gamma\beta},\ \ \Lambda_{\beta\alpha} = (\Lambda_{\alpha\beta})^T,\ \ \Lambda_{\gamma\alpha} = (\Lambda_{\alpha\gamma})^T,
$$

which shows that $M_1 = M_2 = 0$, $M_3 = M_4^T$. Together with (2.39), we obtain that

$$
G'(Y; H) = U(\Lambda_{\alpha\alpha} - \Lambda_{\alpha\beta})V_1^T + \sqrt{2}U\Lambda_{\alpha\gamma}V_2^T.
$$

Suppose instead that G is directionally differentiable at Y with singular values $\sigma_1, \ldots, \sigma_p$. Fix any $U \in \mathcal{O}^p$ and $V \in \mathcal{O}^q$ satisfying $Y = U[\mathrm{diag}[\sigma_1, \ldots, \sigma_p]\ 0]V^T$. For each $i \in \{1, \ldots, p\}$ and each $d_i \in \Re$, let $H := U[\mathrm{diag}[0, \ldots, d_i, \ldots, 0]\ ]V^T$. Since $G'(Y; H) = U[\mathrm{diag}[0, \ldots, g'(\sigma_i; d_i), \ldots, 0]\ 0]V^T$ exists, we get that $g'(\sigma_i, d_i)$ is well defined. □

Lemma 2.3.10. *If g is (strongly) semismooth at σ_i, $i = 1, \ldots, p$. Then, Ψ is (strongly) semismooth at Y.*

Proof. We give below a proof for the strong semismoothness case. The semismoothness case can be derived in a similar way.

By Theorem 2.3.5 and 2.3.9, we know that G is locally Lipschitz continuous and is directionally differentiable at Y. Since g is strongly semismooth at σ_i, $i = 1, \ldots, p$ and in addition g is an odd function, it is also strongly semismooth at $-\sigma_i$, $i = 1, \ldots, p$.

We first show that if $\sigma_1 \geq \ldots \geq \sigma_p > 0$, Ψ is strongly semismooth at Y. From the decomposition of Ψ at $\bar{Y} = Y + H$, we have

$$\begin{aligned}\Psi(\bar{Y}) - \Psi(Y) &= \sum_{j=1}^{2m} g(\mu_j)[P_j(\Xi(\bar{Y})) - P_j(\Xi(Y))] + \\ &\quad + \sum_{j=1}^{2m} \sum_{k=s_j+1}^{s_j+r_j} [g(\lambda_k(\Xi(\bar{Y}))) - g(\mu_j)] e_k(\Xi(\bar{Y}))(e_k(\Xi(\bar{Y})))^T.\end{aligned}$$

Since $P_j(\cdot)$ are twice continuously differentiable near $\Xi(\bar{Y})$, we have

$$\sum_{j=1}^{2m} g(\mu_j)[P_j(\Xi(\bar{Y})) - P_j(\Xi(Y))] = \sum_{j=1}^{2m} g(\mu_j) P_j'(\Xi(\bar{Y}))\Xi(H) + O(\|\Xi(H)\|^2).$$

It follows from Lemma 2.3.3 the eigenvalue function $\lambda_k(\cdot)$ are strongly semismooth and $g(\cdot)$ are strongly semismooth at λ_i. Thus, for $k \in \{s_j + 1, \ldots, s_j + r_j\}$ and $j \in \{1, \ldots, 2m\}$, we have that

$$g(\lambda_k(\Xi(\bar{Y}))) - g(\mu_j) = g'(\lambda_k(\Xi(\bar{Y})), \lambda_k'(\Xi(\bar{Y}), \Xi(H))) + O(\|H\|^2).$$

Since $\|e_k(\Xi(\bar{Y}))(e_k(\Xi(\bar{Y})))^T\|$ are uniformly bounded, we get that

$$\begin{aligned}\Psi(\bar{Y}) - \Psi(Y) &= \sum_{j=1}^{2m} g(\lambda_j) P_j'(\Xi(Y))\Xi(H) \\ &\quad + \sum_{i=1}^{2p} g'(\lambda_i(\Xi(\bar{Y})), \lambda_i'(\Xi(\bar{Y}), \Xi(H))) e_i(\Xi(\bar{Y}))(e_i(\Xi(\bar{Y})))^T + O(\|H\|^2).\end{aligned}$$

By Lemma 2.3.8 we know that for an appropriate choice of $e_i(\Xi(\bar{Y}))$, one has

$$\Psi(\bar{Y}) - \Psi(Y) = \Psi'(\bar{Y}, H) + O(\|H\|^2),$$

which implies that Ψ is strongly semismooth at Y.

