Set-Valued Var. Anal manuscript No.
(will be inserted by the editor)

Variational Analysis of the Ky Fan k-norm

Chao Ding

Received: 20 October 2015 / Accepted: 20 July 2016

Abstract In this paper, we will study some variational properties of the Ky Fan k-norm $\theta = \|\cdot\|_{(k)}$ of matrices, which are closed related to a class of basic nonlinear optimization problems involving the Ky Fan k-norm. In particular, for the basic nonlinear optimization problems, we will introduce the concept of nondegeneracy, strict complementarity and the critical cones associated with the generalized equations. Finally, we present the explicit formulas of the conjugate function of the parabolic second order directional derivative of θ, which will be referred to as the *sigma term* of the second order optimality conditions. The results obtained in this paper provide the necessary theoretical foundations for future work on sensitivity and stability analysis of the nonlinear optimization problems involving the Ky Fan k-norm.

Keywords Ky Fan k-norm · nondegeneracy · critical cone · second order tangent sets

Mathematics Subject Classification (2000) 65K10 · 90C25 · 90C33

1 Introduction

Let $\Re^{m\times n}$ be the vector space of all $m \times n$ real matrices equipped with the inner product $\langle Y, Z\rangle := \mathrm{Tr}(Y^T Z)$ for Y and Z in $\Re^{m\times n}$, where "Tr" denotes the trace, i.e., the sum of the diagonal entries, of a squared matrix. For simplicity, we always assume that $m \le n$. For any given positive integer $1 \le k \le m$, denote $\theta := \|\cdot\|_{(k)}$ the matrix Ky Fan k-norm, i.e., the sum of k largest singular values of matrices. In particular, $\|\cdot\|_{(1)}$ coincides with the spectral norm $\|\cdot\|_2$ of the matrices, i.e., the largest singular value of matrices; $\|\cdot\|_{(m)}$ is the nuclear norm $\|\cdot\|_*$ of matrices, i.e., the sum of singular values of matrices. It is well-known that $\vartheta(Z) := \|Z\|^*_{(k)} = \max\{\|Z\|_2, \|Z\|_*/k\}$ for $Z \in \Re^{m\times n}$ is the dual norm of $\|\cdot\|_{(k)}$ (cf. [1, Exercise IV.1.18]). Since θ is a matrix norm (convex, closed, positively homogeneous and $\theta(0) = 0$), we obtain from [33, Theorem 13.5 & 13.2] that the conjugate function $\theta^* = \delta_{\partial\,\theta(0)}$ is just the indicator function of the subdifferential $\partial\,\theta(0)$ of θ at 0.

This work is supported in part by the National Natural Science Foundation of China (Grant No. 11301515). This work was initiated while C. Ding was with Department of Mathematics, National University of Singapore during 2007 to 2012.

C. Ding
Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, P.R. China.
Tel.: +8610-82541252
E-mail: dingchao@amss.ac.cn

Moreover, it can be verified directly from the definition of dual norm that $\partial\theta(0)$ coincides with the unit ball under the dual norm ϑ, i.e., $\partial\theta(0)=\mathbb{B}_{(k)^*}:=\{S\in\Re^{m\times n}\mid\vartheta(S)\le 1\}$.

Consider the following nonlinear optimization problem involving the Ky Fan k-norm $\theta=\|\cdot\|_{(k)}$

$$\min\left\{f(x)+\theta(g(x))\mid x\in\mathcal{X}\right\}, \tag{1}$$

where $\mathcal{X}$ is a finite dimensional real vector space equipped with a scalar product $\langle\cdot,\cdot\rangle$, $f:\mathcal{X}\to\Re$ is a continuously differentiable real value function, and $g:\mathcal{X}\to\Re^{m\times n}$ is a continuously differentiable function. Since θ is convex and finite everywhere, it is well-known [34, Example 10.8] that for a locally optimal solution $\bar{x}\in\mathcal{X}$ of (1), there always exists a Lagrange multiplier $\overline{S}\in\Re^{m\times n}$, together with $\bar{x}$ satisfying the following first order optimality condition, namely the Karush-Kuhn-Tucker (KKT) condition:

$$\nabla f(\bar{x})+g'(\bar{x})^*\overline{S}=0 \quad\text{and}\quad \overline{S}\in\partial\theta(\overline{X}), \tag{2}$$

where $\overline{X}:=g(\bar{x})$, $\nabla f(\bar{x})\in\mathcal{X}$ is the gradient of f at $\bar{x}$, $g'(\bar{x})^*:\Re^{m\times n}\to\mathcal{X}$ is the adjoint of the derivative mapping $g'(\bar{x})$. Note that if $\mathcal{X}=\Re^{m\times n}$ and $g(x):=x$ is the identity mapping, then the KKT condition (2) becomes the following generalized equation:

$$0\in\nabla f(x)+\partial\theta(x).$$

Note also that if the function θ is replaced by the indicator function δ_K of a set K in a finite dimensional real vector space, then the nonlinear optimization problem (1) becomes

$$\begin{array}{ll}\min & f(x)\\ \text{s.t.} & g(x)\in K.\end{array} \tag{3}$$

During the last three decades, considerable progress has been made in the variational analysis related to the problem (3) [34,12,2,19,24]. In particular, for the general non-polyhedral set K (e.g., the second-order cone and the positively semidefinite (SDP) matrices cone), by employing the well studied properties of the variational inequality $S\in\mathcal{N}_K(x)$, some important properties of (3), such as the constraint nondegeneracy, second order optimality conditions, strong regularity, full stability and calmness, are studied recently by various researchers [2,35,25]. In order to extend those results to the optimization problems involving the Ky Fan k-norm, we need first study the variational properties of (1), especially the properties of the generalized equation $S\in\theta(X)$ and its equivalent dual problem $X\in\theta^*(S)$. Although the optimization problem (1) seems extremely simple, many fundamental and important issues such as the concept of nondegeneracy, the characterizations of critical cones and the second order optimality conditions, are not studied yet in literature. The main purpose of this paper is to build up the necessary variational foundations for the future work on the nonlinear optimization problems involving the Ky Fan k-norm.

Certainly, instead of the *basic* model (1), one can consider its various modifications, e.g., the nonlinear optimization problems involving the Ky Fan k-norm with equality and conic constraints. In particular, the following convex composite matrix optimization problems involving the Ky Fan k-norm frequently arise in various applications such as the matrix norm approximation, matrix completion, rank minimization, graph theory, machine learning, etc [15,36,37,29,4,5,6,7,40,8,3,14,22,11,17,23]:

$$\begin{array}{ll}\min & \frac{1}{2}\langle(X,Y),\mathcal{Q}(X,Y)\rangle+\langle C,(X,Y)\rangle+\theta(X)\\ \text{s.t.} & \mathcal{A}(X,Y)=b,\quad Y\in K,\end{array} \tag{4}$$

where $\mathcal{Y}$ is a finite dimensional real vector space, $\mathcal{Q}:\Re^{m\times n}\times\mathcal{Y}\to\Re^{m\times n}\times\mathcal{Y}$ is a positively semidefinite self-adjoint linear operator, $\mathcal{A}:\Re^{m\times n}\times\mathcal{Y}\to\Re^p$ is a linear operator, $C\in\Re^{m\times n}\times\mathcal{Y}$ and $b\in\Re^p$ are given data, and $K\in\mathcal{Y}$ is a closed convex cone (e.g., the positive orthant, second-order cone of vectors, positive semidefinite matrices cone). As the initial step, in this paper, we will mainly focus on the fundamental model (1), since the obtained variational results will provide the necessary theoretical foundations for the study of more complicate model, e.g., (4). More precisely, we will study the concepts of nondegeneracy and strict complementary to locally optimal solutions of (1). Also, we will define and provide the complete characterizations of the critical cones associated with the generalized equation $S \in \theta(X)$ and its dual problem $X \in \theta^*(S)$. Another important variational property studied in this paper is the conjugate function of the parabolic second order directional derivative of the Ky Fan k-norm θ, which equals to the support function of the second order tangent set of the epigraph of θ. This conjugate function is closely related to the second order optimality conditions of the problem (1). Note that the epigraph of θ is not polyhedral. In general, the conjugate function of the parabolic second order directional derivative of the Ky Fan k-norm θ will not vanish in the corresponding second order optimality conditions, and will be referred to as the *sigma term*, provides the second order information of θ. In this paper, we provide the explicit expression of this sigma term. Consequently, it becomes possible to establish the second order optimality conditions of the problem (1) and study many corresponding sensitivity properties, e.g., the second order optimality conditions and the characterization of strong regularity of the KKT solutions.

The remaining parts of this paper are organized as follows. In Section 2, we introduce some preliminary results on the differential properties of eigenvalue values and vectors of symmetric matrices and singular values and vectors of matrices. In Section 3, we study the properties of the solution of the GE $S \in \partial\theta(X)$, which arises from the KKT condition (2) and its equivalent dual form $X \in \partial\theta^*(S)$. We introduce the nondegeneracy and strict complementarity of (1) in Section 4. In Section 5, we introduce and study the critical cones associated with the GE $S \in \partial\theta(X)$ and $X \in \partial\theta^*(S)$. The second order properties of the Ky Fan k-norm θ are studied in Section 6. We conclude our paper in the final section.

Below are some common notations to be used:

- For any $Z \in \Re^{m\times n}$, we denote by Z_{ij} the (i,j)-th entry of Z.
- For any $Z \in \Re^{m\times n}$, 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 $Z \in \Re^{m\times n}$, 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 $\Re^{m\times n}$ the (i,j)-th entry of $Z := X \circ Y \in \Re^{m\times n}$ is $Z_{ij} = X_{ij}Y_{ij}$.

2 Preliminaries

In this section, we list some useful preliminary results on the eigenvalues of symmetric matrices and the singular values of matrices, which are useful for our subsequent analysis.

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 matrices. 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 $\lambda(X) := (\lambda_1(X), \lambda_2(X), \ldots, \lambda_n(X))^T \in \Re^n$ and $\Lambda(X) := \mathrm{diag}(\lambda(X))$, where for any $x \in \Re^n$, $\mathrm{diag}(x)$ denotes the diagonal matrix whose i-th diagonal entry is x_i, $i = 1,\ldots,n$. Let $P \in \mathcal{O}^n$ be such that

$$X = P\Lambda(X)P^T \,. \tag{5}$$

We denote the set of such matrices P in the eigenvalue decomposition (5) by $\mathcal{O}^n(X)$. Let $\omega_1(X) > \omega_2(X) > \ldots > \omega_r(X)$ be the distinct eigenvalues of X. Define the index sets

$$a_k := \{i \mid \lambda_i(X) = \omega_k(X),\ 1 \leq i \leq n\}, \quad k = 1,\ldots,r. \tag{6}$$

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

$$\begin{aligned}&\lambda_1(X) \geq \ldots \geq \lambda_{i-l_i(X)}(X) > \lambda_{i-l_i(X)+1}(X) = \ldots = \lambda_i(X) = \ldots = \lambda_{i+s_i(X)}(X)\\ &> \lambda_{i+s_i(X)+1}(X) \geq \ldots \geq \lambda_n(X).\end{aligned} \tag{7}$$

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 [13].

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

$$\langle Y, Z\rangle \leq \lambda(Y)^T\lambda(Z)\,. \tag{8}$$

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

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

The following proposition on the directional differentiability of the eigenvalue function $\lambda(\cdot)$ is well known. For example, see [20, Theorem 7] and [38, Proposition 1.4].

Proposition 1 *Let $X \in \mathcal{S}^n$ have the eigenvalue decomposition (5). Then, for any $\mathcal{S}^n \ni H \to 0$, we have*

$$\lambda_i(X+H) - \lambda_i(X) - \lambda_{l_i}(P_{a_k}^T H P_{a_k}) = O(\|H\|^2), \quad i \in \alpha_k,\ k = 1,\ldots,r, \tag{9}$$

where for each $i \in \{1,\ldots,n\}$, l_i is defined in (7). Hence, for any given direction $H \in \mathcal{S}^n$, the eigenvalue function $\lambda_i(\cdot)$ is directionally differentiable at X with $\lambda_i'(X;H) = \lambda_{l_i}(P_{a_k}^T H P_{a_k})$, $i \in a_k$, $k = 1,\ldots,r$.

Let $k \in \{1,\ldots,r\}$ be fixed. For the symmetric matrix $P_{a_k}^T H P_{a_k} \in \mathcal{S}^{|a_k|}$, consider the eigenvalue decomposition

$$P_{a_k}^T H P_{a_k} = R\Lambda(P_{a_k}^T H P_{a_k})R^T, \tag{10}$$

where $R \in \mathcal{O}^{|a_k|}$. Denote the distinct eigenvalues of $P_{a_k}^T H P_{a_k}$ by $\tilde{\mu}_1 > \tilde{\mu}_2 > \ldots > \tilde{\mu}_{\tilde{r}}$. Define

$$\tilde{a}_j := \{i \,|\, \lambda_i(P_{a_k}^T H P_{a_k}) = \tilde{\mu}_j, 1 \leq i \leq |a_k|\}, \quad j = 1,\ldots,\tilde{r}. \tag{11}$$

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

$$\tilde{l}_i := l_{l_i}(P_{a_k}^T H P_{a_k}) \quad \text{and} \quad \tilde{l}_i \in \tilde{a}_{\tilde{k}}, \tag{12}$$

where l_i is defined by (7).

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

$$\lim_{t\downarrow 0} \frac{\Phi(x + th + \frac{1}{2}t^2 w) - \Phi(x) - t\Phi(x;h)}{\frac{1}{2}t^2} \quad \text{exists;}$$

and the above limit is said to be the (parabolic) second order directional derivative of Φ at x along the directions h and w, denoted by $\Phi''(x;h,w)$. The following proposition [38, Proposition 2.2], provides the explicit formula of the (parabolic) second order directional derivative of the eigenvalue function.

Proposition 2 *Let $X \in \mathcal{S}^n$ have the eigenvalue decomposition (5). Then, for any given $H, W \in \mathcal{S}^n$, we have for each $k \in \{1, \ldots, r\}$*

$$\lambda_i''(X; H, W) = \lambda_{\tilde{l}_i}\left(R_{\tilde{a}_{\tilde{k}}}^T P_{a_k}^T \left[W - 2H(X - \lambda_i I_n)^\dagger H\right] P_{a_k} R_{\tilde{a}_{\tilde{k}}}\right), \quad i \in a_k, \tag{13}$$

where $Z^\dagger \in \Re^{n\times n}$ is the Moore-Penrose pseudoinverse of the square matrix $Z \in \Re^{n\times n}$.

Let $X \in \Re^{m\times n}$ be given. Without loss of generality, assume that $m \leq n$. 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. Define $\sigma(X) := (\sigma_1(X), \sigma_2(X), \ldots, \sigma_m(X))^T$ and $\Sigma(X) := \mathrm{diag}(\sigma(X))$. Let $X \in \Re^{m\times n}$ admit the following singular value decomposition (SVD):

$$X = U\left[\Sigma(X) \ \ 0\right] V^T = U\left[\Sigma(X) \ \ 0\right]\left[V_1 \ \ V_2\right]^T = U\Sigma(X)V_1^T, \tag{14}$$

where $U \in \mathcal{O}^m$ and $V = [V_1 \ \ V_2] \in \mathcal{O}^n$ with $V_1 \in \Re^{n\times m}$ and $V_2 \in \Re^{n\times(n-m)}$. The set of such matrices pair (U, V) in the SVD (14) is denoted by $\mathcal{O}^{m,n}(X)$, i.e.,

$$\mathcal{O}^{m,n}(X) := \left\{(U, V) \in \mathcal{O}^m \times \mathcal{O}^n \mid X = U\left[\Sigma(X) \ \ 0\right] V^T\right\}.$$

Define the three index sets a, b and c by

$$a := \{i \mid \sigma_i(X) > 0, \ 1 \leq i \leq m\}, \ b := \{i \mid \sigma_i(X) = 0, \ 1 \leq i \leq m\} \text{ and } c := \{m+1, \ldots, n\}. \tag{15}$$

Let $\nu_1(X) > \nu_2(X) > \ldots > \nu_r(X) > 0$ be the distinct nonzero singular values of X. Without causing any ambiguity, we also use a_k to denote the following index sets

$$a_k := \{i \mid \sigma_i(X) = \nu_k(X), \ 1 \leq i \leq m\}, \quad k = 1, \ldots, r. \tag{16}$$

For the sake of convenience, let $a_{r+1} := b$. 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

$$\begin{aligned}&\sigma_1(X) \geq \ldots \geq \sigma_{i-l_i(X)}(X) > \sigma_{i-l_i(X)+1}(X) = \ldots = \sigma_i(X) = \ldots = \sigma_{i+s_i(X)}(X)\\ &> \sigma_{i+s_i(X)+1}(X) \geq \ldots \geq \sigma_m(X).\end{aligned} \tag{17}$$

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 [27].

Lemma 2 *Let Y and Z be two matrices in $\Re^{m\times n}$. Then*

$$\langle Y, Z\rangle \leq \sigma(Y)^T\sigma(Z), \tag{18}$$

where the equality holds if Y and Z admit a simultaneous ordered singular value decomposition, i.e., there exist orthogonal matrices $U \in \mathcal{O}^m$ and $V \in \mathcal{O}^n$ such that

$$Y = U[\Sigma(Y) \ \ 0]V^T \quad \text{and} \quad Z = U[\Sigma(Z) \ \ 0]V^T.$$

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

$$S(Z) := \frac{1}{2}(Z + Z^T) \quad \text{and} \quad T(Z) := \frac{1}{2}(Z - Z^T) \quad \forall\, Z \in \Re^{p\times p}. \tag{19}$$

The following proposition on the directional derivatives of the singular value functions can be obtained directly from Proposition 1. For more details, see [21, Section 5.1] .

Proposition 3 *Let $X \in \Re^{m\times n}$ have the singular value decomposition (14). For any $\Re^{m\times n} \ni H \to 0$, we have*

$$\sigma_i(X+H) - \sigma_i(X) - \sigma_i'(X;H) = O(\|H\|^2), \quad i = 1,\ldots,m, \tag{20}$$

with

$$\sigma_i'(X;H) = \begin{cases} \lambda_{l_i}\left(S(U_{a_k}^T H V_{a_k})\right) & \text{if } i \in a_k,\ k = 1,\ldots,r, \\ \sigma_{l_i}\left(\begin{bmatrix} U_b^T H V_b & U_b^T H V_2 \end{bmatrix}\right) & \text{if } i \in b, \end{cases} \tag{21}$$

where for each $i \in \{1,\ldots,m\}$, l_i is defined in (17).

Similarly, one can derive the following explicit formulas of the (parabolic) second order directional derivatives of the singular value functions from Proposition 2, directly. For more details, see [42, Theorem 3.1].

