Several classes of stationary points for rank regularized minimization problems

Yulan Liu* and Shaohua Pan†

Abstract

For the rank regularized minimization problem, we introduce several classes of stationary points by the problem itself and its equivalent reformulations including the mathematical program with an equilibrium constraint (MPEC), the global exact penalty of the MPEC, and the surrogate yielded by eliminating the dual part of the exact penalty. A clear relation chart is established among these stationary points, which offers a guidance to choose an appropriate reformulation for seeking a low-rank solution. As a byproduct, for the positive semidefinite (PSD) rank regularized minimization problem, we also provide a weaker condition for a local minimizer of its MPEC reformulation to be the M-stationary point by characterizing the directional limiting normal cone to the graph of the normal cone mapping of the PSD cone.

Keywords: Rank regularized minimization problems; stationary points; matrix MPECs; calmness; directional limiting normal cone

Mathematics Subject Classification(2010): 90C26, 49J52, 49J53

1 Introduction

Let $\mathbb{R}^{m \times n}$ be the linear space of all $m \times n$ $(m \leq n)$ real matrices equipped with the trace inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|_F$, i.e., $\langle X, Y \rangle = \mathrm{tr}(X^{\mathbb{T}} Y)$ for $X, Y \in \mathbb{R}^{m \times n}$. Given a function $f \colon \mathbb{R}^{m \times n} \to \mathbb{R}$, we are interested in the rank regularized problem:

$$\min_{X \in \mathbb{R}^{m \times n}} F(X) := \nu f(X) + \mathrm{rank}(X) + \delta_\Omega(X) \tag{1}$$

where $\nu > 0$ is the regularization parameter and $\Omega \subseteq \mathbb{R}^{m \times n}$ is a closed convex set. Unless otherwise stated, we assume that f is locally Lipschitz and $\widehat{\partial} f(X) = \partial f(X)$ for any $X \in \Omega$, where $\widehat{\partial} f(X)$ and $\partial f(X)$ are the regular and limiting subdifferential of f at X, respectively; see Section 2.1 for their definitions. Such a model is frequently used to seek a low-rank matrix under the scenario where a tight estimation is unavailable for

---

*School of Applied Mathematics, Guangdong University of Technology, Guangzhou.

†Corresponding author (shhpan@scut.edu.cn), School of Mathematics, South China University of Technology, Guangzhou.

the rank of the target matrix, and is found to have a host of applications in a variety of fields such as statistics [26], control and system identification [8,9], signal and image processing [3], finance [30], quantum tomography [12], and so on.

Owing to the combinatorial property of the rank function, the problem (1) is generally NP-hard and it is impossible to achieve a global optimal solution by using an algorithm with polynomial-time complexity. So, it is common to obtain a desirable local optimal even feasible solution by solving a convex relaxation or surrogate problem. Although the nuclear-norm convex relaxation method [7] is very popular, it has a weak ability to promote low-rank solutions and even fails to yielding low-rank solutions in some cases [23]. After recognizing this deficiency, some researchers pay their attentions to the nonconvex surrogates of low-rank optimization problems such as the log-determinant surrogate (see [8,24]) and the Schatten p $(0 < p < 1)$-norm surrogate [15]. As illustrated in [27], the efficiency of nonconvex surrogates depends on its approximation effect.

Recently, by the variational characterization of the rank function, the authors of [1,21] reformulated the rank regularized problem (1) as an equivalent MPEC and derived an equivalent surrogate from its global exact penalty. In order to illustrate this, let $\mathscr{L}$ denote the family of proper lower semi-continuous (lsc) convex functions $\phi\colon \mathbb{R} \to (-\infty, +\infty]$ with

$$\text{int}(\text{dom}\,\phi) \supseteq [0,1],\ \ \phi(1) = 1 > t^* := \underset{0 \le t \le 1}{\arg\min}\,\phi(t)\ \ \text{and}\ \ \phi(t^*) = 0, \tag{2}$$

and for each $\phi \in \mathscr{L}$ let $\psi\colon \mathbb{R} \to (-\infty, +\infty]$ be the associated lsc convex function given by

$$\psi(t) := \begin{cases} \phi(t) & \text{if } t \in [0,1], \\ +\infty & \text{otherwise.} \end{cases} \tag{3}$$

With $\phi \in \mathscr{L}$, the rank regularized problem (1) can be equivalently reformulated as

$$\min_{X,W \in \mathbb{R}^{m \times n}} \nu f(X) + \sum_{i=1}^{m} \phi(\sigma_i(W)) + \delta_{\Omega}(X)$$

$$\text{s.t.}\ \ \|X\|_* - \langle W, X \rangle = 0,\ \|W\| \le 1 \tag{4}$$

which is a matrix MPEC since the constraints $\|X\|_* - \langle W, X \rangle = 0$ and $\|W\| \le 1$ are equivalent to $X \in \mathcal{N}_{\mathbb{B}}(W)$ with $\mathbb{B} := \{Z \in \mathbb{R}^{m \times n} \,|\, \|Z\| \le 1\}$, i.e., the optimality condition of $W \in \arg\max_{Z \in \mathbb{B}} \langle X, Z \rangle$. Under a mild condition, it was shown in [1,21] that the following penalized problem

$$\min_{X,W \in \mathbb{R}^{m \times n}} \nu f(X) + \sum_{i=1}^{m} \phi(\sigma_i(W)) + \rho(\|X\|_* - \langle W, X \rangle)$$

$$\text{s.t.}\ \ X \in \Omega,\ \|W\| \le 1 \tag{5}$$

is a global exact penalty of the MPEC (4) in the sense that there exists $\overline{\rho} > 0$ such that the problem (5) associated to each $\rho \ge \overline{\rho}$ has the same global optimal solution set as (4) does. With the conjugate function $\psi^*(s) := \sup_{t \in \mathbb{R}} \{st - \psi(t)\}$ of ψ, one may eliminate the dual variable W in (5) and get the following equivalent surrogate of the problem (1)

$$\min_{X \in \Omega} \left\{ \nu f(X) + \rho\|X\|_* - \sum_{i=1}^{m} \psi^*(\rho\sigma_i(X)) \right\}. \tag{6}$$

As well known, when an algorithm is applied to nonconvex and nonsmooth optimization problems, one generally expects to achieve a stationary point, while the stationary points of equivalent reformulations may have a big difference. Thus, it is necessary to clarify the relation among the stationary points of (1) defined by its equivalent reformulations. Moreover, such a clarification is prerequisite to describe the landscape of stationary points for the rank regularized problem (1). Motivated by this, in Section 3 we introduce the R(egular)-stationary point, the M-stationary point, the EP-stationary point and the DC-stationary point by the problem (1) itself and its reformulation (4)-(6), respectively, and explore the relation among the four classes of stationary points. Figure 1 in Section 3 shows that the set of M-stationary points is almost same as that of R-stationary points, the latter includes that of EP-stationary points under a rank condition, and the set of EP-stationary points coincides with that of DC-stationary points for some appropriate ϕ. As a byproduct, for the PSD rank regularized minimization problem, we also provide a weaker condition than the one in [5] for a local minimizer of its MPEC reformulation to be the M-stationary point, by the directional limiting normal cone to the graph of the normal cone mapping of the PSD cone $\mathbb{S}_+^n$.

We notice that some active research has been done for the stationary points of zero-norm constrained optimization problems (see, e.g., [2, 10, 28]); for example, Burdakov et al. [2] discussed the relation between the M-stationary point and the S-stationary point of their equivalent MPEC reformulation; and Pan et al. [28] characterizes the first-order optimality condition which actually defines a class of stationary points by the tangent cone to the zero-norm constrained set. To the best of our knowledge, there are few works to study the stationary points of rank regularized optimization problems. For the special case $\Omega \subseteq \mathbb{S}_+^n$, the rank regularized problem (1) can reduce to a mathematical program with semidefinite conic complementarity constraints (MPSCCC) and Ding et al. [5] have established the connection among several class of stationary points for the MPSCCC, which are defined by the equivalent reformulations of the complementarity constraints. However, this work is concerned with the relation among the stationary points defined by different equivalent reformulations of the rank regularized problem (1), and aims to establish a clear relation chart for these stationary points so that the user can be guided to choose an appropriate reformulation to seek a low-rank solution.

2   Notation and preliminaries

Throughout this paper, a hollow capital means a finite dimensional vector space equipped with the inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$. The notation $\mathbb{S}^n$ denotes the vector space of all $n \times n$ real symmetric matrices equipped with the Frobenius norm, and $\mathbb{S}_+^n$ means the set of all positive semidefinite matrices in $\mathbb{S}^n$. Let $\mathbb{O}^{m \times n}$ be the set of $m \times n$ matrices with orthonormal columns and denote $\mathbb{O}^{m \times m}$ by $\mathbb{O}^m$. For a given $X \in \mathbb{R}^{m \times n}$, we denote by $\|X\|_*$ and $\|X\|$ the nuclear norm and the spectral norm of X, respectively, and by $\sigma(X) \in \mathbb{R}^m$ the singular value vector arranged in a nonincreasing order; and write $\mathbb{O}^{m,n}(X) := \{(U, V) \in \mathbb{O}^m \times \mathbb{O}^n \mid X = U\mathrm{Diag}(\sigma(X))V^{\mathbb{T}}\}$. For a given $X \in \mathbb{R}^{m \times n}$ and two index sets $\alpha \subseteq \{1, \ldots, m\}$ and $\beta \subseteq \{1, \ldots, n\}$, $X_{\alpha\beta}$ means the submatrix consists

of those entries X_{ij} with $i \in \alpha$ and $j \in \beta$. We denote by E and e the matrix and the vector of all ones respectively whose dimension are known from the context, and by I an identity matrix whose dimension is known from the context. For a given set S, δ_S denotes the indicator function of S, i.e., $\delta_S(x) = 0$ if $x \in S$, otherwise $\delta_S(x) = +\infty$. For a given vector space $\mathbb{Z}$, $\mathbb{B}_\mathbb{Z}$ denotes the closed unit ball centered at the origin of $\mathbb{Z}$, and $\mathbb{B}_\delta(z)$ means the closed ball of radius δ centered at $z \in \mathbb{Z}$.

2.1 Normal cones and generalized differentials

Let $S \subset \mathbb{Z}$ be a given set. The regular normal cone to S at a point $\overline{z} \in S$ is defined by

$$\widehat{\mathcal{N}}_S(\overline{z}) := \left\{ v \in \mathbb{Z} \mid \limsup_{z \xrightarrow{S} \overline{z}} \frac{\langle v, z - \overline{z} \rangle}{\|z - \overline{z}\|} \le 0 \right\}$$

where the symbol $z \xrightarrow{S} \overline{z}$ signifies $z \to \overline{z}$ with $z \in S$, while the limiting normal cone to S at $\overline{z}$ is defined as the outer limit of $\widehat{\mathcal{N}}_S(z)$ as $z \xrightarrow{S} \overline{z}$, i.e.,

$$\mathcal{N}_S(\overline{z}) := \left\{ v \in \mathbb{Z} \mid \exists z^k \xrightarrow{S} \overline{z}, v^k \to v \ \text{ with } \ v^k \in \widehat{\mathcal{N}}_S(z^k) \right\}. \tag{7}$$

The limiting normal cone $\mathcal{N}_S(\overline{z})$ is generally not convex, but the regular normal $\widehat{\mathcal{N}}_S(\overline{z})$ is always closed convex which is the negative polar of the contingent cone to S at $\overline{x}$:

$$\mathcal{T}_S(\overline{z}) := \left\{ h \in \mathbb{Z} \mid \exists t_k \downarrow 0, \, h^k \to h \text{ with } \overline{z} + t_k h^k \in S \right\}.$$

When S is convex, $\mathcal{N}_S(\overline{z})$ and $\widehat{\mathcal{N}}_S(\overline{z})$ are the normal cone in the sense of convex analysis [31]. The directional limiting normal cone to S at $\overline{z}$ in a direction $u \in \mathbb{X}$ is defined by

$$\mathcal{N}_S(\overline{z}; u) := \left\{ z^* \in \mathbb{Z} \mid \exists t_k \downarrow 0, \, u^k \to u, z^{k*} \to z^* \text{ with } z^{k*} \in \widehat{\mathcal{N}}_S(\overline{z} + t_k u^k) \right\}.$$

By comparing with the definition of $\mathcal{N}_S(\overline{z})$, clearly, $\mathcal{N}_S(\overline{z}; u) \subseteq \mathcal{N}_S(\overline{z})$ for any $u \in \mathbb{X}$.

Let $g: \mathbb{Z} \to [-\infty, +\infty]$ be an extended real-valued lsc function with $g(\overline{z})$ finite. The regular subdifferential of g at $\overline{z}$, denoted by $\widehat{\partial} g(\overline{z})$, is defined as

$$\widehat{\partial} g(\overline{z}) := \left\{ z^* \in \mathbb{X} \mid \liminf_{\substack{z \to \overline{z} \\ z \ne \overline{z}}} \frac{g(z) - g(\overline{z}) - \langle z^*, z - \overline{z} \rangle}{\|z - \overline{z}\|} \ge 0 \right\};$$

and the (limiting) subdifferential of g at $\overline{z}$, denoted by $\partial g(\overline{z})$, is defined as

$$\partial g(\overline{z}) = \left\{ z^* \in \mathbb{X} \mid \exists z^k \xrightarrow{g} z, z^{k,*} \to z^* \text{ such that } z^{k,*} \in \widehat{\partial} g(z^k) \right\}. \tag{8}$$

From [32, Theorem 8.9] we know that there is close relation between the subdifferentials of g at $\overline{z}$ and the normal cones of its epigraph at $(\overline{z}, g(\overline{z}))$. Also, from [32, Exercise 8.14],

$$\widehat{\mathcal{N}}_S(z) = \widehat{\partial} \delta_S(z) \ \text{ and } \ \mathcal{N}_S(z) = \partial \delta_S(z) \ \text{ for } z \in S.$$

In the sequel, we call a point z at which $0 \in \partial g(z)$ (respectively, $0 \in \widehat{\partial} g(z)$) is called a limiting (respectively, regular) critical point of g. By [32, Theorem 10.1], a local minimizer of g is necessarily a regular critical point of g, and then a limiting critical point.

4

2.2 Lipschitz-like properties of multifunctions

Let $\mathcal{F}\colon \mathbb{Z} \rightrightarrows \mathbb{W}$ be a given multifunction. Consider an arbitrary point $(\overline{z}, \overline{w}) \in \mathrm{gph}\mathcal{F}$ at which $\mathcal{F}$ is locally closed, where $\mathrm{gph}\mathcal{F}$ denotes the graph of $\mathcal{F}$. We recall from [6,32] the concepts of the Aubin property, calmness and metric subregularity of $\mathcal{F}$.

Definition 2.1 *The multifunction $\mathcal{F}$ is said to have the Aubin property at $\overline{z}$ for $\overline{w}$ with modulus $\kappa > 0$, if there exist $\varepsilon > 0$ and $\delta > 0$ such that for all $z, z' \in \mathbb{B}_\varepsilon(\overline{z})$,*

$$\mathcal{F}(z) \cap \mathbb{B}_\delta(\overline{w}) \subseteq \mathcal{F}(z') + \kappa\|z - z'\|\mathbb{B}_\mathbb{W}.$$

Definition 2.2 *The multifunction $\mathcal{F}$ is said to be calm at $\overline{z}$ for $\overline{w}$ with modulus $\kappa > 0$ if there exist $\varepsilon > 0$ and $\delta > 0$ such that for all $z \in \mathbb{B}_\varepsilon(\overline{z})$,*

$$\mathcal{F}(z) \cap \mathbb{B}_\delta(\overline{w}) \subseteq \mathcal{F}(\overline{z}) + \kappa\|z - \overline{z}\|\mathbb{B}_\mathbb{W}.$$

If in addition $\mathcal{F}(\overline{z}) \cap \mathbb{B}_\delta(\overline{w}) = \{\overline{w}\}$, $\mathcal{F}$ is said to be isolated calm at $\overline{z}$ for $\overline{w}$.

By [6, Exercise 3H.4], the restriction on $z \in \mathbb{B}_\varepsilon(\overline{z})$ in Definition 2.2 can be removed. It is easily seen that the calmness of $\mathcal{F}$ is a "one-point" variant of the Aubin property, and the calmness of $\mathcal{F}$ at $(\overline{z}, \overline{w}) \in \mathrm{gph}\mathcal{F}$ is implied by its Aubin property or isolated calmness at this point. Notice that the calmness of $\mathcal{F}$ at $\overline{z}$ for $\overline{w} \in \mathcal{F}(\overline{z})$ is equivalent to the metric subregularity of $\mathcal{F}^{-1}$ at $\overline{w}$ for $\overline{z} \in \mathcal{F}^{-1}(\overline{w})$ by [6, Theorem 3H.3].