We next assume that $\sigma_p = 0$. Then g is strongly semismooth at 0. By [3, Proposition 4.10], we know that F is strongly semismooth at $\Xi(Y)$ and hence

$$F(\Xi(Y) + \Xi(H)) - F(\Xi(Y)) = F'(\Xi(Y) + \Xi(H); \Xi(H)) + O(\|\Xi(H)\|^2),$$

which, together with (2.37), yields that

$$\Psi(Y + H) - \Psi(Y) = \Psi'(Y + H; H) + O(\|H\|^2).$$

Thus, Ψ is strongly semismooth at Y. $\square$

Theorem 2.3.11. *If g is (strongly) semismooth at σ_i, $i = 1, \ldots, p$. Then G is (strongly) semismooth at Y.*

Proof. By Lemma 2.3.10, we obtain that

$$\begin{aligned}
\Psi(\bar{Y}) - \Psi(Y) - \Psi'(\bar{Y}, H) &= \begin{bmatrix} 0 & G(\bar{Y}) - G(Y) \\ (G(\bar{Y}) - G(Y))^T & 0 \end{bmatrix} - \\
&\quad - \begin{bmatrix} 0 & G'(\bar{Y}; H) \\ (G'(\bar{Y}; H))^T & 0 \end{bmatrix} = O(\|H\|^2),
\end{aligned}$$

which implies that $G(\bar{Y}) - G(Y) - G'(\bar{Y}; H) = O(\|H\|^2)$ and hence G is strongly semismooth at Y. $\square$

Example 2.3.1. For some given $\tau > 0$, let $g : \Re_+ \to \Re$ be defined by

$$g(t) := (t - \tau)_+.$$

Note that $g(0) = 0$ in this case. We then get that the extended function $g : \Re \to \Re$ has the following form:

$$g(t) = \begin{cases} (t - \tau)_+ & \text{if } t \geq 0, \\ -(-t - \tau)_+ & \text{if } t < 0, \end{cases}$$

that is, $g(t) = (t - \tau)_+ - (-t - \tau)_+$. It can be readily seen that the nonsymmetric matrix-valued function G associated with g becomes the soft thresholding operator. Since g is strongly semismooth everywhere, by Theorem 2.3.11, we can get the result that the soft thresholding operator is strongly semismooth everywhere, which has been shown by Jiang, Sun and Toh [9]. Therefore, our results on the properties of the nonsymmetric matrix-valued function G generalize the results of Jiang, Sun and Toh [9] which considers the case of the soft thresholding operator to general cases.

Chapter 3

Smoothing functions

In this chapter, we will discuss the continuity and differential properties of the smoothing function for the nonsmooth nonsymmetric matrix function. We first give the definition of the smoothing function and then show that the smoothing function inherits the properties of locally Lipschitz continuity, continuous differentiability, directional differentiability and (strongly) semismoothness from the smoothing function h of the real-valued function g.

3.1 Definition

Let $h : \Re_{++} \times \Re_+ \to \Re$ be the smoothing function of $g : \Re_+ \mapsto \Re$. Now we define the smoothing function H of the nonsmooth nonsymmetric matrix value function G as follows:

$$H(\epsilon, Y) := U[\mathrm{diag}[h(\epsilon,\sigma_1), \ldots, h(\epsilon,\sigma_p)] \quad 0]V^T. \tag{3.1}$$

As h is only defined on $\Re_{++} \times \Re_+$, for later discussion we define the extended function $\hat{h} : \Re \to \Re$ by

$$\hat{h}(\epsilon, y) := \begin{cases} h(\epsilon, y) - h(\epsilon, 0) & \text{if } t \geq 0, \\ -(h(\epsilon, -y) - h(\epsilon, 0)) & \text{if } t < 0. \end{cases}$$

We can easily see that $\hat{h}$ is an odd function and

$$H(\epsilon, Y) = U[h(\epsilon, \Sigma)\ 0]V^T = U[\hat{h}(\epsilon, \Sigma)\ 0]V^T + h(\epsilon, 0)Y.$$

For the convenience of discussion, we use h and H to represent $\hat{h}$ and $\hat{H}$, respectively.

In order to study the properties of H, we define the function $\Phi : \Re \times \Re^{p\times q} \to \mathcal{S}^{p+q}$ by

$$\begin{aligned}
\Phi(\epsilon, Y) &:= F(\epsilon, \Xi(Y)) \\
&= Q\mathrm{diag}[h(\epsilon, \sigma_1), \ldots, h(\epsilon, \sigma_p), h(\epsilon, -\sigma_1), \ldots, h(\epsilon, -\sigma_p), h(\epsilon, 0), \ldots, h(\epsilon, 0)]Q^T \\
&= Q[\mathrm{diag}[h(\epsilon, \sigma_1), \ldots, h(\epsilon, \sigma_p), h(\epsilon, -\sigma_1), \ldots, h(\epsilon, -\sigma_p)]\ \ 0]Q^T,
\end{aligned}$$

where Ξ and Q are given by (2.3) and (2.4), respectively. By the same calculation as in (2.7), we can get

$$\Phi(\epsilon, Y) = \begin{bmatrix} 0 & H(\epsilon, Y) \\ (H(\epsilon, Y))^T & 0 \end{bmatrix}. \tag{3.2}$$

3.2 Continuity, differential properties and semismoothness

In this section, we study the continuity, differential and semismooth properties of the smoothing function of the nonsmooth nonsymmetric matrix-valued function.