Proposition 4 *Let $X \in \Re^{m\times n}$ have the singular value decomposition (14). Suppose that the direction $H, W \in \Re^{m\times n}$ are given.*

(i) If $\sigma_i(X) > 0$, then

$$\sigma_i''(X;H,W) = \lambda_{\tilde{l}_i}\left(R_{\tilde{\alpha}_{\tilde{k}}}^T\left(S(U_{a_k}^T W V_{a_k}) - 2\Omega_{a_k}(X,H)\right)R_{\tilde{\alpha}_{\tilde{k}}}\right),$$

where $k \in \{1,\ldots,r\}$ such that $i \in a_k$, $\Omega_{a_k}(X,H) \in \mathcal{S}^m$ is given by

$$\begin{aligned}\Omega_{a_k}(X,H) = {} & (S(U^T H V_1)_{a_k})^T(\Sigma(X) - \nu_k(X)I_m)^\dagger S(U^T H V_1)_{a_k} \\ & + (T(U^T H V_1)_{a_k})^T(-\Sigma(X) - \nu_k(X)I_m)^\dagger T(U^T H V_1)_{a_k} \\ & + \frac{1}{2\nu_k(X)}U_{a_k}^T H V_2 V_2^T H^T U_{a_k},\end{aligned} \tag{22}$$

the matrix $R \in \mathcal{O}^{|a_k|}$ satisfies $S(U_{a_k}^T H V_{a_k}) = R\Lambda(S(U_{a_k}^T H V_{a_k}))R^T$, and $\{\tilde{\alpha}_j\}_{j=1}^{\tilde{r}}$ and $\tilde{l}_i$, $\tilde{k}$ be defined by (11) and (12) respectively for $S(U_{a_k}^T H V_{a_k})$.

(ii) If $\sigma_i(X) = 0$ and $\sigma_{l_i}([U_b^T H V_b \ \ U_b^T H V_2]) > 0$, then

$$\sigma_i''(X;H,W) = \lambda_{\tilde{l}_i}(S(E_{\tilde{a}_{\tilde{k}}}^T[U_b^T Z V_b \ \ U_b^T Z V_2]F_{\tilde{a}_{\tilde{k}}})),$$

where $Z = W - 2HX^\dagger H \in \Re^{m\times n}$, $X^\dagger \in \Re^{n\times m}$ is the Moore-Penrose pseudoinverse of $X \in \Re^{m\times n}$, $E \in \mathcal{O}^{|b|}$, $F = [F_1 \ \ F_2] \in \mathcal{O}^{|b|+(n-m)}$ satisfy

$$[U_b^T H V_b \ \ U_b^T H V_2] = E[\Sigma([U_b^T H V_b \ \ U_b^T H V_2]) \ \ 0]F^T,$$

$\tilde{l}_i \in \{1,\ldots,|\tilde{a}_{\tilde{k}}|\}$ and $\tilde{k} \in \{1,\ldots,\tilde{r}\}$ such that $\tilde{l}_i = l_{l_i}(S(E_{\tilde{a}_{\tilde{k}}}^T[U_b^T Z V_b \ \ U_b^T Z V_2]F_{\tilde{a}_{\tilde{k}}}))$ and $\tilde{l}_i \in \tilde{a}_{\tilde{k}}$, $\tilde{a}_j$, $j = 1,\ldots,\tilde{r}$ are the index sets of $[U_b^T H V_b \ \ U_b^T H V_2]$ defined by

$$\tilde{a}_j := \{i \mid \sigma_i([U_b^T H V_b \ \ U_b^T H V_2]) = \tilde{\nu}_j,\ 1 \le i \le |b|\},$$

and $\tilde{\nu}_1 > \tilde{\nu}_2 > \ldots > \tilde{\nu}_{\tilde{r}}$ are the nonzero distinct singular values of $[U_b^T H V_b \ \ U_b^T H V_2]$.

(iii) If $\sigma_i(X) = 0$ and $\sigma_{l_i}([U_b^T H V_b \ \ U_b^T H V_2]) = 0$, then

$$\sigma_i''(X;H,W) = \sigma_{\tilde{l}_i}\left(E_{\tilde{b}}^T[U_b^T Z V_b \ \ U_b^T Z V_2][F_{\tilde{b}} \ \ F_2]\right),$$

where $Z = W - 2HX^\dagger H \in \Re^{m\times n}$, $\tilde{b} := \{i \mid \sigma_i([U_b^T H V_b \ \ U_b^T H V_2]) = 0,\ 1 \le i \le |b|\}$ and $\tilde{l}_i = l_{l_i}\left(E_{\tilde{b}}^T[U_b^T Z V_b \ \ U_b^T Z V_2][F_{\tilde{b}} \ \ F_2]\right)$ is defined by (17) with respect to $E_{\tilde{b}}^T[U_b^T Z V_b \ \ U_b^T Z V_2][F_{\tilde{b}} \ \ F_2]$.

3 The generalized equations

In this section, we first study some properties of the following simple generalized equation (GE)

$$0 \in -S + \partial\theta(X), \tag{23}$$

which is equivalent to the following dual form

$$0 \in -X + \partial\theta^*(S). \tag{24}$$

Since $\theta^* = \delta_{\mathbb{B}_{(k)^*}}$, it follows from [33, Theorem 23.5] that (23) and (24) are also equivalent to the following complementarity problem

$$(\overline{X}, \theta(\overline{X})) \in \mathcal{K}, \quad (\overline{S}, -1) \in \mathcal{K}^\circ \quad \text{and} \quad \big\langle (\overline{X}, \theta(\overline{X})), (\overline{S}, -1) \big\rangle = 0, \tag{25}$$

where $\mathcal{K}$ is the epigraph of $\theta = \|\cdot\|_{(k)}$, i.e.,

$$\mathcal{K} = \operatorname{epi}\theta = \Big\{ (X, t) \in \Re^{m\times n} \times \Re \mid t \geq \|X\|_{(k)} \Big\} \tag{26}$$

and $\mathcal{K}^\circ$ is the polar cone of $\mathcal{K}$ given by

$$\mathcal{K}^\circ = \bigcup_{\rho \geq 0} \rho(\partial\theta(0), -1) = -\operatorname{epi}\vartheta \quad \text{with } \vartheta = \|\cdot\|^*_{(k)}.$$

On the other hand, it is well-known [26] (see also [33, Theorem 31.5]) that $(\overline{X}, \overline{S})$ is a solution of the GE (23) (or (24)) if and only if

$$\overline{X} - \mathrm{Pr}_\theta(\overline{X} + \overline{S}) = 0 \quad \Longleftrightarrow \quad \overline{S} - \mathrm{Pr}_{\theta^*}(\overline{X} + \overline{S}) = 0,$$

where $\mathrm{Pr}_\theta : \Re^{m\times n} \to \Re^{m\times n}$ is the Moreau-Yosida proximal mapping of θ, and $\mathrm{Pr}_{\theta^*} : \Re^{m\times n} \to \Re^{m\times n}$ is the Moreau-Yosida proximal mapping of θ^*. Denote $X := \overline{X} + \overline{S}$. Let X admit the following singular value decomposition

$$X = \overline{U}\,[\Sigma(X) \;\; 0]\,\overline{V}^T. \tag{27}$$

Let $\sigma = \sigma(X)$, $\overline{\sigma} = \sigma(\overline{X})$ and $\overline{u} = \sigma(\overline{S})$ be the singular values of X, $\overline{X}$ and $\overline{S}$, respectively. Since $\|\cdot\|$ and $\|\cdot\|_{(k)}$ are unitarily invariant, we know from von Neumann's trace inequality (Lemma 2) that

$$\overline{X} = \overline{U}\,[\mathrm{Diag}\,(\overline{\sigma}) \;\; 0]\,\overline{V}^T \quad \text{and} \quad \overline{S} = \overline{U}\,[\mathrm{Diag}\,(\overline{u}) \;\; 0]\,\overline{V}^T \quad \text{with } \overline{\sigma} = g(\sigma) \text{ and } \overline{u} = \sigma - g(\sigma), \tag{28}$$

where $g : \Re^m \to \Re^m$ is the Moreau-Yosida proximal mapping of the vector k-norm (i.e., the sum of the k largest components in absolute value of any vector in $\Re^m$). The properties of the proximal mapping g have been studied recently in [41], e.g., for any given $x \in \Re^m$, the unique optimal solution $g(x) \in \Re^m$ can be computed within $O(m)$ arithmetic operations (see [41, Section 3.1] for details). The following simple observations are useful for our subsequence analysis, which can be obtained directly from the characterization of the subdifferential of $\theta = \|\cdot\|_{(k)}$ (cf. [39,28]).

Lemma 3 *$\overline{\sigma}$ and $\overline{u}$ are the singular values of the solution $(\overline{X}, \overline{S})$ of the GE (23) (or (24)) if and only if $\overline{\sigma}$ and $\overline{u}$ satisfy the following conditions.*

(i) If $\overline{\sigma}_k > 0$, then

$$\overline{u}_\alpha = e_\alpha, \quad 0 \leq \overline{u}_\beta \leq e_\beta, \quad \sum_{i\in\beta} \overline{u}_i = k - k_0 \quad \text{and} \quad \overline{u}_\gamma = 0, \tag{29}$$

where $0 \leq k_0 \leq k-1$ and $k \leq k_1 \leq m$ are two integers such that

$$\overline{\sigma}_1 \geq \ldots \geq \overline{\sigma}_{k_0} > \overline{\sigma}_{k_0+1} = \ldots = \overline{\sigma}_k = \ldots = \overline{\sigma}_{k_1} > \overline{\sigma}_{k_1+1} \geq \ldots \geq \overline{\sigma}_m \geq 0 \tag{30}$$

and

$$\alpha = \{1, \ldots, k_0\}, \quad \beta = \{k_0+1, \ldots, k_1\} \quad \text{and} \quad \gamma = \{k_1+1, \ldots, m\}. \tag{31}$$

(ii) If $\overline{\sigma}_k = 0$, then

$$\overline{u}_\alpha = e_\alpha, \quad 0 \le \overline{u}_\beta \le e_\beta \quad \text{and} \quad \sum_{i\in\beta} \overline{u}_i \le k - k_0, \tag{32}$$

where $0 \le k_0 \le k-1$ is the integer such that

$$\overline{\sigma}_1 \ge \cdots \ge \overline{\sigma}_{k_0} > \overline{\sigma}_{k_0+1} = \ldots = \overline{\sigma}_k = \ldots = \overline{\sigma}_m = 0 \tag{33}$$

and

$$\alpha = \{1, \ldots, k_0\} \quad \text{and} \quad \beta = \{k_0+1, \ldots, m\}. \tag{34}$$

For notational convenience, we use β_1, β_2 and β_3 to denote the index sets

$$\beta_1 := \{i \in \beta \,|\, \overline{u}_i = 1\}, \quad \beta_2 := \{i \in \beta \,|\, 0 < \overline{u}_i < 1\} \quad \text{and} \quad \beta_3 := \{i \in \beta \,|\, \overline{u}_i = 0\}. \tag{35}$$

For $X = \overline{X} + \overline{S}$, let a, b and c be the index sets defined by (15). We use $a_1, \ldots, a_r$ to denote the index sets defined by (16) with respect to X and $a_{r+1} = b$ for the sake of convenience. Thus, by Lemma 3, we know that if $\overline{\sigma}_k > 0$, then there exist integers $r_0 \le r_1 \in \{0, 1, \ldots, r+1\}$, $r_0 \le \widetilde{r}_0 \le r_0 + 1$ and $r_1 - 1 \le \widetilde{r}_1 \le r_1$ such that

$$\alpha = \bigcup_{l=1}^{r_0} a_l, \quad \beta_1 = \bigcup_{l=r_0+1}^{\widetilde{r}_0} a_l, \quad \beta_2 = \bigcup_{l=\widetilde{r}_0+1}^{\widetilde{r}_1} a_l, \quad \beta_3 = \bigcup_{l=\widetilde{r}_1+1}^{r_1} a_l \quad \text{and} \quad \gamma = \bigcup_{l=r_1+1}^{r+1} a_l; \tag{36}$$

if $\overline{\sigma}_k = 0$, then there exist integers $r_0 \in \{0, 1, \ldots, r+1\}$ and $r_0 \le \widetilde{r}_0 \le r_0 + 1$ such that

$$\alpha = \bigcup_{l=1}^{r_0} a_l, \quad \beta_1 = \bigcup_{l=r_0+1}^{\widetilde{r}_0} a_l, \quad \beta_2 = \bigcup_{l=\widetilde{r}_0+1}^{r} a_l \quad \text{and} \quad \beta_3 = b. \tag{37}$$

Moreover, we know that for each $l \in \{1, \ldots, r_0\}$, $\overline{\sigma}_i = \overline{\sigma}_j$ for any $i, j \in a_l$, which implies that we can use $\overline{\nu}_1 > \ldots > \overline{\nu}_{r_0} > 0$ to denote those common values. Similarly, if $\overline{\sigma}_k > 0$, we use $\overline{\mu}_{\widetilde{r}_0+1} > \ldots > \overline{\mu}_{\widetilde{r}_1} > 0$ to denote the corresponding common values of $\overline{u}$; if $\overline{\sigma}_k = 0$, we use $\overline{\mu}_{\widetilde{r}_0+1} > \ldots > \overline{\mu}_r > 0$ to denote the corresponding common values of $\overline{u}$.

4 The nondegeneracy and strict complementarity

In this section, we shall introduce the nondegeneracy and strict complementarity of the optimization problem (1). To do so, let us consider the following conic reformulation of (1):

$$\begin{aligned} &\min\ f(x) + t \\ &\text{s.t.}\ \ (g(x), t) \in \mathcal{K}, \end{aligned} \tag{38}$$

where $\mathcal{K} = \operatorname{epi}\theta$.

Let $(\bar{x}, \bar{t})$ be a feasible point of (38). Denote $\overline{X} = g(\bar{x}) \in \Re^{m\times n}$. Recall the definition [34, Definition 6.1] of the tangent cone $\mathcal{T}_{\mathcal{K}}(\overline{X}, \bar{t})$ of $\mathcal{K}$ at the given point $(\overline{X}, \bar{t}) \in \mathcal{K}$, i.e.,

$$\mathcal{T}_{\mathcal{K}}(\overline{X}, \bar{t}) = \left\{(H, \tau) \in \Re^{m\times n} \times \Re \,|\, \exists\, \rho_n \downarrow 0,\ \operatorname{dist}\left((\overline{X}, \bar{t}) + \rho_n(H, \tau), \mathcal{K}\right) = o(\rho_n)\right\}.$$

For any convex function $\phi : \Re^{m\times n} \to (-\infty, \infty)$, we know from [9, Theorem 2.4.9] that

$$\mathcal{T}_{\operatorname{epi}\phi}(Y, \phi(Y)) = \operatorname{epi}\phi'(Y; \cdot) := \left\{(H, \tau) \in \Re^{m\times n} \times \Re \,|\, \phi'(Y; H) \le \tau\right\}, \quad Y \in \Re^{m\times n}. \tag{39}$$

Therefore, for $\theta = \|\cdot\|_{(k)}$, we know from Proposition 3 that

$$\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X})) = \begin{cases} \{(H,\tau) \mid \mathrm{tr}(\overline{U}_\alpha^T H\overline{V}_\alpha) + \sum_{i=1}^{k-k_0}\lambda_i\left(S(\overline{U}_\beta^T H\overline{V}_\beta)\right) \le \tau\} & \text{if } \sigma_k(\overline{X})>0, \\ \{(H,\tau) \mid \mathrm{tr}(\overline{U}_\alpha^T H\overline{V}_\alpha) + \sum_{i=1}^{k-k_0}\sigma_i\left(\left[\overline{U}_\beta^T H\overline{V}_\beta \ \ \overline{U}_\beta^T H\overline{V}_2\right]\right) \le \tau\} & \text{if } \sigma_k(\overline{X})=0. \end{cases} \tag{40}$$

Define $G : \mathcal{X}\times\Re \to \Re^{m\times n}\times\Re$ by $G(x,t) := (g(x),t)$, $(x,t)\in\mathcal{X}\times\Re$. Robinson's CQ [30] for (38) at a given feasible point $(\bar{x},\bar{t})$ can be written as

$$G'(\bar{x},\bar{t})(\mathcal{X}\times\Re) + \mathcal{T}_{\mathcal{K}}(\overline{X},\bar{t}) = \Re^{m\times n}\times\Re. \tag{41}$$

Proposition 5 *For any $\bar{x}\in\mathcal{X}$, Robinson's CQ* (41) *for* (38) *holds at $(\bar{x},\theta(g(\bar{x})))$.*

Proof. Note that the directional derivative $\theta'(\overline{X};\cdot)$ of the Ky Fan k-norm is finite everywhere. Therefore, the results can be derived directly from (41) and (39). In fact, we only need to show that for any given $(X,t)\in\Re^{m\times n}\times\Re$, there exists $(h,\eta)\in\mathcal{X}\times\Re$ and $(H,\tau)\in\mathcal{T}_{\mathcal{K}}(\overline{X},\bar{t})$ with $\bar{t}=\theta(g(\bar{x}))$ such that

$$(g'(\bar{x})h,\eta) + (H,\tau) = (X,t).$$

Let $H = X$ and $\tau = \theta'(\overline{X};X)$. By choosing $h=0$ and $\eta = t-\tau$, we know that the above equality holds trivially. □

As we mentioned in Section 1, for a locally optimal solution $\bar{x}$ to the optimization problem (1), the corresponding Lagrange multiplier always exists. By (25), we know that there exists a Lagrange multiplier $\overline{S}\in\Re^{m\times n}$ if and only if there exists $\overline{S}\in\Re^{m\times n}$ such that $(\bar{x},\theta(g(\bar{x})),\overline{S},-1)$ is the following KKT solution of (38). On the other hand, it is well-known [43] that for a locally optimal solution of (38), the corresponding set of Lagrange multipliers is nonempty, convex, bounded and compact if and only if Robinson's CQ (41) holds. Therefore, the following proposition follows from Proposition 5, directly.

Proposition 6 *Let $\bar{x}\in\mathcal{X}$ be a locally optimal solution to the problem* (1)*. The set of Lagrange multipliers of* (1) *is a nonempty, convex, bounded and compact subset of $\Re^{m\times n}$.*

Next, let us study the concept of nondegeneracy for the optimization problem (1). For any convex function $\phi:\Re^{m\times n}\to(-\infty,\infty)$ and $Y\in\Re^{m\times n}$, the lineality space of $\mathcal{T}_{\mathrm{epi}\,\phi}(Y,\phi(Y))$, i.e., the largest linear subspace in $\mathcal{T}_{\mathrm{epi}\,\phi}(Y,\phi(Y))$, can be written as

$$\begin{aligned} &\mathrm{lin}\,(\mathcal{T}_{\mathrm{epi}\,\phi}Y,\phi(Y))) = \mathcal{T}_{\mathrm{epi}\,\phi}(Y,\phi(Y))\cap(-\mathcal{T}_{\mathrm{epi}\,\phi}(Y,\phi(Y))) \\ &= \Big\{(H,\tau)\in\Re^{m\times n}\times\Re \,|\, \phi'(Y;H)\le\tau\le-\phi'(Y;-H)\Big\} \\ &= \Big\{(H,\tau)\in\Re^{m\times n}\times\Re \,|\, \phi'(Y;H) = -\phi'(Y;-H) = \tau\Big\}. \end{aligned} \tag{42}$$

The last equation of (42) follows from [33, Theorem 23.1], directly. For the Ky Fan k-norm $\theta = \|\cdot\|_{(k)}$, define the linear subspace $\mathcal{T}^{\mathrm{lin}}(\overline{X})\subseteq\Re^{m\times n}$ by

$$\mathcal{T}^{\mathrm{lin}}(\overline{X}) := \Big\{H\in\Re^{m\times n} \,|\, \theta'(\overline{X};H) = -\theta'(\overline{X};-H)\Big\}. \tag{43}$$

If $\overline{S}\in\partial\,\theta(\overline{X})$, then, by Proposition 3, we have

$$\mathcal{T}^{\mathrm{lin}}(\overline{X}) = \begin{cases} \Big\{H\in\Re^{m\times n} \mid S(\overline{U}_\beta^T H\overline{V}_\beta) = \tau I_{|\beta|} \text{ for some } \tau\in\Re\Big\} & \text{if } \sigma_k(\overline{X})>0, \\ \Big\{H\in\Re^{m\times n} \mid \left[\overline{U}_\beta^T H\overline{V}_\beta \ \ \overline{U}_\beta^T H\overline{V}_2\right] = 0\Big\} & \text{if } \sigma_k(\overline{X})=0, \end{cases} \tag{44}$$

where $\overline{U} \in \mathcal{O}^m$ and $\overline{V} \in \mathcal{O}^n$ are eigenvectors of $X = \overline{X} + \overline{S}$, and the index set β is defined in (31) if $\sigma_k(\overline{X}) > 0$ and in (34) if $\sigma_k(\overline{X}) = 0$.

