SIAM J. OPTIM.
Vol. 30, No. 2, pp. 1756–1775

SEVERAL CLASSES OF STATIONARY POINTS FOR RANK REGULARIZED MINIMIZATION PROBLEMS*

YULAN LIU†, SHUJUN BI‡, 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 difference-of-convex surrogate yielded by eliminating the dual part of the global exact penalty. A clear relation chart is established among these stationary points, which offers guidance to choose an appropriate reformulation for seeking a low-rank solution to this class of problems.

Key words. rank regularized minimization, stationary points, MPECs, DC surrogate

AMS subject classifications. 90C26, 49J52, 49J53

DOI. 10.1137/19M1270987

1. Introduction.

Let $\mathbb{R}^{m\times n}$ be the linear space of all $m \times n$ $(m \le n)$ 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 a regularization parameter and $\Omega \subseteq \mathbb{R}^{m\times n}$ is a closed convex set. Unless otherwise stated, we assume that f is locally Lipschitzian and is regular at each $X \in \Omega$. Such an optimization model is frequently used to seek a low-rank matrix under the scenario where a tight estimation is unavailable for the rank of the target matrix, and it is found to have a host of applications in a variety of fields such as statistics [25], control and system identification [8, 9], signal and image processing [3], machine learning [30], finance [29], 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 optimum by using an algorithm with a polynomial-time complexity. So, a common way is to obtain a desirable local optimum or a feasible solution by solving a convex relaxation or surrogate problem. Although the nuclear-norm convex relaxation method [7] is very popular, the nuclear norm has a weak ability to promote low-rank solutions and even fails to yield low-rank solutions; see the examples in [22]. After recognizing this deficiency, some researchers have turned their attention to nonconvex surrogates of low-rank optimization problems such as the log-determinant surrogate (see [8, 23]) and the Schatten

*Received by the editors June 27, 2019; accepted for publication (in revised form) April 23, 2020; published electronically June 29, 2020.
https://doi.org/10.1137/19M1270987
Funding: This work is supported by the National Natural Science Foundation of China (11971177 and 11701186), and Guangdong Basic and Applied Basic Research Foundation (2020A1515010408).
†School of Applied Mathematics, Guangdong University of Technology, Guangzhou, China (ylliu@gdut.edu.cn).
‡School of Mathematics, South China University of Technology, Guangzhou, China (bishj@scut.edu.cn).
§Corresponding author. School of Mathematics, South China University of Technology, Guangzhou, China (shhpan@scut.edu.cn).

p $(0 < p < 1)$-norm surrogate [15]. As illustrated in [26], the efficiency of nonconvex surrogates depends on its approximation effect.

Recently, by the variational characterization of the rank function, the authors of [1, 20] reformulated the rank regularized problem (1) as an equivalent mathematical program with an equilibrium constraint (MPEC) and achieved a family of equivalent surrogates by establishing its global exact penalty. To illustrate this, let $\mathscr{L}$ denote the family of proper lower semicontinuous (lsc) convex functions $\phi\colon \mathbb{R} \to (-\infty, +\infty]$ satisfying

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

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

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

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

$$\text{(4)} \qquad \begin{aligned} &\min_{X,W\in\mathbb{R}^{m\times n}} \nu f(X) + \textstyle\sum_{i=1}^m \phi(\sigma_i(W)) + \delta_\Omega(X) \\ &\quad \text{s.t. } \ \|X\|_* - \langle W, X\rangle = 0,\ \|W\| \le 1, \end{aligned}$$

which is an MPEC since the constraints $\|X\|_* - \langle W, X\rangle = 0$ and $\|W\| \le 1$ are equivalent to $X \in \mathcal{N}_{\mathbb{B}}(W)$, the optimality condition of $W \in \arg\max_{Z\in\mathbb{B}}\langle X, Z\rangle$. Here, $\|W\|$ means the spectral norm of W and $\mathbb{B}$ is the unit ball on the spectral norm in $\mathbb{R}^{m\times n}$. Under a suitable condition on Ω, it has been shown in [1, 20] that the problem

$$\text{(5)} \qquad \begin{aligned} &\min_{X,W\in\mathbb{R}^{m\times n}} \nu f(X) + \textstyle\sum_{i=1}^m \phi(\sigma_i(W)) + \rho(\|X\|_* - \langle W, X\rangle) \\ &\quad \text{s.t. } \ X \in \Omega,\ \|W\| \le 1 \end{aligned}$$