Theorem 3.2.1. *If h is locally Lipschitz continuous at $(\epsilon, \sigma_1), \ldots, (\epsilon, \sigma_p)$. Then H is locally Lipschitz continuous at (ϵ, Y).*

Proof. Assume that h is locally Lipschitz continuous at $(\epsilon, \sigma_1), \ldots, (\epsilon, \sigma_p)$. Then, it is also locally Lipschitz continuous at $(\epsilon, -\sigma_1), \ldots, (\epsilon, -\sigma_p)$. We show below that Φ is locally Lipschitz continuous at (ϵ, Y).

We first consider the case of $\sigma_1 \geq \ldots \geq \sigma_p > 0$. Since

$$\Phi(\epsilon, Y) = \sum_{i=1}^{2m} h(\epsilon, \mu_i) P_i(\Xi(Y)) + \sum_{i=1}^{2m} \sum_{k=s_i+1}^{s_i+r_i} [h(\epsilon, \lambda_k(\Xi(Y))) - h(\epsilon, \mu_i)] e_k(\Xi(Y)) e_k(\Xi(Y))^T,$$

we obtain that

$$\begin{aligned} \|\Phi(\tau, \bar{Y}) - \Phi(\epsilon, Y)\| \ &\leq \ \sum_{i=1}^{2m} |h(\epsilon, \mu_i)| \|P_i(\Xi(\bar{Y})) - P_i(\Xi(Y))\| \\ &\quad + \sum_{i=1}^{2m} \sum_{k=s_i+1}^{s_i+r_i} |h(\tau, \lambda_k(\Xi(\bar{Y})) - h(\epsilon, \mu_i)| \|e_k(\Xi(\bar{Y})) e_k(\Xi(\bar{Y}))^T\|. \end{aligned}$$

Since $\|e_k(\Xi(\bar{Y})) e_k(\Xi(\bar{Y}))^T\|$ are uniformly bounded, it follows from locally Lipschitz continuity of the eigenvalue function $\lambda_k(\cdot)$ and of $P_k(\cdot)$ that Φ is locally Lipschitz continuous at (ϵ, Y).

We next consider the case of $\sigma_p = 0$. Since h is locally Lipschitz continuous at $(\epsilon, 0)$, from [23, Proposition 3.3], we know that Φ is locally Lipschitz continuous at $(\epsilon, \Xi(Y))$, i.e., there exists $L > 0$ such that

$$\|F(\epsilon + \tau, \Xi(Y) + \Xi(H)) - F(\epsilon, Y)\| \leq L \|(\tau, \Xi(H))\|,$$

which, together with the definition of Φ, implies that

$$\|\Phi(\epsilon + \tau, Y + H) - \Phi(\epsilon, Y)\| \leq \hat{L} \|(\tau, H)\|$$

for some $\hat{L} > 0$. Thus, Φ is locally Lipschitz continuous. It follows from (3.2) that H is locally Lipschitz continuous at (ϵ, Y). $\square$

Theorem 3.2.2. *Given $(\epsilon, Y) \in \Re_{++} \times \Re^{p \times q}$, if h is continuously differentiable at (ϵ, σ_i) $(i = 1, \ldots, p)$, then H is continuously differentiable at (ϵ, Y). Moreover, for any $(\tau, \Delta Y) \in \Re_{++} \times \Re^{p \times q}$, the derivative of H is given by*

$$H'(\epsilon, Y)(\tau, \Delta Y) = H'_Y(\epsilon, Y) \Delta Y + H'_\epsilon(\epsilon, Y) \tau.$$

Proof. Fix $\epsilon > 0$. By Theorem 2.2.6 and 2.2.4, we know that $H(\epsilon, \cdot)$ is continuously differentiable around $Y \in \Re^{p\times q}$ and for any $\Delta Y \in \Re^{p\times q}$,

$$H'_Y(\epsilon, Y)\Delta Y = \frac{1}{2}U[\Omega_{\alpha\alpha} \circ (A^T + A) + \Omega_{\alpha\beta} \circ (A - A^T)]V_1^T + U(\Omega_{\alpha\gamma} \circ B)V_2^T,$$

where $A := U^T \Delta Y V_1 \in \Re^{p\times p}$, $B := U^T \Delta Y V_2 \in \Re^{p\times(q-p)}$ and the matrices $\Omega_{\alpha\alpha}$, $\Omega_{\alpha\beta}$, $\Omega_{\alpha\gamma}$ are defined by

$$(\Omega_{\alpha\alpha})_{ij} := \begin{cases} \dfrac{h(\epsilon,\sigma_i) - r(\epsilon,\sigma_j)}{\sigma_i - \sigma_j} & \text{if } \sigma_i \neq \sigma_j, \\ h'(\epsilon,\sigma_i) & \text{if } \sigma_i = \sigma_j, \end{cases}$$

$$(\Omega_{\alpha\beta})_{ij} := \begin{cases} \dfrac{h(\epsilon,\sigma_i) + h(\epsilon,\sigma_j)}{\sigma_i + \sigma_j} & \text{if } \sigma_i \neq -\sigma_j, \\ h'(\epsilon,0) & \text{if } \sigma_i = -\sigma_j = 0, \end{cases}$$

$$(\Omega_{\alpha\gamma})_{ij} := \begin{cases} \dfrac{h(\epsilon,\sigma_i)}{\sigma_i} & \text{if } \sigma_i \neq 0, \\ h'(\epsilon,0) & \text{if } \sigma_i = 0. \end{cases}$$

