An Introduction to a Class of Matrix Cone Programming

Chao Ding*, Defeng Sun† and Kim-Chuan Toh‡

This version: September 15, 2010

Abstract

In this paper, we define a class of linear conic programming (which we call matrix cone programming or MCP) involving the epigraphs of five commonly used matrix norms and the well studied symmetric cone. MCP has recently found many important applications, for example, in nuclear norm relaxations of affine rank minimization problems. In order to make the defined MCP tractable and meaningful, we must first understand the structure of these epigraphs. So far, only the epigraph of the Frobenius matrix norm, which can be regarded as a second order cone, has been well studied. Here, we take an initial step to study several important properties, including its closed form solution, calm Bouligand-differentiability and strong semismoothness, of the metric projection operator over the epigraph of the l_1, l_∞, spectral or operator, and nuclear matrix norm, respectively. These properties make it possible to apply augmented Lagrangian methods, which have recently received a great deal of interests due to their high efficiency in solving large scale semidefinite programming, to this class of MCP problems. The work done in this paper is far from comprehensive. Rather it is intended as a starting point to call for more insightful research on MCP so that it can serve as a basic tool to solve more challenging convex matrix optimization problems in years to come.

Key Words: matrix cones, metric projectors, conic optimization
AMS subject classifications: 65K05, 90C25, 90C30

1 Introduction

In this section we shall first define several convex matrix cones and then use these cones to introduce a class of matrix cone programming problems that have found important applications in many applied areas.

Let $\mathbb{R}^{m\times n}$ be the linear space of all $m \times n$ real matrices equipped with the inner product $\langle X, Y\rangle := \mathrm{Tr}(X^TY)$ for X and Y in $\mathbb{R}^{m\times n}$, where "Tr" denotes the trace, i.e., the sum of the

*Department of Mathematics, National University of Singapore, Republic of Singapore. Email: dingchao@nus.edu.sg.
†Department of Mathematics and Risk Management Institute, National University of Singapore, Republic of Singapore. Email: matsundf@nus.edu.sg.
‡Department of Mathematics, National University of Singapore, Republic of Singapore. Email: mattohkc@nus.edu.sg.

diagonal entries, of a squared matrix. Let $f \equiv \|\cdot\|$ be any norm function defined on $\mathbb{R}^{m\times n}$. The epigraph of f, denoted by $\operatorname{epi} f$,

$$\operatorname{epi} f := \{(t, X) \in \mathbb{R} \times \mathbb{R}^{m\times n} \mid t \geq f(X)\}$$

is a closed convex cone in $\mathbb{R} \times \mathbb{R}^{m\times n}$. Such a cone will be called a matrix cone for ease of reference. We use $\mathcal{K}$ to represent $\operatorname{epi} f$ or the cross product of several such closed convex cones when we choose f from the following five norms:

(i) $f(\cdot) = \|\cdot\|_F$, the Frobenius norm, i.e., for each $X \in \mathbb{R}^{m\times n}$, $\|X\|_F = (\sum_{i=1}^{m}\sum_{j=1}^{n}|x_{ij}|^2)^{1/2}$;

(ii) $f(\cdot) = \|\cdot\|_\infty$, the l_∞ norm, i.e., for each $X \in \mathbb{R}^{m\times n}$, $\|X\|_\infty = \max\{|x_{ij}| \mid 1 \leq i \leq m, 1 \leq j \leq n\}$;

(iii) $f(\cdot) = \|\cdot\|_1$, the l_1 norm, i.e., for each $X \in \mathbb{R}^{m\times n}$, $\|X\|_1 = \sum_{i=1}^{m}\sum_{j=1}^{n}|x_{ij}|$;

(iv) $f(\cdot) = \|\cdot\|_2$, the spectral or the operator norm, i.e., for each $X \in \mathbb{R}^{m\times n}$, $f(X)$ denotes the largest singular value of X; and

(v) $f(\cdot) = \|\cdot\|_*$, the nuclear norm, i.e., for each $X \in \mathbb{R}^{m\times n}$, $f(X)$ denotes the sum of the singular values of X.

That is, there exists an integer $q \geq 1$ such that

$$\mathcal{K} = \operatorname{epi} f_1 \times \operatorname{epi} f_2 \times \ldots \times \operatorname{epi} f_q\,,$$

where for each $i \geq 1$, f_i is one of the norm functions chosen from (i)-(v) on a matrix space $\mathbb{R}^{m_i\times n_i}$. Denote the Euclidean space $\mathcal{X}$ by

$$\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_q\,,$$

where for each $i \geq 1$, the natural inner product of $\mathcal{X}_i := \mathbb{R} \times \mathbb{R}^{m_i\times n_i}$ is given by

$$\langle (t, X), (\tau, Y)\rangle_{\mathcal{X}_i} := t\tau + \langle X, Y\rangle \quad \forall\, (t, X) \text{ and } (\tau, Y) \in \mathbb{R} \times \mathbb{R}^{m_i\times n_i}\,.$$

Denote the natural inner product of $\mathcal{X}$ by $\langle\cdot,\cdot\rangle_{\mathcal{X}}$. Note that for each $i \geq 1$, except for the case when $f_i(\cdot) = \|\cdot\|_F$, the cone $\operatorname{epi} f_i$ is not self-dual unless $\min\{m_i, n_i\} = 1$. So, in general the above defined closed convex cone $\mathcal{K}$ is not self-dual, i.e., $\mathcal{K} \neq \mathcal{K}^* := \{W \in \mathcal{X} \mid \langle W, Z\rangle_{\mathcal{X}} \geq 0\ \forall\, Z \in \mathcal{K}\}$, the dual cone of $\mathcal{K}$. When $f(\cdot) = \|\cdot\|_F$, $\operatorname{epi} f$ actually turns to be the second order cone (SOC) if we treat a matrix $X \in \mathbb{R}^{m\times n}$ as a vector in $\mathbb{R}^{mn}$ by stacking up the columns of X, from the first to the n-th column, on top of each other. The SOC is a well understood convex cone in the literature and thus is not the focus of this paper. We include it here for the sake of convenience in subsequent discussions.

Let $\mathcal{H}$ be a finite-dimensional real Euclidean space endowed with an inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ and its induced norm $\|\cdot\|_{\mathcal{H}}$. Let $\mathcal{Q} \in \mathcal{H}$ be the cross product of the origin $\{0\}$ and a symmetric cone in lower dimensional subspaces of $\mathcal{H}$. Note that the symmetric cone, which includes the nonnegative orthant, the SOC, and the cone of symmetric and positive semi-definite real matrices, has been completely classified [13]. Let $\mathcal{A} : \mathbb{R}^p \to \mathcal{Y} := \mathcal{H} \times \mathcal{X}$ be a linear operator. Define the natural inner product of $\mathcal{Y}$ by

$$\langle (u,w),(v,z)\rangle := \langle u,v\rangle_{\mathcal{H}} + \langle w,z\rangle_{\mathcal{X}} \quad \forall\,(u,w) \text{ and } (v,z) \in \mathcal{H}\times\mathcal{X}\,.$$

Let $\mathcal{A}^* : \mathcal{Y} \to \mathbb{R}^p$ be the adjoint of $\mathcal{A}$. Let c be a given vector in $\mathbb{R}^p$ and b an element in $\mathcal{Y}$. The matrix cone programming (MCP) we consider in this paper takes the following form

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & \mathcal{A}x \in b + \mathcal{Q}\times\mathcal{K}\,. \end{array} \tag{1}$$

The corresponding Lagrange dual of the MCP can be written as

$$\begin{array}{ll} \max & \langle b, y\rangle \\ \text{s.t.} & \mathcal{A}^* y = c\,, \\ & y \in \mathcal{Q}^* \times \mathcal{K}^*\,, \end{array} \tag{2}$$

where $\mathcal{Q}^*$ represents the dual cone of $\mathcal{Q}$. In applications, many examples can be cast in the form of (1) or (2). Below we list some of them.

Matrix norm approximation. Given matrices $B_0, B_1, \ldots, B_p \in \mathbb{R}^{m\times n}$, the matrix norm approximation problem is to find an affine combination of the matrices which has the minimal spectral norm, i.e.,

$$\min\left\{ \|B_0 + \sum_{k=1}^{p} y_k B_k\|_2 \mid y \in \mathbb{R}^p \right\}. \tag{3}$$

Such problems have been studied in the iterative linear algebra literature, e.g., [15, 47, 48], where the affine combination is a degree-p polynomial function of a given matrix.

It is easy to show that the problem (3) can be cast as a semidefinite programming (SDP) problem whose matrix variable has order $(m+n)\times(m+n)$ [52]. However, such an expansion in the order of the matrix variable implies that it can be very costly, if possible at all, to solve (3) as an SDP problem when m or n is large. Thus it is highly desirable for us to design algorithms that can solve (3) in the original matrix space $\mathbb{R}^{m\times n}$, in particular for the case when $m \ll n$ (assume $m \le n$). We believe that the contributions made in this paper would constitute a key step towards achieving that goal. More specifically, we strongly advocate approaches based on simply writing the problem (3) in the form of (1):

$$\begin{array}{ll} \min & t \\ \text{s.t.} & t \ge \|B_0 + \sum_{k=1}^{p} y_k B_k\|_2\,. \end{array}$$

We note that if for some reasons, a sparse affine combination is desired, one can add a penalty term $\lambda\|y\|_1$ with some $\lambda > 0$ to the objective function in (3) meanwhile to use $\|\cdot\|_2^2$ to replace $\|\cdot\|_2$ to get

$$\min\Big\{\|B_0 + \sum_{k=1}^{p} y_k B_k\|_2^2 + \lambda\|y\|_1 \,|\, y \in \mathbb{R}^p\Big\}. \tag{4}$$

Correspondingly, we can reformulate (4) in terms of the following two MCP forms:

$$\begin{array}{ll} \min & s + \lambda\eta \\ \text{s.t.} & (s+1)/2 \geq \sqrt{((s-1)/2)^2 + t^2}\,, \\ & t \geq \|B_0 + \sum_{k=1}^{p} y_k B_k\|_2\,, \\ & \eta \geq \|y\|_1 \end{array}$$

and

$$\begin{array}{ll} \min & t + \lambda\eta \\ \text{s.t.} & (t+1)/2 \geq \|[(t-1)/2 I_m \quad B_0 + \sum_{k=1}^{p} y_k B_k]\|_2\,, \\ & \eta \geq \|y\|_1\,, \end{array}$$

where I_m is the identity matrix of order m by m.

Matrix completion. Given a matrix $M \in \mathbb{R}^{m\times n}$ with entries in the index set Ω given, the matrix completion problem seeks to find a low-rank matrix X such that $X_{ij} \approx M_{ij}$ for all $(i,j) \in \Omega$. The problem of efficient recovery of a given low-rank matrix has been intensively studied recently. In [2], [3], [16], [22], [35], [36], etc, the authors established the remarkable fact that under suitable incoherence assumptions, an $m \times n$ matrix of rank r can be recovered with high probability from a random uniform sample of $O((m+n)r\text{polylog}(m,n))$ entries by solving the following nuclear norm minimization problem:

$$\min\Big\{\|X\|_* \,|\, X_{ij} = M_{ij} \;\forall\,(i,j) \in \Omega\Big\}.$$

The theoretical breakthrough achieved by Candès et al. has led to the rapid expansion of the nuclear norm minimization approach to model application problems for which the theoretical assumptions may not hold, for example, for problems with noisy data or that the observed samples may not be completely random. Nevertheless, for those application problems, the following model may be considered to accommodate problems with noisy data:

$$\min\Big\{\|P_\Omega(X) - P_\Omega(M)\|_F^2 + \lambda\|X\|_* \,|\, X \in \mathbb{R}^{m\times n}\Big\}, \tag{5}$$

where $P_\Omega(X)$ denotes the vector obtained by extracting the elements of X corresponding to the index set Ω in lexicographical order, and λ is a positive parameter. In the above model, the error term is measured in Frobenius norm. One can of course uses the l_1-norm or the spectral norm if those norms are more appropriate for the applications under consideration. As for the case of the matrix norm approximation, one can easily write (5) in the form of MCP.

Robust matrix completion/Robust PCA. Suppose that $M \in \mathbb{R}^{m\times n}$ is a partially given matrix for which the entries in the index set Ω are observed, but an unknown sparse subset of the observed entries may be grossly corrupted. The problem here seeks to find a low-rank matrix X and a sparse matrix Y such that $M_{ij} \approx X_{ij} + Y_{ij}$ for all $(i,j) \in \Omega$, where the sparse matrix Y attempts to identify the grossly corrupted entries in M, and X attempts to complete the "cleaned" copy of M. This problem has been considered in [4], and it is motivated by earlier results established in [5], [54]. In [4] the following convex optimization problem is solved to recover M:

$$\min\Big\{\|X\|_* + \lambda\|Y\|_1 \,|\, P_\Omega(X) + P_\Omega(Y) = P_\Omega(M)\Big\}, \tag{6}$$

where λ is a positive parameter. In robust subspace segmentation [27], a problem similar to (6) is considered, but the linear constraints are replaced by $M = MX + Y$, and $\|Y\|_1$ is replaced by $\sum_{j=1}^n \|y_j\|_2$, where y_j denotes the j-th column of Y.

In the event that the "cleaned" copy of M itself in (6) is also contaminated with random noise, the following problem could be considered to recover M:

$$\min\Big\{\|P_\Omega(X) + P_\Omega(Y) - P_\Omega(M)\|_F^2 + \rho\Big(\|X\|_* + \lambda\|Y\|_1\Big) \,|\, X, Y \in \mathbb{R}^{m\times n}\Big\}, \tag{7}$$

where ρ is a positive parameter. Again, the Frobenius norm that is used in the first term can be replaced by other norms such as the l_1-norm or the spectral norm if they are more appropriate. In any case, both (6) and (7) can be written in the form of MCP.

Structured low rank matrix approximation. In many applications, one is often faced with the problem of finding a low-rank matrix $X \in \mathbb{R}^{m\times n}$ which approximates a given target matrix M but at the same time it is required to have certain structures (such as being a Hankel matrix) so as to conform to the physical design of the application problem [9]. Suppose that the required structure is encoded in the constraints $\mathcal{A}(X) \in b + \mathcal{Q}$. Then a simple generic formulation of such an approximation problem can take the following form:

$$\min\{\|X - M\|_F \,|\, \mathcal{A}(X) \in b + \mathcal{Q},\ \mathrm{rank}(X) \le r\}. \tag{8}$$

Obviously it is generally NP hard to find the global optimal solution for the above problem. However, given a good starting point, it is quite possible that a local optimization method such as variants of the alternating minimization method may be able to find a local minimizer that is close to being globally optimal. One possible strategy to generate a good starting point for a local optimization method to solve (8) would be to solve the following penalized version of (8):

$$\min\{\|X - M\|_F + \rho \sum_{k=r+1}^{\min\{m,n\}} \sigma_k(X) \,|\, \mathcal{A}(X) \in b + \mathcal{Q}\}, \tag{9}$$

where $\sigma_k(X)$ is the k-th largest singular value of X and $\rho > 0$ is a penalty parameter. The above problem is not convex but we can attempt to solve it via a sequence of convex relaxation problems as proposed in [14] as follows. Starting with $X^0 = 0$ or any feasible matrix X^0 such that $\mathcal{A}(X^0) \in b + \mathcal{Q}$. At the k-th iteration, solve

$$\min\{\lambda\|X - X^k\|_F^2 + \|X - M\|_F + \rho(\|X\|_* - \langle H_k,\, X\rangle) \,|\, \mathcal{A}(X) \in b + \mathcal{Q}\,\} \tag{10}$$

to get X^{k+1}, where λ is a positive parameter and H_k is a sub-gradient of the convex function $\sum_{k=1}^{r} \sigma_k(\cdot)$ at the point X^k. Once again, one may easily write (10) in the form of MCP.

From the examples given in this section, it becomes quite obvious that there is a great demand for efficient and robust algorithms for solving matrix optimization problem of the form (1) or (2), especially for problems that are large scale. The question that one must answer first is if it is possible to design such algorithms at all. One obvious, maybe the biggest, discouraging fact is that for large scale MCP problems, polynomial time interior point methods (IPMs) are powerless due to the fact that the computational cost of each iteration of an IPM becomes prohibitively expensive. This is particularly discouraging given the fact that SDP would not have become so widely investigated and applied in optimization without the invention of polynomial time IPMs. So the answer to the above question appears to be negative. However, during the last few years, we have seen lots of interests in using augmented Lagrangian methods to solve large scale SDP problems. For examples, see [29, 33, 53, 56, 57]. Depending on how the inner subproblems are solved, these methods can be classified into two categories: first order alternating direction based methods [29, 33, 53] and second order semismooth Newton based methods [56, 57]. The efficiency of all these methods depends on the fact that the metric projector over the cone of symmetric and positive semi-definite matrices (in short, SDP cone) admits a closed form solution [40, 19, 50]. Furthermore, the semismooth Newton based method [56, 57] also exploits a crucial property – the strong semismoothness of this metric projector established in [44]. Keeping the progress for solving SDP in mind, we are tempted to apply the augmented Lagrangian methods to solve MCP (1) and (2). Actually, when $\mathcal{K}$ is vacuous, this has been done in the thesis [56] as the metric projector over the symmetric cone has the same desirable properties as the metric projector over the SDP cone [46]. In this paper we shall take an initial step to study the metric projector over epi f, denoted by $\Pi_{\mathrm{epi}\, f}$, with $f = \|\cdot\|_\infty$, $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_*$, respectively. In particular, we shall show that

- for any $(t, X) \in \mathbb{R} \times \mathbb{R}^{m\times n}$, $\Pi_{\mathrm{epi}\, f}(t, X)$ admits a simple closed form solution;
- $\Pi_{\mathrm{epi}\, f}(\cdot,\cdot)$ is calmly B(ouligand)-differentiable at $(t, X) \in \mathbb{R} \times \mathbb{R}^{m\times n}$ and the directional derivative of $\Pi_{\mathrm{epi}\, f}(\cdot,\cdot)$ at (t, X) along any direction in $\mathbb{R} \times \mathbb{R}^{m\times n}$ has an explicit formula; and
- $\Pi_{\mathrm{epi}\, f}(\cdot,\cdot)$ is strongly semismooth at any point in $\mathbb{R} \times \mathbb{R}^{m\times n}$.

The above result, together with the fact that the metric projector over the SOC has already been shown to have the above three properties [8], implies that the metric projector over $\mathcal{K}$ also has the above properties. Thus, these properties, together with the analogous properties of the metric projector over $\mathcal{Q}$, make it possible to apply the aforementioned augmented Lagrangian methods to solve MCP (1) and (2).

The remaining parts of this paper are organized as follows. In Section 2, we give some preliminary results, in particular on matrix functions. Section 3 is devoted to studying the projectors over the epigraphs of the l_1 and l_∞ norms. This also serves as a basis for conducting our study on the projectors over the epigraphs of the spectral and nuclear norms in Section 4. We make our conclusions in the final section.

Below are some common notations to be used:

- For any $Z \in \mathbb{R}^{m\times n}$, we denote by Z_{ij} the (i,j)-th entry of Z.
- For any $Z \in \mathbb{R}^{m\times n}$, we use z_j to represent the jth column of Z, $j = 1,\ldots,n$. Let $\mathcal{J} \subseteq \{1,\ldots,n\}$ be an index set. We use $Z_{\mathcal{J}}$ to denote the sub-matrix of Z obtained by removing all the columns of Z not in $\mathcal{J}$. So for each j, we have $Z_{\{j\}} = z_j$.
- Let $\mathcal{I} \subseteq \{1,\ldots,m\}$ and $\mathcal{J} \subseteq \{1,\ldots,n\}$ be two index sets. For any $Z \in \mathbb{R}^{m\times n}$, we use $Z_{\mathcal{I}\mathcal{J}}$ to denote the $|\mathcal{I}| \times |\mathcal{J}|$ sub-matrix of Z obtained by removing all the rows of Z not in $\mathcal{I}$ and all the columns of Z not in $\mathcal{J}$.
- We use "$\circ$" to denote the Hardamard product between matrices, i.e., for any two matrices X and Y in $\mathbb{R}^{m\times n}$ the (i,j)-th entry of $Z := X \circ Y \in \mathbb{R}^{m\times n}$ is $Z_{ij} = X_{ij}Y_{ij}$.

2 Preliminaries

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 C be a nonempty closed convex set in $\mathcal{Z}$. For any $z \in \mathcal{Z}$, let $\Pi_C(z)$ denote the metric projection of z onto C, which is the unique optimal solution to following convex optimization problem:

$$\begin{array}{ll} \min & \frac{1}{2}\|y - z\|^2 \\ \text{s.t.} & y \in C\,. \end{array}$$

It is well known [55] that $\Pi_C(\cdot)$ is globally Lipschitz continuous with modulus 1. When C is a closed convex cone, we have the following useful cone decomposition proposition due to Moreau [30].

Proposition 2.1 *Let $C \subseteq \mathcal{Z}$ be a nonempty closed convex cone. Denote by $C^o := -C^*$ the polar of C. Then any $z \in \mathcal{Z}$ can be uniquely decomposed into*

$$z = \Pi_C(z) + \Pi_{C^o}(z)\,.$$

By Proposition 2.1, for any nonempty closed convex cone C in $\mathcal{Z}$ we have the following simple observation

$$\Pi_{C^*}(z) = z + \Pi_C(-z) \quad \forall\, z \in \mathcal{Z}\,. \tag{11}$$

Let $\mathcal{O}$ be an open set in $\mathcal{Z}$ and $\mathcal{Z}'$ be another finite dimensional real Euclidean space. Suppose that $\Phi : \mathcal{O} \subseteq \mathcal{Z} \to \mathcal{Z}'$ is a locally Lipschitz continuous function on the open set $\mathcal{O}$. Then, according to Rademacher's theorem, Φ is almost everywhere differentiable (in the sense of Fréchet) in $\mathcal{O}$. Let D_Φ be the set of points in $\mathcal{O}$ where Φ is differentiable. Let $\Phi'(x)$ be the derivative of Φ at $x \in D_\Phi$. Then the B-subdifferential of Φ at $x \in \mathcal{O}$ is denoted by [34]:

$$\partial_B\Phi(x) := \left\{ \lim_{D_\Phi \ni x^k \to x} \Phi'(x^k) \right\}$$

and Clarke's generalized Jacobian of Φ at $x \in \mathcal{O}$ [10] takes the form:

$$\partial \Phi(x) = \mathrm{conv}\{\partial_B \Phi(x)\}\,,$$

where "conv" stands for the convex hull in the usual sense of convex analysis [37].

Definition 2.1 *Let $\Phi : \mathcal{O} \subseteq \mathcal{Z} \to \mathcal{Z}'$ be a locally Lipschitz continuous function on the open set $\mathcal{O}$. The function Φ is said to be G-semismooth at a point $x \in \mathcal{O}$ if for any $y \to x$ and $V \in \partial\Phi(y)$,*

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

The function Φ is said to be strongly G-semismooth at x if for any $y \to x$ and $V \in \partial\Phi(y)$,

$$\Phi(y) - \Phi(x) - V(y-x) = O(||y-x||^2)\,.$$

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

The following result taken from [44, Theorem 3.7] provides a convenient tool for proving the strong G-semismoothness of Lipschitz functions.

Lemma 2.1 *Let $\Phi : \mathcal{O} \subseteq \mathcal{Z} \to \mathcal{Z}'$ be a locally Lipschitz continuous function on the open set $\mathcal{O}$. Then Φ is strongly G-semismooth at $x \in \mathcal{O}$ if and only if for any $D_\Phi \ni y \to x$,*

$$\Phi(y) - \Phi(x) - \Phi'(y)(y-x) = O(||y-x||^2)\,.$$

In the next two subsections, we collect some useful preliminary results on symmetric and non-symmetric matrix functions for studying the projectors over the epigraphs of the spectral and nuclear norms.