For the problem (38), the concept of Robinson's constraint nondegeneracy [31,32] can be specified as follows. The constraint nondegeneracy for (38) holds at the feasible point $(\bar{x}, \bar{t})$ if

$$G'(\bar{x}, \bar{t})(\mathcal{X} \times \Re) + \mathrm{lin}\left(\mathcal{T}_{\mathcal{K}}(\overline{X}, \bar{t})\right) = \Re^{m\times n} \times \Re, \tag{45}$$

where the lineality space $\mathrm{lin}\left(\mathcal{T}_{\mathcal{K}}(\overline{X}, \bar{t})\right)$ is given by (42) with respect to $\theta = \|\cdot\|_{(k)}$.

Proposition 7 *The constraint nondegeneracy* (45) *for* (38) *holds at* $(\bar{x}, \theta(\overline{X}))$ *if and only if*

$$g'(\bar{x})\mathcal{X} + \mathcal{T}^{\mathrm{lin}}(\overline{X}) = \Re^{m\times n}, \tag{46}$$

where $\mathcal{T}^{\mathrm{lin}}(\overline{X}) \in \Re^{m\times n}$ *is the linear subspace defined by* (43)*. Therefore, we say that the nondegeratacy for the problem* (1) *holds at* $\bar{x}$ *if* (46) *holds.*

Proof. For any given $X \in \Re^{m\times n}$, by (45), we know that there exists $h \in \mathcal{X}$, $(H, \eta) \in \mathrm{lin}\left(\mathcal{T}_{\mathcal{K}}(\overline{X}, \bar{t})\right)$ such that

$$(g'(\bar{x})h, -\eta) + (H, \eta) = (X, 0).$$

Since $H \in \mathcal{T}^{\mathrm{lin}}(\overline{X})$, we know that (46) holds.

Conversely, for any $(X, t) \in \Re^{m\times n} \times \Re$, by (46), we know that there exists $h \in \mathcal{X}$ and $H \in \mathcal{T}^{\mathrm{lin}}(\overline{X})$ such that

$$g'(\bar{x})h + H = X.$$

Denote $\tau = \theta'(\overline{X}; H)$. By taking $\eta = t - \tau$, we obtain that

$$(g'(\bar{x})h, \eta) + (H, \tau) = (X, t),$$

which implies that the constraint nondegeneracy (45) holds at $(\bar{x}, \theta(\overline{X}))$. $\square$

Let $\bar{x} \in \mathcal{X}$ be a locally optimal solution of (1). Denote $\overline{X} = g(\bar{x})$. Let β be the index set defined in (31) if $\sigma_k(\overline{X}) > 0$ and in (34) if $\sigma_k(\overline{X}) = 0$. The following definition of the strict complementarity of (1) can be regarded as a generalization of the strict complementarity for the constraint optimization problem (cf. [2, Definition 4.74]).

Definition 1 We say the strict complementarity condition holds at $\bar{x} \in \mathcal{X}$ if there exists $\overline{S} \in \mathrm{ri}\,(\partial\,\theta(\overline{X}))$ such that

$$\nabla f(\bar{x}) + g'(\bar{x})^*\overline{S} = 0. \tag{47}$$

By Lemma 3, one can derive the following proposition easily. For simplicity, we omit the detail proof here.

Proposition 8 *The strict complementarity condition holds at* $\bar{x} \in \mathcal{X}$ *if and only if there exists* $\overline{S} \in \partial\,\theta(\overline{X})$ *such that* (47) *holds and*

(i) if $\sigma_k(\overline{X}) > 0$*, then* $0 < \sigma_\beta(\overline{S}) < e_\beta$*;*
(ii) if $\sigma_k(\overline{X}) = 0$*, then* $\sigma_\beta(\overline{S}) < e_\beta$ *and* $\sum_{i\in\beta}\sigma_i(\overline{S}) < k - k_0$*.*

In the following proposition, we will show that the nondegeneracy condition (46) implies the uniqueness of the Lagrange multiplier.

Proposition 9 *Let* $\bar{x} \in \mathcal{X}$ *be a locally optimal solution of* (1)*. Denote* $\overline{X} = g(\bar{x})$*. If* $\bar{x}$ *is nondegenerate, then* $\overline{S}$ *satisfying* (2) *is unique. Conversely, if* $\overline{S}$ *satisfying* (2) *is unique and the strict complementarity condition holds at* $\bar{x}$*, then* $\bar{x}$ *is nondegenerate.*

Proof. Suppose that $\bar{x}$ is nondegenerate and let $\overline{S}$ and $\overline{S}'$ satisfy (2). Then, we know that $g'(\bar{x})^*(\overline{S}-\overline{S}')=0$, which implies that $\Delta := \overline{S}-\overline{S}' \in \left[g'(\bar{x})\mathcal{X}\right]^{\perp}$. Denote $X=\overline{X}+\overline{S}$ and $X'=\overline{X}+\overline{S}'$. Suppose that X and X' admit the SVD:

$$X=\overline{U}[\Sigma(X)\ \ 0]\overline{V}^T \quad \text{and} \quad X'=\overline{U}'[\Sigma(X')\ \ 0](\overline{V}')^T,$$

where $\overline{U},\overline{U}'\in\mathcal{O}^m$ and $\overline{V},\overline{V}'\in\mathcal{O}^n$. By (28), we know that both $(\overline{U},\overline{V})$ and $(\overline{U}',\overline{V}')$ are eigenvalue vectors of $\overline{X}$. Therefore, it follows from [10, Proposition 5] that if $\sigma_k(\overline{X})>0$, then there exist orthogonal matrices $Q_1\in\mathcal{O}^{|\alpha|}$, $Q_2\in\mathcal{O}^{|\beta|}$, $Q_3\in\mathcal{O}^{|\gamma|}$ and $\mathcal{Q}_3'\in\mathcal{O}^{|\gamma|+n-m}$ such that

$$\overline{U}'=\overline{U}\begin{bmatrix} Q_1 & 0 & 0\\ 0 & Q_2 & 0\\ 0 & 0 & Q_3\end{bmatrix} \quad \text{and} \quad \overline{V}'=\overline{V}\begin{bmatrix} Q_1 & 0 & 0\\ 0 & Q_2 & 0\\ 0 & 0 & Q_3'\end{bmatrix};$$

if $\sigma_k(\overline{X})=0$, then there exist orthogonal matrices $Q_1\in\mathcal{O}^{|\alpha|}$, $Q_2\in\mathcal{O}^{|\beta|}$ and $Q_2'\in\mathcal{O}^{|\beta|+n-m}$ such that

$$\overline{U}'=\overline{U}\begin{bmatrix} Q_1 & 0\\ 0 & Q_2\end{bmatrix} \quad \text{and} \quad \overline{V}'=\overline{V}\begin{bmatrix} Q_1 & 0\\ 0 & Q_2'\end{bmatrix}.$$

Therefore, by (28), we know from Lemma 3 that if $\sigma_k(\overline{X})>0$, then

$$\overline{U}^T\Delta\overline{V}=\begin{bmatrix} 0 & 0 & 0 & 0\\ 0 & S(\overline{U}_\beta^T\Delta\overline{V}_\beta) & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix} \quad \text{with} \quad \operatorname{tr}(S(\overline{U}_\beta^T\Delta\overline{V}_\beta))=0. \tag{48}$$

if $\sigma_k(\overline{X})=0$, then

$$\overline{U}^T\Delta\overline{V}=\begin{bmatrix} 0 & 0 & 0\\ 0 & \overline{U}_\beta^T\Delta\overline{V}_\beta & \overline{U}_\beta^T\Delta\overline{V}_2\end{bmatrix}. \tag{49}$$

Thus, we know from (44) that in both cases,

$$\langle\Delta,H\rangle=\langle\overline{U}^T\Delta\overline{V},\overline{U}^TH\overline{V}\rangle=0 \quad \forall\, H\in\mathcal{T}^{\rm lin}(\overline{X}),$$

which implies that $\Delta\in\left[\mathcal{T}^{\rm lin}(\overline{X})\right]^{\perp}$. Therefore, by (46), we know that $\Delta=0$, i.e., $\overline{S}$ satisfying (2) is unique.

Conversely, since the strict complementarity condition holds at $\bar{x}$, we know that the unique Lagrange multiplier $\overline{S}\in\partial\theta(\overline{X})$ satisfying (i) or (ii) of Proposition 8. Let $X=\overline{X}+\overline{S}$ admit the SVD (27). Suppose that the constraint nondegenerate condition (46) does not hold at $\overline{X}$, i.e., there exists $0\neq H\in\left[g'(\bar{x})\mathcal{X}\right]^{\perp}\cap\left[\mathcal{T}^{\rm lin}(\overline{X})\right]^{\perp}$. Therefore, we know that $g'(\overline{X})^*H=0$. Moreover, by (44), we know that if $\sigma_k(\overline{X})>0$, then (48) holds for H; if $\sigma_k(\overline{X})=0$, then (49) holds for H. Since $g'(\overline{X})^*H=0$, we know that for any ρ,

$$\nabla f(\bar{x})+g'(\overline{X})^*(\overline{S}+\rho H)=0.$$

Moreover, since $\overline{S}$ satisfies (i) and (ii) of Proposition 8, by (48) and (49), we know from Lemma 3 that for $\rho>0$ small enough, $\overline{S}+\rho H\in\partial\theta(\overline{X})$. This contradicts the uniqueness of $\overline{S}$. $\square$

Remark 1 Let $\overline{X}\in\partial\theta^*(\overline{S})$. For the dual norm $\vartheta=\|\cdot\|_{(k)}^*$, since $(\overline{S},-1)\in\mathcal{K}^\circ$, we have

$$\mathcal{T}_{\mathcal{K}^\circ}(\overline{S},-1)=\begin{cases}\left\{(H,\tau)\in\Re^{m\times n}\times\Re \mid \vartheta'(\overline{S};H)\le-\tau\right\} & \text{if } \vartheta(\overline{S})=1,\\ \Re^{m\times n}\times\Re & \text{if } \vartheta(\overline{S})<1.\end{cases}$$

We define the linear subspace $\mathcal{T}_\circ^{\rm lin}(\overline{S}) \subseteq \Re^{m\times n}$ by

$$\mathcal{T}_\circ^{\rm lin}(\overline{S}) := \begin{cases} \left\{H \in \Re^{m\times n} \mid \vartheta'(\overline{S};H) = -\vartheta'(\overline{S};-H) = 0\right\} & \text{if } \vartheta(\overline{S}) = 1, \\ \Re^{m\times n} & \text{if } \vartheta(\overline{S}) < 1. \end{cases} \tag{50}$$

For the case that $\vartheta(\overline{S}) = \max\{\|\overline{S}\|_2, \|\overline{S}\|_*/k\} = 1$, we know from Proposition 3 that if $\|\overline{S}\|_* < k$, then

$$\mathcal{T}_\circ^{\rm lin}(\overline{S}) = \left\{H \in \Re^{m\times n} \mid S([\overline{U}_\alpha \ \ \overline{U}_{\beta_1}]^T H [\overline{V}_\alpha \ \ \overline{V}_{\beta_1}]) = 0\right\}; \tag{51}$$

if $\|\overline{S}\|_* = k$, then

$$\mathcal{T}_\circ^{\rm lin}(\overline{S}) = \left\{H \in \Re^{m\times n} \mid S(\overline{U}_{\alpha\cup\beta_1}^T H \overline{V}_{\alpha\cup\beta_1}) = 0,\ {\rm tr}(\overline{U}_{\beta_2}^T H \overline{V}_{\beta_2}) = 0,\ \overline{U}_{\beta_3\cup\gamma}^T H \overline{V}_{\beta_3\cup\gamma\cup c} = 0\right\}, \tag{52}$$

where $\overline{U} \in \mathcal{O}^m$ and $\overline{V} \in \mathcal{O}^n$ are eigenvectors of $X = \overline{X} + \overline{S}$, the index set β is defined in (31) if $\sigma_k(\overline{X}) > 0$ and in (34) if $\sigma_k(\overline{X}) = 0$, and β_1, β_2 and β_3 are the index sets defined by (35).

5 The critical cones

From now on, let us always assume that $\overline{X} = g(\bar{x})$ and $\overline{S}$ are solutions of the GEs (23) and (24). Therefore, the critical cones associated with the GEs (23) and (24) can be defined correspondingly from the critical cones associated the complementarity problem (25).

Firstly, consider the GE (23). Denote $(X,t) = (\overline{X} + \overline{S}, \theta(\overline{X}) - 1)$. The critical cone of $\mathcal{K}$ at (X,t) associated with the complementarity problem in (25), is defined as

$$\mathcal{C}\left((X,t);\mathcal{K}\right) = \mathcal{T}_{\mathcal{K}}(\overline{X}, \theta(\overline{X})) \cap (\overline{S}, -1)^\perp. \tag{53}$$

Thus, we know from (39) that

$$(H,\tau) \in \mathcal{C}\left((X,t);\mathcal{K}\right) \Longleftrightarrow \begin{cases} H \in \mathcal{C}(X; \partial\,\theta(\overline{X})), \\ \tau = \langle \overline{S}, H\rangle, \end{cases} \tag{54}$$

where $\mathcal{C}(X;\partial\,\theta(\overline{X})) \subseteq \Re^{m\times n}$ is defined by

$$\mathcal{C}(X;\partial\,\theta(\overline{X})) := \left\{H \in \Re^{m\times n} \mid \theta'(\overline{X};H) \le \langle \overline{S}, H\rangle\right\}. \tag{55}$$

Since $\theta'(\overline{X};\cdot)$ is a positively homogeneous convex function with $\theta'(\overline{X};0) = 0$, $\mathcal{C}(X;\partial\,\theta(\overline{X}))$ is indeed a closed convex cone. We call $\mathcal{C}(X;\partial\,\theta(\overline{X}))$ the critical cone of $\partial\,\theta(\overline{X})$ at $X = \overline{X} + \overline{S}$, associated with the GE (23).

Next, we present the following proposition on the characterization of the critical cone $\mathcal{C}(X;\partial\,\theta(\overline{X}))$.

Proposition 10 *Suppose that $(\overline{X}, \overline{S}) \in \Re^{m\times n} \times \Re^{m\times n}$ is a solution of the GE (23). Let $X = \overline{X} + \overline{S}$ admit the SVD (27). Then,*

$$H \in \mathcal{C}(X;\partial\,\theta(\overline{X})) \quad \Longleftrightarrow \quad \theta'(\overline{X};H) = \langle \overline{S}, H\rangle, \tag{56}$$

which is equivalent to the following conditions:

(i) if $\sigma_k(\overline{X}) > 0$, then there exists some $\tau \in \Re$ such that

$$\lambda_{|\beta_1|}(S(\overline{U}_{\beta_1}^T H \overline{V}_{\beta_1})) \ge \tau \ge \lambda_1(S(\overline{U}_{\beta_3}^T H \overline{V}_{\beta_3}))$$

and

$$S(\overline{U}_\beta^T H \overline{V}_\beta) = \begin{bmatrix} S(\overline{U}_{\beta_1}^T H \overline{V}_{\beta_1}) & 0 & 0 \\ 0 & \tau I_{|\beta_2|} & 0 \\ 0 & 0 & S(\overline{U}_{\beta_3}^T H \overline{V}_{\beta_3}) \end{bmatrix};$$

(ii) *if* $\sigma_k(\overline{X}) = 0$ *and* $\|\overline{S}\|_* = k$*, then there exists some* $\tau \geq 0$ *such that*

$$\lambda_{|\beta_1|}(S(\overline{U}_{\beta_1}^T H\overline{V}_{\beta_1})) \geq \tau \geq \sigma_1\left(\begin{bmatrix}\overline{U}_b^T H\overline{V}_b & \overline{U}_b^T H\overline{V}_2\end{bmatrix}\right)$$

and

$$\begin{bmatrix}\overline{U}_\beta^T H\overline{V}_\beta & \overline{U}_\beta^T H\overline{V}_2\end{bmatrix} = \begin{bmatrix} S(\overline{U}_{\beta_1}^T H\overline{V}_{\beta_1}) & 0 & 0 & 0 \\ 0 & \tau I_{|\beta_2|} & 0 & 0 \\ 0 & 0 & \overline{U}_b^T H\overline{V}_b & \overline{U}_b^T H\overline{V}_2 \end{bmatrix};$$

(iii) *if* $\sigma_k(\overline{X}) = 0$ *and* $\|\overline{S}\|_* < k$*, then* $S(\overline{U}_{\beta_1}^T H\overline{V}_{\beta_1}) \succeq 0$ *and*

$$\begin{bmatrix}\overline{U}_\beta^T H\overline{V}_\beta & \overline{U}_\beta^T H\overline{V}_2\end{bmatrix} = \begin{bmatrix} S(\overline{U}_{\beta_1}^T H\overline{V}_{\beta_1}) & 0\,0\,0 \\ 0 & 0\,0\,0 \\ 0 & 0\,0\,0 \end{bmatrix}.$$

Proof. Denote $\overline{\sigma} = \sigma(\overline{X})$ and $\overline{u} = \sigma(\overline{S})$. By

$$\langle \overline{S}, H\rangle = \left\langle \overline{U}^T\overline{S}\overline{V}, \overline{U}^T H\overline{V}\right\rangle = \left\langle [\mathrm{Diag}(\overline{u}) \;\; 0], \overline{U}^T H\overline{V}\right\rangle,$$

we know from Lemma 3 that for any $H \in \Re^{m\times n}$,

$$\langle \overline{S}, H\rangle = \begin{cases} \mathrm{tr}(\overline{U}_\alpha^T H\overline{V}_\alpha) + \left\langle \mathrm{Diag}(\overline{u}_\beta), S(\overline{U}_\beta^T H\overline{V}_\beta)\right\rangle & \text{if } \overline{\sigma}_k > 0, \\ \mathrm{tr}(\overline{U}_\alpha^T H\overline{V}_\alpha) + \left\langle [\mathrm{Diag}(\overline{u}_\beta) \;\; 0], \begin{bmatrix}\overline{U}_\beta^T H\overline{V}_\beta & \overline{U}_\beta^T H\overline{V}_2\end{bmatrix}\right\rangle & \text{if } \overline{\sigma}_k = 0. \end{cases}$$

Thus, by combining with Fan's inequality (Lemma 1) and von Neumann's trace inequality (Lemma 2) , we obtain that for any $H \in \Re^{m\times n}$, if $\overline{\sigma}_k > 0$,

$$\left\langle \mathrm{Diag}(\overline{u}_\beta), S(\widetilde{H}_{\beta\beta})\right\rangle \leq \overline{u}_\beta^T \lambda(S(\widetilde{H}_{\beta\beta})) \leq \sum_{i=1}^{k-k_0} \lambda_i\left(S(\widetilde{H}_{\beta\beta})\right), \tag{57}$$

and if $\overline{\sigma}_k = 0$,

$$\left\langle [\mathrm{Diag}(\overline{u}_\beta) \;\; 0], [\widetilde{H}_{\beta\beta} \;\; \widetilde{H}_{\beta c}]\right\rangle \leq \overline{u}_\beta^T \sigma\left([\widetilde{H}_{\beta\beta} \;\; \widetilde{H}_{\beta c}]\right) \leq \sum_{i=1}^{k-k_0} \sigma_i\left([\widetilde{H}_{\beta\beta} \;\; \widetilde{H}_{\beta c}]\right), \tag{58}$$

where $\widetilde{H} = \overline{U}^T H\overline{V}$. Therefore, we know from (40) that

$$H \in \mathcal{C}(X; \partial\theta(\overline{X})) \Longleftrightarrow \theta'(\overline{X}; H) = \langle \overline{S}, H\rangle \Longleftrightarrow \text{the equalities in (57) and (58) hold.}$$

Consider the following two cases.