For fixed $Y \in \Re^{p\times q}$, since $h(\cdot,\sigma_i)$ $(i = 1,\ldots,p)$ are continuously differentiable on $\Re_{++}$, we know that $H(\cdot, Y)$ is continuously differentiable on $\Re_{++}$ and for any $\tau \in \Re$, we have

$$H'_\epsilon(\epsilon, Y)\tau = \tau U[\mathrm{diag}(h'_\epsilon(\epsilon,\sigma_1),\cdots,h'_\epsilon(\epsilon,\sigma_p)) \ \ 0]V^T.$$

Since $h'_\epsilon(\nu,\sigma_i(Z)) \to h'_\epsilon(\epsilon,\sigma_i(Y))$ as $\nu \to \epsilon, Z \to Y$, we have that $\|H'_\epsilon(\nu, Z) - H'_\epsilon(\epsilon, Y)\| \to 0$, which implies that H'_ϵ is continuous.

The above arguments show that H is differentiable at (ϵ, Y) and

$$H'(\epsilon, Y)(\tau, \Delta Y) = H'_Y(\epsilon, Y)\Delta Y + H'_\epsilon(\epsilon, Y)\tau.$$

Since $H'_Y(\epsilon, Y)$ and $H'_\epsilon(\epsilon, Y)$ are continuous, H' is continuous and thus H is continuously differentiable. $\square$

Let $\mu_1, \ldots, \mu_m$ be the distinct values of $\sigma_1, \ldots, \sigma_p$, $\mu_{m+1}, \ldots, \mu_{2m}$ be the distinct value of $-\sigma_1, \ldots, -\sigma_p$ and $\mu_{2m+1} = 0$ be the value of $\lambda_i(\Xi(Y))$ with $i \geq 2p+1$.

Theorem 3.2.3. *For any $(\epsilon, Y) \in \Re_{++} \times \Re^{p\times q}$, if h is directionally differentiable at (ϵ, σ_i), $i = 1, \ldots, p$, then H is directionally differentiable at (ϵ, Y).*

Proof. Since Y is directionally differentiable at (ϵ, σ_i), it is also directionally differentiable at $(\epsilon, -\sigma_i)$. First we show that Φ is directionally differentiable at (ϵ, Y).

For any $t > 0$, $(\tau, H) \in \Re_{++} \times \Re^{p\times q}$, let $\bar{Y} = Y + tH$ and $\bar{\epsilon} = t\tau + \epsilon$. We first consider the case $\sigma_1 \geq \ldots \geq \sigma_p > 0$. Since

$$h(\bar{\epsilon}, \mu_{2m+1}(\Xi(\bar{Y})) = h(\epsilon, \mu_{2m+1}) = 0$$

and

$$\Phi(\bar{\epsilon}, \bar{Y}) = \sum_{k=1}^{2m} h(\epsilon, \mu_k) P_k(\Xi(\bar{Y})) + \sum_{k=1}^{2p} [h(\bar{\epsilon}, \lambda_k(\Xi(\bar{Y}))) - h(\epsilon, \lambda_k)] e_k(\Xi(\bar{Y})) e_k(\Xi(\bar{Y}))^T,$$

we have

$$\begin{aligned} \Phi(\bar{\epsilon}, \bar{Y}) - \Phi(\epsilon, Y) &= \sum_{k=1}^{2m} h(\epsilon, \mu_k)[P_k(\Xi(\bar{Y})) - P_k(\Xi(Y))] \\ &+ \sum_{k=1}^{2p} [h(\bar{\epsilon}, \lambda_k(\Xi(\bar{Y}))) - h(\epsilon, \lambda_k)] e_k(\Xi(\bar{Y})) e_k(\Xi(\bar{Y}))^T. \end{aligned}$$

Let

$$A = \sum_{k=1}^{2m} h(\epsilon, \mu_k)[P_k(\Xi(\bar{Y})) - P_k(\Xi(Y))]$$

and

$$B = \sum_{k=1}^{2p} [h(\bar{\epsilon}, \lambda_k(\Xi(\bar{Y}))) - h(\epsilon, \lambda_k)] e_k(\Xi(\bar{Y})) e_k(\Xi(\bar{Y}))^T.$$

Then, we easily know that

$$\lim_{t \downarrow 0} t^{-1} A = \sum_{k=1}^{2m} h(\epsilon, \mu_k) P_k'(\Xi(Y)) \Xi(H).$$

Next we calculate the directional derivative of B. Since h is directionally differentiable at (ϵ, λ_k), $k = 1, \dots, 2p$, together with the directionally differentiable of $\lambda_k(\Xi(Y))$, $k = 1, \dots, 2p$, we obtain that for each k,

$$\lim_{t \downarrow 0} t^{-1}(h(\bar{\epsilon}, \lambda_k(\bar{Y})) - h(\epsilon, \lambda_k(Y))) = h'((\epsilon, \lambda_k(Y)); (\tau, \lambda_k'(Y; \Xi(H)))).$$