2.1 The symmetric matrix functions

Let $\mathcal{S}^n$ be the space of all real $n \times n$ symmetric matrices and $\mathcal{O}^n$ be the set of all $n \times n$ orthogonal metrices. Let $X \in \mathcal{S}^n$ be given. We use $\lambda_1(X) \geq \lambda_2(X) \geq \ldots \geq \lambda_n(X)$ to denote the real eigenvalues of X (counting multiplicity) being arranged in non-increasing order. Denote $\lambda(X) := (\lambda_1(X), \lambda_2(X), \ldots, \lambda_n(X))^T \in \mathbb{R}^n$ and $\Lambda(X) := \mathrm{diag}(\lambda(X))$, where for any $x \in \mathbb{R}^n$, $\mathrm{diag}(x)$ denotes the diagonal matrix whose i-th diagonal entry is x_i, $i = 1, \ldots, n$. Let $\overline{P} \in \mathcal{O}^n$ be such that

$$X = \overline{P}\mathrm{diag}(\lambda(X))\overline{P}^T\,. \tag{12}$$

We denote the set of such matrices $\overline{P}$ in the eigenvalue decomposition (12) by $\mathcal{O}^n(X)$. Let $\overline{\mu}_1 > \overline{\mu}_2 > \ldots > \overline{\mu}_r$ be the distinct eigenvalues of X. Define

$$a_k := \{i \,|\, \lambda_i(X) = \overline{\mu}_k,\ 1 \leq i \leq n\}, \quad k = 1, \ldots, r\,. \tag{13}$$

For each $i \in \{1, \ldots, n\}$, we define $l_i(X)$ to be the number of eigenvalues that are equal to $\lambda_i(X)$ but are ranked before i (including i) and $s_i(X)$ to be the number of eigenvalues that are equal to $\lambda_i(X)$ but are ranked after i (excluding i), respectively, i.e., we define $l_i(X)$ and $s_i(X)$ such that

$$\begin{aligned}&\lambda_1(X) \geq \ldots \geq \lambda_{i-l_i(X)}(X) > \lambda_{i-l_i(X)+1}(X) = \ldots = \lambda_i(X) = \ldots = \lambda_{i+s_i(X)}(X) \\ &> \lambda_{i+s_i(X)+1}(X) \geq \ldots \geq \lambda_n(X)\,. \end{aligned} \tag{14}$$

In later discussions, when the dependence of l_i and s_i, $i = 1, \ldots, n$, on X can be seen clearly from the context, we often drop X from these notations.

Next, we list some useful results about the symmetric matrices which are needed in subsequent discussions. The inequality in the following lemma is known as Fan's inequality [12].

Lemma 2.2 *Let Y and Z be two matrices in $\mathcal{S}^n$. Then*

$$\langle Y, Z\rangle \leq \lambda(Y)^T\lambda(Z)\,. \tag{15}$$

The following observation can be checked easily.

Proposition 2.2 *Let $Q \in \mathcal{O}^n$ be an orthogonal matrix such that $Q^T\Lambda(X)Q = \Lambda(X)$. Then, we have*

$$\begin{cases} Q_{a_k a_l} = 0\,, & k, l = 1, \ldots, r,\ k \neq l\,, \quad (16)\\ Q_{a_k a_k}Q^T_{a_k a_k} = Q^T_{a_k a_k}Q_{a_k a_k} = I_{|a_k|}\,, & k = 1, \ldots, r\,. \quad (17)\end{cases}$$

For any subset $\mathcal{A}$ of a finite dimensional Euclidean space $\mathcal{Z}$, let

$$\mathrm{dist}(z, \mathcal{A}) := \inf\{\|z - y\| \mid y \in \mathcal{A}\}\,, \quad z \in \mathcal{Z}\,.$$

The following result, which was stated in [45], was essentially proved in the derivation of Lemma 4.12 in [44].

Proposition 2.3 *For any $H \in \mathcal{S}^n$, let $P \in \mathcal{O}^n$ be an orthogonal matrix such that*

$$P^T(\Lambda(X) + H)P = \mathrm{diag}(\lambda(\Lambda(X) + H))\,.$$

Then, for any $H \to 0$, we have

$$\begin{cases} P_{a_k a_l} = O(\|H\|)\,, & k, l = 1, \ldots, r,\ k \neq l\,, \quad (18)\\ P_{a_k a_k}P^T_{a_k a_k} = I_{|a_k|} + O(\|H\|^2)\,, & k = 1, \ldots, r\,, \quad (19)\\ \mathrm{dist}(P_{a_k a_k},\ \mathcal{O}^{|a_k|}) = O(\|H\|^2)\,, & k = 1, \ldots, r\,. \quad (20)\end{cases}$$

The next proposition follows easily from Proposition 2.3. It has also been proved in [6] based on a so-called "sin(Θ)" theorem in [42, Theorem 3.4].

Proposition 2.4 *For any $H \in \mathcal{S}^n$, let $P \in \mathcal{O}^n$ be an orthogonal matrix such that $X + H = P\mathrm{diag}(\lambda(X+H))P^T$. Then, for any $\mathcal{S}^n \ni H \to 0$, we have*

$$\mathrm{dist}(P, \mathcal{O}^n(X)) = O(\|H\|)\,.$$

The following proposition about the directional differentiability of the eigenvalue function $\lambda(\cdot)$ is well known. For example, see [24, Theorem 7] and [49, Proposition 1.4].

Proposition 2.5 *Let $X \in \mathcal{S}^n$ have the eigenvalue decomposition (12). Then, for any $\mathcal{S}^n \ni H \to 0$, we have*

$$\lambda_i(X+H) - \lambda_i(X) - \lambda_{l_i}(\overline{P}_{a_k}^T H \overline{P}_{a_k}) = O(\|H\|^2), \quad i \in a_k,\ k = 1, \ldots, r\,, \tag{21}$$

where for each $i \in \{1, \ldots, n\}$, l_i is defined in (14). Hence, for any given direction $H \in \mathcal{S}^n$, the eigenvalue function $\lambda_i(\cdot)$ is directionally differentiable at X with $\lambda_i'(X; H) = \lambda_{l_i}(\overline{P}_{a_k}^T H \overline{P}_{a_k})$, $i \in a_k$, $k = 1, \ldots, r$.

Suppose that $X \in \mathcal{S}^n$ has the eigenvalue decomposition (12). Let $f : \mathbb{R} \to \mathbb{R}$ be a scalar function. The corresponding Löwner's symmetric matrix function is defined by [28]

$$F(X) := \overline{P}\,\mathrm{diag}(f(\lambda_1(X)), f(\lambda_2(X)), \ldots, f(\lambda_n(X)))\,\overline{P}^T = \sum_{i=1}^n f(\lambda_i(X))\bar{p}_i\bar{p}_i^T\,. \tag{22}$$

Let $D := \mathrm{diag}(d)$, where $d \in \mathbb{R}^n$ is a given vector. Assume that the scalar function $f(\cdot)$ is differentiable at each d_i with the derivatives $f'(d_i)$, $i = 1, \ldots, n$. Let $f^{[1]}(D) \in \mathcal{S}^n$ be the first divided difference matrix whose (i,j)-th entry is given by

$$(f^{[1]}(D))_{ij} = \begin{cases} \dfrac{f(d_i) - f(d_j)}{d_i - d_j} & \text{if } d_i \neq d_j, \\ f'(d_i) & \text{if } d_i = d_j, \end{cases} \qquad i, j = 1, \ldots, n.$$

The following result for the differentiability of the symmetric matrix function F defined in (22) can be largely derived from [11] or [23]. Actually, Proposition 4.3 of [7] shows that $F(\cdot)$ is differentiable at X if and only if $f(\cdot)$ is differentiable at every eigenvalue of X. This result is also implied in [25, Theorem 3.3] for the case that $f = \nabla h$ for some differentiable function $h : \mathbb{R} \to \mathbb{R}$. Lemma 4 of [6] and Proposition 4.4 of [7] show that $F(\cdot)$ is continuously differentiable at X if and only if $f(\cdot)$ is continuously differentiable at every eigenvalue of X. For the related directional differentiability of F, one may refer to [41] for a nice derivation.

Proposition 2.6 *Let $X \in \mathcal{S}^n$ be given and have the eigenvalue decomposition (12). Then, the symmetric matrix function $F(\cdot)$ is (continuously) differentiable at X if and only for each $i \in \{1, \ldots, n\}$, $f(\cdot)$ is (continuously) differentiable at $\lambda_i(X)$. In this case, the Fréchet derivative of $F(\cdot)$ at X is given by*

$$F'(X)H = \overline{P}\left[f^{[1]}(\Lambda(X)) \circ (\overline{P}^T H \overline{P})\right]\overline{P}^T \quad \forall\, H \in \mathcal{S}^n\,. \tag{23}$$

The following second order differentiability of the symmetric matrix function F can be derived as in [1, Exercise V.3.9].

Proposition 2.7 *Let $X \in \mathcal{S}^n$ have the eigenvalue decomposition (12). If the scalar function $f(\cdot)$ is twice continuously differentiable at each $\lambda_i(X)$, $i = 1, \ldots, n$, then the symmetric matrix function $F(\cdot)$ is twice continuously differentiable at X.*

2.2 The nonsymmetric matrix functions

From now on, without loss of generality, we assume that $m \leq n$. Let $X \in \mathbb{R}^{m \times n}$ be given. We use $\sigma_1(X) \geq \sigma_2(X) \geq \ldots \geq \sigma_m(X)$ to denote the singular values of X (counting multiplicity) being arranged in non-increasing order. Let $\sigma(X) := (\sigma_1(X), \sigma_2(X), \ldots, \sigma_m(X))^T \in \mathbb{R}^m$ and $\Sigma(X) := \mathrm{diag}(\sigma(X))$. Let $X \in \mathbb{R}^{m \times n}$ admit the following singular value decomposition (SVD):

$$X = \overline{U}\,[\Sigma(X) \;\; 0]\,\overline{V}^T = \overline{U}\,[\Sigma(X) \;\; 0]\,\big[\overline{V}_1 \;\; \overline{V}_2\big]^T = \overline{U}\Sigma(X)\overline{V}_1^T\,, \tag{24}$$

where $\overline{U} \in \mathcal{O}^m$ and $\overline{V} = \big[\overline{V}_1 \;\; \overline{V}_2\big] \in \mathcal{O}^n$ with $\overline{V}_1 \in \mathbb{R}^{n \times m}$ and $\overline{V}_2 \in \mathbb{R}^{n \times (n-m)}$. The set of such matrices $(\overline{U}, \overline{V})$ in the SVD (24) is denoted by $\mathcal{O}^{m,n}(X)$, i.e.,

$$\mathcal{O}^{m,n}(X) := \{(U, V) \in \mathbb{R}^{m \times m} \times \mathbb{R}^{n \times n} \,|\, X = U\,[\Sigma(X) \;\; 0]\,V^T\}\,.$$

Define the three index sets a, b and c by

$$a := \{i \,|\, \sigma_i(X) > 0,\; 1 \leq i \leq m\},\; b := \{i \,|\, \sigma_i(X) = 0,\; 1 \leq i \leq m\} \text{ and } c := \{m+1, \ldots, n\}\,. \tag{25}$$

Let $\overline{\mu}_1 > \overline{\mu}_2 > \ldots > \overline{\mu}_r$ be the nonzero distinct singular values of X. Define

$$a_k := \{i \,|\, \sigma_i(X) = \overline{\mu}_k,\; 1 \leq i \leq m\}, \quad k = 1, \ldots, r\,. \tag{26}$$

For each $i \in \{1, \ldots, m\}$, we also define $l_i(X)$ to be the number of singular values that are equal to $\sigma_i(X)$ but are ranked before i (including i) and $s_i(X)$ to be the number of singular values that are equal to $\sigma_i(X)$ but are ranked after i (excluding i), respectively, i.e., we define $l_i(X)$ and $s_i(X)$ such that

$$\begin{aligned} &\sigma_1(X) \geq \ldots \geq \sigma_{i-l_i(X)}(X) > \sigma_{i-l_i(X)+1}(X) = \ldots = \sigma_i(X) = \ldots = \sigma_{i+s_i(X)}(X) \\ &> \sigma_{i+s_i(X)+1}(X) \geq \ldots \geq \sigma_m(X)\,. \end{aligned} \tag{27}$$

In later discussions, when the dependence of l_i and s_i, $i = 1, \ldots, m$, on X can be seen clearly from the context, we often drop X from these notations.

The inequality in the following lemma is known as von Neumann's trace inequality [31].

Lemma 2.3 *Let Y and Z be two matrices in $\mathbb{R}^{m \times n}$. Then*

$$\langle Y, Z \rangle \leq \sigma(Y)^T \sigma(Z)\,. \tag{28}$$

We first consider a useful property about the SVD.

Proposition 2.8 *Let* $\Sigma := \Sigma(X)$*. Then, the two orthogonal matrices* $P \in \mathcal{O}^m$ *and* $W \in \mathcal{O}^n$ *satisfy*

$$P\,[\Sigma \ \ 0] = [\Sigma \ \ 0]\,W \tag{29}$$

if and only if there exist $Q \in \mathcal{O}^{|a|}$, $Q' \in \mathcal{O}^{|b|}$ *and* $Q'' \in \mathcal{O}^{n-|a|}$ *such that*

$$P = \begin{bmatrix} Q & 0 \\ 0 & Q' \end{bmatrix} \quad \text{and} \quad W = \begin{bmatrix} Q & 0 \\ 0 & Q'' \end{bmatrix},$$

where $Q = \mathrm{diag}(Q_1, Q_2, \ldots, Q_r)$ *is a block diagonal orthogonal matrix with the* k*-th diagonal block given by* $Q_k \in \mathcal{O}^{|a_k|}$, $k = 1, \ldots, r$.

Proof. See the Appendix. □

Let $\mathcal{B}(\cdot) : \mathbb{R}^{m\times n} \to \mathcal{S}^{m+n}$ be the linear operator defined by

$$\mathcal{B}(Z) := \begin{bmatrix} 0 & Z \\ Z^T & 0 \end{bmatrix}, \quad Z \in \mathbb{R}^{m\times n}. \tag{30}$$

It is well-known [20, Theorem 7.3.7] that

$$\mathcal{B}(X) = \overline{P} \begin{bmatrix} \Sigma(X) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\Sigma(X) \end{bmatrix} \overline{P}^T, \tag{31}$$

where the orthogonal matrix $\overline{P} \in \mathcal{O}^{m+n}$ is given by

$$\overline{P} = \frac{1}{\sqrt{2}} \begin{bmatrix} \overline{U} & 0 & \overline{U} \\ \overline{V}_1 & \sqrt{2}\,\overline{V}_2 & -\overline{V}_1 \end{bmatrix}.$$

For notational convenience, we define two more linear operators $S : \mathbb{R}^{p\times p} \to \mathcal{S}^p$ and $T : \mathbb{R}^{p\times p} \to \mathbb{R}^{p\times p}$ by

$$S(Z) := \frac{1}{2}(Z + Z^T) \quad \text{and} \quad T(Z) := \frac{1}{2}(Z - Z^T) \quad \forall\, Z \in \mathbb{R}^{p\times p}. \tag{32}$$

Then, by using (31), one can derive the following proposition directly from (21). For more details, see [26, Section 5.1] .

Proposition 2.9 *For any* $\mathbb{R}^{m\times n} \ni H \to 0$*, we have*

$$\sigma_i(X + H) - \sigma_i(X) - \sigma_i'(X; H) = O(\|H\|^2), \quad i = 1, \ldots, m, \tag{33}$$

where

$$\sigma_i'(X; H) = \begin{cases} \lambda_{l_i}\left(S(\overline{U}_{a_k}^T H \overline{V}_{a_k})\right) & \text{if} \quad i \in a_k, \ k = 1, \ldots, r, \\ \sigma_{l_i}\left(\left[\overline{U}_b^T H \overline{V}_b \ \ \overline{U}_b^T H \overline{V}_2\right]\right) & \text{if} \quad i \in b, \end{cases} \tag{34}$$

where for each $i \in \{1, \ldots, m\}$, l_i *is defined in (27).*

Proposition 2.10 *For any* $\mathbb{R}^{m\times n} \ni H \to 0$*, let* $Y := [\Sigma(X) \;\; 0] + H$*. Suppose that* $U \in \mathcal{O}^m$ *and* $V = [V_1 \;\; V_2] \in \mathcal{O}^n$ *with* $V_1 \in \mathbb{R}^{n\times m}$ *and* $V_2 \in \mathbb{R}^{n\times(n-m)}$ *satisfy*

$$[\Sigma(X) \;\; 0] + H = U\,[\Sigma(Y) \;\; 0]\,V^T = U\,[\Sigma(Y) \;\; 0]\,[V_1 \;\; V_2]^T .$$

Then, there exist $Q \in \mathcal{O}^{|a|}$*,* $Q' \in \mathcal{O}^{|b|}$ *and* $Q'' \in \mathcal{O}^{n-|a|}$ *such that*

$$U = \begin{bmatrix} Q & 0 \\ 0 & Q' \end{bmatrix} + O(\|H\|) \quad \text{and} \quad V = \begin{bmatrix} Q & 0 \\ 0 & Q'' \end{bmatrix} + O(\|H\|) , \tag{35}$$

where $Q = \mathrm{diag}(Q_1, Q_2, \ldots, Q_r)$ *is a block diagonal orthogonal matrix with the k-th diagonal block given by* $Q_k \in \mathcal{O}^{|a_k|}$*,* $k = 1, \ldots, r$*. Furthermore, we have*

$$\Sigma(Y)_{a_k a_k} - \Sigma(X)_{a_k a_k} = Q_k^T S(H_{a_k a_k}) Q_k + O(\|H\|^2), \quad k = 1, \ldots, r \tag{36}$$

and

$$[\Sigma(Y)_{bb} - \Sigma(X)_{bb} \;\; 0] = Q'^T\,[H_{bb} \;\; H_{bc}]\,Q'' + O(\|H\|^2) . \tag{37}$$

Proof. See the Appendix. □

Let $g : \mathbb{R}_+ \to \mathbb{R}$ be a scalar function. The corresponding Löwner's nonsymmetric matrix function is defined by

$$G(X) := \overline{U}\,[g(\Sigma(X)) \;\; 0]\,\overline{V}^T = \sum_{i=1}^{m} g(\sigma_i(X))\bar{u}_i\bar{v}_i^T , \tag{38}$$

where $g(\Sigma(X)) := \mathrm{diag}(g(\sigma_1(X)), \ldots, g(\sigma_m(X)))$. For subsequent discussions, we need to extend the values of g to $\mathbb{R}$ as follows

$$g(t) = \begin{cases} g(t) & \text{if} \quad t \geq 0 , \\ -g(-t) & \text{if} \quad t < 0 . \end{cases} \tag{39}$$

It can be checked easily that $g(0) = 0$ is the sufficient and necessary condition for the well definedness of G. So we always assume that $g(0) = 0$.

Next, we consider the differentiability of $G(\cdot)$. Let $F(\cdot) : \mathcal{S}^{m+n} \to \mathcal{S}^{m+n}$ be Löwner's symmetric matrix function with respect to the scalar function g. Define $\Psi : \mathbb{R}^{m\times n} \to \mathcal{S}^{m+n}$ by

$$\Psi(X) := F(\mathcal{B}(X)) \;\; = \;\; \overline{P} \begin{bmatrix} g(\Sigma(X)) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & g(-\Sigma(X)) \end{bmatrix} \overline{P}^T .$$

Then, we have

$$\begin{aligned} \Psi(X) &= \frac{1}{2} \begin{bmatrix} \overline{U} & 0 & \overline{U} \\ \overline{V}_1 & \sqrt{2}\,\overline{V}_2 & -\overline{V}_1 \end{bmatrix} \begin{bmatrix} g(\Sigma(X)) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & g(-\Sigma(X)) \end{bmatrix} \begin{bmatrix} \overline{U}^T & \overline{V}_1^T \\ 0 & \sqrt{2}\,\overline{V}_2^T \\ \overline{U}^T & -\overline{V}_1^T \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} \overline{U}(g(\Sigma(X)) + g(-\Sigma(X)))\overline{U}^T & \overline{U}(g(\Sigma(X)) - g(-\Sigma(X)))\overline{V}_1^T \\ \overline{V}_1(g(\Sigma(X)) - g(-\Sigma(X)))\overline{U}^T & \overline{V}_1(g(\Sigma(X)) + g(-\Sigma(X)))\overline{V}_1^T \end{bmatrix} . \end{aligned}$$

Thus, from (39), we have

$$\Psi(X) = \begin{bmatrix} 0 & G(X) \\ G(X)^T & 0 \end{bmatrix} = \mathcal{B}(G(X)) \,. \tag{40}$$

Therefore, if $F(\cdot)$ is (continuously) differentiable at $\mathcal{B}(X)$, $G(\cdot)$ is also (continuously) differentiable at X with

$$\Psi'(X)H = F'(\mathcal{B}(X))\mathcal{B}(H) = \mathcal{B}(G'(X)H) \quad \forall\, H \in \mathbb{R}^{m\times n} \,. \tag{41}$$

Let $\overline{\mu}_{r+1} := 0$. Then, for each $k \in \{1, \ldots, r\}$, there exists $\delta_k > 0$ such that $|\overline{\mu}_l - \overline{\mu}_k| > \delta_k$ $\forall\, l = 1, \ldots, r+1$ and $l \neq k$. For each $k \in \{1, \ldots, r\}$, define a continuous scalar function $p_k(\cdot) : \mathbb{R} \to \mathbb{R}$ by

$$p_k(t) = \begin{cases} -\dfrac{6}{\delta_k}(t - \overline{\mu}_k - \dfrac{\delta_k}{2}) & \text{if } t \in (\overline{\mu}_k + \frac{\delta_k}{3}, \overline{\mu}_k + \frac{\delta_k}{2}], \\ 1 & \text{if } t \in [\overline{\mu}_k - \frac{\delta_k}{3}, \overline{\mu}_k + \frac{\delta_k}{3}], \\ \dfrac{6}{\delta_k}(t - \overline{\mu}_k + \dfrac{\delta_k}{2}) & \text{if } t \in [\overline{\mu}_k - \frac{\delta_k}{2}, \overline{\mu}_k - \frac{\delta_k}{3}), \\ 0 & \text{otherwise.} \end{cases} \tag{42}$$

Then, we know that $p_k(0) = 0$ for $k = 1, \ldots, r$. Therefore, the corresponding Löwner's non-symmetric matrix function $\mathcal{P}_k(\cdot)$ with respect to $p_k(\cdot)$ is well-defined, i.e., for any $Y \in \mathbb{R}^{m\times n}$,

$$\mathcal{P}_k(Y) = U\,[p_k(\Sigma(Y)) \;\; 0]\,V^T \,, \tag{43}$$

where $p_k(\Sigma(Y)) = \mathrm{diag}(p_k(\sigma_1(Y)), \ldots, p_k(\sigma_m(Y)))$ and $U \in \mathcal{O}^m$ and $V \in \mathcal{O}^n$ are such that $Y = U\,[\Sigma(Y) \;\; 0]\,V^T$. By (42) and (43), we know that there exists an open neighborhood $\mathcal{N}$ of X such that for each $k \in \{1, \ldots, r\}$,

$$\mathcal{P}_k(Y) = \sum_{i\in a_k} u_i v_i^T \quad \forall\, Y \in \mathcal{N}. \tag{44}$$

In order to study the metric projections over $\mathcal{K}$ and $\mathcal{K}^*$, we need to consider the differential properties of $\mathcal{P}_k(\cdot)$, $k = 1, \ldots, r$. Since each $p_k(\cdot)$ is continuously differentiable near 0 and $\pm\sigma_i(X)$, $i = 1, \ldots, m$, we know from Proposition 2.6 that $\mathcal{P}_k(\cdot)$ is also continuously differentiable in $\mathcal{N}$ (shrinking $\mathcal{N}$ if necessary). Let $Y \in \mathcal{N}$ have the following SVD:

$$Y = U\,[\Sigma(Y) \;\; 0]\,V^T$$

with $(U, V) \in \mathcal{O}^{m,n}(Y)$. By further shrinking $\mathcal{N}$ if necessary, we may assume that for any $k, l \in \{1, \ldots, r\}$,

$$\sigma_i(Y) > 0, \quad \sigma_i(Y) \neq \sigma_j(Y) \quad \forall\, i \in a_k,\ j \in a_l \text{ and } k \neq l,$$

Define $\Gamma_k(Y)$ and $\Xi_k(Y) \in \mathbb{R}^{m\times m}$ and $\Upsilon_k(Y) \in \mathbb{R}^{m\times(n-m)}$, $k = 1, \ldots, r$ by

$$(\Gamma_k(Y))_{ij} = \begin{cases} \dfrac{1}{\sigma_i(Y) - \sigma_j(Y)} & \text{if } i \in a_k,\ j \in a_l,\ k \neq l,\ l = 1, \ldots, r+1\,, \\ \dfrac{-1}{\sigma_i(Y) - \sigma_j(Y)} & \text{if } i \in a_l,\ j \in a_k,\ k \neq l,\ l = 1, \ldots, r+1\,, \\ 0 & \text{otherwise}\,, \end{cases} \tag{45}$$

$$(\Xi_k(Y))_{ij} = \begin{cases} \dfrac{1}{\sigma_i(Y)+\sigma_j(Y)} & \text{if } i \in a_k,\ j \in a_l,\ k \neq l,\ l = 1,\ldots,r+1\,, \\ \dfrac{1}{\sigma_i(Y)+\sigma_j(Y)} & \text{if } i \in a_l,\ j \in a_k,\ k \neq l,\ l = 1,\ldots,r+1\,, \\ \dfrac{2}{\sigma_i(Y)+\sigma_j(Y)} & \text{if } i,j \in a_k, \\ 0 & \text{otherwise} \end{cases} \tag{46}$$

and

$$(\Upsilon_k(Y))_{ij} = \begin{cases} \dfrac{1}{\sigma_i(Y)} & \text{if } i \in a_k\,, \\ 0 & \text{otherwise}\,, \end{cases} \qquad j = 1,\ldots,n-m\,. \tag{47}$$

Then, we obtain from (23) and (41) that for each $k \in \{1,\ldots,r\}$ and any $H \in \mathbb{R}^{m\times n}$,

$$\mathcal{P}_k'(Y)H = U[\Gamma_k(Y)\circ S(A) + \Xi_k(Y)\circ T(A)]V_1^T + U(\Upsilon_k(Y)\circ B)V_2^T\,, \tag{48}$$

where $A := U^T H V_1 \in \mathbb{R}^{m\times m}$, $B := U^T H V_2 \in \mathbb{R}^{m\times(n-m)}$, $V = [V_1\ V_2]$ and the two linear operators $S(\cdot)$ and $T(\cdot)$ are defined by (32). Furthermore, for each $k \in \{1,\ldots,r\}$, from the definition of $p_k(\cdot)$, we know that $p_k(\cdot)$ is actually twice continuously differentiable near each $\lambda_i(\mathcal{B}(X))$, $i = 1,\ldots,m+n$. Then, by Proposition 2.7, we know that the corresponding Löwner's symmetric matrix function $F_k(\cdot)$ with respect to p_k is twice continuously differentiable near $\mathcal{B}(X)$. On the other hand, for each $k = 1,\ldots,r$, from (40), we know that

$$\begin{bmatrix} 0 & \mathcal{P}_k(Z) \\ \mathcal{P}_k(Z)^T & 0 \end{bmatrix} = F_k(\mathcal{B}(Z))\,, \quad Z \in \mathbb{R}^{m\times n}\,. \tag{49}$$

Then, we have the following proposition.