Case 1 $\overline{\sigma}_k > 0$. It follows from Lemma 1 that the first equality of (57) holds if and only if $\mathrm{Diag}(\overline{u}_\beta)$ and $S(\widetilde{H}_{\beta\beta})$ admit a simultaneous ordered eigenvalue decomposition, i.e., there exists $R \in \mathcal{O}^{|\beta|}$ such that

$$\mathrm{Diag}(\overline{u}_\beta) = R\mathrm{Diag}(\overline{u}_\beta)R^T \quad \text{and} \quad S(\widetilde{H}_{\beta\beta}) = R\Lambda(S(\widetilde{H}_{\beta\beta}))R^T. \tag{59}$$

Let $r_0 \leq r_1 \in \{0, 1, \ldots, r+1\}$, $r_0 \leq \widetilde{r}_0 \leq r_0 + 1$ and $r_1 - 1 \leq \widetilde{r}_1 \leq r_1$ be the integers such that (36) holds. Therefore, the orthogonal matrix $R \in \mathcal{O}^{|\beta|}$ has the following block diagonal structure:

$$R = \begin{bmatrix} R_1 & 0 & 0 \\ 0 & R_2 & 0 \\ 0 & 0 & R_3 \end{bmatrix} \quad \text{with} \quad R_2 = \begin{bmatrix} R_2^{(1)} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & R_2^{(\widetilde{r}_1 - \widetilde{r}_0)} \end{bmatrix}, \tag{60}$$

where $R_1 \in \mathcal{O}^{|\beta_1|}$, $R_2 \in \mathcal{O}^{|\beta_2|}$, $R_3 \in \mathcal{O}^{|\beta_3|}$ and $R_2^{(l)} \in \mathcal{O}^{|a_{\widetilde{r}_0+l}|}$, $l = 1, \ldots, \widetilde{r}_1 - \widetilde{r}_0$. Thus, (59) holds if and only if $S(\widetilde{H}_{\beta\beta}) \in \mathcal{S}^{|\beta|}$ has the following block diagonal structure:

$$S(\widetilde{H}_{\beta\beta}) = \begin{bmatrix} S(\widetilde{H}_{\beta_1\beta_1}) & 0 & \cdots & 0 & 0 \\ 0 & S(\widetilde{H}_{a_{\widetilde{r}_0+1}a_{\widetilde{r}_0+1}}) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & S(\widetilde{H}_{a_{\widetilde{r}_1}a_{\widetilde{r}_1}}) & 0 \\ 0 & 0 & \cdots & 0 & S(\widetilde{H}_{\beta_3\beta_3}) \end{bmatrix},$$

and the elements of $\Big(\lambda(S(\widetilde{H}_{\beta_1\beta_1})), \lambda(S(\widetilde{H}_{a_{\widetilde{r}_0+1}a_{\widetilde{r}_0+1}})), \ldots, \lambda(S(\widetilde{H}_{a_{\widetilde{r}_1}a_{\widetilde{r}_1}})), \lambda(S(\widetilde{H}_{\beta_3\beta_3}))\Big)$ are in non-increasing order and are the eigenvalues of the symmetric matrix $S(\widetilde{H}_{\beta\beta})$.

On the other hand, by (29), we know that $\overline{u}_{\beta_1} = e_{\beta_1}$, $0 < \overline{u}_{\beta_2} < e_{\beta_2}$, $\overline{u}_{\beta_3} = 0$ and $\langle e_\beta, \overline{u}_\beta \rangle = k - k_0$. Then, we can verify that the second equality of (57) holds if and only if

$$\lambda_i(S(\widetilde{H}_{\beta\beta})) = \lambda_j(S(\widetilde{H}_{\beta\beta})) \quad \forall\, i, j \in \{|\beta_1| + 1, \ldots, |\beta_1| + |\beta_2|\}. \tag{61}$$

In fact, it is clear that (61) implies the second equality of (57) holds. Conversely, without loss of generality, assume that $\beta_2 \neq \emptyset$, then $k - k_0 \in \{|\beta_1| + 1, \ldots, |\beta_1| + |\beta_2|\}$. Suppose that there exists $i \in \{|\beta_1| + 1, \ldots, |\beta_1| + |\beta_2|\}$ but $i \neq k - k_0$ such that $\lambda_i(S(\widetilde{H}_{\beta\beta})) > \lambda_{k-k_0}(S(\widetilde{H}_{\beta\beta}))$ or $\lambda_{k-k_0}(S(\widetilde{H}_{\beta\beta})) > \lambda_i(S(\widetilde{H}_{\beta\beta}))$. Then, since $0 < \overline{u}_{\beta_2} < e_{\beta_2}$ and $\langle e_\beta, \overline{u}_\beta \rangle = k - k_0$, for both cases, we always have

$$\begin{aligned} & \sum_{i=1}^{k-k_0} \lambda_i(S(\widetilde{H}_{\beta\beta})) - \overline{u}_\beta^T \lambda(S(\widetilde{H}_{\beta\beta})) \\ = & \sum_{i=1}^{k-k_0} \lambda_i(S(\widetilde{H}_{\beta\beta}))(1 - (\overline{u}_\beta)_i) - \sum_{i=k-k_0+1}^{|\beta|} (\overline{u}_\beta)_i \lambda_i(S(\widetilde{H}_{\beta\beta})) \\ > & \sum_{i=1}^{k-k_0} \lambda_{k-k_0}(S(\widetilde{H}_{\beta\beta}))(1 - (\overline{u}_\beta)_i) - \sum_{i=k-k_0+1}^{|\beta|} (\overline{u}_\beta)_i \lambda_{k-k_0}(S(\widetilde{H}_{\beta\beta})) \\ = & \lambda_{k-k_0}(S(\widetilde{H}_{\beta\beta})) \left(k - k_0 - \sum_{i=1}^{k-k_0} (\overline{u}_\beta)_i - \sum_{i=k-k_0+1}^{|\beta|} (\overline{u}_\beta)_i \right) = 0, \end{aligned}$$

which implies that the second equality of (57) does not hold, which contradicts the assumption. Therefore, we know that $H \in \mathcal{C}(X; \partial\theta(\overline{X}))$ if and only if (i) holds.

Case 2 $\overline{\sigma}_k = 0$. We know from Lemma 2 that the first equality of (58) holds if and only if $[\mathrm{Diag}(\overline{u}_\beta) \;\; 0]$ and $\big[\widetilde{H}_{\beta\beta} \;\; \widetilde{H}_{\beta c}\big]$ admit a simultaneous ordered SVD, i.e., there exist orthogonal matrices $E \in \mathcal{O}^{|\beta|}$ and $F \in \mathcal{O}^{|\beta|+n-m}$ such that

$$[\mathrm{Diag}(\overline{u}_\beta) \;\; 0] = E[\mathrm{Diag}(\overline{u}_\beta) \;\; 0]F^T \quad \text{and} \quad \big[\widetilde{H}_{\beta\beta} \;\; \widetilde{H}_{\beta c}\big] = E[\Sigma([\widetilde{H}_{\beta\beta} \;\; \widetilde{H}_{\beta c}]) \;\; 0]F^T. \tag{62}$$

Let $r_0 \in \{0, 1, \ldots, r+1\}$ and $r_0 \leq \widetilde{r}_0 \leq r_0 + 1$ be the integers such that (37) holds. Therefore, it follows from [10, Proposition 5] that there exist orthogonal matrices $Q_1 \in \mathcal{O}^{|\beta_1|}$, $Q_2 \in \mathcal{O}^{|\beta_2|}$, $Q_3 \in \mathcal{Q}^{|\beta_3|}$ and $Q_3' \in \mathcal{O}^{|\beta_3|+n-m}$ such that

$$E = \begin{bmatrix} Q_1 & 0 & 0 \\ 0 & Q_2 & 0 \\ 0 & 0 & Q_3 \end{bmatrix} \quad \text{and} \quad F = \begin{bmatrix} Q_1 & 0 & 0 \\ 0 & Q_2 & 0 \\ 0 & 0 & Q_3' \end{bmatrix} \quad \text{with} \quad Q_2 = \begin{bmatrix} Q_2^{(1)} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & Q_2^{(r-\widetilde{r}_0)} \end{bmatrix}, \tag{63}$$

where $Q_2^{(l)} \in \mathcal{O}^{|a_{\tilde{r}_0+l}|}$, $l = 1, \ldots, r - \tilde{r}_0$. Thus, (62) holds if and only if $\big[\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}\big]$ has the following block diagonal structure:

$$\big[\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}\big] = \begin{bmatrix} \widetilde{H}_{a_{r_0+1}a_{r_0+1}} & 0 & \cdots & 0 & 0 & 0 \\ 0 & \widetilde{H}_{a_{\tilde{r}_0+1}a_{\tilde{r}_0+1}} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \widetilde{H}_{a_r a_r} & 0 & 0 \\ 0 & 0 & \cdots & 0 & \widetilde{H}_{bb} & \widetilde{H}_{bc} \end{bmatrix}$$

with $\widetilde{H}_{a_l a_l} \in \mathcal{S}^{|a_l|}$, $l = r_0 + 1, \ldots, r$, and the elements of

$$h := \Big(\lambda(\widetilde{H}_{a_{r_0+1}a_{r_0+1}}), \lambda(\widetilde{H}_{a_{\tilde{r}_0+1}a_{\tilde{r}_0+1}}), \ldots, \lambda(\widetilde{H}_{a_r a_r}), \sigma([\widetilde{H}_{bb} \ \ \widetilde{H}_{bc}])\Big) \in \Re^m$$

are nonnegative and in non-increasing order and $h = \sigma\big([\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}]\big)$.

On the other hand, by (32), we know that $\overline{u}_{\beta_1} = e_{\beta_1}$, $0 < \overline{u}_{\beta_2} < e_{\beta_2}$, $\overline{u}_{\beta_3} = 0$ and $\langle e_\beta, \overline{u}_\beta \rangle \le k - k_0$. Then, we may conclude that the second equality of (58) holds if and only if

$$\begin{cases} \sigma_i\big([\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}]\big) = \sigma_j\big([\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}]\big) \quad \forall\, i, j \in \{|\beta_1| + 1, \ldots, |\beta_1| + |\beta_2|\} & \text{if } \langle e_\beta, \overline{u}_\beta \rangle = k - k_0, \\ \sigma_i\big([\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}]\big) = 0 \quad \forall\, i \in \{|\beta_1| + 1, \ldots, |\beta|\} & \text{if } \langle e_\beta, \overline{u}_\beta \rangle < k - k_0. \end{cases} \tag{64}$$

In fact, it is evident that (64) implies that the second equality of (58) holds. Conversely, consider the following two sub-cases.

Case 2.1 $\|\overline{S}\|_* = k$, i.e., $\langle e_\beta, \overline{u}_\beta \rangle = k - k_0$. Without loss of generality, assume that $\beta_2 \ne \emptyset$, which implies $k - k_0 \in \{|\beta_1| + 1, \ldots, |\beta_1| + |\beta_2|\}$. Suppose that there exists $i \in \{|\beta_1| + 1, \ldots, |\beta_1| + |\beta_2|\}$ but $i \ne k - k_0$ such that $\sigma_i\big([\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}]\big) > \sigma_{k-k_0}\big([\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}]\big)$ or $\sigma_{k-k_0}\big([\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}]\big) > \sigma_i\big([\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}]\big)$. Then, since $0 < \overline{u}_{\beta_2} < e_{\beta_2}$ and $\langle e_\beta, \overline{u}_\beta \rangle = k - k_0$, for both cases, we always have

$$\begin{aligned} & \sum_{i=1}^{k-k_0} \sigma_i\big([\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}]\big) - \overline{u}_\beta^T \sigma\big([\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}]\big) \\ = & \sum_{i=1}^{k-k_0} \sigma_i\big([\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}]\big)(1 - (\overline{u}_\beta)_i) - \sum_{i=k-k_0+1}^{|\beta|} (\overline{u}_\beta)_i \sigma_i\big([\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}]\big) \\ > & \sum_{i=1}^{k-k_0} \sigma_{k-k_0}\big([\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}]\big)(1 - (\overline{u}_\beta)_i) - \sum_{i=k-k_0+1}^{|\beta|} (\overline{u}_\beta)_i \sigma_{k-k_0}\big([\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}]\big) \\ = & \, \sigma_{k-k_0}\big([\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}]\big) \left(k - k_0 - \sum_{i=1}^{k-k_0} (\overline{u}_\beta)_i - \sum_{i=k-k_0+1}^{|\beta|} (\overline{u}_\beta)_i \right) = 0, \end{aligned}$$

which implies that the second equality of (58) does not hold, which contradicts the assumption. Therefore, we know that $H \in \mathcal{C}(X; \partial\theta(\overline{X}))$ if and only if (ii) holds.

Case 2.2 $\|\overline{S}\|_* < k$, i.e., $\langle e_\beta, \overline{u}_\beta \rangle < k - k_0$. We know that $\beta_2 \cup \beta_3 \ne \emptyset$ and $k - k_0 \in \{|\beta_1| + 1, \ldots, |\beta|\}$. Suppose that (64) does not hold. Then, we know that either there exists $i \in \{|\beta_1| + 1, \ldots, |\beta|\}$ such that $i < k - k_0$ and $\sigma_i\big([\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}]\big) > \sigma_{k-k_0}\big([\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}]\big) = 0$ or $\sigma_{k-k_0}\big([\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}]\big) > 0$. For the case that $\sigma_i\big([\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}]\big) > \sigma_{k-k_0}\big([\widetilde{H}_{\beta\beta} \ \ \widetilde{H}_{\beta c}]\big) = 0$, since

$0 < \overline{u}_{\beta_2} < e_{\beta_2}$, we have

$$
\begin{aligned}
&\sum_{i=1}^{k-k_0} \sigma_i\big([\widetilde{H}_{\beta\beta} \quad \widetilde{H}_{\beta c}]\big) - \overline{u}_\beta^T \sigma\big([\widetilde{H}_{\beta\beta} \quad \widetilde{H}_{\beta c}]\big) \\
&= \sum_{i=1}^{k-k_0} \sigma_i\big([\widetilde{H}_{\beta\beta} \quad \widetilde{H}_{\beta c}]\big)(1-(\overline{u}_\beta)_i) - \sum_{i=k-k_0+1}^{|\beta|} (\overline{u}_\beta)_i \sigma_i\big([\widetilde{H}_{\beta\beta} \quad \widetilde{H}_{\beta c}]\big) \\
&> \sum_{i=1}^{k-k_0} \sigma_{k-k_0}\big([\widetilde{H}_{\beta\beta} \quad \widetilde{H}_{\beta c}]\big)(1-(\overline{u}_\beta)_i) - \sum_{i=k-k_0+1}^{|\beta|} (\overline{u}_\beta)_i \sigma_{k-k_0}\big([\widetilde{H}_{\beta\beta} \quad \widetilde{H}_{\beta c}]\big) = 0.
\end{aligned}
$$

For the case that $\sigma_{k-k_0}\big([\widetilde{H}_{\beta\beta} \quad \widetilde{H}_{\beta c}]\big) > 0$, since $\langle e_\beta, \overline{u}_\beta \rangle < k - k_0$, we obtain that

$$
\begin{aligned}
&\sum_{i=1}^{k-k_0} \sigma_i\big([\widetilde{H}_{\beta\beta} \quad \widetilde{H}_{\beta c}]\big) - \overline{u}_\beta^T \sigma\big([\widetilde{H}_{\beta\beta} \quad \widetilde{H}_{\beta c}]\big) \\
&= \sum_{i=1}^{k-k_0} \sigma_i\big([\widetilde{H}_{\beta\beta} \quad \widetilde{H}_{\beta c}]\big)(1-(\overline{u}_\beta)_i) - \sum_{i=k-k_0+1}^{|\beta|} (\overline{u}_\beta)_i \sigma_i\big([\widetilde{H}_{\beta\beta} \quad \widetilde{H}_{\beta c}]\big) \\
&\geq \sum_{i=1}^{k-k_0} \sigma_{k-k_0}\big([\widetilde{H}_{\beta\beta} \quad \widetilde{H}_{\beta c}]\big)(1-(\overline{u}_\beta)_i) - \sum_{i=k-k_0+1}^{|\beta|} (\overline{u}_\beta)_i \sigma_{k-k_0}\big([\widetilde{H}_{\beta\beta} \quad \widetilde{H}_{\beta c}]\big) \\
&= \sigma_{k-k_0}\big([\widetilde{H}_{\beta\beta} \quad \widetilde{H}_{\beta c}]\big) \left(k - k_0 - \sum_{i=1}^{k-k_0} (\overline{u}_\beta)_i - \sum_{i=k-k_0+1}^{|\beta|} (\overline{u}_\beta)_i \right) > 0.
\end{aligned}
$$

Therefore, for both cases, we always conclude that the second equality in (58) does not hold, which contradicts the assumption. Therefore, we know that $H \in \mathcal{C}(X; \partial\theta(\overline{X}))$ if and only if (iii) holds. $\square$

For the given $\overline{S} \in \partial\theta(\overline{X})$, let $\mathrm{aff}(\mathcal{C}(X; \partial\theta(\overline{X}))$ be the affine hull of the critical cone $\mathcal{C}(X; \partial\theta(\overline{X}))$, i.e., the smallest affine space containing $\mathcal{C}(X; \partial\theta(\overline{X}))$. Note that it follows from (56) that $0 \in \mathcal{C}(X; \partial\theta(\overline{X})$. It is easy to see (cf. e.g., [33, Theorem 2.7]) that $\mathrm{aff}(\mathcal{C}(X; \partial\theta(\overline{X})) = \mathcal{C}(X; \partial\theta(\overline{X})) - \mathcal{C}(X; \partial\theta(\overline{X}))$. Therefore, by Proposition 10, one can easily derive the following proposition on the characterization of $\mathrm{aff}(\mathcal{C}(X; \partial\theta(\overline{X}))$. For simplicity, we omit the detail proof here.