The coderivative and graphical derivative of $\mathcal{F}$ are an convenient tool to characterize the Aubin property and the isolated calmness of $\mathcal{F}$, respectively. Recall from [32] that the coderivative of $\mathcal{F}$ at $\overline{z}$ for $\overline{w}$ is the mapping $D^*\mathcal{F}(\overline{z}|\overline{w})\colon \mathbb{W} \rightrightarrows \mathbb{Z}$ defined by

$$u \in D^*\mathcal{F}(\overline{z}|\overline{w})(v) \iff (u, -v) \in \mathcal{N}_{\mathrm{gph}\,\mathcal{F}}(\overline{z}, \overline{w}),$$

and the graphical derivative of $\mathcal{F}$ at $\overline{z}$ for $\overline{w}$ is the mapping $D\mathcal{F}(\overline{z}|\overline{w})\colon \mathbb{Z} \rightrightarrows \mathbb{W}$ given by

$$v \in D\mathcal{F}(\overline{z}|\overline{w})(u) \iff (u, v) \in \mathcal{T}_{\mathrm{gph}\,\mathcal{F}}(\overline{z}, \overline{w}).$$

Lemma 2.1 *(See [25, Theorem 5.7] or [32, Theorem 9.40]) Suppose that $\mathcal{F}$ is locally closed at $(\overline{z}, \overline{w})$. Then $\mathcal{F}$ has the Aubin property at $\overline{z}$ for $\overline{w}$ iff $D^*\mathcal{F}(\overline{z}|\overline{w})(0) = \{0\}$.*

Lemma 2.2 *(See [14, Proposition 2.1] or [18, Proposition 4.1]) Suppose that $\mathcal{F}$ is locally closed at $(\overline{z}, \overline{w})$. Then $\mathcal{F}$ is isolated calm at $\overline{z}$ for $\overline{w}$ iff $D\mathcal{F}(\overline{z}|\overline{w})(0) = \{0\}$.*

2.3 Coderivative of the subdifferential mapping $\partial\|\cdot\|_*$

For a given $X \in \mathbb{R}^{m \times n}$ with SVD as $U[\mathrm{Diag}(\sigma(X))\ \ 0]V^\mathbb{T}$, by [35, Example 2] we have

$$\partial\|X\|_* = \left\{ [U_1\ \ U_2] \begin{bmatrix} I & 0 \\ 0 & Z \end{bmatrix} [V_1\ \ V_2]^\mathbb{T} \mid \|Z\| \leq 1 \right\} \tag{9}$$

where U_1 and V_1 are the submatrix consisting of the first $r = \mathrm{rank}(X)$ columns of U and V, respectively, and U_2 and V_2 are the submatrix consisting of the last $m - r$ columns and

$n-r$ columns of U and V, respectively. In this part we recall from [22] the coderivative of the subdifferential mapping $\partial\|\cdot\|_*$. For this purpose, in the sequel for two positive integers k_1 and k_2 with $k_2 \geq k_1$, we denote by $[k_1, k_2]$ the set $\{k_1, k_1+1, \ldots, k_2\}$. For a given $\overline{Z} \in \mathbb{R}^{m \times n}$, define the following index sets associated to its singular values:

$$\alpha := \{i \in [1, m] \mid \sigma_i(\overline{Z}) > 1\}, \ \ \beta := \{i \in [1, m] \mid \sigma_i(\overline{Z}) = 1\}, \ \ c = [m+1, n], \quad (10\text{a})$$

$$\gamma := \gamma_1 \cup \gamma_0 \text{ for } \gamma_1 := \{i \in [1, m] \mid 0 < \sigma_i(\overline{Z}) < 1\}, \gamma_0 := \{i \in [1, m] \mid \sigma_i(\overline{Z}) = 0\}, \ (10\text{b})$$

and let $\Omega_1, \Omega_2 \in \mathbb{S}^m$ and $\Omega_3 \in \mathbb{R}^{m \times (n-m)}$ be the matrices associated to $\sigma(\overline{Z})$ given by

$$\left(\Omega_1\right)_{ij} := \begin{cases} \frac{\min(1,\sigma_i(\overline{Z})) - \min(1,\sigma_j(\overline{Z}))}{\sigma_i(\overline{Z}) - \sigma_j(\overline{Z})} & \text{if } \sigma_i(\overline{Z}) \neq \sigma_j(\overline{Z}), \\ 0 & \text{otherwise} \end{cases} \quad i,j \in \{1, 2, \ldots, m\}, \quad (11\text{a})$$

$$\left(\Omega_2\right)_{ij} := \begin{cases} \frac{\min(1,\sigma_i(\overline{Z})) + \min(1,\sigma_j(\overline{Z}))}{\sigma_i(\overline{Z}) + \sigma_j(\overline{Z})} & \text{if } \sigma_i(\overline{Z}) + \sigma_j(\overline{Z}) \neq 0, \\ 0 & \text{otherwise} \end{cases} \quad i,j \in \{1, 2, \ldots, m\},$$

$$(11\text{b})$$

$$\left(\Omega_3\right)_{ij} := \begin{cases} \frac{\min(1,\sigma_i(\overline{Z}))}{\sigma_i(\overline{Z})} & \text{if } \sigma_i(\overline{Z}) \neq 0, \\ 0 & \text{otherwise} \end{cases} \quad i \in \{1, \ldots, m\}, j \in \{1, \ldots, n-m\}. \quad (11\text{c})$$

With the matrices $\Omega_1, \Omega_2 \in \mathbb{S}^m$ and $\Omega_3 \in \mathbb{R}^{m \times (n-m)}$, we define the following matrices

$$\Theta_1 := \begin{bmatrix} 0_{\alpha\alpha} & 0_{\alpha\beta} & (\Omega_1)_{\alpha\gamma} \\ 0_{\beta\alpha} & 0_{\beta\beta} & E_{\beta\gamma} \\ (\Omega_1)_{\gamma\alpha} & E_{\gamma\beta} & E_{\gamma\gamma} \end{bmatrix}, \ \Theta_2 := \begin{bmatrix} E_{\alpha\alpha} & E_{\alpha\beta} & E_{\alpha\gamma} - (\Omega_1)_{\alpha\gamma} \\ E_{\beta\alpha} & 0_{\beta\beta} & 0_{\beta\gamma} \\ E_{\gamma\alpha} - (\Omega_1)_{\gamma\alpha} & 0_{\gamma\beta} & 0_{\gamma\gamma} \end{bmatrix},$$

$$\Sigma_1 := \begin{bmatrix} (\Omega_2)_{\alpha\alpha} & (\Omega_2)_{\alpha\beta} & (\Omega_2)_{\alpha\gamma} \\ (\Omega_2)_{\beta\alpha} & 0_{\beta\beta} & E_{\beta\gamma} \\ (\Omega_2)_{\gamma\alpha} & E_{\gamma\beta} & E_{\gamma\gamma} \end{bmatrix}, \Sigma_2 := \begin{bmatrix} E_{\alpha\alpha} - (\Omega_2)_{\alpha\alpha} & E_{\alpha\beta} - (\Omega_2)_{\alpha\beta} & E_{\alpha\gamma} - (\Omega_2)_{\alpha\gamma} \\ E_{\beta\alpha} - (\Omega_2)_{\beta\alpha} & 0_{\beta\beta} & 0_{\beta\gamma} \\ E_{\gamma\alpha} - (\Omega_2)_{\gamma\alpha} & 0_{\gamma\beta} & 0_{\gamma\gamma} \end{bmatrix}.$$

For the index set β, we denote the set of all partitions of β by $\mathscr{P}(\beta)$. Define the set

$$\mathbb{R}_>^{|\beta|} := \{z \in \mathbb{R}^{|\beta|}: \ z_1 \geq \cdots \geq z_{|\beta|} > 0\}.$$

For any $z \in \mathbb{R}_>^{|\beta|}$, let $D(z) \in \mathbb{S}^{|\beta|}$ denote the first generalized divided difference matrix of $h(t) = \min(1, t)$ at z, which is defined as

$$(D(z))_{ij} := \begin{cases} \frac{\min(1,z_i) - \min(1,z_j)}{z_i - z_j} \in [0, 1] & \text{if } z_i \neq z_j, \\ 0 & \text{if } z_i = z_j \geq 1, \\ 1 & \text{otherwise.} \end{cases} \quad (13)$$

Write $\mathcal{U}_{|\beta|} := \{\overline{\Omega} \in \mathbb{S}^{|\beta|}: \ \overline{\Omega} = \lim_{k \to \infty} D(z^k), \ z^k \to e_{|\beta|}, \ z^k \in \mathbb{R}_>^{|\beta|}\}$. For each $\Xi_1 \in \mathcal{U}_{|\beta|}$, by equation (13) there exists a partition $(\beta_+, \beta_0, \beta_-) \in \mathscr{P}(\beta)$ such that

$$\Xi_1 = \begin{bmatrix} 0_{\beta_+\beta_+} & 0_{\beta_+\beta_0} & (\Xi_1)_{\beta_+\beta_-} \\ 0_{\beta_0\beta_+} & 0_{\beta_0\beta_0} & E_{\beta_0\beta_-} \\ (\Xi_1)_{\beta_+\beta_-}^{\mathbb{T}} & E_{\beta_-\beta_0} & E_{\beta_-\beta_-} \end{bmatrix}, \quad (14)$$

6

where each entry of $(\Xi_1)_{\beta_+\beta_-}$ belongs to $[0,1]$. Let Ξ_2 be the matrix associated to Ξ_1:

$$\Xi_2 = \begin{bmatrix} E_{\beta_+\beta_+} & E_{\beta_+\beta_0} & E_{\beta_+\beta_-} - (\Xi_1)_{\beta_+\beta_-} \\ E_{\beta_0\beta_+} & 0_{\beta_0\beta_0} & 0_{\beta_0\beta_-} \\ E_{\beta_-\beta_+} - (\Xi_1)_{\beta_+\beta_-}^{\mathbb{T}} & 0_{\beta_-\beta_0} & 0_{\beta_-\beta_-} \end{bmatrix}. \tag{15}$$

Now we are in a position to give the coderivative of the subdifferential mapping $\partial\|\cdot\|_*$.

Lemma 2.3 *(See [22, Theorem 3.2]) Fix an arbitrary $(X,W) \in \text{gph}\,\partial\|\cdot\|_*$ and let α,β,γ and c be defined by* (10a)-(10b) *with $\overline{Z} = X+W$. Let $(\overline{U},\overline{V}) \in \mathbb{O}^{m,n}(\overline{Z})$ with $\overline{V} = [\overline{V}_1\ \ \overline{V}_2]$ where $\overline{V}_1 \in \mathbb{O}^{n\times m}$ and $\overline{V}_2 \in \mathbb{O}^{n\times(n-m)}$, and for each $H \in \mathbb{R}^{m\times n}$ write $\widetilde{H} = \overline{U}^{\mathbb{T}}H\overline{V}$ and $\widetilde{H}_1 = \overline{U}^{\mathbb{T}}H\overline{V}_1$. Then, $(G,H) \in \mathcal{N}_{\text{gph}\,\partial\|\cdot\|_*}(X,W)$ iff the following relations hold*

$$\Theta_1 \circ \mathcal{S}(\widetilde{H}_1) + \Theta_2 \circ \mathcal{S}(\widetilde{G}_1) + \Sigma_1 \circ \mathcal{X}(\widetilde{H}_1) + \Sigma_2 \circ \mathcal{X}(\widetilde{G}_1) = 0, \tag{16a}$$

$$\widetilde{G}_{\alpha c} + (\Omega_3)_{\alpha c} \circ (\widetilde{H}_{\alpha c} - \widetilde{G}_{\alpha c}) = 0, \ \widetilde{H}_{\beta c} = 0, \ \widetilde{H}_{\gamma c} = 0, \tag{16b}$$

$$(\widetilde{G}_{\beta\beta}, \widetilde{H}_{\beta\beta}) \in \bigcup_{\substack{Q\in\mathbb{O}^{|\beta|} \\ \Xi_1\in\mathcal{U}_{|\beta|}}} \left\{ (M,N) \ \middle| \ \begin{array}{c} \Xi_1 \circ \widehat{N} + \Xi_2 \circ \mathcal{S}(\widehat{M}) + \Xi_2 \circ \mathcal{X}(\widehat{N}) = 0 \\ \text{with } \widehat{N} = Q^{\mathbb{T}}NQ, \ \widehat{M} = Q^{\mathbb{T}}MQ, \\ Q_{\beta_0}^{\mathbb{T}}MQ_{\beta_0} \preceq 0, \ Q_{\beta_0}^{\mathbb{T}}NQ_{\beta_0} \succeq 0 \end{array} \right\} \tag{16c}$$

where $\mathcal{S}\colon \mathbb{R}^{m\times m} \to \mathbb{S}^m$ and $\mathcal{X}\colon \mathbb{R}^{m\times m} \to \mathbb{R}^{m\times m}$ are linear the mappings defined by

$$\mathcal{S}(Y) := (Y+Y^{\mathbb{T}})/2 \ \text{ and } \ \mathcal{X}(Y) := (Y-Y^{\mathbb{T}})/2 \quad \forall Y \in \mathbb{R}^{m\times m}, \tag{17}$$

and the notation "$\circ$" denotes the Hardmard product operator of two matrices.

3 Four classes of stationary points and their relations

To introduce the four classes of stationary points for the problem (1), with each $\phi \in \mathscr{L}$, we write $\widehat{\phi}(t) := \phi(|t|)$ for $t \in \mathbb{R}$ and $\widehat{\Phi}(x) := \sum_{i=1}^m \widehat{\phi}(x_i)$ for $x \in \mathbb{R}^m$; and with the associated ψ, write $\widehat{\psi}(t) := \psi(|t|)$ for $t \in \mathbb{R}$ and $\widehat{\Psi}(x) := \sum_{i=1}^m \widehat{\psi}(x_i)$ for $x \in \mathbb{R}^m$. Clearly, $\widehat{\Phi}$ and $\widehat{\Psi}$ are absolutely symmetric, i.e., $\widehat{\Phi}(Px) = \widehat{\Phi}(x)$ and $\widehat{\Psi}(Px) = \widehat{\Psi}(x)$ for any $m \times m$ signed permutation matrix P. Also, $\widehat{\Phi} \circ \sigma$ is globally Lipschitz continuous over the ball $\mathbb{B}$. The following equivalent relations are often used in the subsequent analysis

$$\|X\|_* - \langle X, W \rangle = 0, \|W\| \leq 1 \Longleftrightarrow W \in \underset{Z\in\mathbb{B}}{\arg\max}\,\langle Z, X \rangle \Longleftrightarrow X \in \mathcal{N}_{\mathbb{B}}(W) \tag{18a}$$

$$\Longleftrightarrow W \in \partial\|\cdot\|_*(X) \Longleftrightarrow (X,W) \in \text{gph}\,\partial\|\cdot\|_* \tag{18b}$$

3.1 R-stationary point

Recall that $\overline{X} \in \mathbb{R}^{m\times n}$ is a regular critical point of F if $0 \in \widehat{\partial}F(\overline{X})$. Since the rank function is regular by [37, Lemma 2.1] and [19, Corollary 7.5], by combining with the assumption on f, from [32, Corollary 10.9] we have $\widehat{\partial}F(\overline{X}) \supseteq \partial f(\overline{X}) + \partial\text{rank}(\overline{X}) + \mathcal{N}_{\Omega}(\overline{X})$. In view of this, we introduce the following R-stationary point of the problem (1).

Definition 3.1 *A matrix $\overline{X} \in \mathbb{R}^{m \times n}$ is called a R-stationary point of the problem* (1) *if*

$$0 \in \nu \partial f(\overline{X}) + \partial \operatorname{rank}(\overline{X}) + \mathcal{N}_\Omega(\overline{X}).$$

Remark 3.1 *Clearly, every R-stationary point of* (1) *is a regular critical point of F. By the given assumption on f and [32, Exercise 10.10], for any $X \in \Omega$ it holds that*

$$\partial F(X) \subset \nu \partial f(X) + \partial (\operatorname{rank} + \delta_\Omega)(X).$$

Thus, when $\partial(\operatorname{rank} + \delta_\Omega)(\overline{X}) \subset \partial \operatorname{rank}(\overline{X}) + \mathcal{N}_\Omega(\overline{X})$, the limiting critical point of F is same as its regular critical point, and coincides with the R-stationary point of (1).

3.2 M-stationary point

By invoking the relation (18b), clearly, the MPEC (4) can be compactly written as

$$\min_{X,W \in \mathbb{R}^{m \times n}} \left\{ \widetilde{F}(X,W) := \nu f(X) + \widehat{\Phi}(\sigma(W)) + \delta_\Omega(X) + \delta_{\operatorname{gph}\partial\|\cdot\|_*}(X,W) \right\}. \tag{19}$$

Moreover, under a suitable constraint qualification (CQ), the following inclusion holds:

$$\partial \widetilde{F}(X,W) \subseteq \left[\nu \partial f(X) + \mathcal{N}_\Omega(X) \right] \times \partial(\widehat{\Phi} \circ \sigma)(W) + \mathcal{N}_{\operatorname{gph}\partial\|\cdot\|_*}(X,W).$$

Motivated by this, we introduce the M-stationary point of the problem (1) as follows.

Definition 3.2 *A matrix $\overline{X} \in \mathbb{R}^{m \times n}$ is called an M-stationary point of the problem* (1) *associated to $\phi \in \mathscr{L}$ if there exist $\overline{W} \in \partial\|\cdot\|_*(\overline{X})$ and $\Delta W \in \partial(\widehat{\Phi} \circ \sigma)(\overline{W})$ such that*

$$0 \in \nu \partial f(\overline{X}) + \mathcal{N}_\Omega(\overline{X}) + D^*\partial\|\cdot\|_*(\overline{X}|\overline{W})(\Delta W). \tag{20}$$

Remark 3.2 *When $\Omega \subseteq \mathbb{S}^n_+$, the rank regularized problem* (1) *can be reformulated as*

$$\min_{X,W \in \mathbb{S}^n} \nu f(X) + \sum_{i=1}^m \phi(\sigma_i(W)) + \delta_\Omega(X)$$
$$\text{s.t.} \quad \langle I - W, X \rangle = 0, \ W \in \mathbb{S}^n_+, \ I - W \in \mathbb{S}^n_+. \tag{21}$$

Notice that $\langle I - W, X \rangle = 0, X \in \mathbb{S}^n_+, W \in \mathbb{S}^n_+$ and $I - W \in \mathbb{S}^n_+$ iff $(X, W - I) \in \operatorname{gph}\mathcal{N}_{\mathbb{S}^n_+}$ and $W \in \mathbb{S}^n_+$. So, for this case, $\overline{X} \in \Omega$ is an M-stationary point if and only if there exist $(\overline{Y}, \Delta Y) \in \mathbb{S}^n_+ \times \mathbb{S}^n$ with $\overline{Y} - I \in \mathcal{N}_{\mathbb{S}^n_+}(\overline{X})$ and $\Delta Y \in \partial(\widehat{\Phi} \circ \sigma)(\overline{Y}) + \mathcal{N}_{\mathbb{S}^n_+}(\overline{Y})$ such that

$$0 \in \nu \partial f(\overline{X}) + \mathcal{N}_\Omega(\overline{X}) + D^*\mathcal{N}_{\mathbb{S}^n_+}(\overline{X}|\overline{Y} - I)(-\Delta Y),$$

or equivalently, there exist $\overline{Y} \in \mathbb{S}^n_+$ with $\overline{Y} - I \in \mathcal{N}_{\mathbb{S}^n_+}(\overline{X})$ and $(\overline{\Gamma}_1, \overline{\Gamma}_2) \in \mathbb{S}^n \times \mathbb{S}^n$ such that

$$\begin{cases} 0 \in \nu \partial f(\overline{X}) + \overline{\Gamma}_1 + \mathcal{N}_\Omega(\overline{X}), & \text{(22a)} \\ 0 \in \partial(\widehat{\Phi} \circ \sigma)(\overline{Y}) - \overline{\Gamma}_2 + \mathcal{N}_{\mathbb{S}^n_+}(\overline{Y}), & \text{(22b)} \\ \overline{\Gamma}_1 \in D^*\mathcal{N}_{\mathbb{S}^n_+}(\overline{X}|\overline{Y} - I)(-\overline{\Gamma}_2). & \text{(22c)} \end{cases}$$

For this class of stationary points, we have the following proposition that is the key to achieve the relation between the M-stationary point and the R-stationary point.