Proposition 2.11 *Let $\mathcal{P}_k(\cdot)$, $k = 1,\ldots,r$ be defined by (43). Then, there exists an open neighborhood $\mathcal{N}$ of X such that for each $k \in \{1,\ldots,r\}$, $\mathcal{P}_k(\cdot)$ is twice continuously differentiable in $\mathcal{N}$.*

Remark 1 *By using the analytic result established in [51] for symmetric functions, one may show that for each $k \in \{1,\ldots,r\}$, $F_k(\cdot)$ is analytic at $\mathcal{B}(X)$. Then from (49), one may derive the conclusion that for each $k \in \{1,\ldots,r\}$, $\mathcal{P}_k(\cdot)$ is analytic at X. Since in this paper we only need the twice continuous differentiability of $\mathcal{P}_k(\cdot)$, $k = 1,\ldots,r$ near X, we will not pursue this analytic property here.*

3 Projections over the epigraphs of the l_∞ and l_1 norms

Since the l_∞ and l_1 norms are entry-wise matrix norms, the epigraphs of the l_∞ and l_1 matrix norms in $\mathbb{R}^{m\times n}$ can be treated as the epigraphs of the l_∞ and l_1 vector norms in $\mathbb{R}^{mn}$, respectively, if we treat a matrix $X \in \mathbb{R}^{m\times n}$ as a vector in $\mathbb{R}^{mn}$. So we only need to study the metric projection operators over the epigraphs of the l_∞ and l_1 vector norms in $\mathbb{R}^{mn}$. Without causing any confusion, we will use $\mathbb{R}^n$, rather than $\mathbb{R}^{mn}$, in our subsequent analysis.

In this section we will mainly focus on the metric projector over the epigraph of the l_∞ norm. The related results of the metric projector over the epigraph of the l_1 norm can be obtained by using (11) accordingly as the epigraph of the l_∞ norm and the epigraph of the l_1 norm are dual to each other under the natural inner product of $\mathbb{R} \times \mathbb{R}^n$. The results obtained in this section are not only of their own interest, but also are crucial for the study of projections over the epigraphs of the spectral and nuclear matrix norms in the next section.

For any $x \in \mathbb{R}^n$, let $x^\downarrow$ be the vector of components of x being arranged in the non-increasing order $x_1^\downarrow \geq \ldots \geq x_n^\downarrow$. Let $\mathrm{sgn}(x)$ be the sign vector of x, i.e., $(\mathrm{sgn})_i(x) = 1$ if $x_i \geq 0$ and -1 otherwise.

Lemma 3.1 *(Hardy, Littlewood, Pólya [18, Theorems 368 & 369]) For $x, y \in \mathbb{R}^n$,*

$$\langle x, y \rangle \leq \langle x^\downarrow, y^\downarrow \rangle \,,$$

where the inequality holds if and only if there exists a permutation π of $\{1, \ldots, n\}$ such that $x_\pi = x^\downarrow$ and $y_\pi = y^\downarrow$.

For any positive constant $\varepsilon > 0$, denote the closed polyhedral convex cone $\mathcal{D}_n^\varepsilon$ by

$$\mathcal{D}_n^\varepsilon := \{(t, x) \in \mathbb{R} \times \mathbb{R}^n \,|\, \varepsilon^{-1} t \geq \; x_i, \; i = 1, \ldots, n\} \,. \tag{50}$$

Let $\Pi_{\mathcal{D}_n^\varepsilon}(\cdot)$ be the metric projector over $\mathcal{D}_n^\varepsilon$ under natural inner product in $\mathbb{R} \times \mathbb{R}^n$. That is, for any $(t, x) \in \mathbb{R} \times \mathbb{R}^n$, $\Pi_{\mathcal{D}_n^\varepsilon}(t, x)$ is the unique optimal solution to the following convex optimization problem

$$\begin{array}{ll} \min & \dfrac{1}{2}\big((\tau - t)^2 + \|y - x\|^2\big) \\ \text{s.t.} & \varepsilon^{-1}\tau \geq y_i, \; i = 1, \ldots, n \,. \end{array} \tag{51}$$

Then we have the following useful result for $\Pi_{\mathcal{D}_n^\varepsilon}(\cdot, \cdot)$.

Proposition 3.1 *Assume that $\varepsilon > 0$ and $(t, x) \in \mathbb{R} \times \mathbb{R}^n$ are given. Let π be a permutation of $\{1, \ldots, n\}$ such that $x^\downarrow = x_\pi$, i.e., $x_i^\downarrow = x_{\pi(i)}$, $i = 1, \ldots, n$ and π^{-1} the inverse of π. For convenience, write $x_0^\downarrow = +\infty$ and $x_{n+1}^\downarrow = -\infty$. Let $\bar{\kappa}$ be the smallest integer $k \in \{0, 1, \ldots, n\}$ such that*

$$x_{k+1}^\downarrow \leq \Big(\sum_{j=1}^{k} x_j^\downarrow + \varepsilon t\Big)/(k + \varepsilon^2) < x_k^\downarrow \,. \tag{52}$$

Define $\bar{y} \in \mathbb{R}^n$ and $\bar{\tau} \in \mathbb{R}$, respectively, by

$$\bar{y}_i := \begin{cases} \Big(\sum_{j=1}^{\bar{\kappa}} x_j^\downarrow + \varepsilon t\Big)/(\bar{\kappa} + \varepsilon^2) & \text{if } 1 \leq i \leq \bar{\kappa} \,, \\ x_i^\downarrow & \text{otherwise} \,, \end{cases} \qquad i = 1, \ldots, n$$

and

$$\bar{\tau} := \varepsilon\Big(\sum_{j=1}^{\bar{\kappa}} x_j^\downarrow + \varepsilon t\Big)/(\bar{\kappa} + \varepsilon^2) \,.$$

Then $\Pi_{\mathcal{D}_n^\varepsilon}(t, x) = (\bar{\tau}, \bar{y}_{\pi^{-1}})$.

Proof. The proof can be obtained in a similar but simpler way to that of Part (i) in Proposition 3.2. We omit it here. $\square$

For any positive constant $\varepsilon > 0$, denote the closed polyhedral convex cone $\mathcal{C}_n^\varepsilon$ by

$$\mathcal{C}_n^\varepsilon := \{(t,x) \in \mathbb{R} \times \mathbb{R}^n \,|\, \varepsilon^{-1} t \geq \|x\|_\infty\}\,. \tag{53}$$

Let $\Pi_{\mathcal{C}_n^\varepsilon}(\cdot,\cdot)$ be the metric projector over $\mathcal{C}_n^\varepsilon$ under the natural inner product in $\mathbb{R} \times \mathbb{R}^n$. That is, for any $(t,x) \in \mathbb{R} \times \mathbb{R}^n$, $\Pi_{\mathcal{C}_n^\varepsilon}(t,x)$ is the unique optimal solution to the following convex optimization problem

$$\begin{array}{ll} \min & \frac{1}{2}\big((\tau - t)^2 + \|y - x\|^2\big) \\ \text{s.t.} & \varepsilon^{-1}\tau \geq \|y\|_\infty\,. \end{array} \tag{54}$$

In the following discussions, we frequently drop n from $\mathcal{C}_n^\varepsilon$ when its size can be found from the context. Also, we will simply use $\mathcal{C}$ to represent $\mathcal{C}^1$.

For any vector $z \in \mathbb{R}^n$, we use $|z|$ to denote the vector in $\mathbb{R}^n$ whose i-th component is $|z_i|$, $i = 1,\ldots,n$. Assume that $\varepsilon > 0$ and $(t,x) \in \mathbb{R} \times \mathbb{R}^n$ are given. Let π be a permutation of $\{1,\ldots,n\}$ such that $|x|^\downarrow = |x|_\pi$, i.e., $|x|_i^\downarrow = |x|_{\pi(i)}$, $i = 1,\ldots,n$ and π^{-1} the inverse of π. Denote $|x|_0^\downarrow = +\infty$ and $|x|_{n+1}^\downarrow = -\infty$. Let $s_0 = 0$ and $s_k = \sum_{i=1}^k |x|_i^\downarrow$, $k = 1,\ldots,n$. Let $\bar{k}$ be the smallest integer $k \in \{0,1,\ldots,n\}$ such that

$$|x|_{k+1}^\downarrow \leq (s_k + \varepsilon t)/(k + \varepsilon^2) < |x|_k^\downarrow\,. \tag{55}$$

Denote

$$\theta^\varepsilon(t,x) := (s_{\bar{k}} + \varepsilon t)/(\bar{k} + \varepsilon^2)\,. \tag{56}$$

Note that if $\bar{k} < n$, then $\theta^\varepsilon(t,x)$ is a nonnegative number by the definition of $\bar{k}$ in (55). But, if $\bar{k} = n$, then $\theta^\varepsilon(t,x)$ can be a negative number. Define three index sets α, β and γ in $\{1,\ldots,n\}$ by

$$\alpha := \{i \,|\, |x_i| > \theta^\varepsilon(t,x)\}, \quad \beta := \{i \,|\, |x_i| = \theta^\varepsilon(t,x)\} \quad \text{and} \quad \gamma := \{i \,|\, |x_i| < \theta^\varepsilon(t,x)\}\,. \tag{57}$$

Define $\bar{x} \in \mathbb{R}^n$ and $\bar{t} \in \mathbb{R}_+$, respectively by

$$\bar{x}_i := \begin{cases} \operatorname{sgn}(x_i)\max\{\theta^\varepsilon(t,x), 0\} & \text{if } i \in \alpha\,, \\ x_i & \text{otherwise}\,, \end{cases} \qquad i = 1,\ldots,n$$

and

$$\bar{t} := \varepsilon \max\{\theta^\varepsilon(t,x), 0\}\,.$$

Then it is easy to see that $(\bar{t},\bar{x}) \in \mathcal{C}^\varepsilon$.

Proposition 3.2 *Assume that $\varepsilon > 0$ and $(t,x) \in \mathbb{R} \times \mathbb{R}^n$ are given.*

(i) The metric projection $\Pi_{\mathcal{C}^\varepsilon}(t,x)$ of (t,x) onto $\mathcal{C}^\varepsilon$ can be computed as follows

$$\Pi_{\mathcal{C}^\varepsilon}(t,x) = (\bar{t},\bar{x})\,. \tag{58}$$

(ii) The continuous mapping $\Pi_{\mathcal{C}^\varepsilon}(\cdot,\cdot)$ is piecewise linear and for any $(\eta,h)\in\mathbb{R}\times\mathbb{R}^n$

$$\Pi_{\mathcal{C}^\varepsilon}(t+\eta,x+h)-\Pi_{\mathcal{C}^\varepsilon}(t,x)=\Pi_{\widehat{\mathcal{C}^\varepsilon}}(\eta,h)\,, \tag{59}$$

where $\widehat{\mathcal{C}^\varepsilon}:=T_{\mathcal{C}^\varepsilon}(\bar t,\bar x)\cap((t,x)-(\bar t,\bar x))^\perp$ is the critical cone of $\mathcal{C}^\varepsilon$ at (t,x) and $T_{\mathcal{C}^\varepsilon}(\bar t,\bar x)$ is the tangent cone of $\mathcal{C}^\varepsilon$ at $(\bar t,\bar x)$. Denote $\delta:=\sqrt{\varepsilon^2+\bar k}$ and $h':=\operatorname{sgn}(x)\circ h$. Let

$$\eta':=\begin{cases}\delta^{-1}(\varepsilon\eta+\sum_{i\in\alpha}h'_i) & \text{if } t\geq-\varepsilon^{-1}\|x\|_1\,,\\ 0 & \text{otherwise}\,.\end{cases}$$

The directional derivative of $\Pi_{\mathcal{C}^\varepsilon}(\cdot,\cdot)$ at (t,x) along the direction $(\eta,h)\in\mathbb{R}\times\mathbb{R}^n$ is given by

$$\Pi'_{\mathcal{C}^\varepsilon}((t,x);(\eta,h))=\Pi_{\widehat{\mathcal{C}^\varepsilon}}(\eta,h)=(\bar\eta,\bar h)\,, \tag{60}$$

where $(\bar\eta,\bar h)\in\mathbb{R}\times\mathbb{R}^n$ satisfies

$$\bar h_i=\operatorname{sgn}(x_i)\varepsilon^{-1}\bar\eta,\quad i\in\alpha\quad\text{and}\quad\bar h_i=h_i,\quad i\in\gamma \tag{61}$$

and

$$\Big(\delta\varepsilon^{-1}\bar\eta,(\operatorname{sgn}(x)\circ\bar h)_\beta\Big)=\begin{cases}\Pi_{\mathcal{D}^\delta_{|\beta|}}(\eta',h'_\beta) & \text{if } t>-\varepsilon^{-1}\|x\|_1\,,\\ \Pi_{\mathcal{C}^\delta_{|\beta|}}(\eta',h'_\beta) & \text{otherwise}\,.\end{cases} \tag{62}$$

Here for the case that $\beta=\emptyset$, we use the convention that $\mathcal{D}^\delta_{|\beta|}:=\mathbb{R}$ and $\mathcal{C}^\delta_{|\beta|}:=\mathbb{R}_+$.

(iii) The mapping $\Pi_{\mathcal{C}^\varepsilon}(\cdot,\cdot)$ is differentiable at (t,x) if and only if $t>\varepsilon||x||_\infty$, or $\varepsilon\|x\|_\infty>t>-\varepsilon^{-1}\|x\|_1$ and $|x|^\downarrow_{\bar k+1}<(s_k+\varepsilon t)/(\bar k+\varepsilon^2)$, or $t<-\varepsilon^{-1}\|x\|_1$.

Proof. (i) It is easy to see that problem (54) can be written equivalently as

$$\begin{array}{ll}\min & \frac{1}{2}\big((\tau-t)^2+\|\,y-|x|\,\|^2\big)\\ \text{s.t.} & \varepsilon^{-1}\tau\geq\|y\|_\infty\end{array} \tag{63}$$

in the sense that $(t^*,y^*)\in\mathbb{R}\times\mathbb{R}^n$ solves problem (63) (note that $y^*\geq0$ in this case) if and only if $(t^*,\operatorname{sgn}(x)\circ y^*)$ solves problem (54). By using Lemma 3.1, we can equivalently reformulate problem (63) as

$$\begin{array}{ll}\min & \frac{1}{2}\big((\tau-t)^2+\|y-|x|^\downarrow\|^2\big)\\ \text{s.t.} & \varepsilon^{-1}\tau\geq||y||_\infty\end{array} \tag{64}$$

in the sense that $(t^*,y^*)\in\mathbb{R}\times\mathbb{R}^n$ solves problem (64) if and only if $(t^*,y^*_{\pi^{-1}})$ solves problem (63).

The Karush-Kuhn-Tucker (KKT) conditions for (64) take the form of

$$\begin{cases}0=\tau-t-\varepsilon^{-1}\mu\,,\\ 0\in y-|x|^\downarrow+\mu\partial||y||_\infty\,,\\ 0\leq(\varepsilon^{-1}\tau-\|y\|_\infty)\perp\mu\geq0\,,\end{cases} \tag{65}$$

where $\mu \in \mathbb{R}_+$ is the corresponding Lagrange multiplier.

We first introduce some notations for the case that (t, x) satisfies $\varepsilon||x||_\infty > t > -\varepsilon^{-1}\|x\|_1$. Since $t < \varepsilon||x||_\infty = \varepsilon|x|^{\downarrow}_1$, the integer $\bar{k} \geq 1$ and for $k = 1$, it holds that

$$\Big(\sum_{j=1}^{k} |x|^{\downarrow}_j + \varepsilon t\Big)/(k + \varepsilon^2) < |x|^{\downarrow}_k \,.$$

Moreover, if for $k = 1$ we have

$$|x|^{\downarrow}_{k+1} \leq \Big(\sum_{j=1}^{k} |x|^{\downarrow}_j + \varepsilon t\Big)/(k + \varepsilon^2)\,,$$

then $\bar{k} = 1$. Otherwise, for $k = 1$ we have

$$|x|^{\downarrow}_{k+1} > \Big(\sum_{j=1}^{k} |x|^{\downarrow}_j + \varepsilon t\Big)/(k + \varepsilon^2)\,,$$

or equivalently,

$$\Big(\sum_{j=1}^{k+1} |x|^{\downarrow}_j + \varepsilon t\Big)/((k+1) + \varepsilon^2) < |x|^{\downarrow}_{k+1} \,.$$

By repeating the above process for $k \geq 2$, we can either find a $k \in \{1, \ldots, n-1\}$ that satisfies (55) or we have

$$\Big(\sum_{j=1}^{n} |x|^{\downarrow}_j + \varepsilon t\Big)/(n + \varepsilon^2) < |x|^{\downarrow}_n \,.$$

In either case, we obtain

$$\Big(\sum_{j=1}^{\bar{k}} |x|^{\downarrow}_j + \varepsilon t\Big)/(\bar{k} + \varepsilon^2) < |x|^{\downarrow}_{\bar{k}}\,,$$

which, together with the definition that $|x|^{\downarrow}_1 \geq \ldots \geq |x|^{\downarrow}_{\bar{k}}$, implies

$$\Big(\sum_{j=1}^{\bar{k}} |x|^{\downarrow}_j + \varepsilon t\Big) < (\bar{k} + \varepsilon^2)|x|^{\downarrow}_{\bar{k}} \leq (\bar{k} + \varepsilon^2)\Big(\sum_{j=1}^{\bar{k}} |x|^{\downarrow}_j\Big)/\bar{k}\,.$$

Thus, we obtain that

$$\bar{k}t < \varepsilon \sum_{j=1}^{\bar{k}} |x|^{\downarrow}_j \,.$$

Define $\bar{y} \in \mathbb{R}^n$ and $\bar{\tau} \in \mathbb{R}_+$, respectively, by

$$\bar{y}_i := \begin{cases} \theta^{\varepsilon}(t, x) & \text{if } 1 \leq i \leq \bar{k}\,, \\ |x|^{\downarrow}_i & \text{otherwise}\,, \end{cases} \qquad i = 1, \ldots, n$$

and

$$\bar{\tau} := \varepsilon\theta^{\varepsilon}(t,x)\,.$$

Let

$$\bar{\mu} := \varepsilon(\bar{\tau}-t) = \varepsilon\Big(\varepsilon\sum_{j=1}^{\bar{k}}|x|_j^{\downarrow} - \bar{k}t\Big)/(\bar{k}+\varepsilon^2)\,.$$

By noting that

$$\bar{y}_1 = \ldots = \bar{y}_{\bar{k}} = \Big(\sum_{j=1}^{\bar{k}}|x|_j^{\downarrow} + \varepsilon t\Big)/(\bar{k}+\varepsilon^2)\,,$$

we have

$$\begin{aligned}\sum_{j=1}^{\bar{k}}(|x|_j^{\downarrow} - \bar{y}_j) &= \sum_{j=1}^{\bar{k}}|x|_j^{\downarrow} - \bar{k}\Big(\sum_{j=1}^{\bar{k}}|x|_j^{\downarrow} + \varepsilon t\Big)/(\bar{k}+\varepsilon^2)\\ &= \varepsilon\Big(\varepsilon\sum_{j=1}^{\bar{k}}|x|_j^{\downarrow} - \bar{k}t\Big)/(\bar{k}+\varepsilon^2) = \bar{\mu}\,.\end{aligned}$$

Then we have

$$\bar{\mu} > 0 \quad \text{and} \quad \sum_{j=1}^{\bar{k}}(|x|_j^{\downarrow} - \bar{y}_j) = \bar{\mu}\,. \tag{66}$$

Define $(t^*, y^*, \mu^*) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}$ by

$$(t^*, y^*, \mu^*) := \begin{cases} (t, |x|^{\downarrow}, 0) & \text{if } t \geq \varepsilon\|x\|_\infty\,,\\ (\bar{\tau}, \bar{y}, \bar{\mu}) & \text{if } \varepsilon\|x\|_\infty > t > -\varepsilon^{-1}\|x\|_1\,,\\ (0, 0, -\varepsilon t) & \text{if } t \leq -\varepsilon^{-1}\|x\|_1\,. \end{cases}$$

Then, by using the facts that $|x|^{\downarrow} \geq \bar{y} \geq 0$ and (66) holds when $\varepsilon||x||_\infty > t > -\varepsilon^{-1}\|x\|_1$, we can readily check that $(t^*, y^*, \mu^*) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}$ satisfies the KKT conditions (65). Consequently, (t^*, y^*) is the unique optimal solution to problem (64). Note that $\alpha = \{\pi^{-1}(i)\,|\, i = 1, \ldots, \bar{k}\}$. Thus, we obtain that $(t^*, \mathrm{sgn}(x) \circ y^*_{\pi^{-1}}) = (\bar{t}, \bar{x})$.

(ii) By noting that $\mathcal{C}^{\varepsilon} = \{(t,x) \in \mathbb{R} \times \mathbb{R}^n \,|\, \varepsilon^{-1}t \geq \|x\|_\infty\} = \{(t,x) \in \mathbb{R} \times \mathbb{R}^n \,|\, t \geq \varepsilon x_i,\ t \geq -\varepsilon x_i,\ i = 1, \ldots, n\}$ is a polyhedral set, we immediately know that $\Pi_{\mathcal{C}^{\varepsilon}}(\cdot)$ is a piecewise linear function. For a short proof, see [39, Chapter 2] or [43, Chapter 5].

Since $\mathcal{C}^{\varepsilon}$ is a polyhedral set, from the results in [17, 32] we know that

$$\Pi'_{\mathcal{C}^{\varepsilon}}((t,x); (\eta, h)) = \Pi_{\widehat{\mathcal{C}^{\varepsilon}}}(\eta, h)\,.$$

Let $f(z) := \|z\|_\infty$, $z \in \mathbb{R}^n$. Then, by using Theorem 2.4.9 in [10], we know that

$$T_{\mathcal{C}^{\varepsilon}}(\varepsilon f(z), z) = \{(\zeta, d) \in \mathbb{R} \times \mathbb{R}^n \,|\, \zeta \geq \varepsilon f'(z; d)\}\,. \tag{67}$$

For any $z \in \mathbb{R}^n$, let

$$I(z) := \{i \,|\, |z_i| = ||z||_\infty,\ i = 1, \ldots, n\}\,.$$

Then, for any $d \in \mathbb{R}^n$,

$$f'(z;d) = \begin{cases} \max\{\text{sgn}(z_i)d_i,\ i \in I(z)\} & \text{if } z \neq 0\,, \\ \|d\|_\infty & \text{if } z = 0\,. \end{cases} \tag{68}$$

We next consider the following five cases:

Case 1: $t > \varepsilon||x||_\infty$. In this case, $(\bar{t}, \bar{x}) = (t, x)$ and $\widehat{\mathcal{C}^\varepsilon} = T_{\mathcal{C}^\varepsilon}(\bar{t}, \bar{x}) = \mathbb{R} \times \mathbb{R}^n$. Thus,

$$\Pi'_{\mathcal{C}^\varepsilon}((t,x);(\eta,h)) = \Pi_{\widehat{\mathcal{C}^\varepsilon}}(\eta, h) = (\eta, h).$$

On the other hand, in this case, we know that $\bar{k} = 0$ and

$$\alpha = \emptyset, \quad \beta = \emptyset \quad \text{and} \quad \gamma = \{1, \ldots, n\}\,.$$

Therefore, $\delta = \varepsilon$ and $\eta' = \eta$. Since $\mathcal{D}^\delta_{|\beta|} = \mathbb{R}$ if $\beta = \emptyset$, we know that $(\bar{\eta}, \bar{h}) = (\eta, h)$. This means that (60) holds.