Thus, $\lim_{t \downarrow 0} t^{-1} B = \sum_{k=1}^{2p} h'((\epsilon, \lambda_k(Y)); (\tau, \lambda_k'(Y; \Xi(H))))e_k(\Xi(Y))e_k(\Xi(Y))^T$. This means that Φ is directionally differentiable at (ϵ, Y) and

$$\begin{aligned} \Phi'((\epsilon, Y); (\tau, H)) &= \sum_{k=1}^{2m} h(\epsilon, \mu_k) P_k'(\Xi(Y))\Xi(H) \\ &+ \sum_{k=1}^{2p} h'((\epsilon, \lambda_k(Y)); (\tau, \lambda_k'(Y; \Xi(H))))e_k(\Xi(Y))e_k(\Xi(Y))^T. \end{aligned}$$

We turn to the case $\sigma_1 \geq \dots \sigma_p = 0$. Then, h is also directionally differentiable at $(\epsilon, 0)$. From [23, Proposition 3.1], we know that F is directionally differentiable at $(\epsilon, \Xi(Y)$. Since

$$\begin{aligned} \Phi'((\epsilon, Y); (\tau, H)) &= \lim_{t \downarrow 0} \frac{\Phi(\epsilon + t\tau, Y + tH) - \Phi(\epsilon, Y)}{t} \\ &= \lim_{t \downarrow 0} \frac{F(\epsilon + t\tau, \Xi(Y) + t\Xi(H)) - F(\epsilon, \Xi(Y))}{t} \\ &= F'((\epsilon, \Xi(Y)); (\tau, \Xi(H))), \end{aligned} \tag{3.3}$$

we obtain that Φ is directionally differentiable at (ϵ, Y). From (3.2), we conclude that H is also directionally differentiable at (ϵ, Y). □

In the following, we show the (strong) semismoothness of H.

Theorem 3.2.4. *If h is (strongly) semismooth at (ϵ, σ_i), $i = 1, \dots, p$. Then H is (strongly) semismooth at (ϵ, Y).*

Proof. We give below a proof for the strong semismoothness case. The semismoothness case can be derived in a similar way.

We first show that Φ is strongly semismooth at (ϵ, Y) by considering two cases.

Case 1: $\sigma_1 \geq \ldots \geq \sigma_p > 0$. For any $(\tau, H) \in \Re_{++} \times \Re^{p\times q}$, we have

$$
\begin{aligned}
&\Phi(\epsilon+\tau, Y+H) - \Phi(\epsilon, Y) \\
&= \sum_{k=1}^{2m} h(\epsilon, \mu_k)[P_k(\Xi(Y+H)) - P_k(\Xi(Y))] \\
&+ \sum_{k=1}^{2p} [h(\tau, \lambda_k(\Xi(Y+H))) - h(\epsilon, \lambda_k(\Xi(Y)))] e_k(\Xi(Y+H)) e_k(\Xi(Y+H))^T.
\end{aligned}
$$

Since $P_j(\cdot)$ are twice continuously differentiable near $\Xi(\bar{Y})$, we have

$$
\sum_{k=1}^{2m} h(\epsilon, \mu_k)[P_k(\Xi(Y+H)) - P_k(\Xi(Y))] = \sum_{k=1}^{2m} h(\epsilon, \mu_k) P_j'(\Xi(\bar{Y}))\Xi(H) + O(\|\Xi(H)\|^2).
$$

It follows from Lemma 2.3.3 and the strong semismoothness of $h(\cdot,\cdot)$ at (ϵ, λ_i) that for $k \in \{1, \ldots, 2p\}$,

$$
h(\epsilon+\tau, \lambda_k(\Xi(\bar{Y}))) - h(\epsilon, \lambda_k) = h'((\epsilon+\tau, \lambda_k(\Xi(\bar{Y}))); \tau, \lambda_k'(\Xi(\bar{Y}), \Xi(H))) + O(\|\Xi(H)\|^2).
$$

Since $\|e_k(\Xi(\bar{Y}))(e_k(\Xi(\bar{Y})))^T\|$ are uniformly bounded, we get that

$$
\begin{aligned}
&\sum_{k=1}^{2p} [h(\tau, \lambda_k(\Xi(Y+H))) - h(\epsilon, \lambda_k(\Xi(Y)))] e_k(\Xi(Y+H)) e_k(\Xi(Y+H))^T \\
= &\sum_{k=1}^{2p} h'((\epsilon+\tau, \lambda_k(\Xi(\bar{Y}))); \tau, \lambda_k'(\Xi(\bar{Y}), \Xi(H))) + O(\|\Xi(H)\|^2).
\end{aligned}
$$

Consequently,

$$
\Phi(\epsilon+\tau, Y+H) - \Phi(\epsilon, Y) = \Phi'((\epsilon+\tau, Y+H); (\tau, H)) + O(\|H\|^2),
$$

which means that Φ is strongly semismooth at (ϵ, Y).