Proposition 3.1 *Let $\mathscr{L}_1$ denote the family of those $\phi \in \mathscr{L}$ that is differentiable on $(0,1]$. If $\overline{X}$ is an M-stationary point of the problem (1) associated to $\phi \in \mathscr{L}_1$, then there exist $\overline{W} \in \partial \|\cdot\|_*(\overline{X})$ and $\Delta\Gamma \in \nu\partial f(\overline{X}) + \mathcal{N}_\Omega(\overline{X})$ such that for the index sets $\alpha, \beta, c, \gamma, \gamma_1$ and γ_0 defined as in (10a)-(10b) with $\overline{Z} = \overline{X} + \overline{W}$ and $(\overline{U}, \overline{V}) \in \mathbb{O}^{m,n}(\overline{Z})$,*

$$\Delta\Gamma = \overline{U} \begin{bmatrix} 0_{\alpha\alpha} & 0_{\alpha\beta} & 0_{\alpha\gamma} & 0_{\alpha c} \\ 0_{\beta\alpha} & (\Delta\widetilde{\Gamma})_{\beta\beta} & (\Delta\widetilde{\Gamma})_{\beta\gamma} & (\Delta\widetilde{\Gamma})_{\beta c} \\ 0_{\gamma\alpha} & (\Delta\widetilde{\Gamma})_{\gamma\beta} & (\Delta\widetilde{\Gamma})_{\gamma\gamma} & (\Delta\widetilde{\Gamma})_{\gamma c} \end{bmatrix} \overline{V}^{\mathbb{T}} \quad \text{for } \Delta\widetilde{\Gamma} = \overline{U}^{\mathbb{T}}\Delta\Gamma\overline{V}, \mathcal{S}[(\Delta\widetilde{\Gamma})_{\beta\beta}] = 0. \quad (23)$$

In particular, if $\phi'(t) \neq 0$ for all $t \in (0,1)$, then $\gamma_1 = \emptyset$; and if $0 \notin \partial\widehat{\phi}(0)$, then $\gamma_0 = \emptyset$.

Proof: Let $\overline{X}$ be an M-stationary point of the problem (1) associated to $\phi \in \mathscr{L}_1$. By Definition 3.2, there exist $\overline{W} \in \partial \|\cdot\|_*(\overline{X})$ and $\Delta W \in \partial(\widehat{\Phi} \circ \sigma)(\overline{W})$ such that (20) holds. So, there exists $\Delta\Gamma \in \nu\partial f(\overline{X}) + \mathcal{N}_\Omega(\overline{X})$ such that $-\Delta\Gamma \in D^*\partial\|\cdot\|_*(\overline{X}|\overline{W})(\Delta W)$. We argue that $\Delta\Gamma$ has the form of (23). Since $(\overline{U}, \overline{V}) \in \mathbb{O}^{m,n}(\overline{Z})$, from $\overline{W} \in \partial\|\cdot\|_*(\overline{X})$,

$$\overline{X} = \overline{U} \begin{bmatrix} \mathrm{Diag}(\sigma_\alpha(\overline{Z}) - e_\alpha) & 0_{\alpha\beta} & 0_{\alpha\gamma} & 0_{\alpha c} \\ 0_{\beta\alpha} & 0_{\beta\beta} & 0_{\beta\gamma} & 0_{\beta c} \\ 0_{\gamma\alpha} & 0_{\gamma\beta} & 0_{\gamma\gamma} & 0_{\gamma c} \end{bmatrix} \overline{V}^{\mathbb{T}}, \quad (24a)$$

$$\overline{W} = \overline{U} \begin{bmatrix} I_{\alpha\alpha} & 0_{\alpha\beta} & 0_{\alpha\gamma} & 0_{\alpha c} \\ 0_{\beta\alpha} & I_{\beta\beta} & 0_{\beta\gamma} & 0_{\beta c} \\ 0_{\gamma\alpha} & 0_{\gamma\beta} & \mathrm{Diag}(\sigma_\gamma(\overline{Z})) & 0_{\gamma c} \end{bmatrix} \overline{V}^{\mathbb{T}}. \quad (24b)$$

Since $\widehat{\Phi}$ is absolutely symmetric and $\Delta W \in \partial(\widehat{\Phi} \circ \sigma)(\overline{W})$, by [20, Corollary 2.5] and equation (24b) there exist $(\widehat{U}, \widehat{V}) \in \mathbb{O}^{m,n}(\overline{W})$ and $\overline{w} \in \partial\widehat{\Phi}(\sigma(\overline{W}))$ such that

$$\Delta W = \widehat{U}[\mathrm{Diag}(\overline{w}) \ \ 0]\widehat{V}^{\mathbb{T}}. \quad (25)$$

From $\overline{w} \in \partial\widehat{\Phi}(\sigma(\overline{W}))$, the definition of $\widehat{\Phi}$ and equation (24b), it follows that

$$\overline{w}_i = \phi'(1) \text{ for } i \in \alpha \cup \beta; \quad \overline{w}_i = \phi'(\sigma_i(\overline{Z})) \text{ for } i \in \gamma_1; \quad \overline{w}_i \in \partial\widehat{\phi}(0) \text{ for } i \in \gamma_0. \quad (26)$$

Without loss of generality, we assume that the matrix $\overline{Z}$ has r distinct singular values belonging to $(0,1)$. Let $\overline{\mu}_1 > \overline{\mu}_2 > \cdots > \overline{\mu}_r$ be the r distinct singular values and write

$$a_k := \{i \in \gamma_1 \mid \sigma_i(\overline{Z}) = \overline{\mu}_k\} \quad \text{for} \quad k = 1, 2, \ldots, r.$$

Since $(\widehat{U}, \widehat{V}) \in \mathbb{O}^{m,n}(\overline{W})$, from equation (24b) and [4, Proposition 5], there exist a block diagonal matrix $\widehat{Q} = \mathrm{Diag}(Q_0, Q_1, \ldots, Q_r)$ with $Q_0 \in \mathbb{O}^{|\alpha|+|\beta|}$ and $Q_k \in \mathbb{O}^{|a_k|}$ for $k = 1, 2, \ldots, r$, and orthogonal matrices $Q' \in \mathbb{O}^{|\gamma_0|}$ and $Q'' \in \mathbb{O}^{|\gamma_0 \cup c|}$ such that

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

Together with equations (25) and (26), it is not difficult to obtain that

$$\Delta W = \overline{U} \begin{bmatrix} \text{Diag}(\overline{w}_{\alpha \cup \beta \cup \gamma_1}) & 0 \\ 0 & Q'\text{Diag}(\overline{w}_{\gamma_0})(Q''_{\gamma_0})^{\mathbb{T}} \end{bmatrix} \overline{V}^{\mathbb{T}},$$

and consequently

$$\Delta \widetilde{W} := \overline{U}^{\mathbb{T}} \Delta W \overline{V} = \begin{bmatrix} \text{Diag}(\overline{w}_{\alpha \cup \beta \cup \gamma_1}) & 0 \\ 0 & Q'\text{Diag}(\overline{w}_{\gamma_0})(Q''_{\gamma_0})^{\mathbb{T}} \end{bmatrix}. \tag{27}$$

Since $(-\Delta\Gamma, -\Delta W) \in \mathcal{N}_{\text{gph}\,\partial\|\cdot\|_*}(\overline{X}, \overline{W})$, by equation (16a)-(16b) of Lemma 2.3, we get

$$\Theta_1 \circ \mathcal{S}(\Delta\widetilde{W}_1) + \Theta_2 \circ \mathcal{S}(\Delta\widetilde{\Gamma}_1) + \Sigma_1 \circ \mathcal{X}(\Delta\widetilde{W}_1) + \Sigma_2 \circ \mathcal{X}(\Delta\widetilde{\Gamma}_1) = 0, \tag{28a}$$

$$(\Delta\widetilde{\Gamma})_{\alpha c} + (\Omega_3)_{\alpha c} \circ [(\Delta\widetilde{W})_{\alpha c} - (\Delta\widetilde{\Gamma})_{\alpha c}] = 0, \quad (\Delta\widetilde{W})_{\beta c} = 0, \quad (\Delta\widetilde{W})_{\gamma c} = 0 \tag{28b}$$

where $\Delta\widetilde{\Gamma}_1 := \overline{U}^{\mathbb{T}} \Delta\Gamma \overline{V}_{\alpha \cup \beta \cup \gamma}, \Delta\widetilde{W}_1 := \overline{U}^{\mathbb{T}} \Delta W \overline{V}_{\alpha \cup \beta \cup \gamma}$, and the matrices $\Theta_1, \Theta_2, \Sigma_1$ and Σ_2 are defined as in Section 2.3. Notice that $[\Delta\widetilde{W}_1]_{\alpha \cup \beta \cup \gamma_1, \alpha \cup \beta \cup \gamma_1}$ is a diagonal matrix by equation (27). Together with (28a)-(28b) and (11a)-(11c), it follows that

$$(\Delta\widetilde{W})_{\alpha c} = 0, \quad (\Delta\widetilde{\Gamma})_{\alpha c} = 0, \quad (\Delta\widetilde{W})_{\gamma\gamma} = 0, \tag{29a}$$

$$[\mathcal{S}(\Delta\widetilde{\Gamma}_1)]_{\alpha\alpha} + (\Sigma_2)_{\alpha\alpha} \circ [\mathcal{X}(\Delta\widetilde{\Gamma}_1)]_{\alpha\alpha} = 0, \tag{29b}$$

$$(\Theta_2)_{\alpha\beta} \circ [\mathcal{S}(\Delta\widetilde{\Gamma}_1)]_{\alpha\beta} + (\Sigma_2)_{\alpha\beta} \circ [\mathcal{X}(\Delta\widetilde{\Gamma}_1)]_{\alpha\beta} = 0, \tag{29c}$$

$$(\Theta_2)_{\beta\alpha} \circ [\mathcal{S}(\Delta\widetilde{\Gamma}_1)]_{\beta\alpha} + (\Sigma_2)_{\beta\alpha} \circ [\mathcal{X}(\Delta\widetilde{\Gamma}_1)]_{\beta\alpha} = 0, \tag{29d}$$

$$(\Theta_2)_{\alpha\gamma} \circ [\mathcal{S}(\Delta\widetilde{\Gamma}_1)]_{\alpha\gamma} + (\Sigma_2)_{\alpha\gamma} \circ [\mathcal{X}(\Delta\widetilde{\Gamma}_1)]_{\alpha\gamma} = 0, \tag{29e}$$

$$(\Theta_2)_{\gamma\alpha} \circ [\mathcal{S}(\Delta\widetilde{\Gamma}_1)]_{\gamma\alpha} + (\Sigma_2)_{\gamma\alpha} \circ [\mathcal{X}(\Delta\widetilde{\Gamma}_1)]_{\gamma\alpha} = 0. \tag{29f}$$

Notice that (29b) is equivalent to $(E + \Sigma_2)_{\alpha\alpha}(\Delta\widetilde{\Gamma}_1)_{\alpha\alpha} + (E - \Sigma_2)_{\alpha\alpha}(\Delta\widetilde{\Gamma}_1^{\mathbb{T}})_{\alpha\alpha} = 0$ which, by the fact that the entries of Σ_2 belongs to $(0,1)$, implies that $(\Delta\widetilde{\Gamma}_1)_{\alpha\alpha} = 0$. Notice that equations (29c) and (29d) can be equivalently written as

$$(\Theta_2 + \Sigma_2)_{\alpha\beta} \circ (\Delta\widetilde{\Gamma}_1)_{\alpha\beta} = (\Sigma_2 - \Theta_2)_{\alpha\beta} \circ (\Delta\widetilde{\Gamma}_1^{\mathbb{T}})_{\alpha\beta}, \tag{30a}$$

$$(\Theta_2 + \Sigma_2)_{\beta\alpha} \circ (\Delta\widetilde{\Gamma}_1)_{\beta\alpha} = (\Sigma_2 - \Theta_2)_{\beta\alpha} \circ (\Delta\widetilde{\Gamma}_1^{\mathbb{T}})_{\beta\alpha}. \tag{30b}$$

Since $[(\Delta\widetilde{\Gamma}_1)_{\beta\alpha}]^{\mathbb{T}} = (\Delta\widetilde{\Gamma}_1^{\mathbb{T}})_{\alpha\beta}$ and $[(\Delta\widetilde{\Gamma}_1^{\mathbb{T}})_{\beta\alpha}]^{\mathbb{T}} = (\Delta\widetilde{\Gamma}_1)_{\alpha\beta}$, by imposing the transpose to the both sides of equality (30b) we immediately obtain that

$$(\Delta\widetilde{\Gamma}_1^{\mathbb{T}})_{\alpha\beta} = [(\Sigma_2 - \Theta_2)_{\alpha\beta} \oslash (\Theta_2 + \Sigma_2)_{\alpha\beta}] \circ (\Delta\widetilde{\Gamma}_1)_{\alpha\beta}$$

where "$\oslash$" denotes the entries division operator of two matrices. Substituting this equality into (30a) yields that $(\Delta\widetilde{\Gamma}_1)_{\alpha\beta} = 0$, and then $(\Delta\widetilde{\Gamma}_1)_{\beta\alpha} = 0$. Similarly, from (29e) and (29f), we can obtain $(\Delta\widetilde{\Gamma}_1)_{\alpha\gamma} = 0$ and $(\Delta\widetilde{\Gamma}_1)_{\beta\gamma} = 0$. Thus,

$$\overline{U}^{\mathbb{T}} \Delta\Gamma \overline{V} = \Delta\widetilde{\Gamma} = \begin{bmatrix} 0_{\alpha\alpha} & 0_{\alpha\beta} & 0_{\alpha\gamma} & 0_{\alpha c} \\ 0_{\beta\alpha} & (\Delta\widetilde{\Gamma})_{\beta\beta} & (\Delta\widetilde{\Gamma})_{\beta\gamma} & (\Delta\widetilde{\Gamma})_{\beta c} \\ 0_{\gamma\alpha} & (\Delta\widetilde{\Gamma})_{\gamma\beta} & (\Delta\widetilde{\Gamma})_{\gamma\gamma} & (\Delta\widetilde{\Gamma})_{\gamma c} \end{bmatrix}.$$

10

Thus, to complete the proof of the first part, we only need to argue that $\mathcal{S}[(\Delta\widetilde{\Gamma})_{\beta\beta}] = 0$. Since $(-\Delta\Gamma, -\Delta W) \in \mathcal{N}_{\mathrm{gph}\,\partial\|\cdot\|_*}(\overline{X}, \overline{W})$, by (16c) there exist $Q \in \mathbb{O}^{|\beta|}$ and $\Xi_1 \in \mathcal{U}_{|\beta|}$ having the form (14) for some partition $(\beta_+, \beta_0, \beta_-)$ of β such that

$$\Xi_1 \circ Q^{\mathbb{T}}(\Delta\widetilde{W})_{\beta\beta}Q + \Xi_2 \circ \mathcal{S}\big[Q^{\mathbb{T}}(\Delta\widetilde{\Gamma})_{\beta\beta}Q\big] + \Xi_2 \circ \mathcal{X}\big[Q^{\mathbb{T}}(\Delta\widetilde{W})_{\beta\beta}Q\big] = 0, \qquad (31)$$

$$Q_{\beta_0}^{\mathbb{T}}(\Delta\widetilde{\Gamma})_{\beta\beta}Q_{\beta_0} \succeq 0, \ Q_{\beta_0}^{\mathbb{T}}(\Delta\widetilde{W})_{\beta\beta}Q_{\beta_0} \preceq 0 \qquad (32)$$

where the matrix Ξ_2 associated with Ξ_1 has the form of (15). From (27) and the first equality in (26), $(\Delta\widetilde{W})_{\beta\beta} = \mathrm{Diag}(\overline{w}_\beta) = \phi'(1)I$. Notice that $\phi'(1) > 0$ by (2). We deduce $\beta_0 = \emptyset$ from the second inequality of (32). Since $\mathcal{X}[Q^{\mathbb{T}}(\Delta\widetilde{W})_{\beta\beta}Q] = 0$, (31) reduces to

$$\Xi_1 \circ (Q^{\mathbb{T}}\mathrm{Diag}(\overline{w}_\beta)Q) + \Xi_2 \circ \mathcal{S}\big[Q^{\mathbb{T}}(\Delta\widetilde{\Gamma})_{\beta\beta}Q\big] = 0.$$

Since $Q^{\mathbb{T}}\mathrm{Diag}(\overline{w}_\beta)Q \succ 0$, by using the expressions of Ξ_1 and Ξ_2 we have $\beta_- = \emptyset$, and then the last equality reduces to $0 = \mathcal{S}\big[Q^{\mathbb{T}}(\Delta\widetilde{\Gamma})_{\beta\beta}Q\big] = \mathcal{S}[(\Delta\widetilde{\Gamma})_{\beta\beta}]$. Thus, we complete the proof of the first part. By combining $(\Delta\widetilde{W})_{\gamma\gamma} = 0$ with (27) and (26), it is easy to see that if $\phi'(t) \neq 0$ for any $t \in (0,1)$, then $\gamma_1 = \emptyset$; and if $0 \notin \partial\widehat{\phi}(0)$, then $\gamma_0 = \emptyset$. $\quad\square$

Now we state the relation between the M-stationary point and the R-stationary point.

Theorem 3.1 *If $\overline{X}$ is an M-stationary point of the problem* (1) *associated to $\phi \in \mathscr{L}_1$, then it is also a R-stationary point. Conversely, if $\overline{X}$ is a R-stationary point of* (1)*, then it is an M-stationary point associated to those $\phi \in \mathscr{L}$ with $0 \in \partial\widehat{\phi}(0)$.*

Proof: Let $\overline{X}$ be an M-stationary point of (1) associated to $\phi \in \mathscr{L}_1$. By Proposition 3.1, there exist $\overline{W} \in \partial\|\cdot\|_*(\overline{X})$ and $\Delta\Gamma \in \nu\partial f(\overline{X}) + \mathcal{N}_\Omega(\overline{X})$ such that for the index sets $\alpha, \beta, c, \gamma, \gamma_1, \gamma_0$ defined as in (10a)-(10b) with $\overline{Z} = \overline{X} + \overline{W}$ and $(\overline{U}, \overline{V}) \in \mathbb{O}^{m,n}(\overline{Z})$, the matrix $\Delta\Gamma$ takes the form of (23). Let $\Delta\widetilde{Z} = \begin{bmatrix} (\Delta\widetilde{\Gamma})_{\beta\beta} & (\Delta\widetilde{\Gamma})_{\beta\gamma} & (\Delta\widetilde{\Gamma})_{\beta c} \\ (\Delta\widetilde{\Gamma})_{\gamma\beta} & (\Delta\widetilde{\Gamma})_{\gamma\gamma} & (\Delta\widetilde{\Gamma})_{\gamma c} \end{bmatrix}$. Take $(P, P') \in \mathbb{O}^{m-|\alpha|, n-|\alpha|}(\Delta\widetilde{Z})$. Write $\widetilde{U} = [\overline{U}_\alpha \ \overline{U}_{\beta\cup\gamma}P]$ and $\widetilde{V} = [\overline{V}_\alpha \ \overline{V}_{\beta\cup\gamma\cup c}P']$. Then,

$$\Delta\Gamma = \widetilde{U}\begin{bmatrix} 0_{\alpha\alpha} & 0_{\alpha,\beta\cup\gamma} & 0_{\alpha c} \\ 0_{\beta\cup\gamma,\alpha} & \mathrm{Diag}(\sigma(\Delta\widetilde{Z})) & 0_{\beta\cup\gamma,c} \end{bmatrix}\widetilde{V}^{\mathbb{T}}.$$

By the definitions of $\widetilde{U}$ and $\widetilde{V}$ and (24a), it is easy to check that $(\widetilde{U}, \widetilde{V}) \in \mathbb{O}^{m,n}(\overline{X})$. Notice that $\mathrm{rank}(\overline{X}) = |\alpha|$. From [17, Theorem 4], we have $-\Delta\Gamma \in \partial\mathrm{rank}(\overline{X})$. Thus, $0 \in \partial\mathrm{rank}(\overline{X}) + \nu\partial f(\overline{X}) + \mathcal{N}_\Omega(\overline{X})$. From Definition 3.1, $\overline{X}$ is a R-stationary point.