Case 2: $t = \varepsilon||x||_\infty$. In this case, $(\bar{t}, \bar{x}) = (t, x)$ and $\widehat{\mathcal{C}^\varepsilon} = T_{\mathcal{C}^\varepsilon}(\bar{t}, \bar{x})$. From (67) and (68) we have

$$\widehat{\mathcal{C}^\varepsilon} = T_{\mathcal{C}^\varepsilon}(\bar{t}, \bar{x}) = \begin{cases} \{(\zeta, d) \in \mathbb{R} \times \mathbb{R}^n \,|\, \varepsilon^{-1}\zeta \geq \text{sgn}(x_i)d_i,\ i \in I(x)\} & \text{if } x \neq 0\,, \\ \mathcal{C}^\varepsilon & \text{if } x = 0\,. \end{cases}$$

In this case, $\bar{k} = 0$ and $\theta^\varepsilon(t, x) = \|x\|_\infty$. We know that

$$\alpha = \emptyset, \quad \beta = I(x) \quad \text{and} \quad \gamma = \{1, \ldots, n\} \setminus I(x)\,.$$

Therefore, since $\delta = \varepsilon$ and $\eta' = \eta$, it can be checked easily that $(\bar{\eta}, \bar{h})$ satisfies the conditions (61) and (62).

Case 3: $\varepsilon||x||_\infty > t > -\varepsilon^{-1}||x||_1$. In this case, $(\bar{t}, \bar{x}) = (\bar{\tau}, \text{sgn}(x) \circ \bar{y}_{\pi^{-1}}) \neq (0, 0)$ and $\text{sgn}(\bar{x}) = \text{sgn}(x)$. Then, from (55) and (58), we know that

$$\bar{I}^0 := \{\pi^{-1}(i) \,|\, i = 1, \ldots, \bar{k}\,\} \subseteq I(\bar{x})$$

and

$$\begin{aligned} &((t, x) - (\bar{t}, \bar{x}))^\perp \\ &= \{(\zeta, d) \in \mathbb{R} \times \mathbb{R}^n \,|\, (t - \bar{t})\zeta + \textstyle\sum_{i \in \bar{I}^0}(x_i - \bar{x}_i)d_i = 0\,\} \\ &= \{(\zeta, d) \in \mathbb{R} \times \mathbb{R}^n \,|\, \textstyle\sum_{j=1}^{\bar{k}}(\bar{y}_j - |x|^\downarrow_j)(\varepsilon^{-1}\zeta) + \sum_{i \in \bar{I}^0}(|x_i| - |\bar{x}_i|)\text{sgn}(x_i)d_i = 0\,\} \\ &= \{(\zeta, d) \in \mathbb{R} \times \mathbb{R}^n \,|\, \textstyle\sum_{i \in \bar{I}^0}(|x_i| - |\bar{x}_i|)(-\varepsilon^{-1}\zeta + \text{sgn}(x_i)d_i) = 0\,\}\,, \end{aligned}$$

which, together with (67), (68), and the facts that $\bar{t} = \varepsilon||\bar{x}||_\infty$ and $|x_i| > |\bar{x}_i|$ for each $i \in \bar{I}^0$, implies that

$$\widehat{\mathcal{C}^\varepsilon} = \{(\zeta, d) \in \mathbb{R} \times \mathbb{R}^n \,|\, \varepsilon^{-1}\zeta = \text{sgn}(x_i)d_i \ \forall\, i \in \bar{I}^0 \text{ and } \varepsilon^{-1}\zeta \geq \text{sgn}(x_i)d_i \ \forall\, i \in I(\bar{x}) \backslash \bar{I}^0\,\}\,.$$

In this case, we know that $\beta = I(x)\backslash\bar{I}^0$. Then, after simple transformations, $\Pi_{\widehat{\mathcal{C}^\varepsilon}}(\eta, h)$ can be computed as in Proposition 3.1, from which we know that $(\bar{\eta}, \bar{h})$ satisfies (61) and (62).

Case 4: $t = -\varepsilon^{-1}||x||_1$ and $(t,x) \neq (0,0)$. In this case, $(\bar{t},\bar{x}) = 0$ and $\widehat{\mathcal{C}^\varepsilon} = T_{\mathcal{C}^\varepsilon}(\bar{t},\bar{x}) \cap (t,x)^\perp = \mathcal{C}^\varepsilon \cap (t,x)^\perp$. Let $\mathrm{supp}(x) := \{i \,|\, x_i \neq 0,\ i = 1,\ldots,n\}$. Then, since

$$(t,x)^\perp = \{(\zeta,d) \in \mathbb{R} \times \mathbb{R}^n \,|\, \varepsilon^{-1}\zeta\|x\|_1 = \langle x,d\rangle\},$$

we have

$$\widehat{\mathcal{C}^\varepsilon} = \mathcal{C}^\varepsilon \cap (t,x)^\perp = \{(\zeta,d) \in \mathbb{R} \times \mathbb{R}^n \,|\, \mathrm{sgn}(x_i)d_i = \varepsilon^{-1}\zeta \geq ||d||_\infty,\ i \in \mathrm{supp}(x)\}.$$

In this case, we know that $\bar{k} = |\mathrm{supp}(x)|$ and $\theta^\varepsilon(t,x) = 0$. Therefore,

$$\alpha = \mathrm{supp}(x), \quad \beta = \{1,\ldots,n\}\backslash\mathrm{supp}(x) \quad \text{and} \quad \gamma = \emptyset.$$

Since for $(\zeta,d) \in \widehat{\mathcal{C}^\varepsilon}$, we have $d_i = \varepsilon^{-1}\zeta$ for any $i \in \alpha$, after simple transformations, we know that $\Pi_{\widehat{\mathcal{C}^\varepsilon}}(\eta,h)$ can be easily computed as in Part (i) of this proposition and $(\bar{\eta},\bar{h})$ also satisfies (61) and (62).

Case 5: $t < -\varepsilon^{-1}||x||_1$. In this case, $(\bar{t},\bar{x}) = 0$ and $\widehat{\mathcal{C}^\varepsilon} = T_{\mathcal{C}^\varepsilon}(\bar{t},\bar{x}) \cap (t,x)^\perp = \{(0,0)\}$. Hence,

$$\Pi'_{\mathcal{C}^\varepsilon}((t,x);(\eta,h)) = (0,0).$$

In this case, we know that

$$\alpha = \{1,\ldots,n\}, \quad \beta = \emptyset \quad \text{and} \quad \gamma = \emptyset.$$

Also, since $\eta' = 0$ and $\mathcal{C}^\delta_{|\beta|} = \mathbb{R}_+$, we know that $\bar{\eta} = 0$ and $\bar{h} = 0$, which means that (60) holds.

(iii) This part follows from the proof of Part (ii) and the fact that $\Pi_{\mathcal{C}^\varepsilon}(\cdot,\cdot)$ is Lipschitz continuous. □

4 Projections over the epigraphs of the spectral and nuclear norms

For any given positive number $\varepsilon > 0$, define the matrix cone $\mathcal{K}^\varepsilon_{m,n}$ by

$$\mathcal{K}^\varepsilon_{m,n} := \{(t,X) \in \mathbb{R} \times \mathbb{R}^{m\times n} \,|\, \varepsilon^{-1}t \geq \|X\|_2\}. \tag{69}$$

For the case that $\varepsilon = 1$, we will simply use $\mathcal{K}_{m,n}$ to represent $\mathcal{K}^1_{m,n}$. That is, $\mathcal{K}_{m,n}$ is the epigraph of the spectral norm $\|\cdot\|_2$ on $\mathbb{R}^{m\times n}$. In the following proposition, we will show that the dual cone of $\mathcal{K}_{m,n}$ is the epigraph of the nuclear norm $\|\cdot\|_*$ and $\mathcal{K}_{m,n}$ is a proper hyperbolic cone. Therefore, we will mainly focus on the metric projector over $\mathcal{K}_{m,n}$. The related properties of the metric projector over the epigraph of the nuclear norm can be readily derived by using (11).

Proposition 4.1 *The dual cone of the* $\mathcal{K}_{m,n}$ *is*

$$\mathcal{K}^*_{m,n} = \{(t,X) \in \mathbb{R} \times \mathbb{R}^{m\times n} \,|\, t \geq \|X\|_*\}.$$

Moreover, $\mathcal{K}_{m,n}$ *is a proper hyperbolic cone.*

Proof. Let

$$C := \{(t, X) \in \mathbb{R} \times \mathbb{R}^{m\times n} \,|\, t \geq \|X\|_* \}\,.$$

Let $(\tau, Y) \in \mathbb{R} \times \mathbb{R}^{m\times n}$ be an arbitrary point in C. Then for any $(t, X) \in \mathcal{K}_{m,n}$, from von Neumann's trace inequality (28)

$$\langle X, -Y\rangle \leq \langle \sigma(X), \sigma(-Y)\rangle$$

and the fact that $\sigma(Y) = \sigma(-Y)$ we obtain that

$$\langle (t, X), (\tau, Y)\rangle = t\tau + \langle X, Y\rangle \geq t\tau - \langle \sigma(X), \sigma(Y)\rangle \geq t\tau - \|X\|_2\|Y\|_* \geq 0\,.$$

This shows that $(\tau, Y) \in \mathcal{K}^*_{m,n}$. Consequently, $C \subseteq \mathcal{K}^*_{m,n}$.

Next, we show that $\mathcal{K}^*_{m,n} \subseteq C$. Suppose on the contrary that this is not true. Then there exists $(\bar{t}, \overline{X}) \in \mathbb{R} \times \mathbb{R}^{m\times n}$ such that $(\bar{t}, \overline{X}) \in \mathcal{K}^*_{m,n}$ but $\bar{t} < \|\overline{X}\|_*$. Assume that $\overline{X}$ has the singular value decomposition

$$\overline{X} = \overline{U}[\overline{\Sigma} \;\; 0]\overline{V}^T,$$

where $(\overline{U}, \overline{V}) \in \mathcal{O}^{m,n}(\overline{X})$ and $\overline{\Sigma}$ is a diagonal matrix whose diagonal entries are the singular values $\bar{\sigma}_1 \geq \ldots \geq \bar{\sigma}_m$ of $\overline{X}$. Let

$$\bar{\tau} = 1 \quad \text{and} \quad \overline{Y} = \overline{U}[-I \;\, 0]\overline{V}^T\,.$$

Then, since $(\bar{t}, \overline{X}) \in \mathcal{K}^*_{m,n}$, we have

$$0 \leq \langle (\bar{t}, \overline{X}), (\bar{\tau}, \overline{Y})\rangle = \bar{t}\bar{\tau} + \langle \overline{X}, \overline{Y}\rangle = \bar{t} - \langle \overline{\Sigma}, I\rangle = \bar{t} - \|\overline{X}\|_* < 0\,.$$

This contradiction implies $\mathcal{K}^*_{m,n} \subseteq C$.

To complete the proof we next show that $\mathcal{K}_{m,n}$ is a proper hyperbolic cone. This can be done by using the polynomial function

$$pol(t, X) := \det\big(\mathrm{diag}(t, tI + \mathcal{B}(X))\big)\,, \quad (t, X) \in \mathbb{R} \times \mathbb{R}^{m\times n}\,,$$

where $\mathcal{B} : \mathbb{R}^{m\times n} \to \mathcal{S}^{m+n}$ is defined by (30) and for any $(t, X) \in \mathbb{R} \times \mathbb{R}^{m\times n}$, the block diagonal matrix $\mathrm{diag}(t, tI + \mathcal{B}(X)) \in \mathcal{S}^{1+m+n}$ is defined by

$$\mathrm{diag}(t, tI + \mathcal{B}(X)) := \begin{bmatrix} t & 0 \\ 0 & tI + \mathcal{B}(X) \end{bmatrix}.$$

Let $\lambda_i(t, X)$, $i = 1, \ldots, m + n + 1$ be the real roots of $pol(\lambda\,(1, 0) - (t, X)) = 0$, where $(1, 0) \in \mathbb{R} \times \mathbb{R}^{m\times n}$. Then $\mathcal{K}_{m,n}$ can be written as

$$\mathcal{K}_{m,n} = \{(t, X) \in \mathbb{R} \times \mathbb{R}^{m\times n} \,|\, \lambda_i(t, X) \geq 0\,,\; i = 1, \ldots, m + n + 1\}\,.$$

That is, $\mathcal{K}_{m,n}$ is a hyperbolic cone of $pol(\cdot)$ with the direction $(1, 0)$. This completes the proof by the fact that $pol(1, 0) = 1 > 0$. $\square$

For any positive constant $\varepsilon > 0$, define the matrix cone $\mathcal{M}_n^\varepsilon$ in $\mathbb{R} \times \mathcal{S}^n$ as the epigraph of the convex function $\varepsilon\lambda_1(\cdot)$, i.e.,

$$\mathcal{M}_n^\varepsilon := \{(t, X) \in \mathbb{R} \times \mathcal{S}^n \,|\, \varepsilon^{-1}t \geq \lambda_1(X)\,\}\,. \tag{70}$$

Let $\Pi_{\mathcal{M}_n^\varepsilon}(\cdot,\cdot)$ be the metric projector over $\mathcal{M}_n^\varepsilon$ under the natural inner product in $\mathbb{R} \times \mathcal{S}^n$. That is, for any $(t, X) \in \mathbb{R} \times \mathcal{S}^n$, $\Pi_{\mathcal{M}_n^\varepsilon}(t, X)$ is the unique optimal solution to the following optimization problem

$$\begin{array}{ll} \min & \frac{1}{2}((\tau - t)^2 + \|Y - X\|_F^2) \\ \text{s.t.} & \varepsilon^{-1}\tau \geq \lambda_1(Y)\,. \end{array} \tag{71}$$

Proposition 4.2 *Assume that $(t, X) \in \mathbb{R} \times \mathcal{S}^n$ is given and X has the eigenvalue decomposition (12). Let $\mathcal{D}_n^\varepsilon$ be the closed convex cone defined in (50). Let $(\bar{t}, \bar{y}) \in \mathbb{R} \times \mathbb{R}^n$ be given by*

$$(\bar{t}, \bar{y}) = \Pi_{\mathcal{D}_n^\varepsilon}(t, \lambda(X))\,,$$

where $\Pi_{\mathcal{D}_n^\varepsilon}(t, \lambda(X))$ can be computed explicitly as in Proposition 3.1. Then,

$$\Pi_{\mathcal{M}_n^\varepsilon}(t, X) = (\bar{t}, \overline{P}\text{diag}(\bar{y})\overline{P}^T)\,. \tag{72}$$

Proof. By the definition of $\Pi_{\mathcal{M}_n^\varepsilon}(\cdot,\cdot)$, we know that $(\bar{t}, \overline{X}) := \Pi_{\mathcal{M}_n^\varepsilon}(t, X)$ is the unique optimal solution to the optimization problem (71). From Fan's inequality (15), we know that

$$\|\, \lambda(Y) - \lambda(X) \,\|^2 \leq \|\, Y - X \,\|_F^2 \quad \forall\, Y \in \mathcal{S}^n\,.$$

Then, from the uniqueness of $(\bar{t}, \overline{X})$, we know that $\overline{X} = \overline{P}\,\text{diag}(\lambda(\overline{X}))\overline{P}^T$ as the Frobenius norm $\|\cdot\|_F$ is unitarily invariant. Therefore, since $(\bar{t}, \bar{y}) \in \mathbb{R} \times \mathbb{R}^n$ is an optimal solution to

$$\begin{array}{ll} \min & \frac{1}{2}((\tau - t)^2 + \|y - \lambda(X)\|^2) \\ \text{s.t.} & \varepsilon^{-1}\tau \geq y_i, \quad i = 1, \ldots, n \end{array}$$

and $\bar{y}_1 \geq \ldots \geq \bar{y}_n$, we conclude that $(\bar{t}, \overline{P}\text{diag}(\bar{y})\overline{P}^T)$ solves (71). Thus, (72) holds. $\square$

Let $\Pi_{\mathcal{K}_{m,n}^\varepsilon}(\cdot,\cdot)$ be the metric projector over $\mathcal{K}_{m,n}^\varepsilon$ under the natural inner product in $\mathbb{R} \times \mathbb{R}^{m\times n}$. That is, for any $(t, X) \in \mathbb{R} \times \mathbb{R}^{m\times n}$, $\Pi_{\mathcal{K}_{m,n}^\varepsilon}(t, X)$ is the unique optimal solution to the following optimization problem

$$\begin{array}{ll} \min & \frac{1}{2}((\tau - t)^2 + \|Y - X\|_F^2) \\ \text{s.t.} & \varepsilon^{-1}\tau \geq \|Y\|_2\,. \end{array} \tag{73}$$

Theorem 4.1 *Assume that $(t, X) \in \mathbb{R} \times \mathbb{R}^{m\times n}$ is given and X has the singular value decomposition (24). Let $\mathcal{C}_m^\varepsilon$ be the closed convex cone defined in (53). Let $(\bar{t}, \bar{y}) \in \mathbb{R} \times \mathbb{R}^n$ be given by*

$$(\bar{t}, \bar{y}) = \Pi_{\mathcal{C}_m^\varepsilon}(t, \sigma(X))\,,$$

where $\Pi_{\mathcal{C}_m^\varepsilon}(t, \sigma(X))$ can be computed explicitly as in Part (i) of Proposition 3.2. Then, we have

$$\Pi_{\mathcal{K}_{m,n}^\varepsilon}(t, X) = (\bar{t}, \overline{U}\,[\text{diag}(\bar{y}) \quad 0]\,\overline{V}^T)\,. \tag{74}$$

Proof. By the definition of $\Pi_{\mathcal{K}^{\varepsilon}_{m,n}}(\cdot,\cdot)$, we know that $(\bar{t}, \overline{X}) := \Pi_{\mathcal{K}^{\varepsilon}_{m,n}}(t, X)$ is the unique optimization to the optimization problem (73). From von Neumann's trace inequality (28), we know that

$$\| \sigma(Y) - \sigma(X) \|^2 \leq \| Y - X \|_F^2 \quad \forall\, Y \in \mathbb{R}^{m\times n} .$$

Then, from the uniqueness of $(\bar{t}, \overline{X})$, we know that $\overline{X} = \overline{U}[\mathrm{diag}(\sigma(\overline{X}))\ \ 0]\overline{V}^T$ as the Frobenius norm $\|\cdot\|_F$ is unitarily invariant. Therefore, since $(\bar{t}, \bar{y}) \in \mathbb{R}_+ \times \mathbb{R}^m$ is an optimal solution to

$$\begin{array}{ll} \min & \dfrac{1}{2}\big((\tau - t)^2 + \|y - \sigma(X)\|^2\big) \\ \text{s.t.} & \varepsilon^{-1}\tau \geq \|y\|_\infty \end{array}$$

and $\bar{y}_1 \geq \ldots \geq \bar{y}_m \geq 0$, we conclude that $(\bar{t}, \overline{U}[\mathrm{diag}(\bar{y})\ 0]\overline{V}^T)$ solves (73). Thus, (74) holds. $\square$

Next, we will consider the (directional) differentiability of the metric projector over $\mathcal{K}_{m,n}$, i.e., $\Pi_{\mathcal{K}_{m,n}}(\cdot,\cdot)$. In the following discussions, we will drop m and n from $\mathcal{K}_{m,n}$ when its dependence on m and n can be seen clearly from the context.

Let $(t, X) \in \mathbb{R} \times \mathbb{R}^{m\times n}$ be given and X have the singular value decomposition (24), i.e.,

$$X = \overline{U}\,[\Sigma(X)\ \ 0]\,\overline{V}^T ,$$

where $\overline{U} \in \mathcal{O}^m$ and $\overline{V} \in \mathcal{O}^n$. As in Section 2.2, we use $\bar{\mu}_1 > \bar{\mu}_2 > \ldots > \bar{\mu}_r$ to denote all the nonzero distinct singular values of X and denote $\bar{\mu}_{r+1} = 0$. For the sake of convenience, we also let $\sigma_0(X) = +\infty$ and $\sigma_{m+1}(X) = -\infty$. Let $s_0 = 0$ and $s_k = \sum_{i=1}^k \sigma_i(X)$, $k = 1, \ldots, m$. Let $\bar{k}$ be the smallest integer $k \in \{0, 1, \ldots, m\}$ such that

$$\sigma_{k+1}(X) \leq (s_k + t)/(k+1) < \sigma_k(X) . \tag{75}$$

Denote $\theta(t, \sigma(X)) \in \mathbb{R}$ by

$$\theta(t, \sigma(X)) := (s_{\bar{k}} + t)/(\bar{k} + 1) . \tag{76}$$

Let α, β and γ be the three index sets in $\{1, \ldots, m\}$ defined by

$$\alpha := \{i \,|\, \sigma_i(X) > \theta(t, \sigma(X))\},\ \beta := \{i \,|\, \sigma_i(X) = \theta(t, \sigma(X))\} \text{ and } \gamma := \{i \,|\, \sigma_i(X) < \theta(t, \sigma(X))\} . \tag{77}$$

Let $\delta := \sqrt{1 + \bar{k}}$. Let $S(\cdot)$ and $T(\cdot)$ be defined by (32). Define $\rho : \mathbb{R} \times \mathbb{R}^{m\times n} \to \mathbb{R}$ as follows

$$\rho(\eta, H) := \left\{ \begin{array}{ll} \delta^{-1}(\eta + \mathrm{Tr}(S(\overline{U}_\alpha^T H \overline{V}_\alpha))) & \text{if } t \geq -\|X\|_* , \\ 0 & \text{otherwise} , \end{array} \right. \quad (\eta, H) \in \mathbb{R} \times \mathbb{R}^{m\times n} . \tag{78}$$

Let $(\tau, Y) \in \mathbb{R} \times \mathbb{R}^{m\times n}$ be given. Suppose that $U \in \mathcal{O}^m$ and $V \in \mathcal{O}^n$ are such that

$$Y = U\,[\Sigma(Y)\ \ 0]\,V^T .$$

For each $k \in \{1, \ldots, r\}$, let $\mathcal{P}_k(Y)$ be defined by (43). Define $g_0(\tau, \sigma(Y)) \in \mathbb{R}$ and $g(\tau, \sigma(Y)) \in \mathbb{R}^m$ by

$$\big(g_0(\tau, \sigma(Y)), g(\tau, \sigma(Y))\big) := \Pi_{\mathcal{C}_m}(\tau, \sigma(Y)) . \tag{79}$$

Let

$$G(\tau, Y) := U[\mathrm{diag}(g(\tau, \sigma(Y))) \;\; 0]V^T \,. \tag{80}$$

Then, from Theorem 4.1, we have

$$(g_0(\tau, \sigma(Y)), G(\tau, Y)) = \Pi_{\mathcal{K}}(\tau, Y) \,. \tag{81}$$

Note that from Proposition 3.2, we know for each $k \in \{1, \ldots, r\}$, $g_i(t, \sigma(X)) = g_j(t, \sigma(X))$ $\forall\, i, j \in a_k$. Therefore, we may define

$$\overline{\nu}_k := g_i(t, \sigma(X)) \quad \text{for an arbitrary } i \in a_k \,, \quad k = 1, \ldots, r \,.$$

Moreover, define

$$G_S(Y) := \sum_{k=1}^{r} \overline{\nu}_k \mathcal{P}_k(Y) \quad \text{and} \quad G_R(\tau, Y) := G(\tau, Y) - G_S(Y) \,. \tag{82}$$

Define $\Omega_1 \in \mathbb{R}^{m\times m}$, $\Omega_2 \in \mathbb{R}^{m\times m}$ and $\Omega_3 \in \mathbb{R}^{m\times(n-m)}$ (depending on X) as follows

$$(\Omega_1)_{ij} := \begin{cases} \dfrac{g_i(t, \sigma(X)) - g_j(t, \sigma(X))}{\sigma_i(X) - \sigma_j(X)} & \text{if } \sigma_i(X) \neq \sigma_j(X) \,, \\ 0 & \text{otherwise} \,, \end{cases} \quad i, j \in \{1, \ldots, m\} \,, \tag{83}$$

$$(\Omega_2)_{ij} := \begin{cases} \dfrac{g_i(t, \sigma(X)) + g_j(t, \sigma(X))}{\sigma_i(X) + \sigma_j(X)} & \text{if } \sigma_i(X) + \sigma_j(X) \neq 0 \,, \\ 0 & \text{otherwise} \,, \end{cases} \quad i, j \in \{1, \ldots, m\} \tag{84}$$

and

$$(\Omega_3)_{ij} := \begin{cases} \dfrac{g_i(t, \sigma(X))}{\sigma_i(X)} & \text{if } \sigma_i(X) \neq 0 \,, \\ 0 & \text{if } \sigma_i(X) = 0 \,, \end{cases} \quad i \in \{1, \ldots, m\}, \quad j \in \{1, \ldots, n-m\} \,. \tag{85}$$

Hence, from Part (i) of Proposition 3.2, we know that the matrices Ω_1, Ω_2 and Ω_3 have the following forms

$$\Omega_1 = \begin{bmatrix} 0 & 0 & (\Omega_1)_{\alpha\gamma} \\ 0 & 0 & E_{\beta\gamma} \\ (\Omega_1)_{\gamma\alpha} & E_{\gamma\beta} & (\Omega_1)_{\gamma\gamma} \end{bmatrix} , \quad \Omega_2 = \begin{bmatrix} (\Omega_2)_{aa} & (\Omega_2)_{ab} \\ (\Omega_2)_{ba} & 0 \end{bmatrix} \quad \text{and} \quad \Omega_3 = \begin{bmatrix} (\Omega_3)_{ac'} \\ 0 \end{bmatrix} , \tag{86}$$

where $E_{\beta\gamma} \in \mathbb{R}^{|\beta|\times|\gamma|}$ and $E_{\gamma\beta} \in \mathbb{R}^{|\gamma|\times|\beta|}$ are two matrices whose entries are all ones and a, b, c are defined in (25) and $c' := \{1, \ldots, n-m\}$.