Proposition 11 *Suppose that $(\overline{X}, \overline{S}) \in \Re^{m\times n} \times \Re^{m\times n}$ is a solution of the GE (23). Let $X = \overline{X} + \overline{S}$ admit the SVD (27). Then, $H \in \mathrm{aff}(\mathcal{C}(X; \partial\theta(\overline{X}))$ if and only if H satisfies the following conditions:*

(i) if $\sigma_k(\overline{X}) > 0$, then there exists some $\tau \in \Re$ such that

$$
S(\overline{U}_\beta^T H \overline{V}_\beta) = \begin{bmatrix} S(\overline{U}_{\beta_1}^T H \overline{V}_{\beta_1}) & 0 & 0 \\ 0 & \tau I_{|\beta_2|} & 0 \\ 0 & 0 & S(\overline{U}_{\beta_3}^T H \overline{V}_{\beta_3}) \end{bmatrix};
$$

(ii) if $\sigma_k(\overline{X}) = 0$ and $\|\overline{S}\|_* = k$, then there exists some $\tau \in \Re$ such that

$$
\left[\overline{U}_\beta^T H \overline{V}_\beta \quad \overline{U}_\beta^T H \overline{V}_2\right] = \begin{bmatrix} S(\overline{U}_{\beta_1}^T H \overline{V}_{\beta_1}) & 0 & 0 & 0 \\ 0 & \tau I_{|\beta_2|} & 0 & 0 \\ 0 & 0 & \overline{U}_b^T H \overline{V}_b & \overline{U}_b^T H \overline{V}_2 \end{bmatrix};
$$

(iii) if $\sigma_k(\overline{X}) = 0$ and $\|\overline{S}\|_* < k$, then

$$\begin{bmatrix}\overline{U}_\beta^T H\overline{V}_\beta & \overline{U}_\beta^T H\overline{V}_2\end{bmatrix} = \begin{bmatrix} S(\overline{U}_{\beta_1}^T H\overline{V}_{\beta_1}) & 0\,0\,0 \\ 0 & 0\,0\,0 \\ 0 & 0\,0\,0 \end{bmatrix}.$$

Next, consider the dual GE (24). The critical cone of $\mathcal{K}^\circ$ at $(X,t) = (\overline{X}+\overline{S}, \theta(\overline{X})-1) \in \Re^{m\times n}\times\Re$, associated with the complementarity problem in (25), is defined as

$$\mathcal{C}\left((X,t);\mathcal{K}^\circ\right) = \mathcal{T}_{\mathcal{K}^\circ}(\overline{S},-1)\cap(\overline{X},\theta(\overline{X}))^\perp. \tag{65}$$

Thus, we know from (39) that

$$(\widehat{H},\tau)\in\mathcal{C}\left((X,t);\mathcal{K}^\circ\right) \Longleftrightarrow \begin{cases} \widehat{H} = H - \overline{U}\begin{bmatrix}\tau I_k & 0\\ 0 & 0\end{bmatrix}\overline{V}^T, \\ H\in\mathcal{C}(X;\partial\theta^*(\overline{S})), \end{cases} \tag{66}$$

where $\mathcal{C}(X;\partial\theta^*(\overline{S}))\subseteq\Re^{m\times n}$ is defined by

$$\mathcal{C}(X;\partial\theta^*(\overline{S})) := \begin{cases}\left\{H\in\Re^{m\times n} \mid \vartheta'(\overline{S};H)\le\langle\overline{X},H\rangle = 0\right\} & \text{if } \vartheta(\overline{S}) = 1,\\ \Re^{m\times n} & \text{if } \vartheta(\overline{S})<1.\end{cases} \tag{67}$$

We call $\mathcal{C}(X;\partial\theta^*(\overline{S}))$ the critical cone of $\partial\theta^*(\overline{S}) = \mathcal{N}_{\mathcal{B}_{(k)^*}}(\overline{S})$ at $X = \overline{X}+\overline{S}$, associated with the dual GE in (24). The following characterization of the critical cone $\mathcal{C}(X;\partial\theta^*(\overline{S}))$ can be obtain similarly as that of $\mathcal{C}(X;\partial\theta(\overline{X}))$. For simplicity, we omit the detail proof here.

Proposition 12 *Suppose that $(\overline{X},\overline{S})\in\Re^{m\times n}\times\Re^{m\times n}$ is a solution of the dual GE (24). Assume that $\vartheta(\overline{S}) = 1$. Let $X = \overline{X}+\overline{S}$ admit the SVD (27). Then,*

$$H\in\mathcal{C}(X;\partial\theta^*(\overline{S})) \quad\Longleftrightarrow\quad \vartheta'(\overline{S};H) = \langle\overline{X},H\rangle = 0,$$

which is equivalent to the following conditions:

(i) if $\sigma_k(\overline{X})>0$, then $\mathrm{tr}(\overline{U}_\beta^T H\overline{V}_\beta) = 0$,

$$S(\overline{U}_{\alpha\cup\beta_1}^T H\overline{V}_{\alpha\cup\beta_1}) = \begin{bmatrix}0 & 0\\ 0 & S(\overline{U}_{\beta_1}^T H\overline{V}_{\beta_1})\end{bmatrix} \quad \text{with} \quad S(\overline{U}_{\beta_1}^T H\overline{V}_{\beta_1})\preceq 0$$

and

$$\begin{bmatrix}\overline{U}_{\beta_3\cup\gamma}^T H\overline{V}_{\beta_3\cup\gamma} & \overline{U}_{\beta_3\cup\gamma}^T H\overline{V}_2\end{bmatrix} = \begin{bmatrix}S(\overline{U}_{\beta_3}^T H\overline{V}_{\beta_3}) & 0\,0\\ 0 & 0\,0\end{bmatrix} \quad\text{with}\quad S(\overline{U}_{\beta_3}^T H\overline{V}_{\beta_3})\succeq 0;$$

(ii) if $\sigma_k(\overline{X}) = 0$ and $\|\overline{S}\|_* < k$, then

$$S(\overline{U}_{\alpha\cup\beta_1}^T H\overline{V}_{\alpha\cup\beta_1}) = \begin{bmatrix}0 & 0\\ 0 & S(\overline{U}_{\beta_1}^T H\overline{V}_{\beta_1})\end{bmatrix} \quad \text{with} \quad S(\overline{U}_{\beta_1}^T H\overline{V}_{\beta_1})\preceq 0;$$

(iii) if $\sigma_k(\overline{X}) = 0$ and $\|\overline{S}\|_* = k$, then $\mathrm{tr}(\overline{U}_{\beta_1\cup\beta_2}^T H\overline{V}_{\beta_1\cup\beta_2}) + \left\|\begin{bmatrix}\overline{U}_b^T H\overline{V}_b & \overline{U}_b^T H\overline{V}_2\end{bmatrix}\right\|_* \le 0$,

$$S(\overline{U}_{\alpha\cup\beta_1}^T H\overline{V}_{\alpha\cup\beta_1}) = \begin{bmatrix}0 & 0\\ 0 & S(\overline{U}_{\beta_1}^T H\overline{V}_{\beta_1})\end{bmatrix} \quad \text{with} \quad S(\overline{U}_{\beta_1}^T H\overline{V}_{\beta_1})\preceq 0.$$

For the given $\overline{X}\in\partial\theta^*(\overline{S})$, let $\mathrm{aff}(\mathcal{C}(X;\partial\theta^*(\overline{S}))$ be the affine hull of the critical cone $\mathcal{C}(X;\partial\theta^*(\overline{S}))$. Therefore, by Proposition 12, we obtain the following characterization of $\mathrm{aff}(\mathcal{C}(X;\partial\theta^*(\overline{S}))$.

Proposition 13 *Suppose that $(\overline{X},\overline{S})\in\Re^{m\times n}\times\Re^{m\times n}$ is a solution of the dual GE (24). Assume that $\vartheta(\overline{S})=1$. Let $X=\overline{X}+\overline{S}$ admit the SVD (27). Then, $H\in\mathrm{aff}(\mathcal{C}(X;\partial\theta^*(\overline{S}))$ if and only if H satisfies the following conditions.*

(i) If $\overline{\sigma}_k>0$, then

$$S(\overline{U}_{\alpha\cup\beta_1}^T H\overline{V}_{\alpha\cup\beta_1})=\begin{bmatrix}0 & 0\\ 0 & S(\overline{U}_{\beta_1}^T H\overline{V}_{\beta_1})\end{bmatrix},\quad \mathrm{tr}(\overline{U}_\beta^T H\overline{V}_\beta)=0. \tag{68}$$

and

$$\begin{bmatrix}\overline{U}_{\beta_3\cup\gamma}^T H\overline{V}_{\beta_3\cup\gamma} & \overline{U}_{\beta_3\cup\gamma}^T H\overline{V}_2\end{bmatrix}=\begin{bmatrix}S(\overline{U}_{\beta_3}^T H\overline{V}_{\beta_3}) & 0 & 0\\ 0 & 0 & 0\end{bmatrix}. \tag{69}$$

(ii) If $\overline{\sigma}_k=0$, then

$$S(\overline{U}_{\alpha\cup\beta_1}^T H\overline{V}_{\alpha\cup\beta_1})=\begin{bmatrix}0 & 0\\ 0 & S(\overline{U}_{\beta_1}^T H\overline{V}_{\beta_1})\end{bmatrix}. \tag{70}$$

6 The second order analysis

In this section, we shall study another important variational property of the Ky Fan k-norm $\theta=\|\cdot\|_{(k)}$, i.e., the conjugate function of the parabolic second order directional derivative of θ, which equals to the support function of the second order tangent set of the epigraph of θ. This conjugate function is closely related to the second order optimality conditions of the problem (1).

For the given $(\overline{X},\theta(\overline{X}))\in\mathcal{K}$, let $\mathcal{T}_{\mathcal{K}}^{i,2}\left((\overline{X},\theta(\overline{X}));(H,\tau)\right)$ and $\mathcal{T}_{\mathcal{K}}^{2}\left((\overline{X},\theta(\overline{X}));(H,\tau)\right)$ be the inner and outer second order tangent sets [2, Definition 3.28] to $\mathcal{K}$ at $(\overline{X},\theta(\overline{X}))\in\mathcal{K}$ along the direction $(H,\tau)\in\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))$, respectively, i.e.,

$$\mathcal{T}_{\mathcal{K}}^{i,2}\left((\overline{X},\theta(\overline{X}));(H,\tau)\right):=\liminf_{\rho\downarrow 0}\frac{\mathcal{K}-(\overline{X},\theta(\overline{X}))-\rho(H,\tau)}{\frac{1}{2}\rho^2}$$

and

$$\mathcal{T}_{\mathcal{K}}^{2}\left((\overline{X},\theta(\overline{X}));(H,\tau)\right):=\limsup_{\rho\downarrow 0}\frac{\mathcal{K}-(\overline{X},\theta(\overline{X}))-\rho(H,\tau)}{\frac{1}{2}\rho^2},$$

where "lim sup" and "lim inf" are the Painlevé-Kuratowski outer and inner limit for sets (cf. [34, Definition 4.1]).

For $\mathcal{T}_{\mathcal{K}}^2:=\mathcal{T}_{\mathcal{K}}^{i,2}\left((\overline{X},\theta(\overline{X}));(H,\tau)\right)$ or $\mathcal{T}_{\mathcal{K}}^{2}\left((\overline{X},\theta(\overline{X}));(H,\tau)\right)$, since $\mathcal{K}$ is convex, we know from [2, Proposition 3.34, (3.62) & (3.63)] that for any $(\overline{X},\theta(\overline{X}))\in\mathcal{K}$ and $(H,\tau)\in\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))$,

$$\mathcal{T}_{\mathcal{K}}^2+\mathcal{T}_{\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))}(H,\tau)\subseteq\mathcal{T}_{\mathcal{K}}^2\subseteq\mathcal{T}_{\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))}(H,\tau), \tag{71}$$

where $\mathcal{T}_{\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))}(H,\tau)$ is the tangent cone of $\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))$ at (H,τ). For any given $(H,\tau)\in\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))$, let us consider the following two cases.

Case 1. $\sum_{i=1}^k\sigma_i'(\overline{X};H)=\tau$, i.e., $(H,\tau)\in\mathrm{bd}\,\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))$. Since $\mathrm{int}\,\mathcal{K}\neq\emptyset$ and the continuous convex function $\theta=\|\cdot\|_{(k)}$ is (parabolically) second order directionally differentiable, we know from [2, Proposition 3.30] that

$$\mathcal{T}_{\mathcal{K}}^{i,2}\left((\overline{X},\theta(\overline{X}));(H,\tau)\right)=\mathcal{T}_{\mathcal{K}}^{2}\left((\overline{X},\theta(\overline{X}));(H,\tau)\right)=\mathrm{epi}\,\theta''(\overline{X};H,\cdot),$$

where $\mathrm{epi}\,\theta''(\overline{X};H,\cdot)$ is the epigraph of the (parabolic) second order directional derivative of θ at $\overline{X}$ along the direction H, which is convex and given by

$$\mathrm{epi}\,\theta''(\overline{X};H,\cdot):=\Big\{(W,\eta)\in\Re^{m\times n}\times\Re\mid\sum_{i=1}^k\sigma_i''(\overline{X};H,W)\leq\eta\Big\}. \tag{72}$$

Case 2. $\sum_{i=1}^{k}\sigma_i'(\overline{X};H) < \tau$, i.e., $(H,\tau)\in \operatorname{int}\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))$. Since $\mathcal{T}_{\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))}(H,\tau)=\Re\times\Re^{m\times n}$, we know from (71) that

$$\mathcal{T}_{\mathcal{K}}^{i,2}\left((\overline{X},\theta(\overline{X}));(H,\tau)\right)=\mathcal{T}_{\mathcal{K}}^{2}\left((\overline{X},\theta(\overline{X}));(H,\tau)\right)=\Re\times\Re^{m\times n}. \tag{73}$$

Therefore, we may denote $\mathcal{T}_{\mathcal{K}}^{2}\left((\overline{X},\theta(\overline{X}));(H,\tau)\right)$ the second order tangent set to $\mathcal{K}$ at $(\overline{X},\theta(\overline{X}))$ along the direction $(H,\tau)\in\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))$.

Next, we shall provide the explicit formula of the support function of the second order tangent set $\mathcal{T}_{\mathcal{K}}^{2}\left((\overline{X},\theta(\overline{X}));(H,\tau)\right)$. Let $(\overline{X},\theta(\overline{X}))\in\mathcal{K}$ be fixed. For any $(H,\tau)\in\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))$, denote $\mathcal{T}^2(H,\tau):=\mathcal{T}_{\mathcal{K}}^{2}\left((\overline{X},\theta(\overline{X}));(H,\tau)\right)$. Consider the support function $\delta^*_{\mathcal{T}^2(\tau,H)}(\cdot,\cdot):\Re\times\Re^{m\times n}\to(-\infty,\infty]$, i.e.,

$$\delta^*_{\mathcal{T}^2(H,\tau)}(S,\zeta)=\sup\left\{\langle S,W\rangle+\zeta\eta \mid (W,\eta)\in\mathcal{T}^2(H,\tau)\right\},\quad (S,\zeta)\in\Re^{m\times n}\times\Re.$$

Claim $\delta^*_{\mathcal{T}^2(H,\tau)}(S,\zeta)\equiv\infty$ if $(S,\zeta)\notin\left(\mathcal{T}_{\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))}(H,\tau)\right)^\circ$.

Proof. Let $(S,\zeta)\notin\left(\mathcal{T}_{\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))}(H,\tau)\right)^\circ$ be arbitrarily given. Since $\mathcal{T}_{\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))}(H,\tau)$ is nonempty, we may assume that there exists $(W^\circ,\eta^\circ)\in\mathcal{T}_{\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))}(H,\tau)$ such that

$$\langle(S,\zeta),(W^\circ,\eta^\circ)\rangle>0.$$

Fix any $(\widetilde{\eta},\widetilde{W})\in\mathcal{T}^2(H,\tau)$. By (71), we have for any $\rho>0$,

$$\rho(W^\circ,\eta^\circ)+(\widetilde{W},\widetilde{\eta})\in\mathcal{T}_{\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))}(H,\tau)+\mathcal{T}^2(H,\tau)\subseteq\mathcal{T}^2(H,\tau).$$

Therefore, we know that

$$\rho\langle(S,\zeta),(W^\circ,\eta^\circ)\rangle+\langle(S,\zeta),(\widetilde{W},\widetilde{\eta})\rangle\le\delta^*((S,\zeta)\mid\mathcal{T}^2(H,\tau)).$$

Since $\langle(S,\zeta),(W^\circ,\eta^\circ)\rangle>0$ and $\rho>0$ can be arbitrarily large, we conclude that $\delta^*_{\mathcal{T}^2(H,\tau)}(S,\zeta)\equiv\infty$ for any $(S,\zeta)\notin\left(\mathcal{T}_{\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))}(H,\tau)\right)^\circ$. □

Since $\mathcal{K}$ is a closed convex cone in $\Re^{m\times n}\times\Re$, it can be verified easily that

$$\mathcal{K}\subseteq\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))\subseteq\mathcal{T}_{\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))}(H,\tau).$$

In particular, we have $\pm(\overline{X},\theta(\overline{X}))\in\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))\subseteq\mathcal{T}_{\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))}(H,\tau)$ and $\pm(H,\tau)\in\mathcal{T}_{\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))}(H,\tau)$. Therefore, we know from the definition of the polar cone that if $(S,\zeta)\in\left(\mathcal{T}_{\mathcal{T}_{\mathcal{K}}(\overline{X},\theta(\overline{X}))}(H,\tau)\right)^\circ$, then

$$(S,\zeta)\in\mathcal{K}^\circ,\quad \left\langle(S,\zeta),(\overline{X},\theta(\overline{X}))\right\rangle=0\quad\text{and}\quad\langle(S,\zeta),(H,\tau)\rangle=0. \tag{74}$$

Hence, by Claim 6, we only need to consider the point $(S,\zeta)\in\Re\times\Re^{m\times n}$ satisfying the condition (74), since otherwise $\delta^*_{\mathcal{T}^2}(S,\zeta)\equiv\infty$. Moreover, instead of considering the general $S\in\Re^{m\times n}$, we only consider the point $\overline{S}$ such that $(\overline{X},\overline{S})\in\Re^{m\times n}\times\Re^{m\times n}$ satisfying the GE (23), i.e., $\overline{S}\in\partial\theta(\overline{X})$, which is equivalent to the complementarity problem in (25).

On the other hand, by the definition of the critical cone (53) of $\mathcal{K}$, it is evident that the given point $(\overline{S},-1)$ satisfies the condition (74) if and only if $(H,\tau)\in\mathcal{C}((X,t);\mathcal{K})$ with $(X,t)=(\overline{X},\theta(\overline{X}))+(\overline{S},-1)$. Thus, by (54) and (56), we know that $(\overline{S},-1)$ satisfies the condition (74) if and only if $H\in\mathcal{C}(X;\partial\theta(\overline{X}))$ (defined by (55)) and $\tau=\langle\overline{S},H\rangle=\sum_{i=1}^{k}\sigma_i'(\overline{X};H)$. Hence, we know from (72) that

$$\mathcal{T}^2(H):=\mathcal{T}^2(H,\tau)=\Big\{(W,\eta)\in\Re^{m\times n}\times\Re \mid \sum_{i=1}^{k}\sigma_i''(\overline{X};H,W)\le\eta\Big\}, \tag{75}$$

where for each i, the second order directional derivative $\sigma_i''(\overline{X}, H, W)$ is given by Proposition 4. Let $X = \overline{X} + \overline{S}$ admit the SVD (27). Let $a_1, \ldots, a_r$ be the index sets defined by (16) with respect to X. Denote $\overline{\sigma} = \sigma(\overline{X})$ and $\overline{u} = \sigma(\overline{S})$. Consider the following two cases.