Now let $\overline{X}$ be a R-stationary point of (1) with $\mathrm{rank}(\overline{X}) = \overline{r}$. Suppose that $\overline{r} > 1$. Take $\phi \in \mathscr{L}$ with $0 \in \partial\widehat{\phi}(0)$. By Definition 3.1, there is $\Delta\Gamma \in \nu\partial f(\overline{X}) + \mathcal{N}_\Omega(\overline{X})$ such that $-\Delta\Gamma \in \partial\mathrm{rank}(\overline{X})$. Along with [17, Theorem 4], there exists $(\overline{U}, \overline{V}) \in \mathbb{O}^{m,n}(\overline{X})$ such that

$$-\Delta\Gamma = \overline{U}[\mathrm{Diag}(\overline{x}) \ 0]\overline{V}^{\mathbb{T}} \quad \text{with} \quad \overline{x}_i = 0 \ \text{ for } \ i = 1, 2, \ldots, \overline{r}.$$

Next we proceed the arguments by $t^* = 0$ and $t^* \neq 0$, where t^* is same as in (2).

11

Case 1: $t^* = 0$. Take $\overline{W} := \overline{U}_1 \overline{V}_1^{\mathbb{T}}$, where $\overline{U}_1$ and $\overline{V}_1$ are the matrix consisting of the first $\overline{r}$ columns of $\overline{U}$ and $\overline{V}$, respectively. Clearly, $\overline{W} \in \partial \| \cdot \|_*(\overline{X})$ and $(\overline{U}, \overline{V}) \in \mathbb{O}^{m,n}(\overline{Z})$ with $\overline{Z} = \overline{X} + \overline{W}$. Let $\alpha, \beta, c, \gamma_0, \gamma_1$ be defined as before. Clearly, $\beta = \emptyset = \gamma_1$. Take

$$\overline{w}_i = \phi'_-(1) \text{ for } i \in \alpha \text{ and } \overline{w}_i = 0 \in \partial \widehat{\phi}(0) \text{ for } i \in \gamma_0. \tag{33}$$

Since ϕ is convex, from [32, Proposition 10.19(i)] it follows that $\overline{w}_i \in \partial \widehat{\phi}(1)$ for $i \in \alpha$. Then $\Delta W = \overline{U}[\text{Diag}(\overline{w}) \ 0]\overline{V}^{\mathbb{T}} \in \partial(\widehat{\Phi} \circ \sigma)(\overline{W})$. Let $\Delta \widetilde{\Gamma} := \overline{U}^{\mathbb{T}} \Delta \Gamma \overline{V}$ and $\Delta \widetilde{W} := \overline{U}^{\mathbb{T}} \Delta W \overline{V}$. Clearly, $\mathcal{X}(\Delta \widetilde{\Gamma}_1) = \mathcal{X}(\Delta \widetilde{W}_1) = 0$ where $\Delta \widetilde{\Gamma}_1 := \overline{U}^{\mathbb{T}} \Delta \Gamma \overline{V}_1$ and $\Delta \widetilde{W}_1 := \overline{U}^{\mathbb{T}} \Delta W \overline{V}_1$ with $\overline{V}_1$ being the matrix consisting of the first m columns of $\overline{V}$. Together with Θ_2 and Σ_2 defined as in Section 2.3, it is immediate to verify that $(-\Delta \widetilde{\Gamma}, -\Delta \widetilde{W})$ satisfies

$$\Theta_1 \circ \mathcal{S}(\Delta \widetilde{W}_1) + \Theta_2 \circ \mathcal{S}(\Delta \widetilde{\Gamma}_1) + \Sigma_1 \circ \mathcal{X}(\Delta \widetilde{W}_1) + \Sigma_2 \circ \mathcal{X}(\Delta \widetilde{\Gamma}_1) = 0,$$
$$(\Delta \widetilde{\Gamma})_{\alpha c} + (\Omega_3)_{\alpha c} \circ [(\Delta \widetilde{W})_{\alpha c} - (\Delta \widetilde{\Gamma})_{\alpha c}] = 0, \ (\Delta \widetilde{W})_{\beta c} = 0, \ (\Delta \widetilde{W})_{\gamma c} = 0.$$

Since $\beta = \emptyset$, from Lemma 2.3 it follows that $(-\Delta \Gamma, -\Delta W) \in \mathcal{N}_{\text{gph} \partial \| \cdot \|_*}(\overline{X}, \overline{W})$, i.e., $-\Delta \Gamma \in D^* \partial \| \cdot \|_*(\overline{X}|\overline{W})(\Delta W)$. By Definition 3.2, $\overline{X}$ is M-stationary associated to ϕ.

Case 2: $t^* \neq 0$. Now $t^* \in (0, 1)$. Take $\overline{W} := \overline{U}_1 \overline{V}_1^{\mathbb{T}} + t^* \overline{U}_2 \overline{V}_2^{\mathbb{T}}$, where $\overline{U}_2$ and $\overline{V}_2$ are the matrix consisting of the last $m - \overline{r}$ and $n - \overline{r}$ columns of $\overline{U}$ and $\overline{V}$, respectively. Clearly, $\overline{W} \in \partial \| \cdot \|_*(\overline{X})$ and $(\overline{U}, \overline{V}) \in \mathbb{O}^{m,n}(\overline{Z})$ with $\overline{Z} = \overline{X} + \overline{W}$. Let α, β, c and $\gamma = \gamma_0 \cup \gamma_1$ be defined as before. Then $\beta = \emptyset$ and $\gamma_0 = \emptyset$. Let $\Delta W = \overline{U}[\text{Diag}(\overline{w}) \ 0]\overline{V}^{\mathbb{T}}$ with

$$\overline{w}_i = \phi'_-(1) \text{ for } i \in \alpha \text{ and } \overline{w}_i = 0 \in \partial \phi(t^*) \text{ for } i \in \gamma_1. \tag{35}$$

Using the same arguments as those for Case 1 can prove that $\overline{X}$ is M-stationary.

When $\overline{r} = 0$, choose $\overline{W} = 0$. Clearly, $\overline{W} \in \partial \| \cdot \|_*(\overline{X})$ since $\overline{X} = 0$. Write $\overline{Z} = \overline{X} + \overline{W}$. Then, $\alpha = \beta = \emptyset = \gamma_1$. Take $\Delta W = 0$. Since $0 \in \partial \widehat{\phi}(0)$, we have $\Delta W \in \partial(\widehat{\Phi} \circ \sigma)(\overline{W})$. Moreover, by Lemma 2.3 it is easy to check that $D^* \partial \| \cdot \|_*(\overline{X}|\overline{W})(\Delta W) = \mathbb{R}^{m \times n}$. Thus, $\overline{X}$ is M-stationary associated to ϕ. The proof is then completed. $\square$

To close this subsection, we provide a condition for a local minimizer of the MPEC (4) associated with $\phi \in \mathscr{L}$ to be an M-stationary point associated to ϕ.

Proposition 3.2 *Let $(\overline{W}, \overline{X})$ be a local minimizer of the MPEC (4) associated to $\phi \in \mathscr{L}$. Then $\overline{X}$ is an M-stationary point of the problem (1) associated to ϕ, provided that*

$$\mathcal{N}_{\text{gph} \mathcal{N}_{\mathbb{B}} \cap (\mathbb{R}^{m \times n} \times \Omega)}(\overline{W}, \overline{X}) \subseteq \mathcal{N}_{\text{gph} \mathcal{N}_{\mathbb{B}}}(\overline{W}, \overline{X}) + \{0\} \times \mathcal{N}_{\Omega}(\overline{X}) \tag{36}$$

where $\mathbb{B} := \{Z \in \mathbb{R}^{m \times n} \mid \|Z\| \leq 1\}$, and if in addition $\phi \in \mathscr{L}_1$, $\overline{X}$ is a R-stationary point.

Proof: By invoking the relation (18a), (W, X) is a feasible point of (4) if and only if $(W, X) \in \text{gph} \mathcal{N}_{\mathbb{B}} \cap (\mathbb{R}^{m \times n} \times \Omega)$. This implies that (4) can be compactly written as

$$\min_{X, W \in \mathbb{R}^{m \times n}} \left\{ \nu f(X) + \widehat{\Phi}(\sigma(W)) + \delta_{\text{gph} \mathcal{N}_{\mathbb{B}} \cap (\mathbb{R}^{m \times n} \times \Omega)}(W, X) \right\}.$$

From the local optimality of $(\overline{W}, \overline{X})$, the assumption on f, the Lipschitz continuity of $\widehat{\Phi} \circ \sigma$ over the ball $\mathbb{B}$, and [32, Theorem 10.1 & Exercise 10.10], it follows that

$$0 \in \partial \widetilde{f}(\overline{W}, \overline{X}) + \partial(\widehat{\Phi} \circ \sigma)(\overline{W}) \times \{0\} + \mathcal{N}_{\mathrm{gph}\,\mathcal{N}_{\mathbb{B}} \cap (\mathbb{R}^{m \times n} \times \Omega)}(\overline{W}, \overline{X})$$

where $\widetilde{f}(W, X) \equiv \nu f(X)$. Together with the inclusion (36) and [32, Exercise 10.10],

$$0 \in \{0\} \times \nu \partial f(\overline{X}) + \mathcal{N}_{\mathrm{gph}\,\mathcal{N}_{\mathbb{B}}}(\overline{W}, \overline{X}) + \{0\} \times \mathcal{N}_{\Omega}(\overline{X}) + \partial(\widehat{\Phi} \circ \sigma)(\overline{W}) \times \{0\} \qquad (37)$$

which is equivalent to saying that there exists $(-\Delta W, \Delta X) \in \mathcal{N}_{\mathrm{gph}\,\mathcal{N}_{\mathbb{B}}}(\overline{W}, \overline{X})$ such that

$$\begin{cases} 0 \in \partial(\widehat{\Phi} \circ \sigma)(\overline{W}) - \Delta W \\ 0 \in \nu \partial f(\overline{X}) + \Delta X + \mathcal{N}_{\Omega}(\overline{X}). \end{cases}$$

Notice that $(-\Delta W, \Delta X) \in \mathcal{N}_{\mathrm{gph}\,\mathcal{N}_{\mathbb{B}}}(\overline{W}, \overline{X})$ if and only if $\Delta X \in D^* \partial \| \cdot \|_*(\overline{X} | \overline{W})(\Delta W)$. So, equation (37) is equivalent to saying that there exists $\Delta W \in \partial(\widehat{\Phi} \circ \sigma)(\overline{W})$ such that

$$0 \in \nu \partial f(\overline{X}) + \mathcal{N}_{\Omega}(\overline{X}) + D^* \partial \| \cdot \|_*(\overline{X} | \overline{W})(\Delta W).$$

In addition, notice that $\overline{X} \in \mathcal{N}_{\mathbb{B}}(\overline{W})$ which is equivalent to $\partial \| \cdot \|_*(\overline{X})$ by (18b). Thus, by Definition 3.2, $\overline{X}$ is an M-stationary point of the problem (1) associated to ϕ. The second part is a direct consequence of Theorem 3.1. The proof is completed. $\qquad \square$

Remark 3.3 (i) *If $\Omega = \mathbb{R}^{m \times n}$, the inclusion (36) automatically holds. If $\Omega \subset \mathbb{R}^{m \times n}$, by [13, Page 211] the inclusion (36) is implied by the calmness of the following multifunction*

$$\mathcal{M}(Y_1, Y_2) := \Big\{ (W, X) \in \mathbb{R}^{m \times n} \times \Omega \mid (W, X) \in \mathrm{gph}\,\mathcal{N}_{\mathbb{B}} - (Y_1, Y_2) \Big\}. \qquad (38)$$

at the origin for $(\overline{W}, \overline{X})$, where $(\overline{W}, \overline{X})$ is an arbitrary feasible point of the MPEC (4).

(ii) *When $\Omega \subseteq \mathbb{S}_+^n$, together with (21) and the Lipschitz continuity of $\widehat{\Phi} \circ \sigma$ in $\mathbb{B}$, in order to achieve the conclusion of Proposition 3.2, we need to replace the inclusion (36) by*

$$\mathcal{N}_C(\overline{W}, \overline{X}) \subseteq \mathcal{N}_{\mathrm{gph}\,\mathcal{N}_{\mathbb{S}_+^n}}(\overline{X}, \overline{W} - I) + \mathcal{N}_{\mathbb{S}_+^n}(\overline{W}) \times \mathcal{N}_{\Omega}(\overline{X})$$

where $C := \{ (W, X) \in \mathbb{S}_+^n \times \Omega \mid (X, W - I) \in \mathrm{gph}\,\mathcal{N}_{\mathbb{S}_+^n} \}$. By invoking [13, Page 211], this inclusion is implied by the calmness of the following multifunction

$$\mathcal{M}(Y_1, Y_2) := \Big\{ (W, X) \in \mathbb{S}_+^n \times \Omega \mid (X, W - I) \in \mathrm{gph}\,\mathcal{N}_{\mathbb{S}_+^n} - (Y_1, Y_2) \Big\} \qquad (39)$$

at the origin for $(\overline{W}, \overline{X})$, where $(\overline{W}, \overline{X})$ is an arbitrary feasible point of the MPEC (21). By the definition of calmness, it is easy to check that the calmness of $\mathcal{M}$ at the origin for $(\overline{W}, \overline{X})$ is implied by that of $\widetilde{\mathcal{M}}$ in (48) with $\Omega_x = \Omega$ and $\Omega_y = \mathbb{S}_+^n$ at the corresponding point, while by Theorem 4.2 the latter holds if for any $0 \neq H = (H_1; H_2) \in \mathcal{T}_{\Omega}(\overline{X}) \times \mathcal{T}_{\mathbb{S}_+^n}(\overline{W})$ such that $(H_1, H_2) \in \mathcal{T}_{\mathrm{gph}\,\mathcal{N}_{\mathbb{S}_+^n}}(\overline{X}, \overline{W} - I)$, the following implication relation holds:

$$\left. \begin{array}{l} \Gamma_1 \in -\mathcal{N}_{\Omega}(\overline{X}; H_1), \quad \Gamma_2 \in -\mathcal{N}_{\mathbb{S}_+^n}(\overline{W}; H_2) \\ (\Gamma_1, \Gamma_2) \in \mathcal{N}_{\mathrm{gph}\,\mathcal{N}_{\mathbb{S}_+^n}}((\overline{X}, \overline{W} - I); (H_1, H_2)) \end{array} \right\} \implies \begin{pmatrix} \Gamma_1 \\ \Gamma_2 \end{pmatrix} = 0. \qquad (40)$$

For the characterization of $\mathcal{N}_{\mathrm{gph}\,\mathcal{N}_{\mathbb{S}_+^n}}((\overline{X}, \overline{W} - I); (H_1, H_2))$, please refer to Appendix.

3.3 EP-stationary points

By the definition of the function $\widehat{\Psi}$, clearly, the problem (5) can be compactly written as

$$\min_{X, W \in \mathbb{R}^{m \times n}} \left\{ \nu f(X) + \delta_\Omega(X) + \widehat{\Psi}(\sigma(W)) + \rho(\|X\|_* - \langle W, X \rangle) \right\}. \tag{41}$$

Based on this equivalent reformulation, we introduce the following stationary point.

Definition 3.3 *A matrix $\overline{X} \in \mathbb{R}^{m \times n}$ is said to be an EP-stationary point of the problem (1) associated to $\phi \in \mathscr{L}$ if there exist a constant $\rho > 0$ and $\overline{W} \in \mathbb{B}$ such that*

$$\rho \overline{X} \in \partial(\widehat{\Psi} \circ \sigma)(\overline{W}) \quad \text{and} \quad 0 \in \nu \partial f(\overline{X}) + \mathcal{N}_\Omega(\overline{X}) + \rho \big[\partial \| \cdot \|_*(\overline{X}) - \overline{W} \big]. \tag{42}$$

Remark 3.4 *By the given assumption on f and the Lipschitz continuity of $\widehat{\Phi} \circ \sigma$ in $\mathbb{B}$, if $(\overline{X}, \overline{W})$ is a limiting critical point of the objective function of (41), it is an EP-stationary point of (1). Thus, every local optimal solution of (5) is an EP-stationary point of (1).*

The following proposition characterizes a key property of the EP-stationary point.