Theorem 4.2 *Assume that $(t, X) \in \mathbb{R} \times \mathbb{R}^{m\times n}$ is given. Let X have the singular value decomposition (24). Then, the metric projector over the matrix cone $\mathcal{K}$, $\Pi_{\mathcal{K}}(\cdot, \cdot)$ is directionally differentiable at (t, X) along any direction. For any $(\eta, H) \in \mathbb{R} \times \mathbb{R}^{m\times n}$, let $A := \overline{U}^T H \overline{V}_1$ and $B := \overline{U}^T H \overline{V}_2$. Then, for given $(\eta, H) \in \mathbb{R} \times \mathbb{R}^{m\times n}$, the directional derivative $\Pi'_{\mathcal{K}}((t, X); (\eta, H))$ can be computed as follows:*

(i) if $t > \|X\|_2$*, then* $\Pi'_{\mathcal{K}}((t,X);(\eta,H)) = (\eta,H)$*;*

(ii) if $\|X\|_2 \geq t > -\|X\|_*$*, then* $\Pi'_{\mathcal{K}}((t,X);(\eta,H)) = (\overline{\eta},\overline{H})$ *with*

$$\overline{\eta} = \delta^{-1}\psi_0^\delta(\eta,H)\,, \tag{87}$$

$$\begin{aligned}\overline{H} &= \overline{U}\begin{bmatrix} \delta^{-1}\psi_0^\delta(\eta,H)I_{|\alpha|} & 0 & (\Omega_1)_{\alpha\gamma}\circ S(A)_{\alpha\gamma} \\ 0 & \Psi^\delta(\eta,H) & S(A)_{\beta\gamma} \\ (\Omega_1)_{\gamma\alpha}\circ S(A)_{\gamma\alpha} & S(A)_{\gamma\beta} & S(A)_{\gamma\gamma}\end{bmatrix}\overline{V}_1^T \\ &\quad + \overline{U}\begin{bmatrix} (\Omega_2)_{aa}\circ T(A)_{aa} & (\Omega_2)_{ab}\circ T(A)_{ab} \\ (\Omega_2)_{ba}\circ T(A)_{ba} & T(A)_{bb}\end{bmatrix}\overline{V}_1^T + \overline{U}\begin{bmatrix} (\Omega_3)_{ac'}\circ B_{ac'} \\ B_{bc'}\end{bmatrix}\overline{V}_2^T\,, \end{aligned}\tag{88}$$

where $\Big(\psi_0^\delta(\eta,H),\Psi^\delta(\eta,H)\Big) \in \mathbb{R}\times\mathcal{S}^{|\beta|}$ *is given by*

$$\Big(\psi_0^\delta(\eta,H),\Psi^\delta(\eta,H)\Big) := \Pi_{\mathcal{M}^\delta_{|\beta|}}(\rho(\eta,H), S(\overline{U}_\beta^T H\overline{V}_\beta))\,. \tag{89}$$

In particular, if $t = \|X\|_2 > 0$*, we have that* $\bar{k} = 0$*,* $\delta = 1$*,* $\alpha = \emptyset$*,* $\rho(\eta,H) = \eta$ *and*

$$\overline{\eta} = \psi_0^\delta(\eta,H), \quad \overline{H} = \overline{U}\begin{bmatrix} \Psi^\delta(\eta,H) + T(A)_{\beta\beta} & A_{\beta\gamma} \\ A_{\gamma\beta} & A_{\gamma\gamma}\end{bmatrix}\overline{V}_1^T + \overline{U}B\overline{V}_2^T\,;$$

(iii) if $t = -\|X\|_*$*, then* $\Pi'_{\mathcal{K}}((t,X);(\eta,H)) = (\overline{\eta},\overline{H})$ *with*

$$\overline{\eta} = \delta^{-1}\psi_0^\delta(\eta,H)\,, \tag{90}$$

$$\overline{H} = \overline{U}\begin{bmatrix} \delta^{-1}\psi_0^\delta(\eta,H)I_{|\alpha|} & 0 \\ 0 & \Psi_1^\delta(\eta,H)\end{bmatrix}\overline{V}_1^T + \overline{U}\begin{bmatrix} 0 \\ \Psi_2^\delta(\eta,H)\end{bmatrix}\overline{V}_2^T\,, \tag{91}$$

where $\psi_0^\delta(\eta,H)\in\mathbb{R}$*,* $\Psi_1^\delta(\eta,H)\in\mathbb{R}^{|\beta|\times|\beta|}$ *and* $\Psi_2^\delta(\eta,H)\in\mathbb{R}^{|\beta|\times(n-m)}$ *are given by*

$$\Big(\psi_0^\delta(\eta,H),\Big[\Psi_1^\delta(\eta,H)\ \ \Psi_2^\delta(\eta,H)\Big]\Big) := \Pi_{\mathcal{K}^\delta_{|\beta|,(n-|\alpha|)}}\Big(\rho(\eta,H),\Big[\overline{U}_\beta^T H V_\beta\ \ \overline{U}_\beta^T H\overline{V}_2\Big]\Big)\,; \tag{92}$$

(iv) if $t < -\|X\|_*$*, then* $\Pi'_{\mathcal{K}}((t,X);(\eta,H)) = (0,0)$*.*

Moreover, $\Pi_{\mathcal{K}}(\cdot,\cdot)$ *is calmly B-differentiable at* (t,X)*, i.e., for any* $(\eta,H)\in\mathbb{R}\times\mathbb{R}^{m\times n}$ *with* $(\eta,H)\to(0,0)$*, we have*

$$\Pi_{\mathcal{K}}(t+\eta,X+H) - \Pi_{\mathcal{K}}(t,X) - \Pi'_{\mathcal{K}}((t,X);(\eta,H)) = O(\|(\eta,H)\|^2)\,. \tag{93}$$

Proof. By Theorem 4.1, we only need to consider the case that $||X||_2 \geq t \geq -\|X\|_*$. For any $(\tau,Y)\in\mathbb{R}\times\mathbb{R}^{m\times n}$, $(g_0(\tau,\sigma(Y)),g(\tau,\sigma(Y)))$ is defined by (79), $G(\tau,Y)$ is defined by (80) and $G_S(Y)$ and $G_R(\tau,Y)$ are defined by (82). Let $(\eta,H)\in\mathbb{R}\times\mathbb{R}^{m\times n}$ be given. We write $(\tau,Y) := (t+\eta,X+H)\in\mathbb{R}\times\mathbb{R}^{m\times n}$. Suppose that $U\in\mathcal{O}^m$ and $V\in\mathcal{O}^n$ are such that

$$Y = U\,[\Sigma(Y)\ \ 0]\,V^T\,. \tag{94}$$

Since $G_S(X) = G(t, X)$, we have

$$G(\tau, Y) - G(t, X) = G_S(Y) - G_S(X) + G_R(\tau, Y).$$

By Proposition 2.11, we know that there exists an open neighborhood $\mathcal{N}$ of X such that for each $k \in \{1, \ldots, r\}$, $\mathcal{P}_k(\cdot)$ is twice continuously differentiable in $\mathcal{N}$. Then, for (η, H) sufficiently close to $(0, 0)$, we know from (48) that

$$\begin{aligned} & G_S(Y) - G_S(X) = \sum_{k=1}^{r} \overline{\nu}_k(\mathcal{P}_k(Y) - \mathcal{P}_k(X)) = \sum_{k=1}^{r} \overline{\nu}_k \mathcal{P}_k'(X)H + O(\|H\|^2) \\ = \ & \overline{U} \left[\sum_{k=1}^{r} \overline{\nu}_k(\Gamma_k(X) \circ S(A) + \Xi_k(X) \circ T(A)) \right] \overline{V}_1^T + \overline{U} \left[\sum_{k=1}^{r} \overline{\nu}_k(\Upsilon_k(X) \circ B) \right] \overline{V}_2^T + O(\|H\|^2) \\ = \ & \overline{U}[\Omega_1 \circ S(A)]\overline{V}_1^T + \overline{U}[\Omega_2 \circ T(A)]\overline{V}_1^T + \overline{U}(\Omega_3 \circ B)\overline{V}_2^T + O(\|H\|^2), \end{aligned} \tag{95}$$

where $A = \overline{U}^T H \overline{V}_1 \in \mathrm{I\!R}^{m \times m}$, $B = \overline{U}^T H \overline{V}_2 \in \mathrm{I\!R}^{m \times (n-m)}$ and $\Omega_1, \Omega_2 \in \mathrm{I\!R}^{m \times m}$ and $\Omega_3 \in \mathrm{I\!R}^{m \times (n-m)}$ are given by (83), (84) and (85), respectively. On the other hand, by (42) and (43), for H sufficiently close to 0, i.e., for Y sufficiently close to X, we have

$$\mathcal{P}_k(Y) = \sum_{i \in a_k} u_i v_i^T, \quad k = 1, \ldots, r.$$

Therefore, we obtain that for $(\tau, Y) \in \mathrm{I\!R} \times \mathcal{N}$ (shrinking $\mathcal{N}$ if necessary),

$$G_R(\tau, Y) = \sum_{k=1}^{r} \sum_{i \in a_k} [g_i(\tau, \sigma(Y)) - g_i(t, \sigma(X))] u_i v_i^T + \sum_{i \in b} g_i(\tau, \sigma(Y)) u_i v_i^T = \sum_{k=1}^{r} \Delta_k + \Delta_{r+1}, \tag{96}$$

where

$$\Delta_k := \sum_{i \in a_k} [g_i(\tau, \sigma(Y)) - \overline{\nu}_k] u_i v_i^T, \quad k = 1, \ldots, r \quad \text{and} \quad \Delta_{r+1} := \sum_{i \in b} g_i(\tau, \sigma(Y)) u_i v_i^T.$$

Firstly, consider the case that $X = [\Sigma(X) \ \ 0]$ and $\overline{U} = I_m$, $\overline{V} = I_n$. Then, from (33) and (34), for (η, H) sufficiently close to $(0, 0)$, we know that

$$\sigma_i(Y) = \sigma_i(X) + \sigma_i'(X; H) + O(\|H\|^2), \quad i = 1, \ldots, m \tag{97}$$

and

$$\sigma_i'(X; H) = \begin{cases} \lambda_{l_i}(S(H_{a_k a_k})) & \text{if} \quad i \in a_k, \ k = 1, \ldots, r, \\ \sigma_{l_i}([H_{bb} \ \ H_{bc}]) & \text{if} \quad i \in b. \end{cases} \tag{98}$$

Since $\Pi_{\mathcal{C}_m}(\cdot, \cdot)$ is Lipschitz continuous on $\mathrm{I\!R} \times \mathrm{I\!R}^m$, we obtain from (59) that

$$\begin{aligned} & \Pi_{\mathcal{C}_m}(\tau, \sigma(Y)) - \Pi_{\mathcal{C}_m}(t, \sigma(X)) \\ = \ & \Pi_{\mathcal{C}_m}(t + \eta, \sigma(X) + \sigma'(X; H) + O(\|H\|^2)) - \Pi_{\mathcal{C}_m}(t, \sigma(X)) \\ = \ & \Pi_{\mathcal{C}_m}(t + \eta, \sigma(X) + \sigma'(X; H)) - \Pi_{\mathcal{C}_m}(t, \sigma(X)) + O(\|(\eta, H)\|^2) \\ = \ & \Pi_{\widehat{\mathcal{C}}_m}(\eta, \sigma'(X; H)) + O(\|(\eta, H)\|^2), \end{aligned} \tag{99}$$

where $\widehat{\mathcal{C}}_m$ is the critical cone of $\mathcal{C}_m$ at $(t, \sigma(X))$. Let $h := \sigma'(X; H) \in \mathbb{R}^m$. Then, from (98), we have

$$h_{a_k} = \lambda(S(H_{a_k a_k})) \in \mathbb{R}^{|a_k|}, \quad k = 1, \ldots, r \tag{100}$$

and

$$h_b = \sigma([H_{bb} \ \ H_{bc}]) \in \mathbb{R}^{|b|} \,. \tag{101}$$

Since $(g_0(t, \sigma(X)), g(t, \sigma(X))) = \Pi_{\mathcal{C}_m}(t, \sigma(X))$, from (99), we obtain that

$$g_0(\tau, \sigma(Y)) - g_0(t, \sigma(X)) = \widehat{\eta} + O(\|(\eta, H)\|^2) \tag{102}$$

and

$$g_i(\tau, \sigma(Y)) - g_i(t, \sigma(X)) = \widehat{h}_i + O(\|(\eta, H)\|^2), \quad i = 1, \ldots, m \,, \tag{103}$$

where

$$(\widehat{\eta}, \widehat{h}) := \Pi_{\widehat{\mathcal{C}}_m}(\eta, h) \,. \tag{104}$$

Hence, since for each $i \in \{1, \ldots, m\}$, $u_i v_i^T$ is uniformly bounded, we obtain that

$$\Delta_k = \sum_{i \in a_k} \widehat{h}_i u_i v_i^T + O(\|(\eta, H)\|^2), \quad k = 1, \ldots, r$$

and

$$\Delta_{r+1} = \sum_{i \in b} \widehat{h}_i u_i v_i^T + O(\|(\eta, H)\|^2) \,.$$

Furthermore, by (35), we know that for each $k \in \{1, \ldots, r\}$, there exists $Q_k \in \mathcal{O}^{|a_k|}$ such that

$$U_{a_k} = \begin{bmatrix} O(\|H\|) \\ Q_k + O(\|H\|) \\ O(\|H\|) \end{bmatrix} \quad \text{and} \quad V_{a_k} = \begin{bmatrix} O(\|H\|) \\ Q_k + O(\|H\|) \\ O(\|H\|) \end{bmatrix} .$$

Note that $\lambda(\cdot)$ and $\sigma(\cdot)$ are both Lipchitz continuous. Since $\Pi_{\widehat{\mathcal{C}}_m}(\cdot, \cdot)$ is Lipschitz continuous on $\mathbb{R} \times \mathbb{R}^m$, from (104), we have

$$\|(\widehat{\eta}, \widehat{h})\| = \|\Pi_{\widehat{\mathcal{C}}_m}(\eta, h)\| = O(\|(\eta, H)\|) \,. \tag{105}$$

Therefore, for each $k \in \{1, \ldots, r\}$, we have

$$\begin{aligned} \Delta_k &= \begin{bmatrix} O(\|(\eta, H)\|^3) & O(\|(\eta, H)\|^2) & O(\|(\eta, H)\|^3) \\ O(\|(\eta, H)\|^2) & Q_k \mathrm{diag}(\widehat{h}_{a_k}) Q_k^T + O(\|(\eta, H)\|^2) & O(\|(\eta, H)\|^2) \\ O(\|(\eta, H)\|^3) & O(\|(\eta, H)\|^2) & O(\|(\eta, H)\|^3) \end{bmatrix} + O(\|(\eta, H)\|^2) \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & Q_k \mathrm{diag}(\widehat{h}_{a_k}) Q_k^T & 0 \\ 0 & 0 & 0 \end{bmatrix} + O(\|(\eta, H)\|^2) \,. \end{aligned} \tag{106}$$

On the other hand, from (36), we know that

$$S(H_{a_k a_k}) = Q_k(\Sigma(Y)_{a_k a_k} - \bar{\mu}_k I_{|a_k|})Q_k^T + O(\|H\|^2), \quad k = 1, \ldots, r\,.$$

Therefore, we obtain from (97) and (100) that

$$\begin{aligned} S(H_{a_k a_k}) &= Q_k \mathrm{diag}(\sigma_i'(X;H) : i \in a_k)Q_k^T + O(\|H\|^2) \\ &= Q_k \mathrm{diag}(h_{a_k})Q_k^T + O(\|H\|^2), \quad k = 1, \ldots, r\,. \end{aligned} \tag{107}$$

Meanwhile, by (35), there exist $W \in \mathcal{O}^{|b|}$ and $Z = [Z_1 \ \ Z_2] \in \mathcal{O}^{n-|a|}$ with $Z_1 \in \mathbb{R}^{(n-|a|)\times|b|}$ and $Z_2 \in \mathbb{R}^{(n-|a|)\times(n-m)}$ such that

$$U_b = \begin{bmatrix} O(\|H\|) \\ W + O(\|H\|) \end{bmatrix} \quad \text{and} \quad [V_b \ \ V_c] = \begin{bmatrix} O(\|H\|) \\ Z + O(\|H\|) \end{bmatrix}.$$

Therefore, from (105), we obtain that

$$\Delta_{r+1} = \begin{bmatrix} 0 & 0 \\ 0 & W\mathrm{diag}(\widehat{h}_b)Z_1^T \end{bmatrix} + O(\|(\eta, H)\|^2)\,. \tag{108}$$

On the other hand, from (37), we know that

$$[H_{bb} \ \ H_{bc}] = W(\Sigma(Y)_{bb} - \bar{\mu}_{r+1} I_{|b|})Z_1^T + O(\|H\|^2)\,.$$

Therefore, since W and Z_1 are uniformly bounded, from (97) and (101), we have

$$[H_{bb} \ \ H_{bc}] = W\mathrm{diag}(\sigma_i'(X;H) : i \in b)Z_1^T + O(\|H\|^2) = W\mathrm{diag}(h_b)Z_1^T + O(\|H\|^2)\,. \tag{109}$$

Hence, by (96), (106) and (108), we obtain that

$$\begin{aligned} & G_R(\tau, Y) \\ = & \begin{bmatrix} Q_1\mathrm{diag}(\widehat{h}_{a_1})Q_1^T & 0 & \cdots & 0 & 0 \\ 0 & Q_2\mathrm{diag}(\widehat{h}_{a_2})Q_2^T & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & Q_r\mathrm{diag}(\widehat{h}_{a_r})Q_r^T & 0 \\ 0 & 0 & \cdots & 0 & W\mathrm{diag}(\widehat{h}_b)Z_1^T \end{bmatrix} + O(\|(\eta, H)\|^2)\,. \end{aligned} \tag{110}$$

Let

$$\eta' := \begin{cases} \delta^{-1}(\eta + \sum_{i\in\alpha} h_i) & \text{if } t \geq -\|X\|_*\,, \\ 0 & \text{otherwise}\,, \end{cases}$$

where $\delta = \sqrt{1+\bar{k}}$. If $t \geq -\|X\|_*$, then by the definition of $\bar{k}$ we can conclude that for any $i \in \alpha$, $\sigma_i(X) > 0$ because in this case $\theta(t, \sigma(X)) \geq 0$. Thus, by (100), we know that for $t \geq -\|X\|_*$,

$$\eta' = \delta^{-1}(\eta + \mathrm{Tr}(S(H_{\alpha\alpha}))) = \rho(\eta, H)\,,$$

where $\rho(\eta, H)$ is defined by (78). By noting that $(\widehat{\eta}, \widehat{h}) = \Pi_{\widehat{\mathcal{C}}_m}(\eta, h)$ and $\sigma(X) \geq 0$, we obtain from Part (ii) of Proposition 3.2 that

$$\widehat{h}_i = \widehat{\eta} \quad \forall\, i \in \alpha, \quad \widehat{h}_i = h_i \quad \forall\, i \in \gamma \tag{111}$$

and

$$(\delta\widehat{\eta}, \widehat{h}_\beta) = \begin{cases} \Pi_{\mathcal{D}^\delta_{|\beta|}}(\eta', h_\beta) & \text{if } t > -\|X\|_*\,, \\ \Pi_{\mathcal{C}^\delta_{|\beta|}}(\eta', h_\beta) & \text{otherwise}\,. \end{cases} \tag{112}$$

Next, we consider the following two cases:

Case 1: $\|X\|_2 \geq t > -\|X\|_*$, i.e., $\|\sigma(X)\|_\infty \geq t > -\|\sigma(X)\|_1$. We first conclude from (76) that for any $i \in \alpha \cup \beta$, $\sigma_i(X) > 0$ because $\theta(t, \sigma(X)) > 0$ in this case. We will separate this case into two subcases.