Case 2: $\sigma_p = 0$. Then, the assumption implies that h is strongly semismooth at $(\epsilon, 0)$. From [23, Theorem 4.2], we know that F is strongly semismooth at $(\epsilon, \Xi(Y))$, that is,

$$
\begin{aligned}
&F(\epsilon+\tau, \Xi(Y)+\Xi(H)) - F(\epsilon, \Xi(Y)) \\
= \; &F'((\epsilon+\tau, \Xi(Y)+\Xi(H)); (\tau, \Xi(H))) + O(\|(\tau, \Xi(H))\|^2).
\end{aligned}
$$

By the definition of Φ, we get that

$$\Phi(\epsilon+\tau, Y+H) - \Phi(\epsilon, Y) = \Phi'((\epsilon+\tau, Y+H); (\tau, H)) + O(\|(\tau, H)\|^2).$$

which yields that Φ is strongly semismooth at (ϵ, Y). From (3.2), we conclude that H is (strongly) semismooth. $\square$

Example 3.2.1. We use Huber smoothing function $h : \Re_{++} \times \Re_{+} \to \Re$ to smooth the soft thresholding operator, which is defined by

$$h(\epsilon, t) = \begin{cases} t & \text{if } t \geq \frac{\epsilon}{2}, \\ \frac{1}{2\epsilon}(t+\frac{\epsilon}{2})^2 & \text{if } -\frac{\epsilon}{2} < t < \frac{\epsilon}{2}, \\ 0 & \text{if } t \leq -\frac{\epsilon}{2}. \end{cases}$$

Then the smoothing function for the soft thresholding operator is

$$h_\tau(\epsilon, t) = \begin{cases} t-\tau & \text{if } t \geq \frac{\epsilon}{2}+\tau, \\ \frac{1}{2\epsilon}(t-\tau+\frac{\epsilon}{2})^2 & \text{if } \tau-\frac{\epsilon}{2} < t < \tau+\frac{\epsilon}{2}, \\ 0 & \text{if } t \leq \tau-\frac{\epsilon}{2}. \end{cases}$$

We define the extended function $\hat{h}_\tau : \Re \to \Re$ by,

$$\hat{h}_\tau(\epsilon, t) := \begin{cases} h(\epsilon, t) - \dfrac{(\tau-\frac{\epsilon}{2})^2}{2\epsilon} & \text{if} \quad t \geq 0, \\ -(h(\epsilon, -t) - \dfrac{(\tau-\frac{\epsilon}{2})^2}{2\epsilon}) & \text{if} \quad t < 0. \end{cases}$$

Since h_τ is strongly semismooth on $\Re_+ \times \Re$, from the above theorem H_τ is strongly semismooth on $\Re_+ \times \Re^{p\times q}$.

Chapter 4

Conclusions

In this thesis, we studied various continuity and differentiability properties of the nonsymmetric matrix-valued function and the smoothing function of the nonsmooth nonsymmetric matrix-valued function. In particular, we showed that the nonsymmetric matrix-valued function G and its smoothing function H inherit the continuity, differentiability, continuous differentiability, locally Lipschitz continuity, directional differentiability and (strongly) semismoothness from the real-valued function g and the smoothing function h of g, respectively. These results can be applied to address some basic issues on the analysis of semismooth/smoothing Newton methods arising from the nonsymmetric matrix optimization problems. These issues are, however, beyond the scope of the thesis. We leave them for future research.

Appendix A

Basic concepts

This appendix reviews some basic properties of vector-valued functions. These properties are continuity, (locally) Lipschitz continuity, directional differentiability, continuous differentiability and (ρ-order) semismoothness.

Throughout this appendix, we assume that $\mathcal{X}$ and $\mathcal{Y}$ are two finite dimensional real vector spaces and $\mathcal{W}$ is an open set in $\mathcal{Y}$. We consider a function $\Theta : \mathcal{W} \to \mathcal{Y}$.

We say that Θ is *continuous* at $x \in \mathcal{W}$ if

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

and Θ is continuous in $\mathcal{W}$ if it is continuous at every $x \in \mathcal{W}$. The function Θ is said to be *locally Lipschitz continuous* at $x \in \mathcal{W}$ if there exists $\kappa > 0$ and $\delta > 0$ such that

$$\|\Theta(y) - \Theta(z)\| \le \kappa \|y - z\|, \quad \forall y \in \mathcal{W} \ \text{ such that } \ \|y - x\| \le \delta, \ \|z - x\| \le \delta;$$

and Θ is locally Lipschitz continuous in $\mathcal{W}$ if it is locally Lipschitz continuous at every $x \in \mathcal{W}$. If δ can be taken to be $+\infty$, Θ is said to be *Lipschitz continuous* with Lipschitz constant κ.

We say that Θ is *directionally differentiable* at $x \in \mathcal{W}$ if for any $h \in \mathcal{X}$,

$$\Theta'(x;h) := \lim_{t \downarrow 0} \frac{\Theta(x+th) - \Theta(x)}{t} \quad \text{exists;}$$

and Θ is directionally differentiable on $\mathcal{W}$ if it is directionally differentiable at every $x \in \mathcal{W}$. The function Θ is said to be (Fréchet) differentiable at $x \in \mathcal{W}$ if there exists a linear operator $\Theta'(x) : \mathcal{W} \to \mathcal{Y}$ such that

$$\Theta(x+h) - \Theta(x) - \Theta'(x)h = o(\|h\|).$$

Moreover, Θ is *continuously differentiable* in $\mathcal{W}$ if Θ is differentiable at every $x \in \mathcal{W}$ and Θ' is continuous.