Proposition 3.3 *Suppose that $\overline{X} \in \mathbb{R}^{m \times n}$ is an EP-stationary point of (1) associated to $\phi \in \mathscr{L}$. Then, there exist $\overline{W} \in \mathbb{B}$ and $(\overline{U}, \overline{V}) \in \mathbb{O}^{m,n}(\overline{W}) \cap \mathbb{O}^{m,n}(\overline{X})$ such that $\operatorname{rank}(\overline{X}) \geq |\{i \mid \sigma_i(\overline{W}) > t^*\}|$, and there exists $\Delta\Gamma \in \nu \partial f(\overline{X}) + \mathcal{N}_\Omega(\overline{X})$ such that*

$$\Delta\Gamma \in \left\{ \overline{U} \begin{bmatrix} 0 & 0 \\ 0 & \rho Z \end{bmatrix} \overline{V}^{\mathbb{T}} \mid Z \in \mathbb{R}^{(m-|\theta_1|) \times (n-|\theta_1|)} \text{ with } \theta_1 = \{i \mid \sigma_i(\overline{W}) = 1\}, \|Z\| < 1 \right\}. \tag{43}$$

Proof: Since $\overline{X}$ is an EP-stationary point of the problem (1), there exist a constant $\rho > 0$ and a matrix $\overline{W} \in \mathbb{B}$ such that the inclusions in (42) hold. Define the index sets

$$\theta_2 := \big\{ i \mid \sigma_i(\overline{W}) \in (0,1) \big\} \quad \text{and} \quad \theta_0 := \big\{ i \mid \sigma_i(\overline{W}) = 0 \big\}.$$

Since $\rho \overline{X} \in \partial(\widehat{\Psi} \circ \sigma)(\overline{W})$, by [20, Corollary 2.5] there exists $(\overline{U}, \overline{V}) \in \mathbb{O}^{m,n}(\overline{W})$ such that

$$\rho \overline{X} = \overline{U} \big[\operatorname{Diag}(\sigma(\overline{X})) \ \ 0 \big] \overline{V}^{\mathbb{T}} \quad \text{and} \quad \sigma(\overline{X}) \in \partial \widehat{\Phi}(\sigma(\overline{W})).$$

Notice that $\sigma_1(X) \geq \cdots \geq \sigma_m(X)$ with $\sigma_i(X) \in \partial\psi(1)$ for $i \in \theta_1$, $\sigma_i(X) \in \partial\psi(\sigma_i(\overline{W}))$ for $i \in \theta_2$ and $\sigma_i(X) \in \partial\widehat{\psi}(0)$ for $i \in \theta_0$. Since $\partial\psi(t) \subset (0, +\infty)$ for any $t > t^*$, we have $\operatorname{rank}(\overline{X}) \geq |\{i \mid \sigma_i(\overline{W}) > t^*\}|$, and the first part follows. Since $(\overline{X}, \overline{W})$ satisfies the second inclusion of (42), there exist $\Delta\Gamma \in \nu \partial f(\overline{X}) + \mathcal{N}_\Omega(\overline{X})$ such that $-\Delta\Gamma \in \rho[\partial\| \cdot \|_*(\overline{X}) - \overline{W}]$. Write $\overline{r} = \operatorname{rank}(\overline{X})$. From the SVD of $\overline{X}$ in the last equation and equation (9), we have

$$\partial\| \cdot \|_*(\overline{X}) = \left\{ \overline{U}_1 \overline{V}_1^{\mathbb{T}} + \overline{U}_2 \Gamma \overline{V}_2^{\mathbb{T}} \mid \|\Gamma\| \leq 1, \Gamma \in \mathbb{R}^{(m-\overline{r}) \times (n-\overline{r})} \right\},$$

where $\overline{U}_1$ and $\overline{V}_1$ are the matrix consisting of the first $\overline{r}$ columns of $\overline{U}$ and $\overline{V}$, respectively, and $\overline{U}_2$ and $\overline{V}_2$ are the matrices consisting of the last $m - \overline{r}$ and $n - \overline{r}$ columns of $\overline{U}$ and

$\overline{V}$, respectively. Together with $-\Delta\Gamma \in \rho[\partial\|\cdot\|_*(\overline{X}) - \overline{W}]$ and $\overline{W} = \overline{U}[\text{Diag}(\sigma(\overline{W}))\ \ 0]\overline{V}^{\mathbb{T}}$, the inclusion in (43) holds. In fact, the matrix Z in the set of (43) has the following form

$$\begin{bmatrix} \text{Diag}(e_{\overline{r}-|\theta_1|}) & 0 \\ 0 & \Gamma \end{bmatrix} - \begin{bmatrix} \text{Diag}(\sigma_{\theta_2'}(\overline{W})) & 0 \\ 0 & \text{Diag}(\sigma_{\theta_0}(\overline{W})) \end{bmatrix}$$

for some $\Gamma \in \mathbb{R}^{(m-\overline{r})\times(n-\overline{r})}$ with $\|\Gamma\| \leq 1$, where $\theta_2' := \{i \in \theta_2 \mid \sigma_i(\overline{W}) \leq t^*\}$. $\square$

Remark 3.5 *If $\overline{X}$ is an EP-stationary point of* (1) *and the associated $\overline{W} \in \mathbb{B}$ is such that $|\theta_1| = \text{rank}(\overline{X})$, then by Definition 3.1 and [17, Theorem 4] $\overline{X}$ is a R-stationary point of* (1). *However, when $\overline{X}$ is a R-stationary point, it is not necessarily EP-stationary.*

3.4 DC-stationary point

With the conjugate $\widehat{\Psi}^*$ of $\widehat{\Psi}$, the surrogate problem (6) can be equivalently written as

$$\min_{X \in \mathbb{R}^{m\times n}} \left\{ \nu f(X) + \delta_\Omega(X) + \rho\|X\|_* - \widehat{\Psi}^*(\rho\sigma(X)) \right\}. \tag{44}$$

By [20, Lemma 2.3], we know that $\widehat{\Psi}^*$ is also absolutely symmetric. Along with its lsc and convexity, from [20, Corollary 2.6] it follows that $\widehat{\Psi}^* \circ \sigma$ is an absolutely symmetric convex function on $\mathbb{R}^m$. Thus, $\delta_\Omega(X) + \rho\|X\|_* - (\widehat{\Psi}^* \circ \sigma)(\rho X)$ is a DC function on $\mathbb{R}^{m\times n}$. In view of this, we present the following DC-stationary point by the reformulation (44).

Definition 3.4 *A matrix $\overline{X} \in \mathbb{R}^{m\times n}$ is called a DC-stationary point of the problem* (1) *associated to $\phi \in \mathscr{L}$ if there exists a constant $\rho > 0$ such that*

$$0 \in \nu\partial f(\overline{X}) + \mathcal{N}_\Omega(\overline{X}) + \rho\partial\|\cdot\|_*(\overline{X}) - \rho\partial(\widehat{\Psi}^* \circ \sigma)(\rho\overline{X}). \tag{45}$$

When f is convex, the problem (44) is a DC program, and now $\overline{X} \in \mathbb{R}^{m\times n}$ is a DC-stationary point if and only if it is a critical point of the objective function of (44) defined by Pang et al. [29]. It is worthwhile to point out that the limiting critical point of the objective function of (44) is a DC-stationary point, but the converse does not hold. For the discussion on the DC-stationary point, the reader may refer to [29]. Here, we focus on the relation between the DC-stationary point and the EP-stationary point.

Theorem 3.2 *Let $\overline{X}$ be a DC-stationary point of* (1) *associated to $\phi \in \mathscr{L}$. Suppose that*

$$\partial\psi(0) = \partial\widehat{\psi}(0) \text{ and } \widehat{\psi}^* \text{ is differentiable on } \mathbb{R}_+ \text{ with } (\psi^*)'(0) = (\widehat{\psi}^*)'(0). \tag{46}$$

Then $\overline{X}$ is an EP-stationary point. Conversely, if $\overline{X}$ is an EP-stationary point associated to $\phi \in \Phi$ with ϕ nondecreasing on $[0, 1]$, then $\overline{X}$ is necessarily a DC-stationary point.

Proof: From the symmetry of $\widehat{\psi}$, it follows that $\widehat{\psi}^*(s) = \psi^*(|s|)$ for any $s \in \mathbb{R}$. Together with the given assumption, we have $(\widehat{\psi}^*)'(s) = (\psi^*)'(s)$ for any $s \geq 0$. By the differentiability of $\widehat{\psi}^*$ on $\mathbb{R}_+$, clearly, $\widehat{\Psi}^*$ is differentiable on $\mathbb{R}_+^m$. Along with its absolute symmetry and convexity, from [20, Theorem 3.1] it follows that $\widehat{\Psi}^* \circ \sigma$ is differentiable in $\mathbb{R}^{m\times n}$,

15

and consequently $\partial(\widehat{\Psi}^* \circ \sigma)(\rho\overline{X}) = \{\nabla(\widehat{\Psi}^* \circ \sigma)(\rho\overline{X})\}$. Since $\overline{X}$ is a DC-stationary point of (1), there exists a constant $\rho > 0$ such that (45) holds. Take $(\overline{U}, \overline{V}) \in \mathbb{O}^{m,n}(\overline{X})$. Let

$$\overline{W} := \overline{U}[\text{Diag}(\overline{w}) \ \ 0]\overline{V}^{\mathbb{T}} \ \ \text{with} \ \ \overline{w}_i = (\psi^*)'(\rho\sigma_i(\overline{X})) \ \ \text{for} \ i = 1, 2, \ldots, m.$$

Since ψ is a closed proper convex function, we have range $\partial\psi^* \subseteq \text{dom}\psi = [0, 1]$ by [31, Section 23], which implies that $\overline{w}_i \in [0, 1]$ for $i = 1, \ldots, m$ and consequently $\|\overline{W}\| \le 1$. Combining $\overline{w}_i = (\psi^*)'(\rho\sigma_i(\overline{X}))$ with [31, Corollary 23.5.1], we obtain

$$\rho\sigma_i(\overline{X}) \in \partial\psi(\overline{w}_i) \subseteq \partial\widehat{\psi}(\overline{w}_i) \ \ \text{for} \ \ i = 1, 2, \ldots, m,$$

where the second inclusion is due to $\overline{w}_i \in [0, 1]$ and $\partial\psi(0) = \partial\widehat{\psi}(0)$. By the definition of $\widehat{\Psi}$, it is not hard to obtain $\rho\overline{X} \in \partial(\widehat{\Psi}\circ\sigma)(\overline{W})$. Thus, by Definition 3.3 and (45), to achieve the first part we only need to argue that $\overline{W} = \nabla(\widehat{\Psi}^*\circ\sigma)(\rho\overline{X})$. Recall that $\overline{w}_i \in (\psi^*)'(\rho\sigma_i(\overline{X}))$ for each i and $(\psi^*)'(s) = (\widehat{\psi}^*)'(s)$ for all $s \ge 0$, we have $\overline{w}_i = (\widehat{\psi}^*)'(\rho\sigma_i(\overline{X}))$ for each i. This along with the expression of $\widehat{\Psi}^*$ means that $\overline{W} = \nabla(\widehat{\Psi}^* \circ \sigma)(\rho\overline{X})$.

Now suppose $\overline{X}$ is a EP-stationary point associated to $\phi \in \mathscr{L}$ with ϕ nondecreasing on $[0, 1]$. Then, there exist $\rho > 0$ and $\overline{W} \in \mathbb{B}$ such that the inclusions in (42) hold. Notice that ψ is nondecreasing and convex. Hence, $\widehat{\psi}$ is convex. Together with its absolute symmetry and convexity, it follows that $\widehat{\Psi}$ is absolutely symmetric and convex. From [20, Corollary 2.5] it follows that $\widehat{\Psi} \circ \sigma$ is convex over $\mathbb{R}^{m \times n}$. From $\rho\overline{X} \in \partial(\widehat{\Psi} \circ \sigma)(\overline{W})$, we get $\overline{W} \in \partial(\widehat{\Psi} \circ \sigma)^*(\rho\overline{X})$. By the von Neumman trace inequality, it is easy to check that $(\widehat{\Psi} \circ \sigma)^* = \widehat{\Psi}^* \circ \sigma$, and then $\overline{W} \in \partial(\widehat{\Psi}^* \circ \sigma)(\rho\overline{X})$. Together with the second inclusion in (42) and Definition 3.4, we conclude that $\overline{X}$ is a DC-stationary point of (1). $\qquad\square$

To sum up the previous discussions, we obtain the relations as shown in Figure 1, where θ_1 is the index set defined as in (43) and $\mathscr{L}_2$ denotes the family of those $\phi \in \mathscr{L}$ that is nondecreasing on $[0, 1]$. We see that the set of R-stationary points is almost same as that of M-stationary points and includes that of EP-stationary points under the rank condition $|\theta_1| = \text{rank}(\overline{X})$, while for some ϕ the set of EP-stationary points coincides with that of DC-stationary points, for example, the following special ϕ.

Example 3.1 *Let* $\phi(t) = \frac{a-1}{a+1}t^2 + \frac{2}{a+1}t \ (a > 1)$ *for* $t \in \mathbb{R}$*. Clearly,* $\phi \in \mathscr{L}_1 \cap \mathscr{L}_2$*. Also,*

$$\psi(t) = \left\{ \begin{array}{ll} \frac{a-1}{a+1}t^2 + \frac{2}{a+1}t & \text{if } 0 \le t \le 1 \\ \infty & \text{otherwise} \end{array} \right. \ \ \text{and} \ \ \widehat{\psi}(t) = \left\{ \begin{array}{ll} \frac{a-1}{a+1}t^2 + \frac{2}{a+1}t & \text{if } 0 \le t \le 1, \\ \frac{a-1}{a+1}t^2 - \frac{2}{a+1}t & \text{if } -1 \le t \le 0, \\ \infty & \text{otherwise.} \end{array} \right.$$

After an elementary calculation, the conjugate ψ^* *and* $\widehat{\psi}^*$ *of* ψ *and* $\widehat{\psi}$ *take the form of*

$$\psi^*(\omega) = \left\{ \begin{array}{ll} 0 & \text{if } \omega \le \frac{2}{a+1}, \\ \frac{((a+1)\omega-2)^2}{4(a^2-1)} & \text{if } \frac{2}{a+1} < \omega \le \frac{2a}{a+1}, \\ \omega - 1 & \text{if } \omega > \frac{2a}{a+1}. \end{array} \right. \ \ \text{and} \ \ \widehat{\psi}^*(\omega) = \psi^*(|\omega|).$$

It is easy to check that ϕ *satisfies the conditions in* (46) *and is nondecreasing in* $[0, 1]$*.*
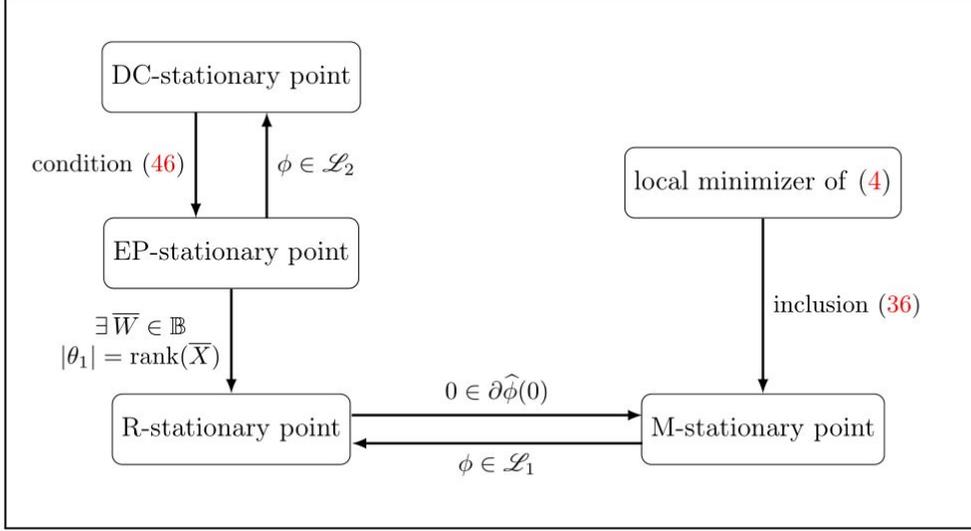
Figure 1: Relations among the four classes of stationary points

4 M-stationary point of MPSCCC

In Section 3.2, the MPEC (4) is the key to characterize the M-stationary point of (1). When $\Omega \subseteq \mathbb{S}^n_+$, it corresponds to (21) which is a special case of the following MPSCCC

$$
\min_{x \in \Omega_x, y \in \Omega_y} \varphi(x, y)
$$
$$
\text{s.t. } (f(x,y), g(x,y)) \in \text{gph}\mathcal{N}_{\mathbb{S}^n_+}, \tag{47}
$$

where $\Omega_x \subseteq \mathbb{X}$ and $\Omega_y \subseteq \mathbb{Y}$ are the closed sets, $\varphi \colon \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$ and $f, g \colon \mathbb{X} \times \mathbb{Y} \to \mathbb{S}^n$ are smooth functions. For this class of problems, since the Robinson CQ does not hold, it is common to seek an M-stationary point which is weaker than the classical KKT point (also called the strong stationary point). In this section, we shall provide a weaker condition for a local minimizer of (47) to be the M-stationary point. For this purpose, we need the multifunction $\widetilde{\mathcal{M}} \colon \mathbb{X} \times \mathbb{Y} \times \mathbb{S}^n \times \mathbb{S}^n \rightrightarrows \mathbb{X} \times \mathbb{Y}$ defined as follows:

$$
\widetilde{\mathcal{M}}(u, v, \xi, \eta) := \Big\{ (x, y) \in \mathbb{X} \times \mathbb{Y} \mid u \in -x + \Omega_x, v \in -y + \Omega_y,
$$
$$
(\xi, \eta) \in -(f(x,y), g(x,y)) + \text{gph}\mathcal{N}_{\mathbb{S}^n_+} \Big\}. \tag{48}
$$

By [5, Proposition 2.1 and Theorem 2.1], it is immediate to have the following result.

Theorem 4.1 *Let $(\overline{x}, \overline{y})$ be a local minimizer of (47). If the perturbed mapping $\widetilde{\mathcal{M}}$ is calm at the origin for $(\overline{x}, \overline{y})$, then $(\overline{x}, \overline{y})$ is an M-stationary point of the problem (47).*

By [11, Corollary 1], one may achieve the calmness of $\widetilde{M}$ at the origin for $(\overline{x}, \overline{y})$ by the directional limiting normal cone to $\text{gph}\mathcal{N}_{\mathbb{S}^n_+}$. That is, the following result holds.

Theorem 4.2 *Consider an arbitrary $(\overline{x}, \overline{y}) \in \widetilde{\mathcal{M}}(0,0,0,0)$. If for any $0 \neq w = (w_1; w_2) \in \mathcal{T}_{\Omega_x}(\overline{x}) \times \mathcal{T}_{\Omega_y}(\overline{y})$ such that $\begin{pmatrix} f'(\overline{x}, \overline{y}) \\ g'(\overline{x}, \overline{y}) \end{pmatrix} w \in \mathcal{T}_{\mathrm{gph}\mathcal{N}_{\mathbb{S}_+^n}}(f(\overline{x}, \overline{y}), g(\overline{x}, \overline{y}))$, the implication holds:*

$$
\left.
\begin{aligned}
d_1 \in \mathcal{N}_{\Omega_x}(\overline{x}; w_1), \ d_1 + \nabla_x f(\overline{x}, \overline{y})\Lambda + \nabla_x g(\overline{x}, \overline{y})\Delta = 0 \\
d_2 \in \mathcal{N}_{\Omega_y}(\overline{y}; w_2), \ d_2 + \nabla_y f(\overline{x}, \overline{y})\Lambda + \nabla_y g(\overline{x}, \overline{y})\Delta = 0 \\
\widetilde{w} = (f'_x(\overline{x}, \overline{y})w_1 + f'_y(\overline{x}, \overline{y})w_2, g'_x(\overline{x}, \overline{y})w_1 + g'_y(\overline{x}, \overline{y})w_2) \\
(\Lambda, \Delta) \in \mathcal{N}_{\mathrm{gph}\mathcal{N}_{\mathbb{S}_+^n}}((f(\overline{x}, \overline{y}), g(\overline{x}, \overline{y})); \widetilde{w})
\end{aligned}
\right\}
\implies
\begin{pmatrix} d_1 \\ d_2 \\ \Lambda \\ \Delta \end{pmatrix} = 0, \qquad (49)
$$

then the multifunction $\widetilde{M}$ is calm at the origin for $(\overline{x}, \overline{y})$.