Case 1.1: $\beta \neq \emptyset$. Then there exists an integer $\bar{r} \in \{0, 1, \ldots, r-1\}$ such that

$$\alpha = \bigcup_{k=1}^{\bar{r}} a_k, \quad \beta = a_{\bar{r}+1} \quad \text{and} \quad \gamma = \bigcup_{k=\bar{r}+2}^{r} a_k \bigcup b\,.$$

From (112), we have

$$(\delta\widehat{\eta}, \widehat{h}_\beta) = \Pi_{\mathcal{D}^\delta_{|\beta|}}(\eta', h_\beta)\,.$$

By Proposition 4.2 and the fact that $\eta' = \rho(\eta, H)$, we know

$$(\delta\widehat{\eta}, Q_\beta \mathrm{diag}(\widehat{h}_\beta) Q_\beta^T) = \Pi_{\mathcal{M}^\delta_{|\beta|}}(\rho(\eta, H), Q_\beta \mathrm{diag}(h_\beta) Q_\beta^T)\,.$$

Note that $\Pi_{\mathcal{M}^\delta_{|\beta|}}(\cdot, \cdot)$ is Lipschitz continuous on $\mathbb{R} \times \mathcal{S}^{|\beta|}$. Then, from (107), we obtain that

$$(\delta\widehat{\eta}, Q_\beta \mathrm{diag}(\widehat{h}_\beta) Q_\beta^T) = \Pi_{\mathcal{M}^\delta_{|\beta|}}(\rho(\eta, H), S(H_{\beta\beta})) + O(\|(\tau, H)\|^2)\,.$$

Therefore, by using the definitions of (87) and (89), we have

$$\widehat{\eta} = \overline{\eta} + O(\|(\tau, H)\|^2) \tag{113}$$

and

$$Q_\beta \mathrm{diag}(\widehat{h}_\beta) Q_\beta^T = \Psi^\delta(\eta, H) + O(\|(\tau, H)\|^2)\,.$$

This, together with (110), (111), (107) and (109), implies

$$G_R(\tau,Y)=\begin{bmatrix}\overline{\eta}I_{|\alpha|} & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & \Psi^{\delta}(\eta,H) & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & S(H_{a_{\overline{r}+2}a_{\overline{r}+2}}) & 0 & 0 & 0 & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & 0 & S(H_{a_ra_r}) & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & H_{bb} & H_{bc}\end{bmatrix}+O(\|(\tau,H)\|^2)\,. \tag{114}$$

Therefore, from (86), (95) and (114), we obtain that

$$\begin{aligned}&G(\tau,Y)-G(t,X)=G_S(Y)-G_S(X)+G_R(\tau,Y)\\ =&\begin{bmatrix}\overline{\eta}I_{|\alpha|} & 0 & (\Omega_1)_{\alpha\gamma}\circ S(H)_{\alpha\gamma} & 0\\ 0 & \Psi^{\delta}(\eta,H) & S(H)_{\beta\gamma} & 0\\ (\Omega_1)_{\gamma\alpha}\circ S(H)_{\gamma\alpha} & S(H)_{\gamma\beta} & S(H)_{\gamma\gamma} & 0\end{bmatrix}\\ &+\begin{bmatrix}(\Omega_2)_{aa}\circ T(H)_{aa} & (\Omega_2)_{ab}\circ T(H)_{ab} & 0\\ (\Omega_2)_{ba}\circ T(H)_{ba} & T(H)_{bb} & 0\end{bmatrix}+\begin{bmatrix}0 & (\Omega_3)_{ac'}\circ H_{ac}\\ 0 & H_{bc}\end{bmatrix}+O(\|(\eta,H)\|^2)\,.\end{aligned} \tag{115}$$

Case 1.2: $\beta=\emptyset$. Then there exists $\overline{r}\in\{1,\ldots,r-1\}$ such that

$$\alpha=\bigcup_{k=1}^{\overline{r}}a_k,\quad \beta=\emptyset\quad\text{and}\quad \gamma=\bigcup_{k=\overline{r}+1}^{r}a_k\bigcup b\,.$$

Since $\mathcal{D}^{\delta}_{|\beta|}=\mathbb{R}$, we know from (112) that $\widehat{\eta}=\delta^{-1}\eta'$. Also, since $\mathcal{M}^{\delta}_{|\beta|}=\mathbb{R}$, we have

$$\overline{\eta}=\delta^{-1}\psi_0^{\delta}(\eta,H)=\delta^{-1}\eta'=\widehat{\eta}\,. \tag{116}$$

Then, from (110), (111), (107) and (109), we obtain that

$$G_R(\tau,Y)=\begin{bmatrix}\overline{\eta}I_{|\alpha|} & 0 & 0 & 0 & 0 & 0\\ 0 & S(H_{a_{\overline{r}+1}a_{\overline{r}+1}}) & 0 & 0 & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & S(H_{a_ra_r}) & 0 & 0\\ 0 & 0 & 0 & 0 & H_{bb} & H_{bc}\end{bmatrix}+O(\|(\tau,H)\|^2)\,.$$

This, together with (86) and (95), implies

$$\begin{aligned}&G(\tau,Y)-G(t,X)=G_S(Y)-G_S(X)+G_R(\tau,Y)\\ =&\begin{bmatrix}\overline{\eta}I_{|\alpha|} & (\Omega_1)_{\alpha\gamma}\circ S(H)_{\alpha\gamma} & 0\\ (\Omega_1)_{\gamma\alpha}\circ S(H)_{\gamma\alpha} & S(H)_{\gamma\gamma} & 0\end{bmatrix}+\begin{bmatrix}(\Omega_2)_{aa}\circ T(H)_{aa} & (\Omega_2)_{ab}\circ T(H)_{ab} & 0\\ (\Omega_2)_{ba}\circ T(H)_{ba} & T(H)_{bb} & 0\end{bmatrix}\\ &+\begin{bmatrix}0 & (\Omega_3)_{ac'}\circ H_{ac}\\ 0 & H_{bc}\end{bmatrix}+O(\|(\eta,H)\|^2)\,.\end{aligned} \tag{117}$$

Case 2: $t = -\|X\|_*$, i.e., $t = -\|\sigma(X)\|_1$. In this case, $\theta(t,\sigma(X)) = 0$. Therefore, we have

$$\alpha = a = \{i \,|\, \sigma_i(X) > 0\}, \quad \beta = b = \{i \,|\, \sigma_i(X) = 0\} \quad \text{and} \quad \gamma = \emptyset\,.$$

Then, from (112), we have

$$(\delta\widehat{\eta}, \widehat{h}_\beta) = \Pi_{\mathcal{C}^\delta_{|\beta|}}(\eta', h_\beta)\,.$$

From Theorem 4.1 and the fact that $\eta' = \rho(\eta, H)$, we know that

$$(\delta\widehat{\eta}, W\mathrm{diag}(\widehat{h}_\beta)Z_1^T) = \Pi_{\mathcal{K}^\delta_{|\beta|,(n-|\alpha|)}}(\rho(\eta,H), W\mathrm{diag}(h_\beta)Z_1^T)\,.$$

By noting that $\Pi_{\mathcal{K}^\delta_{|\beta|,(n-|\alpha|)}}(\cdot,\cdot)$ is Lipschitz continuous on $\mathbb{R}\times\mathbb{R}^{|\beta|\times(n-|\alpha|)}$, we obtain from (109) that

$$(\delta\widehat{\eta}, W\mathrm{diag}(\widehat{h}_\beta)Z_1^T) = \Pi_{\mathcal{K}^\delta_{|\beta|,(n-|\alpha|)}}(\rho(\eta,H), [H_{\beta\beta} \;\; H_{\beta c}]) + O(\|(\tau,H)\|^2)\,.$$

Then, by using the definitions of (90) and (92), we obtain that

$$\widehat{\eta} = \bar{\eta} + O(\|(\tau,H)\|^2) \tag{118}$$

and

$$W\mathrm{diag}(\widehat{h}_\beta)Z_1^T = \left[\Psi_1^\delta(\eta,H) \;\; \Psi_2^\delta(\eta,H)\right] + O(\|(\tau,H)\|^2)\,,$$

which, together with (110), (111), (107) and (109), implies

$$G_R(\tau,Y) = \begin{bmatrix} \bar{\eta}I_{|\alpha|} & 0 & 0 \\ 0 & \Psi_1^\delta(\eta,H) & \Psi_2^\delta(\eta,H) \end{bmatrix} + O(\|(\tau,H)\|^2)\,.$$

From (58) and the fact that $\theta(t,\sigma(X)) = 0$, we have

$$g_i(t,\sigma(X)) = \theta(t,\sigma(X)) = 0, \quad i = 1,\ldots,m\,.$$

Thus, by using (95) and the fact that in this case, $\Omega_1 = 0$, $\Omega_2 = 0$ and $\Omega_3 = 0$ we obtain that

$$\begin{aligned} G(\tau,Y) - G(t,X) &= G_S(Y) - G_S(X) + G_R(\tau,Y) \\ &= \begin{bmatrix} \bar{\eta}I_{|\alpha|} & 0 & 0 \\ 0 & \Psi_1^\delta(\eta,H) & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \Psi_2^\delta(\eta,H) \end{bmatrix} + O(\|(\eta,H)\|^2)\,. \end{aligned} \tag{119}$$

Next, consider the general case for $X \in \mathbb{R}^{m\times n}$. Re-write (94) as

$$[\Sigma(X) \;\; 0] + \overline{U}^T H \overline{V} = \overline{U}^T U \,[\Sigma(X+H) \;\; 0]\, V^T \overline{V}\,.$$

Let $\widetilde{U} := \overline{U}^T U$, $\widetilde{V} := \overline{V}^T V$ and $\widetilde{H} := \overline{U}^T H \overline{V} = \left[\overline{U}^T H \overline{V}_1 \;\; \overline{U}^T H \overline{V}_2\right] = [A \;\; B]$. Let

$$\widetilde{X} := [\Sigma(X) \;\; 0] \quad \text{and} \quad \widetilde{Y} := [\Sigma(X) \;\; 0] + \widetilde{H} = \widetilde{U}\,[\Sigma(X+H) \;\; 0]\,\widetilde{V}^T\,.$$

Then, we have

$$G(\tau,Y) - G(t,X) = \overline{U}\left[G(\tau,\widetilde{Y}) - G(t,\widetilde{X})\right]\overline{V}^T\,.$$

Since $\Sigma(\widetilde{X}) = \Sigma(X)$ and $\widetilde{X} = [\Sigma(X) \;\; 0]$, we know from (102), (113), (116), (115) and (117) that if $\|X\|_2 \geq t > -\|X\|_*$, then for any $(\eta, H) \in \mathbb{R} \times \mathbb{R}^{m\times n}$ with $(\eta, H) \to 0$,

$$g_0(\tau, \sigma(Y)) - g_0(t, \sigma(X)) = \overline{\eta} + O(\|(\eta, H)\|^2)$$

and

$$\begin{aligned} & G(\tau, Y) - G(t, X) \\ = \; & \overline{U} \begin{bmatrix} \overline{\eta} I_{|\alpha|} & 0 & (\Omega_1)_{\alpha\gamma} \circ S(A)_{\alpha\gamma} \\ 0 & \Psi^\delta(\eta, H) & S(A)_{\beta\gamma} \\ (\Omega_1)_{\gamma\alpha} \circ S(A)_{\gamma\alpha} & S(A)_{\gamma\beta} & S(A)_{\gamma\gamma} \end{bmatrix} \overline{V}_1^T \\ & + \overline{U} \begin{bmatrix} (\Omega_2)_{aa} \circ T(A)_{aa} & (\Omega_2)_{ab} \circ T(A)_{ab} \\ (\Omega_2)_{ba} \circ T(A)_{ba} & T(A)_{bb} \end{bmatrix} \overline{V}_1^T + \overline{U} \begin{bmatrix} (\Omega_3)_{ac'} \circ B_{ac'} \\ B_{bc'} \end{bmatrix} \overline{V}_2^T + O(\|(\eta, H)\|^2) \,, \end{aligned}$$

where $(\psi_0^\delta(\eta, H), \Psi^\delta(\eta, H)) \in \mathbb{R} \times \mathcal{S}^{|\beta|}$ is given by (89). Similarly, we know from (102), (118) and (119) that if $t = -\|X\|_*$, then for any $(\eta, H) \in \mathbb{R} \times \mathbb{R}^{m\times n}$ with $(\eta, H) \to 0$,

$$g_0(\tau, \sigma(Y)) - g_0(t, \sigma(X)) = \overline{\eta} + O(\|(\eta, H)\|^2)$$

and

$$G(\tau, Y) - G(t, X) = \overline{U} \begin{bmatrix} \overline{\eta} I_{|\alpha|} & 0 \\ 0 & \Psi_1^\delta(\eta, H) \end{bmatrix} \overline{V}_1^T + \overline{U} \begin{bmatrix} 0 \\ \Psi_2^\delta(\eta, H) \end{bmatrix} \overline{V}_2^T + O(\|(\eta, H)\|^2) \,,$$

where $\psi_0^\delta(\eta, H) \in \mathbb{R}$, $\Psi_1^\delta(\eta, H) \in \mathbb{R}^{|\beta|\times|\beta|}$ and $\Psi_2^\delta(\eta, H) \in \mathbb{R}^{|\beta|\times(n-m)}$ are given by (92).

Finally, from (81) and the above analysis we have shown that $\Pi_{\mathcal{K}}(\cdot, \cdot)$ is directionally differentiable at (t, X), the directional derivative of $\Pi_{\mathcal{K}}(\cdot, \cdot)$ at (t, X) along any direction $(\eta, H) \in \mathbb{R} \times \mathbb{R}^{m\times n}$ is given by Parts (i)-(iv) in this theorem and for $(\eta, H) \in \mathbb{R} \times \mathbb{R}^{m\times n}$ with $(\eta, H) \to 0$, (93) holds. $\square$

We characterize the differentiability of the metric projector $\Pi_{\mathcal{K}}(\cdot, \cdot)$ in the following theorem.

Theorem 4.3 *Let $\rho : \mathbb{R} \times \mathbb{R}^{m\times n} \to \mathbb{R}$ be the linear operator defined by (78). The metric projector $\Pi_{\mathcal{K}}(\cdot, \cdot)$ is differentiable at $(t, X) \in \mathbb{R} \times \mathbb{R}^{m\times n}$ if and only if (t, X) satisfies one of the following three conditions:*

(i) $t > \|X\|_2$;

(ii) $\|X\|_2 > t > -\|X\|_*$ *but* $\sigma_{\bar{k}+1}(X) < \theta(t, \sigma(X))$, *where* $\bar{k}$ *and* $\theta(t, \sigma(X))$ *are defined by (75) and (76), respectively;*

(iii) $t < -\|X\|_*$.

In this case, for any $(\eta, H) \in \mathbb{R} \times \mathbb{R}^{m\times n}$, $\Pi'_{\mathcal{K}}(t, X)(\eta, H) = (\overline{\eta}, \overline{H})$, where under condition (i), $(\overline{\eta}, \overline{H}) = (\eta, H)$; under condition (ii),

$$\overline{\eta} = \delta^{-1} \rho(\eta, H) \tag{120}$$

and

$$\overline{H} = \overline{U}\begin{bmatrix} \delta^{-1}\rho(\eta,H)I_{|\alpha|} & (\Omega_1)_{\alpha\gamma}\circ S(A)_{\alpha\gamma} \\ (\Omega_1)_{\gamma\alpha}\circ S(A)_{\gamma\alpha} & S(A)_{\gamma\gamma}\end{bmatrix}\overline{V}_1^T$$
$$+\overline{U}\begin{bmatrix} (\Omega_2)_{aa}\circ T(A)_{aa} & (\Omega_2)_{ab}\circ T(A)_{ab} \\ (\Omega_2)_{ba}\circ T(A)_{ba} & T(A)_{bb}\end{bmatrix}\overline{V}_1^T + \overline{U}\begin{bmatrix} (\Omega_3)_{ac'}\circ B_{ac'} \\ B_{bc'}\end{bmatrix}\overline{V}_2^T \quad (121)$$

with $A := \overline{U}^T H\overline{V}_1$, $B := \overline{U}^T H\overline{V}_2^T$*; and under condition (iii),* $(\overline{\eta},\overline{H}) = (0,0)$.

Proof. " $\Longleftarrow$ " Suppose that $(t,X)\in \mathbb{R}\times\mathbb{R}^{m\times n}$ satisfies one of the three conditions (i), (ii) and (iii). Since $\Pi_{\mathcal{K}}(\cdot,\cdot)$ is Lipschitz continuous on $\mathbb{R}\times\mathbb{R}^{m\times n}$, we know that the Gâteaux differentiability and Fréchet differentiability of $\Pi_{\mathcal{K}}(\cdot,\cdot)$ coincide [10]. Therefore, we only need to show that $\Pi_{\mathcal{K}}(\cdot,\cdot)$ is Gâteaux differentiable at (t,X). From Theorem 4.2, we know that $\Pi_{\mathcal{K}}(\cdot,\cdot)$ is directionally differentiable everywhere and the operator $\Pi'_{\mathcal{K}}((t,X);(\cdot,\cdot)) : \mathbb{R}\times\mathbb{R}^{m\times n}\to \mathbb{R}\times\mathbb{R}^{m\times n}$ is linear. Therefore, $\Pi_{\mathcal{K}}(\cdot,\cdot)$ is Gâteaux differentiable at (t,X).

" $\Longrightarrow$ " Let $(t,X)\in\mathbb{R}\times\mathbb{R}^{m\times n}$ be given. We only need to show that $\Pi_{\mathcal{K}}(\cdot,\cdot)$ is not differentiable in the following two cases:

Case 1: $\|X\|_2 = t > 0$ or $\|X\|_2 > t > -\|X\|_*$ and $\sigma_{\bar{k}+1}(X) = \theta(t,\sigma(X))$. In this case, we know that $\beta\neq\emptyset$ and for any $(\eta,H)\in\mathbb{R}\times\mathbb{R}^{m\times n}$, $\rho(\eta,H)$ is given by

$$\rho(\eta,H) = \delta^{-1}(\eta + \mathrm{Tr}(S(\overline{U}_\alpha^T H\overline{V}_\alpha)))\,.$$

Let

$$\Phi(\eta,H) := \Pi_{\mathcal{M}^\delta_{|\beta|}}(\rho(\eta,H), S(\overline{U}_\beta^T H\overline{V}_\beta)) \in \mathbb{R}\times\mathcal{S}^{|\beta|}\,,\quad (\eta,H)\in\mathbb{R}\times\mathbb{R}^{m\times n}\,.$$

Choose $(\eta,H) = (1,0)\in\mathbb{R}\times\mathbb{R}^{m\times n}$. Then, from Propositions 3.1 and 4.2 we know that

$$\Phi(\eta,H) = \Pi_{\mathcal{M}^\delta_{|\beta|}}(\delta^{-1},0) = (\delta^{-1},0)$$

and

$$\Phi(-(\eta,H)) = \Phi(-1,0) = \Pi_{\mathcal{M}^\delta_{|\beta|}}(-\delta^{-1},0) = -(|\beta|+\delta^2)^{-1}(\delta, I_{|\beta|}) \neq -\Phi(\eta,H).$$

Therefore, $\Phi(\cdot,\cdot)$ is not a linear mapping. This, together with Part (ii) in Theorem 4.2, means that $\Pi_{\mathcal{K}}(\cdot,\cdot)$ is not Gâteaux differentiable at (t,X).

Case 2: $t = -\|X\|_*$. In this case, we know that $\theta(t,\sigma(X)) = 0$. Therefore, we have

$$\alpha = a = \{i\,|\,\sigma_i(X) > 0\},\quad \beta = b = \{i\,|\,\sigma_i(X) = 0\}\quad \text{and}\quad \gamma = \emptyset\,.$$

Also, in this case, we have

$$\rho(\eta,H) = \delta^{-1}(\eta + \mathrm{Tr}(S(\overline{U}_\alpha^T H\overline{V}_\alpha))),\quad (\eta,H)\in\mathbb{R}\times\mathbb{R}^{m\times n}\,.$$

Let

$$\Phi(\eta,H) := \Pi_{\mathcal{K}^\delta_{|\beta|,(n-|a|)}}\Big(\rho(\eta,H), \Big[\overline{U}_\beta^T HV_\beta \;\; \overline{U}_\beta^T H\overline{V}_2\Big]\Big)\,,\quad (\eta,H)\in\mathbb{R}\times\mathbb{R}^{m\times n}\,.$$

Fix $(\eta,H) = (1,0)\in\mathbb{R}\times\mathbb{R}^{m\times n}$. Then, from Theorem 4.1, we know that $\Phi(\eta,H) = (\delta^{-1},0)\in \mathbb{R}\times\mathbb{R}^{|\beta|\times(n-|a|)}$ and $\Phi(-(\eta,H)) = (0,0)\in\mathbb{R}\times\mathbb{R}^{|\beta|\times(n-|a|)}$. Therefore, $\Phi(\cdot,\cdot)$ is not a linear mapping. Consequently, we know from Part (iii) in Theorem 4.2, that $\Pi_{\mathcal{K}}(\cdot,\cdot)$ is not Gâteaux differentiable at (t,X).

From the fact that $\beta=\emptyset$ if (t,X) satisfies condition (ii), we can conclude from Theorem 4.2 that (120) and (121) hold. Moreover, the formulas for $(\overline{\eta},\overline{H})$ under conditions (i) and (iii) follow directly from Theorem 4.2. The proof is completed. $\square$

Theorem 4.4 *The metric projector $\Pi_{\mathcal{K}}(\cdot,\cdot)$ is strongly G-semismooth at $(t,X)\in\mathbb{R}\times\mathbb{R}^{m\times n}$.*

Proof. Denote the set of points in $\mathbb{R}\times\mathbb{R}^{m\times n}$ where $\Pi_{\mathcal{K}}(\cdot,\cdot)$ is differentiable by $D_{\Pi_{\mathcal{K}}}$. By Lemma 2.1, in order to show that $\Pi_{\mathcal{K}}(\cdot,\cdot)$ is strongly G-semismooth at $(t,X)\in\mathbb{R}\times\mathbb{R}^{m\times n}$ we only need to show that for any $(\tau,Y)\in D_{\Pi_{\mathcal{K}}}$ converging to (t,X),

$$\Pi_{\mathcal{K}}(\tau,Y)-\Pi_{\mathcal{K}}(t,X)-\Pi_{\mathcal{K}}'(\tau,Y)(\eta,H)=O(\|\eta,H\|^2)\,, \tag{122}$$

where $(\eta,H):=(\tau,Y)-(t,X)\in\mathbb{R}\times\mathbb{R}^{m\times n}$. When $t>\|X\|_2$ or $t<-\|X\|_*$, according to Theorem 4.3, $\Pi_{\mathcal{K}}(\cdot,\cdot)$ is locally a linear function near (t,X) and thus (122) holds. From now on we always assume that (t,X) satisfies $\|X\|_2\ge t\ge-\|X\|_*$.

Recall that for any $(\tau,Y)\in\mathbb{R}\times\mathbb{R}^{m\times n}$, $(g_0(\tau,\sigma(Y)),g(\tau,\sigma(Y)))$ is defined by (79), $G(\tau,Y)$ is defined by (80) and $G_S(Y)$ and $G_R(\tau,Y)$ are defined by (82). Since $G_S(X)=G(t,X)$, we have

$$G(\tau,Y)-G(t,X)=G_S(Y)-G_S(X)+G_R(\tau,Y)\quad\forall\,(\tau,Y)\in\mathbb{R}\times\mathbb{R}^{m\times n}\,.$$

Suppose that $U\in\mathcal{O}^m$ and $V\in\mathcal{O}^n$ (depending on Y) are such that

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

By Proposition 2.11, we know that there exists an open neighborhood $\mathcal{N}$ of X in $\mathbb{R}^{m\times n}$ such that for each $k\in\{1,\ldots,r\}$, $\mathcal{P}_k(\cdot)$ is twice continuously differentiable in $\mathcal{N}$. By taking a smaller $\mathcal{N}$ if necessary, we assume that for any $Y\in\mathcal{N}$ and $k,l\in\{1,\ldots,r\}$,

$$\sigma_i(Y)>0,\quad \sigma_i(Y)\neq\sigma_j(Y)\quad\forall\,i\in a_k,\ j\in a_l\text{ and }k\neq l\,. \tag{123}$$

Then, from (48), we obtain that for any $Y\in\mathcal{N}$,

$$\begin{aligned}G_S(Y)-G_S(X)&=\sum_{k=1}^{r}\overline{\nu}_k(\mathcal{P}_k(Y)-\mathcal{P}_k(X))=\sum_{k=1}^{r}\overline{\nu}_k\mathcal{P}_k'(Y)H+O(\|H\|^2)\\&=\sum_{k=1}^{r}\overline{\nu}_k(U[\Gamma_k\circ S(A)]V_1^T+U[\Xi_k\circ T(A)]V_1^T+U[\Upsilon_k\circ B]V_2^T)+O(\|H\|^2)\,,\end{aligned}$$

where $A:=U^THV_1\in\mathbb{R}^{m\times m}$ and $B:=U^THV_2\in\mathbb{R}^{m\times(n-m)}$; and for $k\in\{1,\ldots,r\}$, $\Gamma_k\in\mathbb{R}^{m\times m}$, $\Xi_k\in\mathbb{R}^{m\times m}$ and $\Upsilon_k\in\mathbb{R}^{m\times(n-m)}$ are given in (45), (46) and (47), respectively. Since $\Pi_{\mathcal{C}_m}(\cdot,\cdot)$ is globally Lipschitz continuous on $\mathbb{R}\times\mathbb{R}^m$, we know that for any $(\tau,Y)\in\mathbb{R}\times\mathbb{R}^{m\times n}$ converging to (t,X),

$$g_i(\tau,\sigma(Y))=\overline{\nu}_k+O(\|(\eta,H)\|)\quad\forall\,i\in a_k,\quad k=1,\ldots,r\,.$$

Therefore, since $U \in \mathcal{O}^m$ and $V \in \mathcal{O}^n$ are uniformly bounded, there exists an open neighborhood $\widehat{\mathcal{N}}$ of (t, X) in $\mathbb{R} \times \mathbb{R}^{m\times n}$ such that for any $(\tau, Y) \in \widehat{\mathcal{N}}$,

$$G_S(Y) - G_S(X) = U[\Gamma' \circ S(A)]V_1^T + U[\Xi' \circ T(A)]V_1^T + U[\Upsilon' \circ B]V_2^T + O(\|(\eta, H)\|^2)\,, \quad (124)$$

where $\Gamma' \in \mathbb{R}^{m\times m}$, $\Xi' \in \mathbb{R}^{m\times m}$ and $\Upsilon' \in \mathbb{R}^{m\times(n-m)}$ are given, respectively, by

$$(\Gamma')_{ij} = \begin{cases} \dfrac{g_i(\tau,\sigma(Y)) - g_j(\tau,\sigma(Y))}{\sigma_i(Y) - \sigma_j(Y)} & \text{if } i \in a_k,\ j \in a_l \text{ and } l \neq k\,, \\ 0 & \text{otherwise}\,, \end{cases} \qquad k, l = 1, \ldots, r+1\,,$$

$$(\Xi')_{ij} = \begin{cases} \dfrac{g_i(\tau,\sigma(Y)) + g_j(\tau,\sigma(Y))}{\sigma_i(Y) + \sigma_j(Y)} & \text{if } i \notin b \quad \text{or} \quad j \notin b\,, \\ 0 & \text{otherwise} \end{cases}$$

and

$$(\Upsilon')_{ij} = \begin{cases} \dfrac{g_i(\tau,\sigma(Y))}{\sigma_i(Y)} & \text{if } i \in a_k,\ k = 1, \ldots, r\,, \\ 0 & \text{if } i \in b\,, \end{cases} \qquad j = 1, \ldots, n-m\,.$$