If Θ is a locally Lipschitz continuous function in $\mathcal{W}$. Then, by Rademacher's theorem [17, Chapter 9.J] we know that Θ is almost everywhere differentiable in $\mathcal{W}$. Let $\mathcal{W}_\Theta$ denote the set of points in $\mathcal{W}$ where Θ is differentiable. Then, the *Clarke's generalized Jacobian* of Θ at $x \in \mathcal{W}$ is defined by (cf. [5])

$$\partial\Theta(x) := \text{conv}\{\partial_B\Theta(x)\},$$

where "conv" denotes the convex hull and the *B-subdifferential* $\partial_B\Theta(x)$, defined by Qi in [15], is given by

$$\partial_B\Theta(y) := \left\{ V \,:\, V = \lim_{j \to \infty} \Theta'(x^j)\,,\ x^j \to x\,,\ x^j \in \mathcal{W}_\Theta \right\}.$$

The concept of semismoothness was first introduced by Mifflin ([14]) for functionals and was extended to vector-valued functions by Qi and Sun ([16]).

Definition A.0.1. Assume that Θ is a locally Lipschitz continuous function on $\mathcal{W}$. We say that Θ is *semismooth* at a point $x \in \mathcal{W}$ if

(i) Θ is directionally differentiable at x; and

(ii) for any $y \to x$ and $V \in \partial\Theta(y)$,

$$\Theta(y) - \Theta(x) - V(y-x) = o(\|y-x\|)\,.$$

The function Θ is said to be *ρ-order semismooth* at $x \in \mathcal{W}$ if Θ is semismooth at Θ and, for any $y \to x$ and $V \in \partial\Theta(y)$, one has

$$\Theta(y) - \Theta(x) - V(y - x) = O(\|y - x\|^{1+\rho}) \,.$$

We say that Θ is strongly semismooth at $x \in \mathcal{W}$ if it is 1-order semismooth at $x \in \mathcal{W}$.

The following result, originally shown by Sun and Sun [19], will be needed in our analysis.

Proposition A.0.5. [19, Theorem 3.7] *Suppose Θ is locally Lipschitz continuous and directionally differentiable in a neighborhood of $x \in \mathcal{W}$. Then, for any $0 < \rho < \infty$, the following two statements are equivalent:*

(I) For any $h \in \mathcal{X}$ and any $V \in \partial\Theta(x + h)$,

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

(II) For any $h \in \mathcal{X}$ such that Θ is differentiable at $x + h$,

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

Appendix B

Properties of symmetric matrix-valued functions

This appendix contains some results related to the properties of symmetric matrix-valued functions, which will be used to analyze the properties of nonsymmetric matrix-valued functions.

Let $X \in \mathcal{S}^n$ have the eigenvalue decomposition of the form:

$$X = P\mathrm{diag}[\lambda_1, \ldots, \lambda_n]P^T, \tag{B.1}$$

where P is an orthogonal matrix and $\mathrm{diag}[\lambda_1, \ldots, \lambda_n]$ denotes the $n \times n$ diagonal matrix with its ith diagonal entry λ_i. Then, for any function $f : \Re \to \Re$, we can define a matrix-valued function $F : \mathcal{S}^n \to \mathcal{S}^n$ (cf. [1, 8]), associated with f, by

$$F(X) := P\mathrm{diag}[f(\lambda_1), \ldots, f(\lambda_n)]P^T. \tag{B.2}$$

By [1, Chapter V], we know that $F(X)$ is well defined (independent of the ordering of the eigenvalues and the choice of the eigenvectors).

From [3], we know that F inherits the properties of continuity, (locally) Lipschitz continuity, directional differentiability, differentiability, continuous differentiability and semismoothness of f. For the convenience of our proof, we list below the related results in [3] and the references therein.

Proposition B.0.6. *The function F is continuous at $X \in \mathcal{S}^n$ with eigenvalues $\lambda_1, \ldots, \lambda_n$ if and only if f is continuous at $\lambda_1, \ldots, \lambda_n$.*

Proposition B.0.7. *For any $f : \Re \to \Re$, the following results hold:*

(a) F is directionally differentiable at an $X \in \mathcal{S}$ with eigenvalues $\lambda_1, \cdots, \lambda_n$ if and only if f is directionally differentiable at $\lambda_1, \cdots, \lambda_n$. Moreover, for any nonzero $H \in \mathcal{S}^n$,

$$F'(X; H) = P(F^{[1]}(\lambda; P^T HP)P^T$$

for some orthogonal matrix such that $(P^T HP)_{ij} = 0$ whenever $\lambda_i = \lambda_j$ and $i \neq j$.

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

(b) F is directionally differentiable if and only if f is directionally differentiable.