Remark 4.1 **(i)** *Notice that $(\Delta x, \Delta y) \in D\widetilde{\mathcal{M}}((0,0,0,0)|(\overline{x}, \overline{y}))(\Delta u, \Delta v, \Delta \xi, \Delta \eta)$ iff*

$$
\begin{cases}
\Delta \omega_1 := f'_x(\overline{x}, \overline{y})\Delta x + f'_y(\overline{x}, \overline{y})\Delta y + \Delta \xi, & (50a) \\
\Delta \omega_2 := g'_x(\overline{x}, \overline{y})\Delta x + g'_y(\overline{x}, \overline{y})\Delta y + \Delta \eta, & (50b) \\
\Delta u + \Delta x \in \mathcal{T}_{\Omega_x}(\overline{x}), \ \Delta v + \Delta y \in \mathcal{T}_{\Omega_y}(\overline{y}), & (50c) \\
\Delta \omega_2 \in D\mathcal{N}_{\mathbb{S}_+^n}(f(\overline{x}, \overline{y})|g(\overline{x}, \overline{y}))(\Delta \omega_1). & (50d)
\end{cases}
$$

Together with Lemma 2.2, there is no nonzero $w = (w_1; w_2) \in (\mathcal{T}_{\Omega_x}(\overline{x}) \times \mathcal{T}_{\Omega_y}(\overline{y}))$ such that $\begin{pmatrix} f'(\overline{x}, \overline{y}) \\ g'(\overline{x}, \overline{y}) \end{pmatrix} w \in \mathcal{T}_{\mathrm{gph}\mathcal{N}_{\mathbb{S}_+^n}}(f(\overline{x}, \overline{y}), g(\overline{x}, \overline{y}))$ if and only if $\widetilde{\mathcal{M}}$ is isolated calmness at the origin for $(\overline{x}, \overline{y})$. Thus, Theorem 4.2 is stating that if $\widetilde{\mathcal{M}}$ is not isolated calm and the implication in (49) holds, then $\widetilde{\mathcal{M}}$ is necessarily calm at the origin for $(\overline{x}, \overline{y})$.

(ii) *Notice that $(\Delta u, \Delta v, \Delta \xi, \Delta \eta) \in D^*\widetilde{\mathcal{M}}((\overline{x}, \overline{y})|(0,0,0,0))(-\Delta x, -\Delta y)$ if and only if*

$$
\begin{cases}
\Delta x \in \nabla_x f(\overline{x}, \overline{y})\Delta \xi + \nabla_x g(\overline{x}, \overline{y})\Delta \eta + \Delta u, \ \Delta u \in \mathcal{N}_{\Omega_x}(\overline{x}), & (51a) \\
\Delta y \in \nabla_y f(\overline{x}, \overline{y})\Delta \xi + \nabla_y g(\overline{x}, \overline{y})\Delta \eta + \Delta v, \ \Delta v \in \mathcal{N}_{\Omega_y}(\overline{y}), & (51b) \\
\Delta \xi \in D^*\mathcal{N}_{\mathbb{S}_+^n}(f(\overline{x}, \overline{y})|g(\overline{x}, \overline{y}))(-\Delta \eta). & (51c)
\end{cases}
$$

Together with Lemma 2.1, the Aubin property of $\widetilde{\mathcal{M}}$ is equivalent to the implication

$$
\left.
\begin{aligned}
0 \in \nabla_x f(\overline{x}, \overline{y})\Delta \xi + \nabla_x g(\overline{x}, \overline{y})\Delta \eta + \Delta u \\
0 \in \nabla_y f(\overline{x}, \overline{y})\Delta \xi + \nabla_y g(\overline{x}, \overline{y})\Delta \eta + \Delta v \\
\Delta u \in \mathcal{N}_{\Omega_x}(\overline{x}), \Delta v \in \mathcal{N}_{\Omega_y}(\overline{y}) \\
(\Delta \xi, \Delta \eta) \in \mathcal{N}_{\mathrm{gph}\mathcal{N}_{\mathbb{S}_+^n}}(f(\overline{x}, \overline{y}), g(\overline{x}, \overline{y}))
\end{aligned}
\right\}
\implies (\Delta u, \Delta v, \Delta \xi, \Delta \eta) = 0. \qquad (52)
$$

Since $\mathcal{N}_{\mathrm{gph}\mathcal{N}_{\mathbb{S}_+^n}}((f(\overline{x}, \overline{y}), g(\overline{x}, \overline{y})); (d_1, d_2)) \subset \mathcal{N}_{\mathrm{gph}\mathcal{N}_{\mathbb{S}_+^n}}(f(\overline{x}, \overline{y}), g(\overline{x}, \overline{y}))$ for $(d_1, d_2) \in \mathbb{S}^n \times \mathbb{S}^n$, the implication in (49) is weaker than the one in (52) which is precisely the M-stationary point condition given in [5, Theorem 6.1(i)]. For the characterization of the directional normal cone $\mathcal{N}_{\mathrm{gph}\mathcal{N}_{\mathbb{S}_+^n}}((f(\overline{x}, \overline{y}), g(\overline{x}, \overline{y})); (d_1, d_2))$, the reader refers to Appendix.

To close this section, we illustrate Theorem 4.2 by the following special example

$$\min_{x\in\mathbb{R}^3, y\in\mathbb{R}^3} \|x\|^2 + \|y\|^3$$
$$\text{s.t.} \quad (\mathcal{A}(x)+C, \mathcal{A}(y)+D) \in \mathrm{gph}\mathcal{N}_{\mathbb{S}^3_+}, \tag{53}$$

where $C = \mathrm{Diag}(1,0,0)$, $D = \mathrm{Diag}(0,0,-1)$, and $\mathcal{A}\colon \mathbb{R}^3 \to \mathbb{S}^3$ is the linear mapping

$$\mathcal{A}(x) := \begin{pmatrix} x_1 & x_3 & x_2 \\ x_3 & x_2 & x_1 \\ x_2 & x_1 & x_3 \end{pmatrix} \quad \forall x \in \mathbb{R}^3.$$

Consider $\overline{x} = (0,0,0)^{\mathbb{T}}$ and $\overline{y} = (0,0,0)^{\mathbb{T}}$. Write $X := f(\overline{x}, \overline{y})$ and $Y := g(\overline{x}, \overline{y})$. Clearly, $(X, Y) = (C, D) \in \mathrm{gph}\mathbb{S}^3_+$. Moreover, the index sets α, β and γ defined by (56) with $A = X + Y$ satisfy $\alpha = \{1\}, \beta = \{2\}$ and $\gamma = \{3\}$. Fix an arbitrary $0 \neq w = (w_1; w_2) \in \mathbb{R}^3 \times \mathbb{R}^3$ with $w_1 = (w_{11}, w_{12}, w_{13})^{\mathbb{T}}$ and $w_2 = (w_{21}, w_{22}, w_{23})^{\mathbb{T}}$ such that $(G, H) \in \mathcal{T}_{\mathrm{gph}\mathcal{N}_{\mathbb{S}^n_+}}(f(\overline{x}, \overline{y}), g(\overline{x}, \overline{y}))$, where $G = f'_x(\overline{x}, \overline{y})w_1 + f'_y(\overline{x}, \overline{y})w_2$ and $H = g'_x(\overline{x}, \overline{y})w_1 + g'_y(\overline{x}, \overline{y})w_2$. Since $(G, H) \in \mathcal{T}_{\mathrm{gph}\mathcal{N}_{\mathbb{S}^n_+}}(f(\overline{x}, \overline{y}), g(\overline{x}, \overline{y}))$, by the expressions of f and g it is not hard to obtain

$$G = \begin{pmatrix} 0 & 0 & w_{12} \\ 0 & w_{12} & 0 \\ w_{12} & 0 & 0 \end{pmatrix} \quad \text{and} \quad H = \begin{pmatrix} 0 & 0 & w_{22} \\ 0 & w_{22} & 0 \\ w_{22} & 0 & 0 \end{pmatrix}.$$

with $0 \leq w_{12} \perp w_{22} \leq 0$ and $w_{12} + w_{22} \neq 0$. Then $B := P_\beta^{\mathbb{T}}(G+H)P_\beta = w_{12} + w_{22}$. Let $(d_1, d_2) \in \mathbb{R}^3 \times \mathbb{R}^3$ and $(\Lambda, \Delta) \in \mathbb{S}^3 \times \mathbb{S}^3$ satisfy the conditions on the left hand side of (49). Since $\Omega_x = \Omega_y = \mathbb{R}^3$, we have $(d_1, d_2) = 0$. Thus,

$$\nabla_x f(\overline{x}, \overline{y})\Lambda + \nabla_x g(\overline{x}, \overline{y})\Delta = \begin{pmatrix} \Lambda_{11} + 2\Lambda_{23} \\ \Lambda_{22} + 2\Lambda_{13} \\ 2\Lambda_{12} + \Lambda_{33} \end{pmatrix} = 0, \tag{54}$$

$$\nabla_y f(\overline{x}, \overline{y})\Lambda + \nabla_y g(\overline{x}, \overline{y})\Delta = \begin{pmatrix} \Delta_{11} + 2\Delta_{23} \\ \Delta_{22} + 2\Delta_{13} \\ 2\Delta_{12} + \Delta_{33} \end{pmatrix} = 0. \tag{55}$$

Case 1: $w_{12} > 0$. Now we have $w_{22} = 0$, and the index sets π, δ and ν defined by (57) with B satisfy $\pi = \{1\}, \delta = \emptyset$ and $\nu = \emptyset$. From Theorem 1 in Appendix, it follows that $(\Lambda, \Delta) \in \mathcal{N}_{\mathrm{gph}\mathcal{N}_{\mathbb{S}^3_+}}((X, Y); (G, H))$ if and only if

$$\Lambda = \begin{pmatrix} 0 & 0 & \Lambda_{13} \\ 0 & 0 & \Lambda_{23} \\ \Lambda_{13} & \Lambda_{23} & \Lambda_{33} \end{pmatrix} \quad \text{and} \quad \Delta = \begin{pmatrix} \Delta_{11} & \Delta_{12} & \Delta_{13} \\ \Delta_{12} & \Delta_{22} & 0 \\ \Delta_{13} & 0 & 0 \end{pmatrix} \quad \text{with } \Lambda_{13} + \Delta_{13} = 0.$$

Together with (54) and (55), we get $\Lambda = 0$ and $\Delta = 0$. Thus, $(d_1, d_2, \Lambda, \Delta) = (0, 0, 0, 0)$.

Case 2: $w_{12} = 0$. Now we have $w_{22} < 0$, and consequently the index sets π, δ and ν defined by (57) with B satisfy $\pi = \emptyset, \delta = \emptyset$ and $\nu = \{1\}$. From Theorem 1 in Appendix, it follows that $(\Lambda, \Delta) \in \mathcal{N}_{\mathrm{gph}\mathcal{N}_{\mathbb{S}^3_+}}((X, Y); (G, H))$ if and only if

$$\Lambda = \begin{pmatrix} 0 & 0 & \Lambda_{13} \\ 0 & \Lambda_{22} & \Lambda_{23} \\ \Lambda_{13} & \Lambda_{23} & \Lambda_{33} \end{pmatrix} \text{ and } \Delta = \begin{pmatrix} \Delta_{11} & \Delta_{12} & \Delta_{13} \\ \Delta_{12} & 0 & 0 \\ \Delta_{13} & 0 & 0 \end{pmatrix} \text{ with } \Lambda_{13} + \Delta_{13} = 0.$$

Together with (54) and (55), we get $\Lambda = 0$ and $\Delta = 0$. Thus, $(d_1, d_2, \Lambda, \Delta) = (0, 0, 0, 0)$.

The above arguments show that the implication (49) holds, and then the condition in Theorem 4.2 is satisfied. Thus, the global minimizer $(\overline{x}, \overline{y})$ is a M-stationary point of (53), but by [5, Theorem 6.1(i)] we can not judge whether $(\overline{x}, \overline{y})$ is a M-stationary or not.

References

[1] S. J. BI AND S. H. PAN, *Multistage convex relaxation approach to rank regularized minimizataion problems based on equivalent mathematical program with a generalized complementarity constraint*, SIAM Journal on Control and Optimization, 55(2017): 2493-2518.

[2] O. P. BURDAKOV, C. KANZOW AND A. SCHWARTZ, *Mathematical programs with cardinality constraints: reformulation by complementarity-type conditions and a regularization method*, SIAM Journal of Optimization, 26(2016): 397-425.

[3] M. A. DAVENPORT AND J. ROMBERG, *An overview of low-rank matrix recovery from incomplete observations*, IEEE Journal of Selected Topics in Signal Processing, 10(2016): 608-622.

[4] C. DING, D. F. SUN AND K. C. TOH, *An introduction to a class of matrix cone programming*, Mathematical Programming, 144(2014): 141-179.

[5] C. DING, D. F. SUN AND J. J. YE, *First order optimality conditions for mathematical programs with semidefinite cone complementarity constraints*, Mathematical Programming, 144(2014): 141-179.

[6] A. L. DONTCHEV AND R. T. ROCKAFELLAR, *Implicit Functions and Solution Mappings*, Springer Monographs in Mathematics, LLC, New York, 2009.

[7] M. FAZEL, *Matrix Rank Minimization with Applications*, Stanford University, PhD thesis, 2002.

[8] M. FAZEL, H. HINDI AND S. BOYD, *Log-det heuirstic for matrix rank minimization with applications to Hankel and Euclidean distance matrices*, American Control Conference, 2003. Proceedings of the 2003, 3: 2156-2162.

[9] M. FAZEL, T. K. PONG, D. F. SUN AND P. TSENG, *Hankel matrix rank minimization with applications to system identification and realization*, SIAM Journal on Matrix Analysis, 34(2013): 946-977.

[10] M. L. FLEGEL AND C. KANZOW, *On M-stationary points for mathematical programs with equilibrium constraints*, Journal of Mathematical Analysis and Applications, 310(2005): 286-302.

[11] H. GFRERER AND D. KLATTE, *Lipschitz and Holder stability of optimization problems and generalized equations*, Mathematical Programming, 158(2016): 35-75.

[12] D. GROSS, *Recovering low-rank matrices from few coefficients in any basis*, IEEE Transactions on Information Theory, 57(2011): 1548-1566.

[13] A. D. IOFFE AND J. V. OUTRATA, *On metric and calmness qualification conditions in subdifferential calculus*, Set-Valued Analysis, 16(2008): 199–227.

[14] A. J. KING AND R. T. ROCKAFELLAR, *Sensitivity analysis for nonsmooth generalized equations*, Mathematical Programming, 55(1992): 193-212.

[15] M. J. LAI, Y. Y. XU AND W. T. YIN, *Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization*, SIAM Journal on Numerical Analysis, 51 (2013): 927-957.

[16] P. LANCASTER, *On eigenvalues of matrices dependent on a parameter*, Numerische Mathematik, 6(1964): 377-387.

[17] H. Y. LE, *Generalized subdifferentials of the rank function*, Optimization Letters, 7(2013):731-743.

[18] A. B. LEVY, *Implicit multifunction theorems for the sensitivity analysis of variational conditions*, Mathematical Programming, 74(1996): 333-350.

[19] A. S. LEWIS AND H. SENDOV, *Nonsmooth analysis of singular values*, Set-Valued Analysis, 13(2005): 213-241.

[20] A. S. LEWIS, *The convex analysis of unitarily invariant matrix functions*, Jounral of Convex Analysis, 2(1995): 173-183.

[21] Y. L. LIU, S. J. BI and S. H. PAN, *Equivalent Lipschitz surrogates for zero-norm and rank optimization problems*, Journal of Global Optimization, 72(2018): 679-704.

[22] Y. L. LIU and S. H. PAN, *Regular and limiting normal cones to the graph of the subdifferential mapping of the nuclear norm*, Set-Valued and Variational Analysis, 27(2019): 71-85.

[23] W. M. Mao, S. H. Pan and D. F. Sun, *A rank-corrected procedure for matrix completion with fixed basis coefficients*, Mathematical Programming, 159(2016): 289–338.

[24] K. Mohan and M. Fazel, *Iterative reweighted algorithm for matrix rank minimization*, Journal of Machine Learning Research, 13(2012): 3441-3473.

[25] B. S. Mordukhovich, *Complete characterization of openness, metric regularity, and Lipschitzian properties of multifunctions*, Transactions of the American Mathematical Society, 340(1993): 1-35.

[26] S. Negahban and M. J. Wainwright, *Estimation of (near) low-rank matrices with noise and high-dimensional scaling*, Annals of Statistics, 39(2011): 1068-1097.

[27] F. Nie, H. Huang and C. Ding, *Low-rank matrix recovery via efficient schatten p-norm minimization*, In Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, 2012: 655-661.

[28] L. L. Pan, Z. Y. Luo and N. H. Xiu, *Restricted robinson constraint qualification and optimality for cardinality-constrained cone Programming*, Journal of Optimziation Theory and Application, 175(2017): 104-118.

[29] J. S. Pang, M. Razaviyayn and A. Alvarado, *Computing B-stationary points of nonsmooth dc programs*, Mathematics of Operations Research, 42(2017): 95-118.

[30] R. Pietersz and P. J. F. Groenen, *Rank reduction of correlation matrices by majorization*, Quantitative Finance, 4(2004): 649-662.

[31] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

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

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

[34] M. Torki, *Second-order directional derivatives of all eigenvalues of a symmetric matrix*, Nonlinear analysis, 46(2001): 1133-1150.

[35] G. A. Watson, *Characterization of the subdifferential of some matrix norms*, Linear Algebra and Applications, 170(1992): 33-45.

[36] J. Wu, L. W. Zhang and Y. Zhang, *Mathematical programs with semidefinite cone complementary constraints: constraint qualifications and optimality conditions*, Set-Valued and Variational Analysis, 22(2014): 155-187.

[37] Y. Q. Wu, S. H. Pan and S. J. Bi, *KL property of exponent 1/2 for zero-norm regularized quadratic function on sphere and its application*, arXiv:1811.04371, 2018.

Appendix: Directional limiting normal cone to $\mathrm{gph}\mathcal{N}_{\mathbb{S}^n_+}$

In this part we characterize the directional limiting normal cone to $\mathrm{gph}\mathcal{N}_{\mathbb{S}^n_+}$. For this purpose, for a given $A \in \mathbb{S}^n$, we denote by $\lambda(A) \in \mathbb{R}^n$ the eigenvalue vector of A arranged in a nonincreasing order and write $\mathbb{O}^n(A) := \{P \in \mathbb{O}^n \mid A = P\mathrm{Diag}(\lambda(A))P^{\mathbb{T}}\}$.