Let $(\tau, Y) \in D_{\Pi_{\mathcal{K}}} \cap \widehat{\mathcal{N}}$. Note that by replacing (t, X) with (τ, Y), we can also use (75) to define an index integer $\bar{k}$ for (τ, Y). We denote this index integer by $\bar{k}'$ to distinguish the index integer for (t, X). If $\beta \neq \emptyset$, then since $\|X\|_2 \geq t \geq -\|X\|_*$, from (75) and (76) we know that

$$\sigma_{\bar{k}+1}(X) = \theta(t, \sigma(X)) < \sigma_{\bar{k}}(X)\,.$$

Therefore, since for any $k \in \beta$, $\sigma_k(X) = \sigma_{\bar{k}+1}(X)$, we have

$$\sigma_{\bar{k}+|\beta|+1}(X) < (\sum_{i=1}^{k} \sigma_i(X) + t)/(k+1) = \theta(t, \sigma(X)) < \sigma_{\bar{k}}(X) \quad \forall\, k \in \beta\,.$$

If $\beta = \emptyset$, we have

$$\sigma_{\bar{k}+|\beta|+1}(X) < \theta(t, \sigma(X)) < \sigma_{\bar{k}}(X)\,.$$

Therefore, in both cases, by the continuity of the singular value function $\sigma(\cdot)$, we may assume that the integer $\bar{k}'$ lies in $\{\bar{k}, \bar{k}+1, \ldots, \bar{k}+|\beta|\}$, i.e., there exists an integer $j \in \{0, 1, \ldots, |\beta|\}$ such that $\bar{k}' = \bar{k} + j$. Define the corresponding index sets in $\{1, \ldots, m\}$ for (τ, Y) by

$$\alpha' := \{i \,|\, \sigma_i(Y) > \theta(\tau, \sigma(Y))\},\ \ \beta' := \{i \,|\, \sigma_i(Y) = \theta(\tau, \sigma(Y))\} \text{ and } \gamma' := \{i \,|\, \sigma_i(Y) < \theta(\tau, \sigma(Y))\}$$

and

$$a' := \{i \,|\, \sigma_i(Y) > 0\} \quad \text{and} \quad b' := \{i \,|\, \sigma_i(Y) = 0\}\,.$$

Since $(\tau, Y) \in D_{\Pi_{\mathcal{K}}} \cap \widehat{\mathcal{N}}$, from Theorem 4.3 we know that $\beta' = \emptyset$. Meanwhile, by (123), we have

$$\alpha' \supseteq \alpha, \quad \gamma' \supseteq \gamma, \quad a' \supseteq a \quad \text{and} \quad b' \subseteq b\,. \quad (125)$$

Let $\delta' := \sqrt{1 + \bar{k}'}$ and $\rho' \in \mathbb{R}$ be defined by

$$\rho' := \begin{cases} \delta'^{-1}(\eta + \mathrm{Tr}(S(U_{\alpha'}^T H V_{\alpha'}))) & \text{if } \tau \geq -\|Y\|_*\,, \\ 0 & \text{otherwise}\,. \end{cases} \quad (126)$$

Define $\Omega_1' \in \mathbb{R}^{m\times m}$, $\Omega_2' \in \mathbb{R}^{m\times m}$ and $\Omega_3' \in \mathbb{R}^{m\times(n-m)}$ by (83), (84) and (85), respectively with (t,X) being replaced by (τ, Y). Therefore, from Theorem 4.3 we know that

$$\begin{aligned} & G'(\tau,Y)(\eta,H) \\ = \; & U\begin{bmatrix} \delta'^{-1}\rho' I_{|\alpha'|} & (\Omega_1')_{\alpha'\gamma'}\circ S(A)_{\alpha'\gamma'} \\ (\Omega_1')_{\gamma'\alpha'}\circ S(A)_{\gamma'\alpha'} & S(A)_{\gamma'\gamma'} \end{bmatrix} V_1^T \\ & + U\begin{bmatrix} (\Omega_2')_{a'a'}\circ T(A)_{a'a'} & (\Omega_2')_{a'b'}\circ T(A)_{a'b'} \\ (\Omega_2')_{b'a'}\circ T(A)_{b'a'} & T(A)_{b'b'} \end{bmatrix} V_1^T + U\begin{bmatrix} (\Omega_3')_{a'c'}\circ B_{a'c'} \\ B_{b'c'} \end{bmatrix} V_2^T \,, \end{aligned} \tag{127}$$

where $A := U^T H V_1$, $B := U^T H V_2^T$ and $c' = \{1,\ldots,n-m\}$. Let

$$\widehat{R}(\eta,H) := G'(\tau,Y)(\eta,H) - (G_S(Y) - G_S(X))\,.$$

From the formula of $\Pi_{\mathcal{C}_m}(\tau,\sigma(Y))$ in (58), we know that

$$g_i(\tau,\sigma(Y)) = g_j(\tau,\sigma(Y)) \quad \forall\, i,j\in\alpha' \quad \text{and} \quad g_i(\tau,\sigma(Y)) = \sigma_i(Y) \quad \forall\, i\in\gamma'\,.$$

Therefore, by (124) and (127), we obtain from (125) that there exist $R_k(\eta,H) \in \mathbb{R}^{|a_k|\times|a_k|}$, $k=1,\ldots,r$ and $R_{r+1}(\eta,H)\in\mathbb{R}^{|b|\times(n-|a|)}$ such that

$$\widehat{R}(\eta,H) = U\begin{bmatrix} R_1(\eta,H) & 0 & \cdots & 0 & 0 \\ 0 & R_2(\eta,H) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & R_r(\eta,H) & 0 \\ 0 & 0 & \cdots & 0 & R_{r+1}(\eta,H) \end{bmatrix} V^T + O(\|(\eta,H)\|^2)\,, \tag{128}$$

where the formulas of $R_i(\eta,H)$, $i=1,\ldots,r+1$ are determined by the following two cases:

Case 1: $\|X\|_2 \geq t > -\|X\|_*$. In this case, we know that $\theta(t,\sigma(X)) > 0$ and there exists $\bar{r}\in\{0,1,\ldots,r\}$ such that

$$\alpha = \bigcup_{k=1}^{\bar{r}} a_k, \quad \beta = a_{\bar{r}+1} \text{ (or } \emptyset) \quad \text{and} \quad \gamma = \bigcup_{k=r'}^{r} a_k \bigcup b\,,$$

where $r' = \bar{r}+2$ if $\beta\neq\emptyset$ and $r' = \bar{r}+1$ if $\beta=\emptyset$. Since there exists an integer $j\in\{0,1,\ldots,|\beta|\}$ such that $\bar{k}' = \bar{k}+j$, we can define two index sets

$$\beta_1 := \{\bar{k}+1,\ldots,\bar{k}+j\} \quad \text{and} \quad \beta_2 := \{\bar{k}+j+1,\ldots,\bar{k}+|a_{\bar{r}+1}|\}\,.$$

Therefore, by noting that $\alpha' = \alpha\cup\beta_1$, $\gamma' = \beta_2\cup\gamma$ and $\beta_1=\emptyset$ if $\beta=\emptyset$, we obtain from (124) and (127) that

$$\left\{\begin{array}{rcl} R_k(\eta,H) &=& \delta'^{-1}\rho' I_{|a_k|}, \quad k=1,\ldots,\bar{r}\,, \\ R_{\bar{r}+1}(\eta,H) &=& \begin{bmatrix} \delta'^{-1}\rho' I_{|\beta_1|} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & (\Omega_1')_{\beta_1\beta_2} \\ (\Omega_1')_{\beta_2\beta_1} & E \end{bmatrix} \circ S(A_{a_{\bar{r}+1}a_{\bar{r}+1}})\,, \\ R_k(\eta,H) &=& S(A_{a_k a_k}), \quad k=\bar{r}+2,\ldots,r\,, \\ R_{r+1}(\eta,H) &=& [A_{bb} \;\; B_{bc'}]\,, \end{array}\right. \tag{129}$$

where E is a $(|a_{\bar{r}+1}|-j)$ by $(|a_{\bar{r}+1}|-j)$ matrix whose entries are all ones.

Case 2: $t=-\|X\|_*$. In this case, we know that $\theta(t,\sigma(X))=0$. Therefore, we have

$$\alpha=\bigcup_{k=1}^{r} a_k = a, \quad \beta=b \quad \text{and} \quad \gamma=\emptyset\,.$$

Also, since there exists an integer $j\in\{0,1,\ldots,|\beta|\}$ such that $\bar{k}'=\bar{k}+j$, we can define two index sets

$$\beta_1:=\{\bar{k}+1,\ldots,\bar{k}+j\} \quad \text{and} \quad \beta_2:=\{\bar{k}+j+1,\ldots,\bar{k}+|b|\}\,.$$

Therefore, since $\alpha'=\alpha\cup\beta_1$ and $\gamma'=\beta_2\cup\gamma$, we obtain from (124) and (127) that

$$\left\{\begin{array}{rcl} R_k(\eta,H) &=& \delta'^{-1}\rho' I_{|a_k|}, \quad k=1,\ldots,r\,, \\ R_{r+1}(\eta,H) &=& \left[\left(\begin{array}{cc} \delta'^{-1}\rho' I_{|\beta_1|} & 0 \\ 0 & 0 \end{array}\right)+\left(\begin{array}{cc} 0 & (\Omega_1')_{\beta_1\beta_2} \\ (\Omega_1')_{\beta_2\beta_1} & E \end{array}\right)\circ A_{bb} \;\; B_{bc'}\right]\,, \end{array}\right. \tag{130}$$

where E is a $(|b|-j)$ by $(|b|-j)$ matrix whose entries are all ones.

Consider the singular value decomposition of X, i.e., $X=\overline{U}\,[\Sigma(X)\;\;0]\,\overline{V}^T$, where $\overline{U}\in\mathcal{O}^m$ and $\overline{V}\in\mathcal{O}^n$. Then, we have

$$[\Sigma(X)\;\;0]+\overline{U}^T H\overline{V}=\overline{U}^T U\,[\Sigma(Y)\;\;0]\,V^T\overline{V}\,.$$

Let $\widetilde{H}:=\overline{U}^T H\overline{V}$, $\widetilde{U}:=\overline{U}^T U$ and $\widetilde{V}:=\overline{V}^T V$. Then,

$$U^T HV=\widetilde{U}^T\overline{U}^T H\overline{V}\widetilde{V}=\widetilde{U}^T\widetilde{H}\widetilde{V}\,.$$

From (35), we know that there exist $Q_k\in\mathcal{O}^{|a_k|}$, $k=1,\ldots,r$ and $Q'\in\mathcal{O}^{|b|}$, $Q''\in\mathcal{O}^{n-|a|}$ such that

$$A_{a_ka_k}=U_{a_k}^T HV_{a_k}=\widetilde{U}_{a_k}^T\widetilde{H}\widetilde{V}_{a_k}=Q_k^T\widetilde{H}_{a_ka_k}Q_k+O(\|H\|^2), \quad k=1,\ldots,r$$

and

$$[A_{bb}\;\;B_{bc'}]=\left[U_b^T HV_b\;\;U_b^T HV_2\right]=\left[\widetilde{U}_b^T\widetilde{H}\widetilde{V}_b\;\;\widetilde{U}_b^T\widetilde{H}\widetilde{V}_2\right]=Q'^T\left[\widetilde{H}_{bb}\;\;\widetilde{H}_{bc}\right]Q''+O(\|H\|^2)\,.$$

Then, from (36) and (37) in Proposition 2.10, we obtain that

$$S(A_{a_ka_k})=Q_k^T S(\widetilde{H}_{a_ka_k})Q_k+O(\|H\|^2)=\Sigma(Y)_{a_ka_k}-\Sigma(X)_{a_ka_k}+O(\|H\|^2)\,, \quad k=1,\ldots,r$$

and

$$[A_{bb}\;\;B_{bc'}]=Q'^T\left[\widetilde{H}_{bb}\;\;\widetilde{H}_{bc}\right]Q''+O(\|H\|^2)=[\Sigma(Y)_{bb}-\Sigma(X)_{bb}\;\;0]+O(\|H\|^2)\,.$$

Let $h:=\sigma'(Y;H)$. Since $\sigma(\cdot)$ is strongly semismooth [45], we know that

$$\begin{aligned} S(A_{a_ka_k}) &= \operatorname{diag}(\sigma_i'(Y;H):\; i\in a_k)+O(\|H\|^2) \\ &= \operatorname{diag}(h_{a_k})+O(\|H\|^2)\,, \quad k=1,\ldots,r \end{aligned} \tag{131}$$

and

$$
\begin{aligned}
[A_{bb} \;\; B_{bc'}] &= \left[\mathrm{diag}(\sigma_i'(Y;H):\; i\in b) \;\; 0\right] + O(\|H\|^2) \\
&= [\mathrm{diag}(h_b) \;\; 0] + O(\|H\|^2)\,. \qquad (132)
\end{aligned}
$$

Therefore, by noting that in each case $\alpha' = \alpha\cup\beta_1$ and $\gamma' = \beta_2\cup\gamma$ and that $0 \le (\Omega_1')_{i,j} \le 1$ for any $i\in\beta_1$ and $j\in\beta_2$, we obtain from (128), (129), (130), (131) and (132) that

$$
\widehat{R}(\eta,H) = U\begin{bmatrix} \delta'^{-1}\rho' I_{|\alpha'|} & 0 & 0 \\ 0 & \mathrm{diag}(h_{\gamma'}) & 0 \end{bmatrix} V^T + O(\|(\eta,H)\|^2)\,. \qquad (133)
$$

On the other hand, by (42) and (43), for Y sufficiently close to X, we have

$$
\mathcal{P}_k(Y) = \sum_{i\in a_k} u_i v_i^T, \quad k = 1,\ldots,r\,.
$$

Therefore, we obtain that for any $(\tau,Y)\in D_{\Pi_{\mathcal{K}}}\cap\widehat{\mathcal{N}}$ (shrinking $\widehat{\mathcal{N}}$ if necessary),

$$
G_R(\tau,Y) = G(\tau,Y) - G_S(Y) = \sum_{k=1}^{r}\sum_{i\in a_k}[g_i(\tau,\sigma(Y)) - g_i(t,\sigma(X))]u_i v_i^T + \sum_{i\in b} g_i(\tau,\sigma(Y))u_i v_i^T\,.
$$

Note that from Part (iii) of Proposition 3.2 and Theorem 4.3, we know that $\Pi_{\mathcal{K}}(\cdot,\cdot)$ is differentiable at (τ,Y) if and only if $\Pi_{\mathcal{C}_m}(\cdot,\cdot)$ is differentiable at $(\tau,\sigma(Y))$. Since the continuous mapping $\Pi_{\mathcal{C}}(\cdot,\cdot)$ is piecewise linear, it is strongly G-semismooth at $(t,\sigma(X))$. Meanwhile, we know that the singular value function $\sigma(\cdot)$ is strongly semismooth at X. Therefore, we obtain that for any $(\tau,Y)\in D_{\Pi_{\mathcal{K}}}\cap\widehat{\mathcal{N}}$ (shrinking $\widehat{\mathcal{N}}$ if necessary),

$$
\begin{aligned}
\Pi_{\mathcal{C}_m}(\tau,\sigma(Y)) - \Pi_{\mathcal{C}_m}(t,\sigma(X)) &= \Pi_{\mathcal{C}_m}'(\tau,\sigma(Y))(\eta,\sigma(Y)-\sigma(X)) + O(\|(\eta,H)^2\|) \\
&= \Pi_{\mathcal{C}_m}'(\tau,\sigma(Y))(\eta,\sigma'(Y;H) + O(\|H\|^2)) + O(\|(\eta,H)^2\|) \\
&= \Pi_{\mathcal{C}_m}'(\tau,\sigma(Y))(\eta,\sigma'(Y;H)) + O(\|(\eta,H)^2\|)\,.
\end{aligned}
$$

Let

$$
(\phi_0(\eta,h),\phi(\eta,h)) := \Pi_{\mathcal{C}_m}'(\tau,\sigma(Y))(\eta,h)\,.
$$

Then, we have

$$
g_0(\tau,\sigma(Y)) - g_0(t,\sigma(X)) = \phi_0(\eta,h) + O(\|(\eta,H)\|^2) \qquad (134)
$$

and

$$
g_i(\tau,\sigma(Y)) - g_i(t,\sigma(X)) = \phi_i(\eta,h) + O(\|(\eta,H)\|^2), \quad i=1,\ldots,m\,.
$$

Since $U \in \mathcal{O}^m$ and $V \in \mathcal{O}^n$ are uniformly bounded, we know that

$$
\begin{aligned}
G_R(\tau, Y) &= U \begin{bmatrix} \phi_1(\eta,h) & 0 & \cdots & 0 \\ 0 & \phi_2(\eta,h) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_m(\eta,h) \end{bmatrix} V_1^T + O(\|(\eta,H)\|^2) \\
&= U \begin{bmatrix} \phi_1(\eta,h) & 0 & \cdots & 0 & 0 \\ 0 & \phi_2(\eta,h) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \phi_m(\eta,h) & 0 \end{bmatrix} V^T + O(\|(\eta,H)\|^2)\,.
\end{aligned}
$$

From Part (ii) of Proposition 3.2, we have

$$\phi_0(\eta,h) = \delta'^{-1}\rho' \tag{135}$$

and

$$\phi_i(\eta,h) = \phi_0(\eta,h) \quad \forall\, 1 \le i \le \bar{k}' \quad \text{and} \quad \phi_i(\eta,h) = h_i \quad \forall\, \bar{k}'+1 \le i \le m\,.$$

Thus, from (133), we obtain that

$$\widehat{R}(\eta,H) = G_R(\tau,Y) + O(\|(\eta,H)\|^2)\,. \tag{136}$$

That is, for any for any $(\tau,Y) \in D_{\Pi_{\mathcal{K}}}$ converging to (t,X),

$$
\begin{aligned}
G(\tau,Y) - G(t,X) - G'(\tau,Y)(\eta,H) &= G_S(Y) - G_S(X) - G'(\tau,Y)(\eta,H) + G_R(\tau,Y) \\
&= -\widehat{R}(\eta,H) + G_R(\tau,Y) \\
&= O(\|(\eta,H)\|^2)\,,
\end{aligned}
$$

which, together with (81), (135), (120) and (134), shows that (122) holds. □

5 Conclusions

In this paper, we have identified a class of matrix cone programming involving the epigraphs of the l_1, l_∞, Frobenius, spectral and nuclear norms that has many important applications. In order to make this class of problems tractable via variants of the augmented Lagrange method, we have made efforts to establish several key properties including the closed form solution, calm B-differentiability and strong semismoothness of the metric projection operator over the epigraph of the l_1, l_∞, spectral, and nuclear matrix norm, respectively. These results, together with the known analogous ones for symmetric cones, will constitute the backbone for using augmented Lagrangian methods to solve large scale problems of practical significance. Our next step is to develop numerical algorithms and software along this line. The work done in this paper on matrix cone programming is by no means complete. There are many unanswered questions. For example, besides the analytic solution and the first order differentiability of the metric projector over the epigraphs of the spectral and nuclear matrix norms, the research on the second order properties of these non-polyhedral closed convex sets is certainly of paramount necessity for understanding second order optimality conditions of matrix cone programming. Another direction is to consider convex matrix cones beyond epigraphs of matrix norms such as the epigraph of the convex function that is defined as the sum of the first several largest singular values of a matrix (or the Ky Fan k-norm). It is our firm belief that a better understanding of the inherent structures of these matrix cones rather than projecting them into higher dimensional spaces will lead to more efficient optimization methods for solving matrix cone programming.

References

[1] Bhatia, R.: *Matrix Analysis*, Springer-Verlag, New York, 1997.

[2] Candès, E.J. and Recht, B.: *Exact matrix completion via convex optimization*, Foundations of Computational Mathematics 9 (2008) 717–772.

[3] Candès, E.J. and Tao, T.: *The power of convex relaxation: Near-optimal matrix completion*, IEEE Transactions on Information Theory 56 (2009) 2053–2080.

[4] Candès, E.J., Li, X., Ma, Y. and Wright, J.: *Robust principal component analysis?*, Preprint available at `http://www-stat.stanford.edu/~candes/papers/RobustPCA.pdf`.

[5] Chandrasekaran, V., Sanghavi, S., Parrilo, P.A. and Willsky, A.: *Rank-sparsity incoherence for matrix decomposition*, Preprint available at `http://arxiv.org/abs/0906.2220`.

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

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

[8] Chen, X.D., Sun, D.F. and Sun, J.: *Complementarity functions and numerical experiments for second-order-cone complementarity problems*, Computational Optimization and Applications 25 (2003) 39–56.

[9] Chu, M., Funderlic, R. and Plemmons, R.: *Structured low rank approximation*, Linear Algebra and its Applications 366 (2003) 157–172.

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

[11] Donoghue, W.F.: *Monotone Matrix Functions and Analytic Continuation*, Springer, New York, 1974.

[12] Fan, K.: *On a theorem of Weyl concerning eigenvalues of linear transformations*, Proceedings of the National Academy of Sciences of U.S.A. 35 (1949) 652–655.

[13] Faraut, J. and Korányi, A.: *Analysis on Symmetric Cones*, Clarendon Press, Oxford, 1994.

[14] Gao, Y. and Sun, D.F.: *A majorized penalty approach for calibrating rank constrained correlation matrix problems*, Preprint available at http://www.math.nus.edu.sg/∼matsundf/MajorPen.pdf.

[15] Greenbaum, A. and Trefethen, L.N.: *GMRES/CR and Arnoldi/Lanczos as matrix approximation problems*, SIAM Journal on Scientific Computing 15 (1994) 359–368.

[16] Gross, D.: *Recovering low-rank matrices from few coefficients in any basis*, Preprint available at http://arxiv.org/abs/0910.1879v4.

[17] Haraux, A.: *How to differentiate the projection on a convex set in Hilbert space. Some applications to variational inequalities*, Journal of the Mathematical Society of Japan 29 (1977) 615–631.

[18] Hardy, G.H., Littlewood, J.E. and Pólya, G.: *Inequalities*, 2nd edition, Cambridge University Press, 1952.

[19] Higham, N.J.: *Computing a nearest symmetric positive semidefinite matrix*, Linear Algebra and Its Applications 103 (1988) 103–118.

[20] Horn, R.A. and Johnson, C.R.: *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[21] Horn, R.A. and Johnson, C.R.: *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.

[22] Keshavan, R.H., Montanari, A. and Oh, S.: *Matrix completion from a few entries*, Preprint available at http://arxiv.org/abs/0901.3150.

[23] Korányi, A.: *Monotone functions on formally real Jordan algebras*, Mathematische Annalen 269 (1984) 73–76.

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

[25] Lewis, A.S. and Sendov, H.S.: *Twice differentiable spectral functions*, SIAM Journal on Matrix Analysis and Applications 23 (2001) 368–386.

[26] Lewis, A.S. and Sendov, H.S.: *Nonsmooth analysis of singular values. Part II: applications*, Set-Valued Analysis 13 (2005) 243–264.

[27] Liu, G., Lin, Z. and Yu, Y.: *Robust subspace segmentation by low-rank representation*, Proceedings of the 26^{th} International Conference on Machine Learning, Haifa, Israel, 2010.

[28] Löwner, K.: *Über monotone matrixfunktionen*, Mathematische Zeitschrift 38 (1934) 177–216.

[29] Malick, J., Povh, J., Rendl, F. and Wiegele, A.: *Regularization methods for semidefinite programming*, SIAM Journal on Optimization 20 (2009) 336–356.

[30] Moreau, J.J.: *Décomposition orthogonale d'un espace hilbertien selon deux cones mutuellement polaires*, Comptes Rendus de l'Académie des Sciences 255 (1962) 238–240.

[31] von Neumann, J.: *Some matrix-inequalities and metrization of matric-space*, Tomsk University Review 1 (1937) 286–300. In: *Collected Works*, Pergamon, Oxford, 1962, Volume IV, 205-218.