Proposition B.0.8. *The function F is locally Lipschitz continuous at $X \in \mathcal{S}^n$ with eigenvalues $\lambda_1, \ldots, \lambda_n$ if and only if f is locally Lipschitz continuous at $\lambda_1, \ldots, \lambda_n$.*

Proposition B.0.9. *The function F is differentiable at $X \in \mathcal{S}^n$ with eigenvalues $\lambda_1, \ldots, \lambda_n$ if and only if f is differentiable at $\lambda_1, \ldots, \lambda_n$. Moreover, $F'(X)$ is given by*

$$F'(X)H = P(f^{[1]}(\lambda) \circ (P^T HP))P^T, \quad \forall H \in \mathcal{S}^n$$

for some orthogonal matrix such that $X = P\mathrm{diag}[\lambda_1, \cdots, \lambda_n]P^T$, where $f^{[1]}(\lambda)$ is the symmetric matrix defined by

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

Proposition B.0.10. *Let* $f : \Re \to \Re$ *be locally Lipschitz continuous. Then, for any* $X \in \mathcal{S}^n$*, the generalized Jacobian* $\partial_B F(X)$ *is well defined and nonempty. Moreover, for any* $V \in \partial_B F(X)$*, we have*

$$VH = P(\Lambda \circ (P^T H P))P^T \quad \forall h \in \mathcal{S}^n$$

for some orthogonal matrix P *such that* $X = P\mathrm{diag}[\lambda_1, \ldots, \lambda_n]P^T$*, where "*$\circ$*" denotes the Hardmard product of two matrices and the matrix* Λ *is defined by*

$$\Lambda_{ij} = \begin{cases} \dfrac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j} & \text{if } \lambda_i \neq \lambda_j, \\ \in \partial f(\lambda_i) & \text{if } \lambda_i = \lambda_j. \end{cases}$$

Proposition B.0.11. *The function* F *is semismooth if and only if* f *is semismooth. Moreover, if* f *is* ρ*-order semismooth* ($0 < \rho < \infty$)*, then* F *is* $\min\{1, \rho\}$*-order semismooth.*

Bibliography

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

[2] J.-F. Cai, E.J. Candès and Z.W. Shen, *A singular value thresholding algorithm for matrix completion*, preprint available at `http://arxiv.org/abs/0810.3286`.

[3] X. Chen, H.-D. Qi, and P. Tseng, *Analysis of nonsmooth symmetric-matrix-valued functions with applications to semidefinite complementarity problems*, SIAM Journal on Optimization, 13 (2003), pp. 960–985.

[4] X. Chen and P. Tseng, *Non-Interior continuation methods for solving semidefinite complementarity problems*, Mathematical Programming 95 (2003), pp. 431–474.

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

[6] A. Fischer, *Solution of monotone complementarity problems with locally Lipschitzian functions*, Mathematical Programming 76 (1997), pp. 513–532.

[7] G.H. Golub, C.F. Van Loan, Matrix Computations. 3rd edn, The Johns Hopkins University Press, Baltimore, USA, 1996.

[8] R.A. Horn and C.R. Johnson, Topics in Matrix Analysis, Cambridge Universtiy Press, Cambridge, United Kindom, 1991.

[9] K.-F. Jiang, D.F. Sun, and K.-C. Toh, *A proximal point method for matrix least squares problem with nuclear norm regularization*, Manuscript, National University of Singapore, 2009.

[10] P. Lancaster, *On eigenvalues of matrices dependent on a parameter*, Numerische Mathematik 6 (1964), pp. 377–387.

[11] Y.-J. Liu, D.F. Sun, and K.-C. Toh, *An Implementable Proximal Point Algorithmic Framework for Nuclear Norm Minimization*, Technical Report, National University of Singapore, July 2009.

[12] Y.Y. Luo, Inexact smoothing Newton method for solving matrix nuclear norm problems, Master Thesis, National University of Singapore, in preparation.

[13] S.Q. Ma, D. Goldfarb and L.F. Chen, *Fixed point and Bregman iterative methods for matrix rank minimization*, Preprint, Columbia University, 2008.

[14] R. Mifflin, *Semismoothness and semiconvex functions in constrained optimization*, SIAM Journal on Control Optimization 15 (1977), pp. 959–972.

[15] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Mathematics of Operations Research 18 (1993), pp. 227–244.

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

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

[18] A. Shapiro, *On differentiability of symmetric matrix valued functions*, Published electronically in Optimization Online, 2002.

[19] D.F. Sun and J. Sun, *Semismooth matrix valued functions*, Mathematics of Operations Research 27 (2002), pp. 150–169.

[20] D.F. Sun and J. Sun, *Strong semismoothness of eigenvalues of symmetric matrices and its application to inverse eigenvalues problems* , SIAM Journal on Numerical Analysis 40 (2003), pp. 2352–2367.

[21] D.F. Sun and J. Sun, *Strong semismoothness of the Fischer-Burmeister SDC and SOC complementarity functions*, Mathematical Programming, 103 (2005), pp. 575–581.

[22] K.-C. Toh and S.W. Yun, *An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems*, Technical Report, National University of Singapore, April 2009.

[23] J.Y. Zhao, *The smoothing function of the nonsmooth matrix valued function, Master Thesis, National University of Singapore, 2004.*