Case 1. $\overline{\sigma}_k > 0$. Let α, β and γ be the index sets defined by (31) and β_1, β_2 and β_3 be the index sets defined by (35). Let $r_0 \leq r_1 \in \{0, 1, \ldots, r+1\}$, $r_0 \leq \widetilde{r}_0 \leq r_0 + 1$ and $r_1 - 1 \leq \widetilde{r}_1 \leq r_1$ be the integers such that (36) holds. For each $l \in \{1, \ldots, r_0\}$, since $\overline{\sigma}_i = \overline{\sigma}_{i'}$ for any $i, i' \in a_l$, we use $\overline{\nu}_l$ to denote the common value. By (54) and (56) , we know that there exists an orthogonal matrix $R \in \mathcal{O}^{|\beta|}$ such that (59) holds, i.e., $\mathrm{Diag}(\overline{u}_\beta)$ and $S(\overline{U}_\beta^T H \overline{V}_\beta)$ admit a simultaneous ordered eigenvalue decomposition. Therefore, R has the block diagonal structure (60). Hence, we know from the part (i) of Proposition 4 that $(W, \eta) \in \mathcal{T}^2(H)$ if and only if

$$\begin{aligned}\sum_{i=1}^{k} \sigma_i''(\overline{X}; H, W) &= \sum_{l=1}^{r_0} \mathrm{tr}\,(S(\overline{U}_{a_l}^T W \overline{V}_{a_l})) - 2\sum_{l=1}^{r_0} \mathrm{tr}\left(\Omega_{a_l}(\overline{X}, H)\right) \\ &\quad + \mathrm{tr}\left(R_1^T\left(S(\overline{U}_\beta^T W \overline{V}_\beta) - 2\Omega_\beta(\overline{X}, H)\right)R_1\right) \\ &\quad + \sum_{i=|\beta_1|+1}^{k-k_0} \lambda_i\left(R_2^T\left(S(\overline{U}_\beta^T W \overline{V}_\beta) - 2\Omega_\beta(\overline{X}, H)\right)R_2\right) \leq \eta, \end{aligned} \tag{76}$$

where $\Omega_{a_l}(\overline{X}, H) \in \mathcal{S}^m$, $l = 1, \ldots, r_0$ and $\Omega_\beta(\overline{X}, H) \in \mathcal{S}^m$ are given by (22) with respect to $\overline{X}$, $R_1 \in \mathcal{O}(S(\overline{U}_{\beta_1}^T H \overline{V}_{\beta_1}))$ and $R_2 \in \mathcal{O}(S(\overline{U}_{\beta_2}^T H \overline{V}_{\beta_2}))$. Meanwhile, since $\overline{u}_\alpha = e_\alpha$, we have for any $(W, \eta) \in \mathcal{T}^2(H)$,

$$\begin{aligned}-\eta + \langle \overline{S}, W\rangle &= -\eta + \left\langle \overline{U}^T \overline{S}\overline{V}, \overline{U}^T W \overline{V}\right\rangle \\ &= -\eta + \left\langle \begin{bmatrix} \mathrm{Diag}(\overline{u}_\alpha) & 0 \\ 0 & \mathrm{Diag}(\overline{u}_\beta)\end{bmatrix}, \begin{bmatrix} S(\overline{U}_\alpha^T W \overline{V}_\alpha) & 0 \\ 0 & S(\overline{U}_\beta^T W \overline{V}_\beta)\end{bmatrix}\right\rangle \\ &= -\eta + \sum_{l=1}^{r_0} \mathrm{tr}\,(S(\overline{U}_{a_l}^T W \overline{V}_{a_l}^T)) + \left\langle \mathrm{Diag}(\overline{u}_\beta), S(\overline{U}_\beta^T W \overline{V}_\beta^T)\right\rangle \\ &= \Xi(W, \eta) + 2\sum_{l=1}^{r_0} \mathrm{tr}\left(\Omega_{a_l}(\overline{X}, H)\right) + \left\langle \mathrm{Diag}(\overline{u}_\beta), 2\Omega_\beta(\overline{X}, H)\right\rangle, \end{aligned} \tag{77}$$

where

$$\begin{aligned}\Xi(W, \eta) &= -\eta + \sum_{l=1}^{r_0} \mathrm{tr}\,(S(\overline{U}_{a_l}^T W \overline{V}_{a_l}^T)) - 2\sum_{l=1}^{r_0} \mathrm{tr}\left(\Omega_{a_l}(\overline{X}, H)\right) \\ &\quad + \left\langle \mathrm{Diag}(\overline{u}_\beta), S(\overline{U}_\beta^T W \overline{V}_\beta^T) - 2\Omega_\beta(\overline{X}, H)\right\rangle. \end{aligned} \tag{78}$$

Next, we shall show that

$$\max\left\{\Xi(W, \eta) \,|\, (W, \eta) \in \mathcal{T}^2(H)\right\} = 0. \tag{79}$$

In fact, since $0 \leq \overline{u}_\beta \leq e_\beta$ and $\langle e_\beta, \overline{u}\rangle = k - k_0$, we know from Lemma 1 (Fan's inequality) that the last term of (78) satisfies

$$\begin{aligned}&\left\langle \mathrm{Diag}(\overline{u}_\beta), S(\overline{U}_\beta^T W \overline{V}_\beta^T) - 2\Omega_\beta(\overline{X}, H)\right\rangle = \left\langle \mathrm{Diag}(\overline{u}_\beta), R^T\left(S(\overline{U}_\beta^T W \overline{V}_\beta^T) - 2\Omega_\beta(\overline{X}, H)\right)R\right\rangle \\ &\leq \mathrm{tr}\left(R_1^T\left(S(\overline{U}_\beta^T W \overline{V}_\beta) - 2\Omega_\beta(\overline{X}, H)\right)R_1\right) + \left\langle \overline{u}_{\beta_2}, \lambda\left(R_2^T\left(S(\overline{U}_\beta^T W \overline{V}_\beta) - 2\Omega_\beta(\overline{X}, H)\right)R_2\right)\right\rangle \\ &\leq \mathrm{tr}\left(R_1^T\left(S(\overline{U}_\beta^T W \overline{V}_\beta) - 2\Omega_\beta(\overline{X}, H)\right)R_1\right) + \sum_{i=|\beta_1|+1}^{k-k_0} \lambda_i\left(R_2^T\left(S(\overline{U}_\beta^T W \overline{V}_\beta) - 2\Omega_\beta(\overline{X}, H)\right)R_2\right).\end{aligned}$$

Therefore, together with (76) and (78), we obtain that for any $(W,\eta)\in\mathcal{T}^2(H)$, $\varXi(W,\eta)\le 0$. Also, it is easy to check that there exists $(W^*,\eta^*)\in\mathcal{T}^2(H)$ such $\varXi(W^*,\eta^*)=0$.

By combining (77) and (79), we obtain that

$$\begin{aligned}\delta^*_{\mathcal{T}^2(H)}(\overline{S},-1)&=\sup\left\{\langle\overline{S},W\rangle-\eta\mid(W,\eta)\in\mathcal{T}^2(H)\right\}\\&=\sum_{l=1}^{r_0}\operatorname{tr}\left(2\varOmega_{a_l}(\overline{X},H)\right)+\left\langle\operatorname{Diag}(\overline{u}_\beta),2\varOmega_\beta(\overline{X},H)\right\rangle.\end{aligned}\tag{80}$$

Case 2. $\overline{\sigma}_k=0$. Let α and β be the index sets defined by (34) and β_1, β_2 and β_3 be the index sets defined by (35). Let $r_0\in\{0,1,\ldots,r+1\}$ and $r_0\le\widetilde{r}_0\le r_0+1$ be the integers such that (37) holds. For each $l\in\{1,\ldots,r_0\}$, since $\overline{\sigma}_i=\overline{\sigma}_{i'}$ for any $i,i'\in a_l$, we still use $\overline{\nu}_l$ to denote the common value. By (54) and (56), we know that there exist orthogonal matrices $E\in\mathcal{O}^{|\beta|}$ and $F\in\mathcal{O}^{|\beta|+n-m}$ such that (62) holds, i.e., $[\operatorname{Diag}(\overline{u}_\beta)\ \ 0]$ and $[\overline{U}_\beta^T H\overline{V}_\beta\ \ \overline{U}_\beta^T H\overline{V}_2]$ admit a simultaneous ordered SVD, which implies that E and F have the block diagonal structure (63). Therefore, we know from the part (ii) and (iii) of Proposition 4 that $(W,\eta)\in\mathcal{T}^2(H)$ if and only if (W,η) satisfies the following conditions: if $\sigma_{k-k_0}\left([\overline{U}_\beta^T H\overline{V}_\beta\ \ \overline{U}_\beta^T H\overline{V}_2]\right)>0$, then

$$\begin{aligned}&\sum_{i=1}^{k}\sigma_i''(\overline{X};H,W)\\&=\sum_{l=1}^{r_0}\operatorname{tr}\left(S(\overline{U}_{a_l}^T W\overline{V}_{a_l}^T)\right)-2\sum_{l=1}^{r_0}\operatorname{tr}\left(\varOmega_{a_l}(\overline{X},H)\right)\\&\quad+\operatorname{tr}\left(S\big(Q_1^T[\overline{U}_\beta^T(W-2H\,\overline{X}^\dagger\,H)\overline{V}_\beta\ \ \overline{U}_\beta^T(W-2H\,\overline{X}^\dagger\,H)\overline{V}_2]Q_1\big)\right)\\&\quad+\sum_{i=|\beta_1|+1}^{k-k_0}\lambda_i\left(S\big(Q_2^T[\overline{U}_\beta^T(W-2H\,\overline{X}^\dagger\,H)\overline{V}_\beta\ \ \overline{U}_\beta^T(W-2H\,\overline{X}^\dagger\,H)\overline{V}_2]Q_2\big)\right)\le\eta;\end{aligned}\tag{81}$$

if $\sigma_{k-k_0}\left([\overline{U}_\beta^T H\overline{V}_\beta\ \ \overline{U}_\beta^T H\overline{V}_2]\right)=0$, then

$$\begin{aligned}&\sum_{i=1}^{k}\sigma_i''(\overline{X};H,W)\\&=\sum_{l=1}^{r_0}\operatorname{tr}\left(S(\overline{U}_{a_l}^T W\overline{V}_{a_l}^T)\right)-2\sum_{l=1}^{r_0}\operatorname{tr}\left(\varOmega_{a_l}(\overline{X},H)\right)\\&\quad+\operatorname{tr}\left(S\big(Q_1^T[\overline{U}_\beta^T(W-2H\,\overline{X}^\dagger\,H)\overline{V}_\beta\ \ \overline{U}_\beta^T(W-2H\,\overline{X}^\dagger\,H)\overline{V}_2]Q_1\big)\right)\\&\quad+\sum_{i=|\beta_1|+1}^{k-k_0}\sigma_i\left({Q_2'}^T[\overline{U}_\beta^T(W-2H\,\overline{X}^\dagger\,H)\overline{V}_\beta\ \ \overline{U}_\beta^T(W-2H\,\overline{X}^\dagger\,H)\overline{V}_2]Q_2''\right)\le\eta,\end{aligned}\tag{82}$$

where $\varOmega_{a_l}(\overline{X},H)\in\mathcal{S}^m$, $l=1,\ldots,r_0$ are given by (22) with respect to $\overline{X}$, $Q_1\in\mathcal{O}^{|\beta_1|}$, $Q_2\in\mathcal{O}^{|\beta_2|}$, $Q_3\in\mathcal{O}^{|b|}$ and $Q_3'\in\mathcal{O}^{|b|+n-m}$ are given by (63), $Q_2'\in\mathcal{O}^{|\beta_2|+|\beta_3|}$ and $Q_2''\in\mathcal{O}^{|\beta_2|+|b|+n-m}$ are defined by

$$Q_2'=\begin{bmatrix}Q_2&0\\0&Q_3\end{bmatrix}\quad\text{and}\quad Q_2''=\begin{bmatrix}Q_2&0\\0&Q_3'\end{bmatrix}.$$

Meanwhile, since $\overline{u}_\alpha = e_\alpha$, we have for any $(W,\eta) \in \mathcal{T}^2(H)$,

$$
\begin{aligned}
-\eta + \langle \overline{S}, W\rangle &= -\eta + \left\langle \overline{U}^T\overline{S}\overline{V}, \overline{U}^T W\overline{V}\right\rangle \\
&= -\eta + \left\langle \begin{bmatrix} \mathrm{Diag}(\overline{u}_\alpha) & 0 & 0 \\ 0 & \mathrm{Diag}(\overline{u}_\beta) & 0 \end{bmatrix}, \begin{bmatrix} S(\overline{U}_\alpha^T W\overline{V}_\alpha) & 0 & 0 \\ 0 & \overline{U}_\beta^T W\overline{V}_\beta & \overline{U}_\beta^T W\overline{V}_2 \end{bmatrix} \right\rangle \\
&= -\eta + \sum_{l=1}^{r_0} \mathrm{tr}\,(S(\overline{U}_{a_l}^T W\overline{V}_{a_l}^T)) + \left\langle [\mathrm{Diag}(\overline{u}_\beta) \ \ 0], [\overline{U}_\beta^T W\overline{V}_\beta \ \ \overline{U}_\beta^T W\overline{V}_2]\right\rangle \\
&= \Xi(W,\eta) + 2\sum_{l=1}^{r_0} \mathrm{tr}\left(\Omega_{a_l}(\overline{X}, H)\right) \\
&\quad + \left\langle [\mathrm{Diag}(\overline{u}_\beta) \ \ 0], 2[\overline{U}_\beta^T H\,\overline{X}^\dagger\, H\overline{V}_\beta \ \ \overline{U}_\beta^T H\,\overline{X}^\dagger\, H\overline{V}_2]\right\rangle, \qquad (83)
\end{aligned}
$$

where

$$
\begin{aligned}
\Xi(W,\eta) = -\eta &+ \sum_{l=1}^{r_0} \mathrm{tr}\,(S(\overline{U}_{a_l}^T W\overline{V}_{a_l}^T)) - 2\sum_{l=1}^{r_0} \mathrm{tr}\left(\Omega_{a_l}(\overline{X}, H)\right) \\
&+ \left\langle [\mathrm{Diag}(\overline{u}_\beta) \ \ 0], [\overline{U}_\beta^T (W - 2H\,\overline{X}^\dagger\, H)\overline{V}_\beta \ \ \overline{U}_\beta^T (W - 2H\,\overline{X}^\dagger\, H)\overline{V}_2]\right\rangle. \qquad (84)
\end{aligned}
$$

Similarly, we are able to show that

$$
\max\left\{\Xi(W,\eta) \,|\, (W,\eta) \in \mathcal{T}^2(H)\right\} = 0. \qquad (85)
$$

In fact, if $\sigma_{k-k_0}\left([\overline{U}_\beta^T H\overline{V}_\beta \ \ \overline{U}_\beta^T H\overline{V}_2]\right) > 0$, then since $0 \le \overline{u}_\beta \le e_\beta$ and $\langle e_\beta, \overline{u}_\beta\rangle \le k - k_0$, we know from Lemma 1 (Fan's inequality) that the last term of (84) satisfies

$$
\begin{aligned}
&\left\langle [\mathrm{Diag}(\overline{u}_\beta) \ \ 0], [\overline{U}_\beta^T (W - 2H\,\overline{X}^\dagger\, H)\overline{V}_\beta \ \ \overline{U}_\beta^T (W - 2H\,\overline{X}^\dagger\, H)\overline{V}_2]\right\rangle \\
&= \left\langle [\mathrm{Diag}(\overline{u}_\beta) \ \ 0], E^T[\overline{U}_\beta^T (W - 2H\,\overline{X}^\dagger\, H)\overline{V}_\beta \ \ \overline{U}_\beta^T (W - 2H\,\overline{X}^\dagger\, H)\overline{V}_2]F\right\rangle \\
&\le \mathrm{tr}\,\left(S\big(Q_1^T[\overline{U}_\beta^T (W - 2H\,\overline{X}^\dagger\, H)\overline{V}_\beta \ \ \overline{U}_\beta^T (W - 2H\,\overline{X}^\dagger\, H)\overline{V}_2]Q_1\big)\right) \\
&\quad + \sum_{i=|\beta_1|+1}^{k-k_0} \lambda_i\left(S\big(Q_2^T[\overline{U}_\beta^T (W - 2H\,\overline{X}^\dagger\, H)\overline{V}_\beta \ \ \overline{U}_\beta^T (W - 2H\,\overline{X}^\dagger\, H)\overline{V}_2]Q_2\big)\right).
\end{aligned}
$$

Thus, together with (81) and (84), we obtain that $\Xi(W,\eta) \le 0$ for any $(W,\eta) \in \mathcal{T}^2(H)$. If $\sigma_{k-k_0}\left([\overline{U}_\beta^T H\overline{V}_\beta \ \ \overline{U}_\beta^T H\overline{V}_2]\right) = 0$, then by Lemma 2 (von Neumann's trace inequality), we know that

$$
\begin{aligned}
&\left\langle [\mathrm{Diag}(\overline{u}_\beta) \ \ 0], [\overline{U}_\beta^T (W - 2H\,\overline{X}^\dagger\, H)\overline{V}_\beta \ \ \overline{U}_\beta^T (W - 2H\,\overline{X}^\dagger\, H)\overline{V}_2]\right\rangle \\
&= \left\langle [\mathrm{Diag}(\overline{u}_\beta) \ \ 0], E^T[\overline{U}_\beta^T (W - 2H\,\overline{X}^\dagger\, H)\overline{V}_\beta \ \ \overline{U}_\beta^T (W - 2H\,\overline{X}^\dagger\, H)\overline{V}_2]F\right\rangle \\
&\le \mathrm{tr}\,\left(S\big(Q_1^T[\overline{U}_\beta^T (W - 2H\,\overline{X}^\dagger\, H)\overline{V}_\beta \ \ \overline{U}_\beta^T (W - 2H\,\overline{X}^\dagger\, H)\overline{V}_2]Q_1\big)\right) \\
&\quad + \sum_{i=|\beta_1|+1}^{k-k_0} \sigma_i\left({Q_2'}^T[\overline{U}_\beta^T (W - 2H\,\overline{X}^\dagger\, H)\overline{V}_\beta \ \ \overline{U}_\beta^T (W - 2H\,\overline{X}^\dagger\, H)\overline{V}_2]Q_2''\right).
\end{aligned}
$$

Together with (82) and (84), we conclude that $\Xi(W,\eta) \leq 0$ for any $(W,\eta) \in \mathcal{T}^2(H)$. Moreover, it is easy to check that in both case there exists $(W^*,\eta^*) \in \mathcal{T}^2(H)$ such that $\Xi(\eta^*, W^*) = 0$ (e.g., $W^* = 2H\overline{X}^\dagger H \in \Re^{m\times n}$ and $\eta^* = \sum_{i=1}^k \sigma_i''(\overline{X}; H, W^*)$).

By combining (83) and (85), we obtain that

$$\begin{aligned}\delta^*_{\mathcal{T}^2(H)}(\overline{S},-1) &= \sup\left\{\langle \overline{S}, W\rangle - \eta \mid (W,\eta) \in \mathcal{T}^2\right\} \\ &= 2\sum_{l=1}^{r_0} \operatorname{tr}\left(\Omega_{a_l}(\overline{X},H)\right) + \left\langle \operatorname{Diag}(\overline{u}_\beta), 2\overline{U}_\beta^T H\,\overline{X}^\dagger\, H\overline{V}_\beta\right\rangle. \end{aligned} \tag{86}$$

We summarize the above results on the support function $\delta^*_{\mathcal{T}^2(H)}$ of the second order tangent set $\mathcal{T}^2(H)$ in the following proposition.