[32] Pang, J.S.: *Newton's method for B-differentiable equations*, Mathematics of Operations Research 15 (1990) pp. 149-160.

[33] Povh, J., Rendl, F. and Wiegele, A.: *A boundary point method to solve semidefinite programs*, Computing, 78 (2006) 277-286.

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

[35] Recht, B.: *A Simpler Approach to Matrix Completion*, Preprint available at `http://pages.cs.wisc.edu/~brecht/publications.html`.

[36] Recht, B., Fazel, M. and Parrilo, P.A.: *Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization*, SIAM Review to appear.

[37] Rockafellar, R.T.: *Convex Analyis*, Princeton University Press, Princeton, 1970.

[38] Rockafellar, R.T.: *Conjugate Duality and Optimization*, SIAM, Philadelphia, 1974.

[39] Scholtes, S.: *Introduction to Piecewise Differentiable Equations*, Habilitation Thesis, Institut für Statistik und Mathematische Wirtschaftstheorie, Universit"at Karlsruhe, 1994.

[40] Schwertman, N.C. and Allen, D.M.: *Smoothing an indefinite variance-covariance matrix*, Journal of Statistical Computation and Simulation 9 (1979) 183–194.

[41] Shapiro, A.: *On differentiability of symmetric matrix valued functions*, Optimization Online, 2002.

[42] Stewart, G.W. and Sun, J.G.: *Matrix Perturbation Theory*, Academic Press, New York, 1990.

[43] Sun, D.F.: *Algorithms and Convergence Analysis for Nonsmooth Optimization and Nonsmooth Equations*, PhD Thesis, Institute of Applied Mathematics, Chinese Academy of Sciences, China, December 1994.

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

[45] Sun, D.F. and Sun, J.: *Strong semismoothness of eigenvalues of symmetric matrices and its applications in inverse eigenvalue problems*, SIAM Journal on Numerical Analysis 40 (2003) 2352–2367.

[46] Sun, D.F. and Sun, J.: *Löwner's Operator and spectral functions in Euclidean Jordan algebras*, Mathematics of Operations Research 33 (2008) 421–445.

[47] Toh, K.C.: *GMRES vs. ideal GMRES*, SIAM J. of Matrix Analysis and Applications 18 (1997) 30–36.

[48] Toh, K.C. and Trefethen, L.N.: *The Chebyshev polynomials of a matrix*, SIAM J. Matrix Analysis and Applications 20 (1998) 400–419.

[49] Torki, M.: *Second-order directional derivatives of all eigenvalues of a symmetric matrix*, Nonlinear Analysis, Ser. A Theory, Methods 46 (2001) 1133–1150.

[50] Tseng, P.: *Merit functions for semi-definite complementarity problems*, Mathematical Programming 83 (1998) 159–185.

[51] Tsing, N.K., Fan, M.K.H. and Verriest, E.I.: *On analyticity of functions involving eigenvalues*, Linear Algebra and Applications 207 (1994) 159–180.

[52] Vandenberghe, L. and Boyd, S.: *Semidefinite programming*, SIAM Review 38 (1996) 49–95.

[53] Wen, Z., Goldfarb, D. and Yin, W.: *Alternating direction augmented Lagrangian methods for semidefinite programming*, Rice University CAAM Technical Report TR09-42, 2009.

[54] Wright, J., Ma, Y., Ganesh, A., and Rao, S.: *Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices via Convex Optimization*, Submitted to the Journal of the ACM, 2009.

[55] Zarantonello, E.H.: *Projections on convex sets in Hilbert space and spectral theory I and II*, Contributions to Nonlinear Functional Analysis (E. H. Zarantonello, ed.), Academic Press, New York, 1971, 237–424.

[56] Zhao, X.Y.: *A semismooth Newton-CG augmented Lagrangian method for large scale linear and convex quadratic SDPs*, PhD thesis, Department of Mathematics, National University of Singapore, 2009.

[57] Zhao, X.Y., Sun, D.F. and Toh, K.C.: *A Newton-CG augmented Lagrangian method for semidefinite programming*, SIAM Journal on Optimization 20 (2010) 1737–1765.

6 Appendix

Proof of Proposition 2.8: "$\Longleftarrow$" Obvious.

"$\Longrightarrow$" Define $\Sigma_+ := \Sigma_{aa}$. Let $\bar{a} := \{1, \ldots, n\} \setminus a$. From (29), we obtain that

$$\begin{bmatrix} P_{aa} & P_{ab} \\ P_{ba} & P_{bb} \end{bmatrix} \begin{bmatrix} \Sigma_+ & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \Sigma_+ & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} W_{aa} & W_{a\bar{a}} \\ W_{\bar{a}a} & W_{\bar{a}\bar{a}} \end{bmatrix},$$

which, implies

$$P_{aa}\Sigma_+ = \Sigma_+ W_{aa}, \quad \Sigma_+ W_{a\bar{a}} = 0 \quad \text{and} \quad P_{ba}\Sigma_+ = 0\,.$$

Since Σ_+ is nonsingular, we know that $W_{a\bar{a}} = 0$ and $P_{ba} = 0$. Then, since W and P are two orthogonal matrices, we also have

$$P^T \begin{bmatrix} \Sigma_+ & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \Sigma_+ & 0 \\ 0 & 0 \end{bmatrix} W^T ,$$

which, implies $W_{\bar{a}a} = 0$ and $P_{ab} = 0$. Therefore, we know that

$$P = \begin{bmatrix} P_{aa} & 0 \\ 0 & P_{bb} \end{bmatrix} \quad \text{and} \quad W = \begin{bmatrix} W_{aa} & 0 \\ 0 & W_{\bar{a}\bar{a}} \end{bmatrix} ,$$

where $W_{aa}, P_{aa} \in \mathcal{O}^{|a|}$, $P_{bb} \in \mathcal{O}^{m-|a|}$ and $W_{\bar{a}\bar{a}} \in \mathcal{O}^{n-|a|}$. By noting that

$$\Sigma_+ = \begin{bmatrix} \overline{\mu}_1 I_{|a_1|} & 0 & \cdots & 0 \\ 0 & \overline{\mu}_2 I_{|a_2|} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \overline{\mu}_r I_{|a_r|} \end{bmatrix} ,$$

from $P_{aa}\Sigma_+ = \Sigma_+ W_{aa}$, we obtain that

$$\begin{bmatrix} \overline{\mu}_1 P_{a_1a_1} & \overline{\mu}_2 P_{a_1a_2} & \cdots & \overline{\mu}_r P_{a_1a_r} \\ \overline{\mu}_1 P_{a_2a_1} & \overline{\mu}_2 P_{a_2a_2} & \cdots & \overline{\mu}_r P_{a_2a_r} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{\mu}_1 P_{a_ra_1} & \overline{\mu}_2 P_{a_ra_2} & \cdots & \overline{\mu}_r P_{a_ra_r} \end{bmatrix} = \begin{bmatrix} \overline{\mu}_1 W_{a_1a_1} & \overline{\mu}_1 W_{a_1a_2} & \cdots & \overline{\mu}_1 W_{a_1a_r} \\ \overline{\mu}_2 W_{a_2a_1} & \overline{\mu}_2 W_{a_2a_2} & \cdots & \overline{\mu}_2 W_{a_2a_r} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{\mu}_r W_{a_ra_1} & \overline{\mu}_r W_{a_ra_2} & \cdots & \overline{\mu}_r W_{a_ra_r} \end{bmatrix} . \tag{137}$$

By using the fact that $\overline{\mu}_k > 0$, $k = 1, \ldots, r$, we obtain from (137) that

$$\begin{cases} P_{a_ka_k} = W_{a_ka_k}, & k = 1, \ldots, r , \quad (138) \\ P_{a_ka_l} = \overline{\mu}_l^{-1}\overline{\mu}_k W_{a_ka_l}, & k, l = 1, \ldots, r, \ k \neq l . \quad (139) \end{cases}$$

Next, we shall show by induction that for each $k \in \{1, \ldots, r\}$,

$$P_{a_ka_l} = W_{a_ka_l} = 0 \quad \text{and} \quad P_{a_la_k} = W_{a_la_k} = 0 \ \ \forall\, l = 1, \ldots, r, \ l \neq k . \tag{140}$$

First for $k = 1$, since P and W are orthogonal matrices, we have

$$I_{|a_1|} = \sum_{l=1}^{r} P_{a_1a_l} P_{a_1a_l}^T = \sum_{l=1}^{r} W_{a_1a_l} W_{a_1a_l}^T .$$

Therefore, by further using (138) and (139), we obtain that

$$\sum_{l=2}^{r} (1 - (\overline{\mu}_l^{-1}\overline{\mu}_1)^2) W_{a_1a_l} W_{a_1a_l}^T = 0 .$$

Since for each $l \in \{2, 3, \ldots, r\}$, $\overline{\mu}_l^{-1}\overline{\mu}_1 > 1$ and $W_{a_1 a_l} W_{a_1 a_l}^T$ is symmetric and positive semidefinite, we can easily conclude that

$$W_{a_1 a_l} = 0 \quad \forall\, l = 2, 3, \ldots, r \quad \text{and} \quad W_{a_1 a_1}^{-1} = W_{a_1 a_1}^T \,.$$

From the condition that $W^T W = I_m$, we also have

$$I_{|a_1|} = W_{a_1 a_1}^T W_{a_1 a_1} + \sum_{l=2}^{r} W_{a_l a_1}^T W_{a_l a_1} \,.$$

Then, $W_{a_1 a_1}^T W_{a_1 a_1} = I_{|a_1|}$ implies that

$$\sum_{l=2}^{r} W_{a_l a_1}^T W_{a_l a_1} = 0 \,.$$

Therefore, we have $W_{a_l a_1} = 0$, for each $l \in \{2, 3, \ldots, r\}$. By (139), we know that (140) holds for $k = 1$.

Now, suppose that for some $p \in \{1, \ldots, r-1\}$, (140) holds for any $k \leq p$. We will show that (140) also holds for $k = p+1$. Since P and W are orthogonal matrices, from the induction assumption we know that

$$I_{|a_{p+1}|} = \sum_{l=p+1}^{r} P_{a_{p+1} a_l} P_{a_{p+1} a_l}^T = \sum_{l=p+1}^{r} W_{a_{p+1} a_l} W_{a_{p+1} a_l}^T \,.$$

From (138) and (139), we obtain that

$$\sum_{l=p+2}^{r} (1 - (\overline{\mu}_l^{-1}\overline{\mu}_{p+1})^2) W_{a_{p+1} a_l} W_{a_{p+1} a_l}^T = 0 \,.$$

Since $\overline{\mu}_l^{-1}\overline{\mu}_{p+1} > 1$ for each $l \in \{p+2, \ldots, r\}$, it can then be checked easily that

$$W_{a_{p+1} a_l} = 0 \quad \forall\, l \in \{p+2, \ldots, r\} \quad \text{and} \quad W_{a_{p+1} a_{p+1}}^{-1} = W_{a_{p+1} a_{p+1}}^T \,.$$

So we have

$$I_{|a_{p+1}|} = W_{a_{p+1} a_{p+1}}^T W_{a_{p+1} a_{p+1}} + \sum_{l=p+2}^{r} W_{a_l a_{p+1}}^T W_{a_l a_{p+1}} \,,$$

which, together with $W_{a_{p+1} a_{p+1}}^T W_{a_{p+1} a_{p+1}} = I_{|a_{p+1}|}$, implies that

$$\sum_{l=p+2}^{r} W_{a_l a_{p+1}}^T W_{a_l a_{p+1}} = 0 \,.$$

Therefore, we have $W_{a_l a_{p+1}} = 0$ for all $l \in \{p+2, \ldots, r\}$. From (139), we know that (140) holds for $k = p+1$.

Since (140) holds for all $k \in \{1, \ldots, r\}$, we obtain from (138) that $P_{aa} = W_{aa}$. Let $Q := P_{aa} = W_{aa}$, $Q' := P_{bb}$ and $Q'' := W_{\bar{a}\bar{a}}$. Then,

$$P = \begin{bmatrix} Q & 0 \\ 0 & Q' \end{bmatrix} \quad \text{and} \quad W = \begin{bmatrix} Q & 0 \\ 0 & Q'' \end{bmatrix},$$

where $Q = \mathrm{diag}(Q_1, Q_2, \ldots, Q_r)$ is a block diagonal orthogonal matrix with the k-th diagonal block given by $Q_k = P_{a_k a_k} \in \mathcal{O}^{|a_k|}$, $k = 1, \ldots, r$. The proof is completed. □

Proof of Proposition 2.10: Let $\widehat{X} := [\Sigma(X) \ \ 0]$. Let $H \in \mathbb{R}^{m \times n}$ be given. We use $I_p^{\uparrow}$ to denote the p by p anti-diagonal matrix whose anti-diagonal entries are all ones and other entries are zeros. Denote

$$U_a^{\uparrow} = U_a I_{|a|}^{\uparrow} \quad \text{and} \quad V_a^{\uparrow} = V_a I_{|a|}^{\uparrow}.$$

Let

$$P^{\uparrow} := \frac{1}{\sqrt{2}} \begin{bmatrix} U_a & U_b & 0 & U_b & U_a^{\uparrow} \\ V_a & V_b & \sqrt{2}\, V_2 & -V_b & -V_a^{\uparrow} \end{bmatrix} \in \mathbb{R}^{(m+n) \times (m+n)}. \tag{141}$$

Then, from (31), we have

$$\mathcal{B}(Y) = \mathcal{B}(\widehat{X}) + \mathcal{B}(H) = P^{\uparrow} \Lambda(\mathcal{B}(Y)) (P^{\uparrow})^T.$$

By Proposition 2.4, we know that for any $H \to 0$, there exists $P' \in \mathcal{O}^{m+n}(\mathcal{B}(\widehat{X}))$ such that

$$P^{\uparrow} - P' = O(\|\mathcal{B}(H)\|) = O(\|H\|). \tag{142}$$

On the other hand, suppose that $\widehat{U} \in \mathcal{O}^m$ and $\widehat{V} \in \mathcal{O}^n$ are two arbitrary orthogonal matrices such that

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

From Proposition 2.8, we know that

$$\widehat{U}_a = \begin{bmatrix} \widehat{U}_{aa} \\ 0 \end{bmatrix} \quad \text{and} \quad \widehat{V}_a = \begin{bmatrix} \widehat{U}_{aa} \\ 0 \end{bmatrix}, \tag{143}$$

where $\widehat{U}_{aa} = \mathrm{diag}(\widehat{U}_{a_1 a_1}, \widehat{U}_{a_2 a_2}, \ldots, \widehat{U}_{a_r a_r})$ is a block diagonal orthogonal matrix with the k-th diagonal block given by $\widehat{U}_{a_k a_k} \in \mathcal{O}^{|a_k|}$, $k = 1, \ldots, r$. Let

$$\widehat{P}^{\uparrow} := \frac{1}{\sqrt{2}} \begin{bmatrix} \widehat{U}_a & \widehat{U}_b & 0 & \widehat{U}_b & \widehat{U}_a^{\uparrow} \\ \widehat{V}_a & \widehat{V}_b & \sqrt{2}\, \widehat{V}_2 & -\widehat{V}_b & -\widehat{V}_a^{\uparrow} \end{bmatrix} \in \mathbb{R}^{(m+n) \times (m+n)},$$

where

$$\widehat{U}_a^{\uparrow} = \widehat{U}_a I_{|a|}^{\uparrow} \quad \text{and} \quad \widehat{V}_a^{\uparrow} = \widehat{V}_a I_{|a|}^{\uparrow}.$$

Then, from (31), we know that the orthogonal matrix $\widehat{P}^{\uparrow} \in \mathcal{O}^{m+n}(\mathcal{B}(\widehat{X}))$. By Proposition 2.2, we know that there exist orthogonal matrices $N_k, N_k' \in \mathcal{O}^{|a_k|}$, $k = 1, \ldots, r$ and $M \in \mathcal{O}^{2|b|+n-m}$ such that

$$P' = \widehat{P}^{\uparrow} \mathrm{diag}(N_1, \ldots, N_r, M, N_r', \ldots, N_1').$$

Therefore, from (142), we obtain that

$$\begin{bmatrix} U_a \\ V_a \end{bmatrix} = \begin{bmatrix} \widehat{U}_a \mathrm{diag}(N_1, N_2, \ldots, N_r) \\ \widehat{V}_a \mathrm{diag}(N_1, N_2, \ldots, N_r) \end{bmatrix} + O(\|H\|) \,. \tag{144}$$

Denote

$$Q := \widehat{U}_{aa} \mathrm{diag}(N_1, N_2, \ldots, N_r) \,.$$

Then, we know that $Q = \mathrm{diag}(Q_1, Q_2, \ldots, Q_r)$ is a block diagonal orthogonal matrix with the k-th diagonal block given by $Q_k = \widehat{U}_{a_k a_k} N_k \in \mathcal{O}^{|a_k|}$, $k = 1, \ldots, r$. Thus, from (143) and (144), we obtain that

$$U_a = \begin{bmatrix} Q \\ 0 \end{bmatrix} + O(\|H\|) \quad \text{and} \quad V_a = \begin{bmatrix} Q \\ 0 \end{bmatrix} + O(\|H\|) \,.$$

Since U and Q are orthogonal matrices, from $0 = U_a^T U_b = Q^T U_{ab} + O(\|H\|)$, we obtain that

$$U_{ab} = O(\|H\|) \,.$$

Therefore, we have

$$I_{|b|} = U_{ab}^T U_{ab} + U_{bb}^T U_{bb} = U_{bb}^T U_{bb} + O(\|H\|^2) \,.$$

By considering the singular value decomposition of U_{bb}, we know that there exists an orthogonal matrix $Q' \in \mathcal{O}^{|b|}$ such that

$$U_{bb} = Q' + O(\|H\|^2) \,.$$

Similarly, since V and Q are orthogonal matrices, from $0 = V_a^T V_{\bar{a}} = Q^T V_{a\bar{a}} + O(\|H\|)$, we know that

$$V_{a\bar{a}} = O(\|H\|) \,,$$

where $\bar{a} = \{1, \ldots, n\} \setminus a$. Therefore, we have

$$I_{|\bar{a}|} = V_{a\bar{a}}^T V_{a\bar{a}} + V_{\bar{a}\bar{a}}^T V_{\bar{a}\bar{a}} = V_{\bar{a}\bar{a}}^T V_{\bar{a}\bar{a}} + O(\|H\|^2) \,.$$

By considering the singular value decomposition of $V_{\bar{a}\bar{a}}$, we know that there exists an orthogonal matrix $Q'' \in \mathcal{O}^{n-|a|}$ such that

$$V_{\bar{a}\bar{a}} = Q'' + O(\|H\|^2) \,.$$

Thus,

$$U = \begin{bmatrix} Q & 0 \\ 0 & Q' \end{bmatrix} + O(\|H\|) \quad \text{and} \quad V = \begin{bmatrix} Q & 0 \\ 0 & Q'' \end{bmatrix} + O(\|H\|) \,. \tag{145}$$

Hence, (35) is proved.

From $\mathcal{B}(\widehat{X}) + \mathcal{B}(H) = P^{\uparrow} \Lambda(\mathcal{B}(Y))(P^{\uparrow})^T$ and $\widehat{P}^{\uparrow} \in \mathcal{O}^{m+n}(\mathcal{B}(\widehat{X}))$, we obtain that

$$\Lambda(\mathcal{B}(\widehat{X})) + (\widehat{P}^{\uparrow})^T \mathcal{B}(H) \widehat{P}^{\uparrow} = (\widehat{P}^{\uparrow})^T P^{\uparrow} \Lambda(\mathcal{B}(Y))(P^{\uparrow})^T \widehat{P}^{\uparrow} \,. \tag{146}$$

Let $\widetilde{P} := (\widehat{P}^{\uparrow})^T P^{\uparrow}$ and $\widetilde{\mathcal{B}}(H) := (\widehat{P}^{\uparrow})^T \mathcal{B}(H) \widehat{P}^{\uparrow}$. Then, we can re-write (146) as

$$\widetilde{P}^T (\Lambda(\mathcal{B}(\widehat{X})) + \widetilde{\mathcal{B}}(H)) \widetilde{P} = \Lambda(\mathcal{B}(Y)) \,. \tag{147}$$

By comparing both sides of (147), we obtain that

$$\widetilde{P}_{a_k}^T \Lambda(\mathcal{B}(\widehat{X}))\widetilde{P}_{a_k} + (P_{a_k}^{\uparrow})^T \mathcal{B}(H) P_{a_k}^{\uparrow} = \Lambda(B(Y))_{a_k a_k}, \quad k = 1, \ldots, r \,. \tag{148}$$

From (18) in Proposition 2.3, we know that

$$\widetilde{P}_{a_k}^T \Lambda(\mathcal{B}(\widehat{X}))\widetilde{P}_{a_k} = \widetilde{P}_{a_k a_k}^T \Lambda(\mathcal{B}(\widehat{X}))_{a_k a_k} \widetilde{P}_{a_k a_k} + O(\|H\|^2) \,.$$

By noting that for each $k \in \{1, \ldots, r\}$, $\Lambda(\mathcal{B}(\widehat{X}))_{a_k a_k} = \Sigma(X)_{a_k a_k} = \overline{\mu}_k I_{|a_k|}$ and $\Lambda(B(Y))_{a_k a_k} = \Sigma(Y)_{a_k a_k}$, we obtain from (148) that

$$\overline{\mu}_k \widetilde{P}_{a_k a_k}^T \widetilde{P}_{a_k a_k} + (P_{a_k}^{\uparrow})^T \mathcal{B}(H) P_{a_k}^{\uparrow} = \Sigma(Y)_{a_k a_k} + O(\|H\|^2), \quad k = 1, \ldots, r \,.$$

By (19) in Proposition 2.3, we know that $\widetilde{P}_{a_k a_k}^T \widetilde{P}_{a_k a_k} = I_{|a_k|} + O(\|H\|^2)$, $k = 1, \ldots, r$. Therefore, from (141), we obtain that

$$S(U_{a_k}^T H V_{a_k}) = \Sigma(Y)_{a_k a_k} - \overline{\mu}_k I_{|a_k|} + O(\|H\|^2) = \Sigma(Y)_{a_k a_k} - \Sigma(X)_{a_k a_k} + O(\|H\|^2), \quad k = 1, \ldots, r \,.$$

By (145), we know that

$$U_{a_k}^T H V_{a_k} = Q_k^T H_{a_k a_k} Q_k + O(\|H\|^2) \,.$$

Therefore, we have

$$Q_k^T S(H_{a_k a_k}) Q_k = \Sigma(Y)_{a_k a_k} - \Sigma(X)_{a_k a_k} + O(\|H\|^2), \quad k = 1, \ldots, r \,.$$

Hence (36) is proved.

Next, we shall show that (37) holds. Since $[\Sigma(X) \;\; 0] + H = U\, [\Sigma(Y) \;\; 0]\, V^T$, we know that

$$U_b^T ([\Sigma(X) \;\; 0] + H) V_{\bar{a}} = [\Sigma(Y)_{bb} \;\; 0] \,. \tag{149}$$

Again, from (145), we know that

$$U_b = \begin{bmatrix} O(\|H\|) \\ U_{bb} \end{bmatrix} \quad \text{and} \quad V_{\bar{a}} = \begin{bmatrix} O(\|H\|) \\ V_{\bar{a}\bar{a}} \end{bmatrix} .$$

By comparing both sides of (149), we obtain that

$$U_{bb}^T\, [\Sigma(X)_{bb} \;\; 0]\, V_{\bar{a}\bar{a}} + U_{bb}^T\, [H_{bb} \;\; H_{bc}]\, V_{\bar{a}\bar{a}} + O(\|H\|^2) = [\Sigma(Y)_{bb} \;\; 0] \,.$$

Since $\Sigma(X)_{bb} = 0$, we have

$$U_{bb}^T\, [H_{bb} \;\; H_{bc}]\, V_{\bar{a}\bar{a}} = [\Sigma(Y)_{bb} - \Sigma(X)_{bb} \;\; 0] + O(\|H\|^2) \,.$$

From (145), we know that

$$U_{bb}^T\, [H_{bb} \;\; H_{bc}]\, V_{\bar{a}\bar{a}} = Q'^T\, [H_{bb} \;\; H_{bc}]\, Q'' + O(\|H\|^2) \,.$$

Therefore,

$$Q'^T\, [H_{bb} \;\; H_{bc}]\, Q'' = [\Sigma(Y)_{bb} - \Sigma(X)_{bb} \;\; 0] + O(\|H\|^2) \,.$$

Hence (37) is proved. The proof is completed. $\square$