Fix an arbitrary $(X, Y) \in \mathrm{gph}\mathcal{N}_{\mathbb{S}^n_+}$ and write $A = X + Y$. Suppose that A has the eigenvalue decomposition $A = P\mathrm{Diag}(\lambda(A))P^{\mathbb{T}}$ where $P \in \mathbb{O}^n(A)$. Define the index sets

$$\alpha := \{i \mid \lambda_i(A) > 0\}, \ \beta := \{i \mid \lambda_i(A) = 0\} \ \text{ and } \ \gamma := \{i \mid \lambda_i(A) < 0\}. \tag{56}$$

For given $G, H \in \mathbb{S}^n$, by the eigenvalues of $B := P_\beta^{\mathbb{T}}(G+H)P_\beta$ we define the index sets:

$$\pi := \{i \in [1, |\beta|] \mid \lambda_i(B) > 0\}, \delta := \{i \in [1, |\beta|] \mid \lambda_i(B) = 0\}, \nu := \{i \in [1, |\beta|] \mid \lambda_i(B) < 0\}. \tag{57}$$

For the index set δ, we denote the set of all partitions of δ by $\mathscr{P}(\delta)$. Define the set

$$\mathbb{R}^{|\delta|}_{\gtrsim} := \{z \in \mathbb{R}^{|\delta|} : \ z_1 \geq \cdots \geq z_{|\delta|}\}.$$

For any $z \in \mathbb{R}^{|\delta|}_{\gtrsim}$, let $D(z) \in \mathbb{S}^{|\delta|}$ denote the first generalized divided difference matrix of $h(t) = \max(t, 0)$ at z, which is defined as

$$(D(z))_{ij} := \begin{cases} \frac{\max(0, z_i) - \max(0, z_j)}{z_i - z_j} \in [0, 1] & \text{if } z_i \neq z_j, \\ 1 & \text{if } z_i = z_j > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{58}$$

Write $\mathcal{U}_{|\beta|} := \{\overline{\Omega} \in \mathbb{S}^{|\delta|} : \ \overline{\Omega} = \lim_{k \to \infty} D(z^k), \ z^k \to 0, \ z^k \in \mathbb{R}^{|\delta|}_{\gtrsim}\}$. Let $\widehat{\Xi}_1 \in \mathcal{U}_{|\delta|}$. By the definition of $\mathcal{U}_{|\delta|}$, there exists a partition $(\delta_+, \delta_0, \delta_-) \in \mathscr{P}(\delta)$ such that

$$\widehat{\Xi}_1 = \begin{bmatrix} E_{\delta_+\delta_+} & E_{\delta_+\delta_0} & (\widehat{\Xi}_1)_{\delta_+\delta_-} \\ E_{\delta_0\delta_+} & 0 & 0 \\ (\widehat{\Xi}_1)_{\delta_+\delta_-}^{\mathbb{T}} & 0 & 0 \end{bmatrix} \in \mathbb{S}^{|\delta|}$$

where every entry of $(\widehat{\Xi}_1)_{\delta_+\delta_-}$ belongs to $[0, 1]$. We also write $\widehat{\Xi}_2 = E - \widehat{\Xi}_1$, i.e.,

$$\widehat{\Xi}_2 := \begin{bmatrix} 0 & 0 & E_{\delta_+\delta_-} - (\widehat{\Xi}_1)_{\delta_+\delta_-} \\ 0 & 0 & E_{\delta_0\delta_-} \\ (E_{\delta_+\delta_-} - (\widehat{\Xi}_1)_{\delta_+\delta_-})^{\mathbb{T}} & E_{\delta_-\delta_0} & E_{\delta_-\delta_-} \end{bmatrix}.$$

Next we provide a characterization for the direction limiting normal cone to $\mathrm{gph}\mathcal{N}_{\mathbb{S}^n_+}$.

Theorem 1 *Consider an arbitrary point $(X, Y) \in \mathrm{gph}\mathcal{N}_{\mathbb{S}^n_+}$. Let $A = X + Y$ have the eigenvalue decomposition as above with α, β, γ defined by (56). Then, for any given*

$(G, H) \in \mathbb{S}^n \times \mathbb{S}^n$ *with* π, δ *and* ν *defined as in* (57) *for* $B := P_\beta^{\mathbb{T}}(G+H)P_\beta$, *it holds that* $(X^*, Y^*) \in \mathcal{N}_{\mathrm{gph}\mathcal{N}_{\mathbb{S}^n_+}}((X, Y); (G, H))$ *if and only if* (G, H) *and* (X^*, Y^*) *satisfy*

$$
\begin{cases}
G = P \begin{bmatrix} \widetilde{G}_{\alpha\alpha} & \widetilde{G}_{\alpha\beta} & \widetilde{G}_{\alpha\gamma} \\ \widetilde{G}_{\alpha\beta}^{\mathbb{T}} & \widetilde{G}_{\beta\beta} & 0 \\ \widetilde{G}_{\alpha\gamma}^{\mathbb{T}} & 0 & 0 \end{bmatrix} P^{\mathbb{T}}, \quad H = P \begin{bmatrix} 0 & 0 & \widetilde{H}_{\alpha\gamma} \\ 0 & \widetilde{H}_{\beta\beta} & \widetilde{H}_{\beta\gamma} \\ \widetilde{H}_{\alpha\gamma}^{\mathbb{T}} & \widetilde{H}_{\beta\gamma}^{\mathbb{T}} & \widetilde{H}_{\gamma\gamma} \end{bmatrix} P^{\mathbb{T}}, & \text{(59a)} \\[18pt]
(E_{\alpha\gamma} - \Sigma_{\alpha\gamma}) \circ \widetilde{G}_{\alpha\gamma} + \Sigma_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} = 0, \ \ \widetilde{G}_{\beta\beta} = \Pi_{\mathbb{S}^{|\beta|}_+}(\widetilde{G}_{\beta\beta} + \widetilde{H}_{\beta\beta}) & \text{(59b)}
\end{cases}
$$

and

$$
\begin{cases}
X^* = P \begin{bmatrix} 0 & 0 & \widetilde{X}^*_{\alpha\gamma} \\ 0 & \widetilde{X}^*_{\beta\beta} & \widetilde{X}^*_{\beta\gamma} \\ (\widetilde{X}^*_{\alpha\gamma})^{\mathbb{T}} & (\widetilde{X}^*_{\beta\gamma})^{\mathbb{T}} & \widetilde{X}^*_{\gamma\gamma} \end{bmatrix} P^{\mathbb{T}}, \ \ Y^* = P \begin{bmatrix} \widetilde{Y}^*_{\alpha\alpha} & \widetilde{Y}^*_{\alpha\beta} & \widetilde{Y}^*_{\alpha\gamma} \\ (\widetilde{Y}^*_{\alpha\beta})^{\mathbb{T}} & \widetilde{Y}^*_{\beta\beta} & 0 \\ (\widetilde{Y}^*_{\alpha\gamma})^{\mathbb{T}} & 0 & 0 \end{bmatrix} P^{\mathbb{T}}, & \text{(60a)} \\[18pt]
\Sigma_{\alpha\gamma} \circ \widetilde{X}^*_{\alpha\gamma} + (E_{\alpha\gamma} - \Sigma_{\alpha\gamma}) \circ (\widetilde{Y}^*_{\alpha\gamma}) = 0, \ \ (\widetilde{X}^*_{\beta\beta}, \widetilde{Y}^*_{\beta\beta}) \in \mathcal{N}_{\mathrm{gph}\mathcal{N}_{\mathbb{S}^{|\beta|}_+}}(0, 0) \text{ with} & \text{(60b)} \\[10pt]
\mathcal{N}_{\mathrm{gph}\mathcal{N}_{\mathbb{S}^{|\beta|}_+}}(0, 0) = \bigcup_{\substack{Q \in \mathbb{O}^{|\beta|} \\ \widehat{\Xi}_1 \in \mathcal{U}_{|\delta|}}} \left\{ (U^*, V^*) \ \middle| \ \begin{array}{l} \Xi_1 \circ (Q^{\mathbb{T}}U^*Q) + \Xi_2 \circ (Q^{\mathbb{T}}V^*Q) = 0, \\ Q_{\delta_0}^{\mathbb{T}} U^* Q_{\delta_0} \preceq 0, \ Q_{\delta_0}^{\mathbb{T}} V^* Q_{\delta_0} \succeq 0 \end{array} \right\} & \text{(60c)}
\end{cases}
$$

where

$$
\Xi_1 := \begin{bmatrix} E_{\pi\pi} & E_{\pi\delta} & \Gamma_{\pi\nu} \\ E_{\delta\pi} & \widehat{\Xi}_1 & 0 \\ (\Gamma_{\pi\nu})^{\mathbb{T}} & 0 & 0 \end{bmatrix} \quad \text{and} \quad \Xi_2 := \begin{bmatrix} 0_{\pi\pi} & 0_{\pi\delta} & E_{\pi\nu} - \Gamma_{\pi\nu} \\ 0_{\delta\pi} & \widehat{\Xi}_2 & E_{\delta\pi} \\ (E_{\pi\nu} - \Gamma_{\pi\nu})^{\mathbb{T}} & E_{\nu\delta} & E_{\nu\nu} \end{bmatrix}
$$

with $\Gamma_{ij} = \frac{\max(0, \lambda_i(B)) - \max(0, \lambda_j(B))}{\lambda_i(B) - \lambda_j(B)}$ *for each* $(i, j) \in \pi \times \nu$.

Proof: "$\Longrightarrow$". Since $(X^*, Y^*) \in \mathcal{N}_{\mathrm{gph}\mathcal{N}_{\mathbb{S}^n_+}}((X, Y); (G, H))$, there exist sequences $t_k \downarrow 0$ and $(G^k, H^k, X^k, Y^k) \to (G, H, X^*, Y^*)$ with $(X^k, Y^k) \in \widehat{\mathcal{N}}_{\mathrm{gph}\mathcal{N}_{\mathbb{S}^n_+}}(X + t_k G^k, Y + t_k H^k)$ for each k. Since $(X + t_k G^k, Y + t_k H^k) \in \mathrm{gph}\,\mathcal{N}_{\mathbb{S}^n_+}$ for each k, it holds that

$$
X + t_k G^k = \Pi_{\mathbb{S}^n_+}(X + Y + t_k(G^k + H^k)).
$$

Notice that $X = \Pi_{\mathbb{S}^n_+}(X+Y)$ since $(X, Y) \in \mathrm{gph}\,\mathcal{N}_{\mathbb{S}^n_+}$ and the projection operator $\Pi_{\mathbb{S}^n_+}(\cdot)$ is directionally differentiable everywhere in the Hadamard sense. Taking the limit to the last equality, we obtain $G = \Pi'_{\mathbb{S}^n_+}(X+Y; G+H)$. By the expression of $\Pi'_{\mathbb{S}^n_+}(X+Y; G+H)$ in [33], it follows that G and H satisfy (59a)-(59b). Write $A^k := (X+Y) + t_k(G^k+H^k)$. For each k, let A^k have the spectral decomposition $A^k = (P^k)^{\mathbb{T}}\mathrm{Diag}(\lambda(A^k))P^k$ with $P^k \in \mathbb{O}^n(A^k)$. Since $\lambda(A) = \lim_{k\to\infty} \lambda(A^k)$, we have $\lambda_i(A^k) > 0$ for $i \in \alpha$ and $\lambda_i(A^k) < 0$ for $i \in \gamma$ when k is sufficiently large, and $\lim_{k\to\infty} \lambda_i(A^k) = 0$ for $i \in \beta$. Since $\{P^k\}$ is bounded, we may assume (if necessary taking a subsequence) that $\{P^k\}$ converges to $\widehat{P} \in \mathbb{O}^n(A)$. Since $P \in \mathbb{O}^n(A)$, there exists $Q \in \mathbb{O}^{|\beta|}$ such that $\widehat{P} = [P_\alpha \ \ P_\beta Q \ \ P_\gamma]$.

24

We assume (if necessary taking a subsequence) that there is a partition $(\beta_+, \beta_0, \beta_-)$ of β such that

$$\lambda_i(A^k) > 0 \ \ \forall i \in \beta_+, \ \lambda_i(A^k) = 0 \ \ \forall i \in \beta_0 \ \text{ and } \ \lambda_i(A^k) < 0 \ \ \forall i \in \beta_- \ \ \text{ for each } k.$$

In addition, from [16, Theorem 7] or [34, Proposition 1.4], it follows that

$$\lambda_i(A^k) = t_k \lambda_{l_i}\big[P_\beta^{\mathbb{T}}(G^k + H^k)P_\beta\big] + o(t_k) \quad \forall i \in \beta \tag{61}$$

where l_i is the number of eigenvalues that are equal to $\lambda_i(A)$ but are ranked before i (including i). Write $B^k := P_\beta^{\mathbb{T}}(G^k + H^k)P_\beta$ for each k. Since $\lambda(B) = \lim_{k\to\infty}\lambda(B^k)$, we have $\lambda_i(B^k) > 0$ for $i \in \pi$ and $\lambda_i(B^k) < 0$ for $i \in \nu$ when k is sufficiently large, and $\lim_{k\to\infty}\lambda_i(B^k) = 0$ for $i \in \delta$. By further taking a subsequence if necessary, we may assume that there exists a partition $(\delta_+, \delta_0, \delta_-)$ of δ such that for each k,

$$\lambda_i(B^k) > 0 \ \ \forall i \in \delta_+, \ \lambda_i(B^k) = 0 \ \ \forall i \in \delta_0 \ \text{ and } \ \lambda_i(B^k) < 0 \ \ \forall i \in \delta_-.$$

This means that $\pi \cup \delta_+ = \beta_+ - |\alpha|$ and $\nu \cup \delta_- = \beta_- - |\alpha|$, and then $\delta_0 = \beta_0 - |\alpha|$. For convenience, we write $\overline{\pi} = \pi + |\alpha|$, $\overline{\delta} = \delta + |\alpha|$ and $\overline{\nu} = \nu + |\alpha|$. Then, we have

$$\{i \mid \lambda_i(A^k) > 0\} = \alpha \cup \overline{\pi} \cup \overline{\delta}_+, \ \{i \mid \lambda_i(A^k) = 0\} = \overline{\delta}_0, \ \{i \mid \lambda_i(A^k) < 0\} = \overline{\delta}_- \cup \overline{\nu} \cup \gamma.$$

Since $(X^k, -Y^k) \in \widehat{\mathcal{N}}_{\mathrm{gph}\,\mathcal{N}_{\mathbb{S}_+^n}}(X + t_k G^k, Y + t_k H^k)$, by [36, Corollary 3.2] it follows that

$$\Theta_1^k \circ \widetilde{X}^k + \Theta_2^k \circ (-\widetilde{Y}^k) = 0, \ \widetilde{X}^k_{\overline{\delta}_0\overline{\delta}_0} \succeq 0 \ \text{ and } \ \widetilde{Y}^k_{\overline{\delta}_0\overline{\delta}_0} \succeq 0 \tag{62}$$

with $\widetilde{X}^k = (P^k)^{\mathbb{T}}X^k P^k$ and $\widetilde{Y}^k = (P^k)^{\mathbb{T}}Y^k P^k$, where Θ_1^k and Θ_2^k take the following form

$$\Theta_1^k = \begin{bmatrix} E_{\alpha\alpha} & E_{\alpha\overline{\pi}} & E_{\alpha\overline{\delta}_+} & E_{\alpha\overline{\delta}_0} & \Sigma^k_{\alpha\overline{\delta}_-} & \Sigma^k_{\alpha\overline{\nu}} & \Sigma^k_{\alpha\gamma} \\ E_{\overline{\pi}\alpha} & E_{\overline{\pi}\overline{\pi}} & E_{\overline{\pi}\overline{\delta}_+} & E_{\overline{\pi}\overline{\delta}_0} & \Sigma^k_{\overline{\pi}\overline{\delta}_-} & \Sigma^k_{\overline{\pi}\overline{\nu}} & \Sigma^k_{\overline{\pi}\gamma} \\ E_{\overline{\delta}_+\alpha} & E_{\overline{\delta}_+\overline{\pi}} & E_{\overline{\delta}_+\overline{\delta}_+} & E_{\overline{\delta}_+\overline{\delta}_0} & \Sigma^k_{\overline{\delta}_+\overline{\delta}_-} & \Sigma^k_{\overline{\delta}_+\overline{\nu}} & \Sigma^k_{\overline{\delta}_+\gamma} \\ E_{\overline{\delta}_0\alpha} & E_{\overline{\delta}_0\overline{\pi}} & E_{\overline{\delta}_0\overline{\delta}_+} & 0 & 0 & 0 & 0 \\ (\Sigma^k_{\alpha\overline{\delta}_-})^{\mathbb{T}} & (\Sigma^k_{\overline{\pi}\overline{\delta}_-})^{\mathbb{T}} & (\Sigma^k_{\overline{\delta}_+\overline{\delta}_-})^{\mathbb{T}} & 0 & 0 & 0 & 0 \\ (\Sigma^k_{\alpha\overline{\nu}})^{\mathbb{T}} & (\Sigma^k_{\overline{\pi}\overline{\nu}})^{\mathbb{T}} & (\Sigma^k_{\overline{\delta}_+\overline{\nu}})^{\mathbb{T}} & 0 & 0 & 0 & 0 \\ (\Sigma^k_{\alpha\gamma})^{\mathbb{T}} & (\Sigma^k_{\overline{\pi}\gamma})^{\mathbb{T}} & (\Sigma^k_{\overline{\delta}_+\gamma})^{\mathbb{T}} & 0 & 0 & 0 & 0 \end{bmatrix} \tag{63}$$

and

$$\Theta_2^k = \begin{bmatrix} 0_{\alpha\alpha} & 0_{\alpha\overline{\pi}} & 0_{\alpha\overline{\delta}_+} & 0_{\alpha\overline{\delta}_0} & \widetilde{\Sigma}^k_{\alpha\overline{\delta}_-} & \widetilde{\Sigma}^k_{\alpha\overline{\nu}} & \widetilde{\Sigma}^k_{\alpha\gamma} \\ 0_{\overline{\pi}\alpha} & 0_{\overline{\pi}\overline{\pi}} & 0_{\overline{\pi}\overline{\delta}_+} & 0_{\overline{\pi}\overline{\delta}_0} & \widetilde{\Sigma}^k_{\overline{\pi}\overline{\delta}_-} & \widetilde{\Sigma}^k_{\overline{\pi}\overline{\nu}} & \widetilde{\Sigma}^k_{\overline{\pi}\gamma} \\ 0_{\overline{\delta}_+\alpha} & 0_{\overline{\delta}_+\overline{\pi}} & 0_{\overline{\delta}_+\overline{\delta}_+} & 0_{\overline{\delta}_+\overline{\delta}_0} & \widetilde{\Sigma}^k_{\overline{\delta}_+\overline{\delta}_-} & \widetilde{\Sigma}^k_{\overline{\delta}_+\overline{\nu}} & \widetilde{\Sigma}^k_{\overline{\delta}_+\gamma} \\ 0_{\overline{\delta}_0\alpha} & 0_{\overline{\delta}_0\overline{\pi}} & 0_{\overline{\delta}_0\overline{\delta}_+} & 0_{\overline{\delta}_0\overline{\delta}_0} & E_{\overline{\delta}_0\overline{\delta}_-} & E_{\overline{\delta}_0\overline{\nu}} & E_{\overline{\delta}_0\gamma} \\ (\widetilde{\Sigma}^k_{\alpha\overline{\delta}_-})^{\mathbb{T}} & (\widetilde{\Sigma}^k_{\overline{\pi}\overline{\delta}_-})^{\mathbb{T}} & (\widetilde{\Sigma}^k_{\overline{\delta}_+\overline{\delta}_-})^{\mathbb{T}} & E_{\overline{\delta}_-\overline{\delta}_0} & E_{\overline{\delta}_-\overline{\delta}_-} & E_{\overline{\delta}_-\overline{\nu}} & E_{\overline{\delta}_-\gamma} \\ (\widetilde{\Sigma}^k_{\alpha\overline{\nu}})^{\mathbb{T}} & (\widetilde{\Sigma}^k_{\overline{\pi}\overline{\nu}})^{\mathbb{T}} & (\widetilde{\Sigma}^k_{\overline{\delta}_+\overline{\nu}})^{\mathbb{T}} & E_{\overline{\nu}\overline{\delta}_0} & E_{\overline{\nu}\overline{\delta}_-} & E_{\overline{\nu}\overline{\nu}} & E_{\overline{\nu}\gamma} \\ (\widetilde{\Sigma}^k_{\alpha\gamma})^{\mathbb{T}} & (\widetilde{\Sigma}^k_{\overline{\pi}\gamma})^{\mathbb{T}} & (\widetilde{\Sigma}^k_{\overline{\delta}_+\gamma})^{\mathbb{T}} & E_{\gamma\overline{\delta}_0} & E_{\gamma\overline{\delta}_-} & E_{\gamma\overline{\nu}} & E_{\gamma\gamma} \end{bmatrix} \tag{64}$$