Proposition 14 *Let $(\overline{X},\overline{S}) \in \Re^{m\times n}\times\Re^{m\times n}$ be a solution of the GE (23), i.e., $\overline{S} \in \partial\theta(\overline{X})$. Let $X = \overline{X} + \overline{S}$ admit the SVD (27). Denote $\overline{\sigma} = \sigma(\overline{X})$ and $\overline{u} = \sigma(\overline{S})$. For any $H \in \mathcal{C}(X, \partial\theta(\overline{X}))$, let $\mathcal{T}^2(H) \in \Re^{m\times n}\times\Re$ be the second order tangent set defined by (75), and $\Omega_{a_l}(\overline{X},H) \in \mathcal{S}^m$, $l = 1,\ldots,r_0$ and $\Omega_\beta(\overline{X},H) \in \mathcal{S}^m$ be the matrices given by (22) with respect to $\overline{X}$. Then, the support function of $\mathcal{T}^2(H)$ at $(\overline{S},-1)$ is given as follows.*

(i) If $\overline{\sigma}_k > 0$, then

$$\delta^*_{\mathcal{T}^2(H)}(\overline{S},-1) = \sum_{l=1}^{r_0} \operatorname{tr}\left(2\Omega_{a_l}(\overline{X},H)\right) + \left\langle \operatorname{Diag}(\overline{u}_\beta), 2\Omega_\beta(\overline{X},H)\right\rangle.$$

(ii) If $\overline{\sigma}_k = 0$, then

$$\delta^*_{\mathcal{T}^2(H)}(\overline{S},-1) = \sum_{l=1}^{r_0} \operatorname{tr}\left(2\Omega_{a_l}(\overline{X},H)\right) + \left\langle \operatorname{Diag}(\overline{u}_\beta), 2\overline{U}_\beta^T H\,\overline{X}^\dagger\, H\overline{V}_\beta\right\rangle.$$

Remark 2 By (75), we know that for the given $\overline{S} \in \partial\theta(\overline{X})$ and $H \in \mathcal{C}(X, \partial\theta(\overline{X}))$, the second order tangent set $\mathcal{T}^2(H)$ is the epigraph of the closed convex function $\psi := \theta''(\overline{X}; H, \cdot) : \Re^{m\times n} \to \Re$. Then, the support function of $\mathcal{T}^2(H)$ at $(\overline{S},-1)$ obtained in Proposition 14 equals to the conjugate function value of ψ at $\overline{S}$, i.e.,

$$\psi^*(\overline{S}) := \sup\{\langle W, \overline{S}\rangle - \psi(W) \mid W \in \Re^{m\times n}\} = \delta^*_{\mathcal{T}^2(H)}(\overline{S},-1).$$

Definition 2 For any given $\overline{X} \in \Re^{m\times n}$, define the function $\Upsilon_{\overline{X}} : \partial\theta(\overline{X}) \times \Re^{m\times n} \to \Re$ by for any $\overline{S} \in \partial\theta(\overline{X})$ and $H \in \Re^{m\times n}$, if $\overline{\sigma}_k > 0$, then

$$\Upsilon_{\overline{X}}\left(\overline{S}, H\right) := \sum_{l=1}^{r_0} \operatorname{tr}\left(2\Omega_{a_l}(\overline{X},H)\right) + \left\langle \operatorname{Diag}(\overline{u}_\beta), 2\Omega_\beta(\overline{X},H)\right\rangle,$$

if $\overline{\sigma}_k = 0$, then

$$\Upsilon_{\overline{X}}\left(\overline{S}, H\right) := \sum_{l=1}^{r_0} \operatorname{tr}\left(2\Omega_{a_l}(\overline{X},H)\right) + \left\langle \operatorname{Diag}(\overline{u}_\beta), 2\overline{U}_\beta^T H\,\overline{X}^\dagger\, H\overline{V}_\beta\right\rangle,$$

where $\overline{\sigma} = \sigma(\overline{X})$, $\overline{u} = \sigma(\overline{S})$, and $\Omega_{a_l}(\overline{X},H) \in \mathcal{S}^m$, $l = 1,\ldots,r_0$ and $\Omega_\beta(\overline{X},H) \in \mathcal{S}^m$ are given by (22) with respect to $\overline{X}$.

Similarly, for the dual GE (24), by employing the similar arguments, we are able to derive the general results on the support function values corresponding to the second order tangent sets of the polar cone $\mathcal{K}^\circ$. In particular, we are interesting in the support function value of the following the special second order tangent set $\mathcal{T}_\circ^2(H)$ at $H \in \mathcal{C}(X; \partial\theta^*(\overline{S}))$, which is defined by

$$\mathcal{T}_\circ^2(H) := \mathcal{T}_{\mathcal{K}^\circ}^2((\overline{S}, -1); (H, 0)) = \begin{cases} -\mathrm{epi}\,\vartheta''(\overline{S}; H, \cdot) & \text{if } \vartheta(\overline{S}) = 1, \\ \Re^{m\times n} \times \Re & \text{if } \vartheta(\overline{S}) < 1, \end{cases}$$

where $\vartheta = \|\cdot\|_{(k)}^*$ is the dual norm of the Ky Fan k-norm. For simplicity, we omit the detail proof here.

Proposition 15 *Let $(\overline{X}, \overline{S}) \in \Re^{m\times n} \times \Re^{m\times n}$ be a solution of the dual GE (24). Suppose that $X = \overline{X} + \overline{S}$ has the SVD (27). Denote $\overline{\sigma} = \sigma(\overline{X})$ and $\overline{u} = \sigma(\overline{S})$. For any $H \in \mathcal{C}(X; \partial\theta^*(\overline{S}))$, let $\Omega_{\alpha\cup\beta_1}(\overline{S}, H) \in \mathcal{S}^{|\alpha|+|\beta_1|}$ and $\Omega_{a_l}(\overline{S}, H) \in \mathcal{S}^{|a_l|}$, $l = \widetilde{r}_0 + 1, \ldots, \widetilde{r}_1$ be the matrices defined by (22) with respect to $\overline{S}$. Then, the support function of $\mathcal{T}_\circ^2(H)$ at $(\overline{X}, \theta(\overline{X}))$ is given as follows.*

(i) If $\overline{\sigma}_k > 0$, then

$$\begin{aligned} \delta^*_{\mathcal{T}_\circ^2(H)}(\overline{X}, \theta(\overline{X})) = & \sum_{l=1}^{r_0} \overline{\nu}_l \mathrm{tr}\left(2\left(\Omega_{\alpha\cup\beta_1}(\overline{S}, H)\right)_{a_l a_l}\right) + \overline{\sigma}_k \mathrm{tr}\left(2\left(\Omega_{\alpha\cup\beta_1}(\overline{S}, H)\right)_{\beta_1\beta_1}\right) \\ & + \overline{\sigma}_k \sum_{l=\widetilde{r}_0+1}^{\widetilde{r}_1} \mathrm{tr}\left(2\Omega_{a_l}(\overline{S}, H)\right) + \overline{\sigma}_k \mathrm{tr}\left(2\overline{U}_{\beta_3}^T H \overline{S}^\dagger H \overline{V}_{\beta_3}\right) \\ & + \left\langle \mathrm{Diag}(\overline{\sigma}_\gamma), 2\overline{U}_\gamma^T H \overline{S}^\dagger H \overline{V}_\gamma \right\rangle. \end{aligned}$$

(ii) If $\overline{\sigma}_k = 0$, then

$$\delta^*_{\mathcal{T}_\circ^2(H)}(\overline{X}, \theta(\overline{X})) = \sum_{l=1}^{r_0} \overline{\nu}_l \mathrm{tr}\left(2\left(\Omega_{\alpha\cup\beta_1}(\overline{S}, H)\right)_{a_l a_l}\right).$$

Definition 3 For any given $\overline{S} \in \Re^{m\times n}$, define the function $\Upsilon_{\overline{S}}^\circ : \partial\theta^*(\overline{S}) \times \Re^{m\times n} \to \Re$ by for any $\overline{X} \in \partial\theta^*(\overline{S})$ and $H \in \Re^{m\times n}$, if $\sigma_k(\overline{X}) > 0$, then

$$\begin{aligned} \Upsilon_{\overline{S}}^\circ\left(\overline{X}, H\right) := & \sum_{l=1}^{r_0} \overline{\nu}_l \mathrm{tr}\left(2\left(\Omega_{\alpha\cup\beta_1}(\overline{S}, H)\right)_{a_l a_l}\right) + \overline{\sigma}_k \mathrm{tr}\left(2\left(\Omega_{\alpha\cup\beta_1}(\overline{S}, H)\right)_{\beta_1\beta_1}\right) \\ & + \overline{\sigma}_k \sum_{l=\widetilde{r}_0+1}^{\widetilde{r}_1} \mathrm{tr}\left(2\Omega_{a_l}(\overline{S}, H)\right) + \overline{\sigma}_k \mathrm{tr}\left(2\overline{U}_{\beta_3}^T H \overline{S}^\dagger H \overline{V}_{\beta_3}\right) \\ & + \left\langle \mathrm{Diag}(\overline{\sigma}_\gamma), 2\overline{U}_\gamma^T H \overline{S}^\dagger H \overline{V}_\gamma \right\rangle, \end{aligned}$$

if $\sigma_k(\overline{X}) = 0$, then

$$\Upsilon_{\overline{S}}^\circ\left(\overline{X}, H\right) := \sum_{l=1}^{r_0} \overline{\nu}_l \mathrm{tr}\left(2\left(\Omega_{\alpha\cup\beta_1}(\overline{S}, H)\right)_{a_l a_l}\right),$$

where $\overline{\sigma} = \sigma(\overline{X})$, $\overline{u} = \sigma(\overline{S})$, and $\Omega_{\alpha\cup\beta_1}(\overline{S}, H) \in \mathcal{S}^{|\alpha|+|\beta_1|}$ and $\Omega_{a_l}(\overline{S}, H) \in \mathcal{S}^{|a_l|}$, $l = \widetilde{r}_0 + 1, \ldots, \widetilde{r}_1$ are given by (22) with respect to $\overline{S}$.

It seems that the functions $\Upsilon_{\overline{X}}$ and $\Upsilon_{\overline{S}}^\circ$ are quite complicate from the definitions. However, one can easily compute the values by elementary calculations. Moreover, we have the following interesting proposition on the defined functions $\Upsilon_{\overline{X}}$ and $\Upsilon_{\overline{S}}^\circ$.

Proposition 16 *Let $\overline{S} \in \partial\theta(\overline{X})$ (or equivalently $\overline{X} \in \partial\theta^*(\overline{S})$) be given. Then, for any $H \in \Re^{m\times n}$, $\Upsilon_{\overline{X}}(\overline{S}, H) \le 0$, $\Upsilon^{\circ}_{\overline{S}}\big(\overline{X}, H)\big) \le 0$. Moreover, we have*

$$\Upsilon_{\overline{X}}(\overline{S}, H) = 0 \quad \Longleftrightarrow \quad \Upsilon^{\circ}_{\overline{S}}\big(\overline{X}, H\big) = 0,$$

which is equivalent to the following conditions.

(i) If $\sigma_k(\overline{X}) > 0$, then

$$\begin{cases} \begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta_1} & \widetilde{H}_{\alpha\beta_2} \\ \widetilde{H}_{\beta_1\alpha} & \widetilde{H}_{\beta_1\beta_1} & \widetilde{H}_{\beta_1\beta_2} \\ \widetilde{H}_{\beta_2\alpha} & \widetilde{H}_{\beta_2\beta_1} & \widetilde{H}_{\beta_2\beta_2} \end{bmatrix} \in \mathcal{S}^{|\alpha|+|\beta_1|+|\beta_2|}, \\ \widetilde{H}_{\beta_1\beta_3} = (\widetilde{H}_{\beta_3\beta_1})^T, \quad \widetilde{H}_{\beta_2\beta_3} = (\widetilde{H}_{\beta_3\beta_2})^T \\ \widetilde{H}_{\alpha\beta_2} = (\widetilde{H}_{\beta_2\alpha})^T = 0, \quad \widetilde{H}_{\alpha\beta_3} = (\widetilde{H}_{\beta_3\alpha})^T = 0, \\ \widetilde{H}_{\alpha\gamma} = (\widetilde{H}_{\gamma\alpha})^T = 0, \\ \widetilde{H}_{\beta_1\gamma} = (\widetilde{H}_{\gamma\beta_1})^T = 0, \quad \widetilde{H}_{\beta_2\gamma} = (\widetilde{H}_{\gamma\beta_2})^T = 0, \\ \widetilde{H}_{\alpha c} = 0, \quad \widetilde{H}_{\beta_1 c} = 0, \quad \widetilde{H}_{\beta_2 c} = 0, \end{cases} \tag{87}$$

where $\widetilde{H} = \overline{U}^T H \overline{V}$, and the index sets α, β, γ, and β_i, $i = 1, 2, 3$ are defined by (31) and (35).

(ii) If $\sigma_k(\overline{X}) = 0$, then

$$\begin{cases} \widetilde{H}_{\alpha\alpha} \in \mathcal{S}^{|\alpha|}, \quad \widetilde{H}_{\alpha\beta_1} = (\widetilde{H}_{\beta_1\alpha})^T \\ \widetilde{H}_{\alpha\beta_2} = (\widetilde{H}_{\beta_2\alpha})^T = 0, \quad \widetilde{H}_{\alpha\beta_3} = (\widetilde{H}_{\beta_3\alpha})^T = 0, \\ \widetilde{H}_{\alpha c} = 0, \end{cases} \tag{88}$$

where $\widetilde{H} = \overline{U}^T H \overline{V}$, and the index sets α, β, and β_i, $i = 1, 2, 3$ are defined by (34) and (35).

Proof. Let $X = \overline{X} + \overline{S}$ admit the SVD (27). Denote $\overline{\sigma} = \sigma(\overline{X})$ and $\overline{u} = \sigma(\overline{S})$. Let $\widetilde{H}_1 = \overline{U}^T H \overline{V}_1$ and $\widetilde{H}_2 = \overline{U}^T H \overline{V}_2$. Consider the following two cases.

Case 1. $\overline{\sigma}_k > 0$. By (22) and the definition of the pseudoinverse, we obtain that

$$\begin{aligned} \mathrm{tr}\left(2\Omega_{a_l}(\overline{X}, H)\right) &= \sum_{\substack{l'=1 \\ l' \neq l}}^{r+1} \frac{2}{\bar{\nu}_{l'} - \bar{\nu}_l} \|S(\widetilde{H}_1)_{a_l a_{l'}}\|^2 + \sum_{l'=1}^{r+1} \frac{2}{-\bar{\nu}_{l'} - \bar{\nu}_l} \|T(\widetilde{H}_1)_{a_l a_{l'}}\|^2 \\ &\quad + \frac{1}{-\bar{\nu}_l} \|(\widetilde{H}_2)_{a_l}\|^2, \quad l = 1, \ldots, r_0 \end{aligned}$$

and

$$\begin{aligned} \bar{\mu}_l \mathrm{tr}\left(2\Omega_{a_l}(\overline{X}, H)\right) &= \sum_{l'=1}^{r_0} \frac{2\bar{\mu}_l}{\bar{\nu}_{l'} - \overline{\sigma}_k} \|S(\widetilde{H}_1)_{a_l a_{l'}}\|^2 + \sum_{l'=r_1+1}^{r+1} \frac{2\bar{\mu}_l}{\bar{\nu}_{l'} - \overline{\sigma}_k} \|S(\widetilde{H}_1)_{a_l a_{l'}}\|^2 \\ &\quad + \sum_{l'=1}^{r+1} \frac{2\bar{\mu}_l}{-\bar{\nu}_{l'} - \overline{\sigma}_k} \|T(\widetilde{H}_1)_{a_l a_{l'}}\|^2 + \frac{\bar{\mu}_l}{-\overline{\sigma}_k} \|(\widetilde{H}_2)_{a_l}\|^2, \quad l = r_0 + 1, \ldots, \widetilde{r}_1. \end{aligned}$$

Thus, since $\overline{u}_i = 0$ if $i \in \beta_3$, we have

$$\left\langle \mathrm{Diag}(\overline{u}_\beta), 2\Omega_\beta(\overline{X}, H) \right\rangle = \sum_{l=r_0+1}^{\widetilde{r}_1} \bar{\mu}_l \mathrm{tr}\left(2\Omega_{a_l}(\overline{X}, H)\right).$$

Therefore, we obtain the following explicit formula of $\Upsilon_{\overline{X}}(\overline{S},H)$:

$$\begin{aligned}\Upsilon_{\overline{X}}(\overline{S},H) = &\sum_{l=1}^{r_0}\sum_{l'=\widetilde{r}_0+1}^{\widetilde{r}_1}\frac{2(1-\bar{\mu}_{l'})}{\overline{\sigma}_k-\bar{\nu}_l}\|S(\widetilde{H}_1)_{a_la_{l'}}\|^2 \\ &+\sum_{l=1}^{r_0}\sum_{l'=\widetilde{r}_1+1}^{r+1}\frac{2}{\bar{\nu}_{l'}-\bar{\nu}_l}\|S(\widetilde{H}_1)_{a_la_{l'}}\|^2+\sum_{l=\widetilde{r}_0+1}^{\widetilde{r}_1}\sum_{l'=r_1+1}^{r+1}\frac{2\bar{\mu}_l}{\bar{\nu}_{l'}-\overline{\sigma}_k}\|S(\widetilde{H}_1)_{a_la_{l'}}\|^2 \\ &+\sum_{l=1}^{r_0}\sum_{l'=1}^{r+1}\frac{2}{-\bar{\nu}_{l'}-\bar{\nu}_l}\|T(\widetilde{H}_1)_{a_la_{l'}}\|^2+\sum_{l=r_0+1}^{\widetilde{r}_1}\sum_{l'=1}^{r+1}\frac{2\bar{\mu}_l}{-\bar{\nu}_{l'}-\overline{\sigma}_k}\|T(\widetilde{H}_1)_{a_la_{l'}}\|^2 \\ &+\sum_{l=1}^{r_0}\frac{-1}{\bar{\nu}_l}\|(\widetilde{H}_2)_{a_l}\|^2+\sum_{l=r_0+1}^{\widetilde{r}_1}\frac{-\bar{\mu}_l}{\overline{\sigma}_k}\|(\widetilde{H}_2)_{a_l}\|^2. \end{aligned} \tag{89}$$

Since

$$\begin{cases}\overline{\sigma}_k<\bar{\nu}_l,\ l=1,\ldots,r_0,\\ \bar{\nu}_l<\overline{\sigma}_k,\ l=r_1+1,\ldots,r+1,\\ \bar{\mu}_l<1,\quad l=\widetilde{r}_0+1,\ldots,\widetilde{r}_1,\\ \bar{\nu}_{l'}<\bar{\nu}_l,\ l=1,\ldots,r_0,\ l'=\widetilde{r}_1+1,\ldots,r+1,\\ \bar{\nu}_l>0,\quad l=1,\ldots,\widetilde{r}_1,\end{cases} \tag{90}$$

it is easy to see that all the coefficients of the quadric terms of (89) are negative, which implies that $\Upsilon_{\overline{X}}(\overline{S},H)\le 0$ and $\Upsilon_{\overline{X}}(\overline{S},H)=0$ if and only if $H\in\Re^{m\times n}$ satisfies the conditions (87).