25

with $\Sigma_{ij}^k := \frac{\max(0,\lambda_i(A^k))-\max(0,\lambda_j(A^k))}{\lambda_i(A^k)-\lambda_j(A^k)}$ and $\widetilde{\Sigma}_{ij}^k := E_{ij}^k - \Sigma_{ij}^k$. It is immediate to have that

$$\widetilde{X}^k \to \begin{bmatrix} \widetilde{X}_{\alpha\alpha}^* & \widetilde{X}_{\alpha\beta}^* Q & \widetilde{X}_{\alpha\gamma}^* \\ Q^{\mathbb{T}}(\widetilde{X}_{\alpha\beta}^*)^{\mathbb{T}} & Q^{\mathbb{T}}\widetilde{X}_{\beta\beta}^* Q & Q^{\mathbb{T}}\widetilde{X}_{\beta\gamma}^* \\ (\widetilde{X}_{\alpha\gamma}^*)^{\mathbb{T}} & (\widetilde{X}_{\beta\gamma}^*)^{\mathbb{T}} Q & \widetilde{X}_{\gamma\gamma}^* \end{bmatrix} \text{ and } \widetilde{Y}^k \to \begin{bmatrix} \widetilde{Y}_{\alpha\alpha}^* & \widetilde{Y}_{\alpha\beta}^* Q & \widetilde{Y}_{\alpha\gamma}^* \\ Q^{\mathbb{T}}(\widetilde{Y}_{\alpha\beta}^*)^{\mathbb{T}} & Q^{\mathbb{T}}\widetilde{Y}_{\beta\beta}^* Q & Q^{\mathbb{T}}\widetilde{Y}_{\beta\gamma}^* \\ (\widetilde{Y}_{\alpha\gamma}^*)^{\mathbb{T}} & (\widetilde{Y}_{\beta\gamma}^*)^{\mathbb{T}} Q & \widetilde{Y}_{\gamma\gamma}^* \end{bmatrix}.$$

By the expression of Σ_{ij}^k for $(i,j) \in (\alpha \cup \overline{\pi} \cup \overline{\delta}_+) \times (\overline{\delta}_- \cup \overline{\nu} \cup \gamma)$ and equation (61),

$$\lim_{k\to\infty} \Sigma_{\alpha\overline{\delta}_-}^k = E_{\alpha\overline{\delta}_-}, \ \lim_{k\to\infty} \Sigma_{\alpha\overline{\nu}}^k = E_{\alpha\overline{\nu}}, \ \lim_{k\to\infty} \Sigma_{\alpha\gamma}^k = \Sigma_{\alpha\gamma}, \ \lim_{k\to\infty} \Sigma_{\overline{\delta}_+\overline{\nu}}^k = 0_{\overline{\delta}_+\overline{\nu}},$$

$$\lim_{k\to\infty} \Sigma_{\overline{\pi}\overline{\delta}_-}^k = E_{\overline{\pi}\overline{\delta}_-}, \ \lim_{k\to\infty} \Sigma_{\overline{\pi}\nu}^k = \Sigma_{\overline{\pi}\nu}, \ \lim_{k\to\infty} \Sigma_{\overline{\pi}\gamma}^k = 0_{\overline{\pi}\gamma}, \ \lim_{k\to\infty} \Sigma_{\overline{\delta}_+\gamma}^k = 0_{\overline{\delta}_+\gamma},$$

where $\Sigma_{ij} = \frac{\max(0,\lambda_{i-|\alpha|}(B))-\max(0,\lambda_{j-|\alpha|}(B))}{\lambda_{i-|\alpha|}(B)-\lambda_{j-|\alpha|}(B)}$ for $(i,j) \in \overline{\pi} \times \overline{\nu}$. Thus, there exists $\widehat{\Xi}_1 \in \mathcal{U}_{|\delta|}$ such that $\widehat{\Xi}_1$ and the associated matrices $\widehat{\Xi}_2$, Ξ_1 and Ξ_2 satisfy

$$\lim_{k\to\infty} \Theta_1^k = \Theta_1 + \begin{bmatrix} 0_{\alpha\alpha} & 0_{\alpha\beta} & 0_{\alpha\gamma} \\ 0_{\beta\alpha} & \Xi_1 & 0 \\ 0_{\alpha\gamma} & 0 & 0 \end{bmatrix} \text{ and } \lim_{k\to\infty} \Theta_2^k = \Theta_2 + \begin{bmatrix} 0_{\alpha\alpha} & 0_{\alpha\beta} & 0_{\alpha\gamma} \\ 0_{\beta\alpha} & \Xi_2 & 0 \\ 0_{\alpha\gamma} & 0 & 0 \end{bmatrix}.$$

Taking the limit $k \to \infty$ to (62), we have that (X^*, Y^*) satisfies the condition (60a)-(60b).

"$\Longleftarrow$". Let (G, H) and (X^*, Y^*) satisfy (59a)-(59b) and (60a)-(60b), respectively. We shall prove that there exist sequences $t_k \downarrow 0$ and $(G^k, H^k, X^k, Y^k) \to (G, H, X^*, Y^*)$ with $(X^k, -Y^k) \in \widehat{\mathcal{N}}_{\text{gph}\,\mathcal{N}_{\mathbb{S}_+^n}}(X + t_k G^k, Y + t_k H^k)$ for each k. Let B have the spectral decomposition $B = U\text{Diag}(\lambda(B))U^{\mathbb{T}}$ and write $\widehat{P} = [P_\alpha \ P_\beta U \ P_\gamma]$. Since $(\widetilde{X}_{\beta\beta}^*, -\widetilde{Y}_{\beta\beta}^*) \in \mathcal{N}_{\text{gph}\,\mathcal{N}_{\mathbb{S}_+^{|\beta|}}}(0,0)$, there exist an orthogonal matrix $Q \in \mathbb{O}^{|\beta|}$ and $\widehat{\Xi}_1 \in \mathcal{U}_{|\delta|}$ such that

$$\Xi_1 \circ Q^{\mathbb{T}}\widetilde{X}_{\beta\beta}^* Q + \Xi_2 \circ (-Q^{\mathbb{T}}\widetilde{Y}_{\beta\beta}^* Q) = 0, \ Q_{\delta_0}^{\mathbb{T}}\widetilde{X}_{\beta\beta}^* Q_{\delta_0} \succeq 0 \ \text{ and } \ Q_{\delta_0}^{\mathbb{T}}\widetilde{Y}_{\beta\beta}^* Q_{\delta_0} \succeq 0. \quad (65)$$

Since $\widehat{\Xi}_1 \in \mathcal{U}_{|\delta|}$, we know that there exists a sequence $\{z^k\} \in \mathbb{R}_{\gtrsim}^{|\delta|}$ converging to 0 such that $\widehat{\Xi}_1 = \lim_{k\to\infty} D(z^k)$. Without loss of generality, we can assume that there exists a partition $(\delta_+, \delta_0, \delta_-) \in \mathscr{P}(\delta)$ such that for all k,

$$z_i^k > 0 \ \ \forall i \in \delta_+, \ \ z_i^k = 0 \ \ \forall i \in \delta_0 \ \text{ and } \ z_i^k < 0 \ \ \forall i \in \delta_-.$$

For each k, with $x^k = [\lambda_\pi(B); z_{\delta_+}^k; 0_{\delta_0}; 0_{\delta_-}; 0_\nu]$ and $y^k = [0_\pi; 0_{\delta_+}; 0_{\delta_0}; z_{\delta_-}^k; \lambda_\nu(B)]$, define

$$\widehat{G}^k = \widehat{P} \begin{bmatrix} \widetilde{G}_{\alpha\alpha} & \widetilde{G}_{\alpha\beta} & \widetilde{G}_{\alpha\gamma} \\ \widetilde{G}_{\alpha\beta}^{\mathbb{T}} & \text{Diag}(x^k) & 0 \\ \widetilde{G}_{\alpha\gamma}^{\mathbb{T}} & 0 & 0 \end{bmatrix} \widehat{P}^{\mathbb{T}} \text{ and } \widehat{H}^k = \widehat{P} \begin{bmatrix} 0_{\alpha\alpha} & 0_{\alpha\beta} & \widetilde{H}_{\alpha\gamma} \\ 0_{\beta\alpha} & \text{Diag}(y^k) & \widetilde{H}_{\beta\gamma} \\ \widetilde{H}_{\alpha\gamma}^{\mathbb{T}} & \widetilde{H}_{\beta\gamma}^{\mathbb{T}} & \widetilde{H}_{\gamma\gamma} \end{bmatrix} \widehat{P}^{\mathbb{T}}.$$

Clearly, for each k, $(\widehat{G}^k, \widehat{H}^k)$ satisfies equation (59a)-(59b), which is equivalent to saying that $(\widehat{G}^k, \widehat{H}^k) \in \mathcal{T}_{\text{gph}\,\mathcal{N}_{\mathbb{S}_+^n}}(X, Y)$. Thus, for each k, there exist $t_{k_j} \downarrow 0$ and $(\widehat{G}^{k_j}, \widehat{H}^{k_j}) \to$

26

$(\widehat{G}^k, \widehat{H}^k)$ as $j \to \infty$ such that $(X, Y) + t_{k_j}(\widehat{G}^{k_j}, \widehat{H}^{k_j}) \in \text{gph} \, \mathcal{N}_{\mathbb{S}^n_+}$ for each j. By this, we can find sequences $t_k \downarrow 0$ and $(G^k, H^k) \to (G, H)$ such that $(X + t_k G^k, Y + t_k H^k) \in \text{gph} \, \mathcal{N}_{\mathbb{S}^n_+}$ for each k. For each k, write $A^k = X + Y + t_k(G^k + H^k)$ and define $\Theta_1^k \in \mathbb{S}^n$ and $\Theta_2^k \in \mathbb{S}^n$ as in (63) and (64), respectively, except that $\Sigma^k_{\alpha\gamma}$ is replaced by $\Sigma_{\alpha\gamma}$. By the proof of the previous necessity, for all sufficiently large k (if necessary taking a subsequence of $\{A^k\}$),

$$\{i \,|\, \lambda_i(A^k) > 0\} = \alpha \cup \overline{\pi} \cup \overline{\delta}_+, \ \{i \,|\, \lambda_i(A^k) = 0\} = \overline{\delta}_0, \ \{i \,|\, \lambda_i(A^k) < 0\} = \overline{\delta}_- \cup \overline{\nu} \cup \gamma$$

where $\overline{\pi}$ and $\overline{\nu}$ are the same as before and $\overline{\delta} = \delta + |\alpha|$. Next for each k we shall define the matrices $\widetilde{X}^k \in \mathbb{S}^n$ and $\widetilde{Y}^k \in \mathbb{S}^n$. Let $i, j \in \{1, 2, \ldots, n\}$. If (i, j) and (j, i)

Case 1: (i, j) or $(j, i) \in \alpha \times (\overline{\delta}_- \cup \overline{\nu})$. In this case, we have $\widetilde{X}^*_{ij} = 0$ by (60a). Define

$$\widetilde{Y}^k_{ij} \equiv \widetilde{Y}^*_{ij} \quad \text{and} \quad \widetilde{X}^k_{ij} = \frac{1 - \Sigma^k_{ij}}{\Sigma^k_{ij}} \widetilde{Y}^k_{ij}. \tag{66}$$

Since $\Sigma^k_{ij} \to 1$ in this case, it immediately follows that $(\widetilde{X}^k_{ij}, \widetilde{Y}^k_{ij}) \to (\widetilde{X}^*_{ij}, \widetilde{Y}^*_{ij})$.

Case 2: (i, j) or $(j, i) \in \overline{\pi} \times \gamma$. Now we have $\widetilde{Y}^*_{ij} = 0$ by equation (60a). Define

$$\widetilde{X}^k_{ij} \equiv \widetilde{X}^*_{ij} \quad \text{and} \quad \widetilde{Y}^k_{ij} = \frac{\Sigma^k_{ij}}{1 - \Sigma^k_{ij}} \widetilde{X}^k_{ij}. \tag{67}$$

Notice that $\Sigma^k_{ij} \to 0$ in this case. It immediately follows that $(\widetilde{X}^k_{ij}, \widetilde{Y}^k_{ij}) \to (\widetilde{X}^*_{ij}, \widetilde{Y}^*_{ij})$.

Case 3: (i, j) or $(j, i) \in \overline{\pi} \times \overline{\nu}$. Now $\Sigma^k_{ij} = \frac{\lambda_i(A^k)}{\lambda_i(A^k)^k - \lambda_j(A^k)} = \frac{\lambda_{l_i}(G^k + H^k)}{\lambda_{l_i}(G^k + H^k) - \lambda_{l_j}(G^k + H^k)}$ for each k. Together with the definition of Ξ_1, $\Sigma^k_{ij} \to (\Xi_1)_{i'j'}$ with $i' = i - |\alpha|$ and $j' = j - |\alpha|$.

Subcase 3.1: $(\Xi_1)_{ij} \neq 1$. Then $\Sigma^k_{ij} \neq 1$ for all sufficiently large k. We define

$$\widetilde{X}^k_{ij} \equiv Q_i^{\mathbb{T}} \widetilde{X}^*_{\beta\beta} Q_j \quad \text{and} \quad \widetilde{Y}^k_{ij} = \frac{\Sigma^k_{ij}}{1 - \Sigma^k_{ij}} \widetilde{X}^k_{ij}. \tag{68}$$

From equation (65), it follows that $\widetilde{Y}^k_{ij} \to \frac{(\Xi_1)_{ij}}{1 - (\Xi_1)_{ij}} Q_i^{\mathbb{T}} \widetilde{X}^*_{\beta\beta} Q_j = Q_i^{\mathbb{T}} \widetilde{Y}^*_{\beta\beta} Q_j$.

Subcase 3.2: $(\Xi_1)_{ij} = 1$. Since $\Sigma^k_{ij} \neq 0$ for all sufficiently large k, we define

$$\widetilde{Y}^k_{ij} \equiv Q_i^{\mathbb{T}} \widetilde{Y}^*_{\beta\beta} Q_j \quad \text{and} \quad \widetilde{X}^k_{ij} = \frac{1 - \Sigma^k_{ij}}{\Sigma^k_{ij}} \widetilde{Y}^k_{ij}. \tag{69}$$

From equation (65), it follows that $\widetilde{X}^k_{ij} \to \frac{1 - (\Xi_1)_{ij}}{(\Xi_1)_{ij}} Q_i^{\mathbb{T}} \widetilde{Y}^*_{\beta\beta} Q_j = Q_i^{\mathbb{T}} \widetilde{X}^*_{\beta\beta} Q_j$.

Case 4: (i, j) or $(j, i) \in (\beta \times \beta) \backslash (\overline{\pi} \times \overline{\nu})$. In this case we define

$$\widetilde{X}^k_{ij} = Q_i^{\mathbb{T}} \widetilde{X}^*_{\beta\beta} Q_j \quad \text{and} \quad \widetilde{Y}^k_{ij} \equiv Q_i^{\mathbb{T}} \widetilde{Y}^*_{\beta\beta} Q_j. \tag{70}$$

27

Case 5: (i,j) or $(j,i) \notin (\alpha \times \overline{\nu}) \cup (\overline{\pi} \times \gamma) \cup (\beta \times \beta)$. In this case we define

$$\widetilde{X}_{ij}^k \equiv \widetilde{X}_{\beta\beta}^* \quad \text{and} \quad \widetilde{Y}_{ij}^k \equiv \widetilde{Y}_{\beta\beta}^*. \tag{71}$$

Now for each k we define $X^k = \widehat{P}\widetilde{X}^k\widehat{P}^{\mathbb{T}}$ and $Y^k = \widehat{P}\widetilde{Y}^k\widehat{P}^{\mathbb{T}}$. Then, from (66)-(71), it follows that $(\widehat{P}^{\mathbb{T}}X^k\widehat{P}, \widehat{P}^{\mathbb{T}}Y^k\widehat{P}) = (\widetilde{X}^k, \widetilde{Y}^k) \to (X^*, Y^*)$ as $k \to \infty$, and moreover,

$$\Theta_1^k \circ (\widehat{P}^{\mathbb{T}}X^k\widehat{P}) + \Theta_2^k \circ (-\widehat{P}^{\mathbb{T}}Y^k\widehat{P}) = 0, \quad k = 1, 2, \ldots,$$

Moreover, from equations (70) and the last inequalities in (65), it follows that

$$Q_{\beta_0}^{\mathbb{T}}\widetilde{X}^kQ_{\beta_0} \equiv Q_{\beta_0}^{\mathbb{T}}\widetilde{X}_{\beta\beta}^*Q_{\beta_0} \succeq 0 \quad \text{and} \quad Q_{\beta_0}^{\mathbb{T}}\widetilde{Y}^kQ_{\beta_0} \equiv Q_{\beta_0}^{\mathbb{T}}\widetilde{Y}_{\beta\beta}^*Q_{\beta_0} \succeq 0, \quad k = 1, 2, \ldots.$$

By [36, Corollary 3.2], we have that $(-X^k, Y^k) \in \widehat{\mathcal{N}}_{\mathrm{gph}\,\mathcal{N}_{\mathbb{S}_+^n}}(\overline{X}^k, \overline{Y}^k)$ for each k with $(X^*, Y^*) = \lim_{k \to \infty}(X^k, Y^k)$. To sum up, there exist $t_k \downarrow 0$ and $(G^k, H^k, X^k, Y^k) \to (G^*, H^*, X^*, Y^*)$ with $(-X^k, Y^k) \in \widehat{\mathcal{N}}_{\mathrm{gph}\,\mathcal{N}_{\mathbb{S}_+^n}}(\overline{X} + t_kG^k, \overline{Y} + t_kH^k)$ for each k. $\qquad\square$