Meanwhile, by (22) and the pseudoinverse, we obtain that

$$\begin{aligned}\bar{\nu}_l\mathrm{tr}\left(2\left(\Omega_{\alpha\cup\beta_1}(\overline{S},H)\right)_{a_la_l}\right) = &\sum_{l'=\widetilde{r}_0+1}^{\widetilde{r}_1}\frac{2\bar{\nu}_l}{\bar{\mu}_{l'}-1}\|S(\widetilde{H}_1)_{a_la_{l'}}\|^2+\sum_{l'=\widetilde{r}_1+1}^{r+1}\frac{2\bar{\nu}_l}{-1}\|S(\widetilde{H}_1)_{a_la_{l'}}\|^2 \\ &+\sum_{l'=1}^{r+1}\frac{2\bar{\nu}_l}{-\bar{\mu}_{l'}-1}\|T(\widetilde{H}_1)_{a_la_{l'}}\|^2+\frac{\bar{\nu}_l}{-1}\|(\widetilde{H}_2)_{a_l}\|^2,\quad l=1,\ldots,\widetilde{r}_0\end{aligned}$$

and

$$\begin{aligned}&\overline{\sigma}_k\mathrm{tr}\left(2\Omega_{a_l}(\overline{S},H)\right) \\ &=\sum_{l'=1}^{\widetilde{r}_0}\frac{2\overline{\sigma}_k}{1-\bar{\mu}_l}\|S(\widetilde{H}_1)_{a_la_{l'}}\|^2+\sum_{\substack{l'=\widetilde{r}_0+1\\ l'\neq l}}^{\widetilde{r}_1}\frac{2\overline{\sigma}_k}{\bar{\mu}_{l'}-\bar{\mu}_l}\|S(\widetilde{H}_1)_{a_la_{l'}}\|^2+\sum_{l'=\widetilde{r}_1+1}^{r+1}\frac{2\overline{\sigma}_k}{-\bar{\mu}_l}\|S(\widetilde{H}_1)_{a_la_{l'}}\|^2 \\ &\quad+\sum_{l'=1}^{r+1}\frac{2\overline{\sigma}_k}{-\bar{\mu}_{l'}-\bar{\mu}_l}\|T(\widetilde{H}_1)_{a_la_{l'}}\|^2+\frac{\overline{\sigma}_k}{-\bar{\mu}_l}\|(\widetilde{H}_2)_{a_l}\|^2,\quad l=\widetilde{r}_0+1,\ldots,\widetilde{r}_1.\end{aligned}$$

Note that for any $A,B\in\Re^{p\times q}$, $\mathrm{tr}(A^TB)=\|\frac{A+B}{2}\|^2-\|\frac{A-B}{2}\|^2$. Thus, we have for $l=\widetilde{r}_1+1,\ldots,r_1$,

$$\begin{aligned}&\overline{\sigma}_k\mathrm{tr}\left(2\overline{U}_{a_l}^TH\overline{S}^\dagger H\overline{V}_{a_l}\right) \\ &=\sum_{l'=1}^{\widetilde{r}_0}\frac{2\overline{\sigma}_k}{1}\left(\|S(\widetilde{H}_1)_{a_la_{l'}}\|^2-\|T(\widetilde{H}_1)_{a_la_{l'}}\|^2\right)+\sum_{l'=\widetilde{r}_0+1}^{\widetilde{r}_1}\frac{2\overline{\sigma}_k}{\bar{\mu}_{l'}}\left(\|S(\widetilde{H}_1)_{a_la_{l'}}\|^2-\|T(\widetilde{H}_1)_{a_la_{l'}}\|^2\right),\end{aligned}$$

and for $l = r_1+1,\ldots,r$,

$$\begin{aligned}
&\bar{\nu}_l \mathrm{tr}\left(2\overline{U}_{a_l}^T H\overline{S}^\dagger H\overline{V}_{a_l}\right)\\
&= \sum_{l'=1}^{\widetilde{r}_0} \frac{2\bar{\nu}_l}{1}\left(\|S(\widetilde{H}_1)_{a_l a_{l'}}\|^2 - \|T(\widetilde{H}_1)_{a_l a_{l'}}\|^2\right) + \sum_{l'=\widetilde{r}_0+1}^{\widetilde{r}_1} \frac{2\bar{\nu}_l}{\bar{\mu}_{l'}}\left(\|S(\widetilde{H}_1)_{a_l a_{l'}}\|^2 - \|T(\widetilde{H}_1)_{a_l a_{l'}}\|^2\right).
\end{aligned}$$

By noting that $\overline{\sigma}_i = 0$ if $i \in b$, we have

$$\begin{aligned}
&\overline{\sigma}_k \mathrm{tr}\left(2\overline{U}_{\beta_3}^T H\overline{S}^\dagger H\overline{V}_{\beta_3}\right) + \left\langle \mathrm{Diag}(\overline{\sigma}_\gamma), 2\overline{U}_\gamma^T H\overline{S}^\dagger H\overline{V}_\gamma \right\rangle\\
&= \sum_{l=\widetilde{r}_1+1}^{r_1} \overline{\sigma}_k \mathrm{tr}\left(2\overline{U}_{a_l}^T H\overline{S}^\dagger H\overline{V}_{a_l}\right) + \sum_{l=r_1+1}^{r} \bar{\nu}_l \mathrm{tr}\left(2\overline{U}_{a_l}^T H\overline{S}^\dagger H\overline{V}_{a_l}\right).
\end{aligned}$$

Therefore, we obtain the following explicit formula of $\Upsilon_{\overline{S}}^\circ\left(\overline{X}, H\right)$:

$$\begin{aligned}
&\Upsilon_{\overline{S}}^\circ\left(\overline{X}, H\right)\\
&= \sum_{l=1}^{r_0}\sum_{l'=\widetilde{r}_0+1}^{\widetilde{r}_1}\left(\frac{2\bar{\nu}_l}{\bar{\mu}_{l'}-1} + \frac{2\overline{\sigma}_k}{1-\bar{\mu}_{l'}}\right)\|S(\widetilde{H}_1)_{a_l a_{l'}}\|^2\\
&\quad + \sum_{l=1}^{r_0}\sum_{l'=\widetilde{r}_1+1}^{r_1} 2(\overline{\sigma}_k - \bar{\nu}_l)\|S(\widetilde{H}_1)_{a_l a_{l'}}\|^2 + \sum_{l=1}^{r_0}\sum_{l'=r_1+1}^{r+1}\frac{2\bar{\nu}_l}{-1}\|S(\widetilde{H}_1)_{a_l a_{l'}}\|^2\\
&\quad + \sum_{l=r_0+1}^{\widetilde{r}_1}\sum_{l'=r_1+1}^{r+1}\frac{2\overline{\sigma}_k}{-\bar{\mu}_l}\|S(\widetilde{H}_1)_{a_l a_{l'}}\|^2 + \sum_{l=1}^{\widetilde{r}_0}\sum_{l'=1}^{\widetilde{r}_1}\frac{2\bar{\nu}_l}{-\bar{\mu}_{l'}-1}\|T(\widetilde{H}_1)_{a_l a_{l'}}\|^2\\
&\quad + \sum_{l=1}^{\widetilde{r}_0}\sum_{l'=\widetilde{r}_1+1}^{r_1} 2\left(\frac{-\bar{\nu}_l}{\bar{\mu}_{l'}+1} - \overline{\sigma}_k\right)\|T(\widetilde{H}_1)_{a_l a_{l'}}\|^2 + \sum_{l=1}^{\widetilde{r}_0}\sum_{l'=r_1+1}^{r} 2\left(\frac{-\bar{\nu}_l}{\bar{\mu}_{l'}+1} - \bar{\nu}_{l'}\right)\|T(\widetilde{H}_1)_{a_l a_{l'}}\|^2\\
&\quad + \sum_{l=\widetilde{r}_0+1}^{\widetilde{r}_1} -2\bar{\nu}_l\|T(\widetilde{H}_1)_{a_l a_{r+1}}\|^2 + \sum_{l=1}^{\widetilde{r}_0} -\bar{\nu}_l\|(\widetilde{H}_2)_{a_l}\|^2 + \sum_{l=\widetilde{r}_0+1}^{\widetilde{r}_1}\frac{\overline{\sigma}_k}{-\bar{\mu}_l}\|(\widetilde{H}_2)_{a_l}\|^2. \qquad (91)
\end{aligned}$$

Again, it follows from (90) that all the coefficients of the quadric terms of (91) are negative, which implies that $\Upsilon_{\overline{S}}^\circ\left(\overline{X}, H\right) \le 0$ and $\Upsilon_{\overline{S}}^\circ\left(\overline{X}, H\right) = 0$ if and only if $H \in \Re^{m\times n}$ satisfies the conditions (87).

Case 2. $\overline{\sigma}_k = 0$. By the similar arguments, we are able to show that for any $H \in \Re^{m\times n}$,

$$\begin{aligned}
\Upsilon_{\overline{X}}(\overline{S}, H) &= \sum_{l=1}^{r_0}\sum_{l'=\widetilde{r}_0+1}^{\widetilde{r}_1} 2\frac{\bar{\mu}_{l'}-1}{\bar{\nu}_l}\|S(\widetilde{H}_1)_{a_l a_{l'}}\|^2 + \sum_{l=1}^{r_0}\sum_{l'=\widetilde{r}_1+1}^{r+1}\frac{2}{-\bar{\nu}_l}\|S(\widetilde{H}_1)_{a_l a_{l'}}\|^2\\
&\quad + \sum_{l=1}^{r_0}\sum_{l'=1}^{r_0}\frac{2}{-\bar{\nu}_{l'}-\bar{\nu}_l}\|T(\widetilde{H}_1)_{a_l a_{l'}}\|^2 + \sum_{l=1}^{r_0}\sum_{l'=r_0+1}^{\widetilde{r}_0}\frac{4}{-\bar{\nu}_l}\|T(\widetilde{H}_1)_{a_l a_{l'}}\|^2\\
&\quad + \sum_{l=1}^{r_0}\sum_{l'=\widetilde{r}_0+1}^{\widetilde{r}_1} 2\frac{\bar{\mu}_l+1}{-\bar{\nu}_l}\|T(\widetilde{H}_1)_{a_l a_{l'}}\|^2 + \sum_{l=1}^{r_0}\sum_{l'=\widetilde{r}_1+1}^{r+1}\frac{2}{-\bar{\nu}_l}\|T(\widetilde{H}_1)_{a_l a_{l'}}\|^2\\
&\quad + \sum_{l=1}^{r_0}\frac{1}{-\bar{\nu}_l}\|(\widetilde{H}_2)_{a_l}\|^2 \qquad (92)
\end{aligned}$$

and

$$\Upsilon_{\overline{S}}^{\circ}\left(\overline{X},H\right)=\sum_{l=1}^{r_0}\sum_{l'=\widetilde{r}_0+1}^{\widetilde{r}_1}\frac{2\bar{\nu}_l}{\bar{\mu}_{l'}-1}\|S(\widetilde{H}_1)_{a_la_{l'}}\|^2+\sum_{l=1}^{r_0}\sum_{l'=\widetilde{r}_1+1}^{r+1}\frac{2\bar{\nu}_l}{-1}\|S(\widetilde{H}_1)_{a_la_{l'}}\|^2$$

$$+\sum_{l=1}^{r_0}\sum_{l'=1}^{r+1}\frac{2\bar{\nu}_l}{-\bar{\mu}_{l'}-1}\|T(\widetilde{H}_1)_{a_la_{l'}}\|^2+\sum_{l=1}^{r_0}\frac{\bar{\nu}_l}{-1}\|(\widetilde{H}_2)_{a_l}\|^2. \tag{93}$$

Thus, it follows from the fact (90) that all the coefficients of the quadric terms of (92) and (93) are negative, which implies that both $\Upsilon_{\overline{X}}(\overline{S},H)\le 0$, $\Upsilon_{\overline{S}}^{\circ}\left(\overline{X},H\right))\le 0$ and

$$\Upsilon_{\overline{X}}(\overline{S},H)=0 \quad\Longleftrightarrow\quad \Upsilon_{\overline{S}}^{\circ}\left(\overline{X},H\right)=0,$$

which is equivalent to the conditions (88). □

7 Conclusions

In this paper, we studied some important variational properties of the Ky Fan k-norm of matrices related to the nonlinear optimization problem involving the Ky Fan k-norm, which frequently arises and plays a crucial role in various applications. In particular, we introduced and study the concepts of nondegeneracy, strict complementary and critical cone to the locally optimal solutions of the basic nonlinear optimization model (1). Moreover, we provide the explicit formula of the conjugate function of the parabolic second order directional derivative of the Ky Fan k-norm, which provides the necessary second order information for the study of nonlinear optimization problem involving the Ky Fan k-norm. The variational results obtained in this paper can be applied immediately to the study of various perturbation and sensitivity properties, e.g., the second order optimality conditions, strong regularity, full stability and calmness of the general Ky Fan k-norm related optimization problems.

References

1. Bhatia, R.: Matrix Analysis. Springer Science & Business Media (1997).
2. Bonnans, J.F., Shapiro, A.:, Perturbation Analysis of Optimization Problems, Springer, New York, 2000.
3. Boyd, S., Diaconis, P., Xiao, L.: Fastest Mixing Markov Chain on a Graph. SIAM Review 46, 667–689 (2004).
4. Candès, E., Recht, B.: Exact Matrix Completion via Convex Optimization. Foundations of Computational Mathematics 9, 717–772 (2009).
5. Candès, E., Tao, T.: The Power of Convex Relaxation: Near-Optimal Matrix Completion. IEEE Transactions on Information Theory 56, 2053–2080 (2010).
6. Candès, E., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? Journal of the ACM 58, article No. 11 (2011).
7. Chandrasekaran, V., Sanghavi, S., Parrilo, P.A., Willsky, A.: Rank-sparsity incoherence for matrix decomposition. SIAM Journal of Optimization 21, 572–596 (2011).
8. Chu, M., Funderlic, R., Plemmons,R.: Structured low rank approximation. Linear Algebra and its Applications 366, 157–172 (2003).
9. Clarke, F.H.: Optimization and Nonsmooth Analysis, John Wiley & Sons, New York, 1983.
10. Ding, C., Sun, D., Toh, K.-C.: An introduction to a class of matrix cone programming. Mathematical Programming 144, 141–179 (2014).
11. Dobrynin, V.: On the rank of a matrix associated with a graph. Discrete Mathematics 276, 169–175 (2004).
12. Facchinei, F., Pang, J.S., eds. Finite-dimensional variational inequalities and complementarity problems. Vol. 1. Springer Science & Business Media, 2003.
13. Fan, K.: On a theorem of Weyl concerning eigenvalues of linear transformations. Proceedings of the National Academy of Sciences of U.S.A. 35, 652–655 (1949).
14. Y. Gao, Y., Sun, D.: A majorized penalty approach for calibrating rank constrained correlation matrix problems. Preprint available at `http://www.math.nus.edu.sg/~matsundf/MajorPen_May5.pdf` (2010).

15. Greenbaum, A., Trefethen, L.N.: GMRES/CR and Arnoldi/Lanczos as matrix approximation problems. SIAM Journal on Scientific Computing 15, 359–368 (1994).
16. Gross, D.: Recovering low-rank matrices from few coefficients in any basis. IEEE Transactions on Information Theory 57, 1548–1566 (2009).
17. Kotlov, A., Lovász, L., Vempala, S.: The Colin de Verdière number and sphere representations of a graph. Combinatorica 17, 483–521 (1997).
18. Keshavan, R.H., Montanari, A., Oh, S.: Matrix completion from noisy entries. Journal of Machine Learning Research 11, 20572078 (2010).
19. Klatte, D., Kummer, B.: Nonsmooth Equations in Optimization: regularity, calculus, methods, and applications. Kluwer Academic Publishers, (2002).
20. Lancaster, P.: On eigenvalues of matrices dependent on a parameter, Numerische Mathematik 6, 377–387 (1964).
21. Lewis, A.S., Sendov, H.S.: Nonsmooth analysis of singular values. Part II: applications. Set-Valued Analysis 13, 243–264 (2005).
22. Lovász, L.: On the Shannon capacity of a graph. IEEE Transactions on Information Theory 25, 1–7 (1979).
23. Miao, W.M., Sun, D., Pan, S.: A rank-corrected procedure for matrix completion with fixed basis coefficients. Preprint available at `http://arxiv.org/abs/1210.3709` (2012).
24. Mordukhovich, B.S.: Variational Analysis and Generalized Differentiation I: Basic Theory. Springer Berlin Heidelberg (2006).
25. Mordukhovich, B.S., Rockafellar, R.T., Sarabi, M.E.: Characterizations of Full Stability in Constrained Optimization. SIAM Journal of Optimization 23, 1810–1849 (2013).
26. Moreau, J.-J.: Proximité et dualité dans un espace hilbertien. Bulletin de la Société Mathématique de France 93, 1067–1070 (1965).
27. von Neumann, J.: Some matrix-inequalities and metrization of matric-space, Tomsk University Review 1, 286–300 (1937). In: Collected Works, Pergamon, Oxford, 1962, Volume IV, 205-218.
28. Overton, M., Womersley, R.: Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices. Mathematical Programming 62, 321–357 (1993).
29. Recht, B., Fazel, M., Parrilo, P.A.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Review 52, 471–501 (2010).
30. Robinson, S.M.: First order conditions for general nonlinear optimization. SIAM Journal on Applied Mathematics 30, 597–607 (1976).
31. Robinson, S.M.: Local structure of feasible sets in nonlinear programming. II: Nondegeneracy. Mathematical Programming Study 22, 217–230 (1984).
32. Robinson, S.M.: Local structure of feasible sets in nonlinear programming. III: Stability and sensitivity. Mathematical Programming Study 30, 45–66 (1987).
33. Rockafellar, R.T.: Convex Analysis. Princeton University Press (1970).
34. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Springer Berlin Heidelberg, Berlin, Heidelberg (1998).
35. Sun, D.: The Strong Second-Order Sufficient Condition and Constraint Nondegeneracy in Nonlinear Semidefinite Programming and Their Implications. Mathematics of Operations Research 31, 761–776 (2006).
36. Toh, K.-C.: GMRES vs. ideal GMRES. SIAM Journal on Matrix Analysis and Applications 18, 30–36 (1997).
37. Toh, K.-C. and Trefethen, L.N.: *The Chebyshev polynomials of a matrix*. SIAM Journal on Matrix Analysis and Applications 20, 400–419 (1998).
38. Torki, M.: Second-order directional derivatives of all eigenvalues of a symmetric matrix. Nonlinear Analysis 46, 1133–1150 (2001).
39. Watson, G.A.: Linear best approximation using a class of polyhedral norms. Numerical Algorithms 2, 321–336 (1992).
40. Wright, J., Ma, Y., Ganesh, A., Rao, S.: Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization. In Y. Bengio, D. Schuurmans, J. Lafferty and C. Williams, editors, Advances in Neural Information Processing Systems 22, (2009).
41. Wu, B., Ding, C., Sun, D., Toh, K.-C.: On the Moreau–Yosida regularization of the vector k-norm related functions. SIAM Journal on Optimization 24, 766–794 (2014).
42. Zhang, L., Zhang, N., Xiao, X.: On the Second-order Directional Derivatives of Singular Values of Matrices and Symmetric Matrix-valued Functions. Set-Valued and Variational Analysis 21(3), 557–586 (2013).
43. Zowe, J., Kurcyusz, S.: Regularity and stability for the mathematical programming problem in Banach spaces, Applied Mathematics and Optimization 5, 49–62 (1979).