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

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

It is well known that an algorithm for nonconvex nonsmooth optimization problems is generally expected to yield a stationary point, and the stationary points of equivalent reformulations may have a big difference. Thus, it is of great value to clarify the relation among the stationary points of (1) defined by its equivalent reformulations. Moreover, such a clarification is a prerequisite to describing the landscape of stationary points for (1). Inspired by this, by the problem (1) itself and its reformulation (4)–(6), in section 3 we introduce the R(egular)-stationary point, the M-stationary point, the EP(exact penalty)-stationary point, and the DC-stationary point, respectively, and explore the relation among the four classes of stationary points. Figure 1 shows that the set of M-stationary points is almost the same as that of R-stationary points; the latter includes those EP-stationary points satisfying the equilibrium constraint, and the set of EP-stationary points coincides with that of DC-stationary points for those $\phi \in \mathscr{L}$ that are nondecreasing on $[0,1]$.

We notice that some active research has been done for the stationary points of zero-norm constrained optimization problems (see, e.g., [2, 10, 27]); for example, Burdakov, Kanzow, and Schwartz [2] discussed the relation between the M-stationary point and the S-stationary point of their equivalent MPEC reformulation, and Pan, Luo, and Xiu [27] characterize 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 that study the stationary points of rank regularized optimization problems. For the special case $\Omega \subseteq \mathbb{S}^n_+$, the rank regularized problem (1) reduces to a mathematical program with semidefinite conic complementarity constraints (MPSCCC) and Ding, Sun, and Ye [5] have established the connection among several classes 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 $\|\cdot\|_F$. 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; we write $\mathbb{O}^{m,n}(X) := \{(U,V) \in \mathbb{O}^m \times \mathbb{O}^n \,|\, X = U[\mathrm{Diag}(\sigma(X)) \quad 0]V^{\mathbb{T}}\}$; and for given 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 $j \in \beta$, and X_β means the submatrix consisting of those columns $X_{\cdot j}$ with $j \in \beta$. We denote by E and e the matrix and the vector of all ones and by I an identity matrix, respectively, whose dimensions are known from the context. For a given set S, δ_S denotes the indicator function of S, i.e., $\delta_S(x) = 0$ if $x \in S$, and otherwise $\delta_S(x) = +\infty$. For a 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

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

and the limiting normal cone to S at $\overline{z}$ is defined as the outer limit of $\widehat{\mathcal{N}}_S(z)$ as $z \xrightarrow[S]{} \overline{z}$:

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

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 $u \in \mathbb{Z}$ is defined by

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

Obviously, $\mathcal{N}_S(\overline{z}; u) = \emptyset$ if $u \notin \mathcal{T}_S(\overline{z})$, $\mathcal{N}_S(\overline{z}; u) \subseteq \mathcal{N}_S(\overline{z})$ and $\mathcal{N}_S(\overline{z}; 0) = \mathcal{N}_S(\overline{z})$. When S is convex and $u \in \mathcal{T}_S(\overline{z})$, it holds that $\mathcal{N}_S(\overline{z}; u) = \mathcal{N}_S(\overline{z}) \cap [\![u]\!]^{\perp} = \mathcal{N}_{\mathcal{T}_S(\overline{z})}(u)$.

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

$$\widehat{\partial} g(\overline{z}) := \bigg\{ z^* \in \mathbb{X} \;\Big|\; \liminf_{\substack{z \to \overline{z} \\ z \neq \overline{z}}} \frac{g(z) - g(\overline{z}) - \langle z^*, z - \overline{z} \rangle}{\|z - \overline{z}\|} \geq 0 \bigg\},$$

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

$$\partial g(\overline{z}) := \Big\{ z^* \in \mathbb{X} \,|\, \exists\, z^k \xrightarrow[g]{} z, z^{k,*} \to z^* \text{ such that } z^{k,*} \in \widehat{\partial} g(z^k) \Big\}. \tag{8}$$

From [32, Theorem 8.9] we know that there is a 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 what follows, we call a point z at which $0 \in \partial g(z)$ (respectively, $0 \in \widehat{\partial} g(z)$) 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 of g.

2.2. Lipschitz-like properties of multifunctions. Let $\mathcal{F} : \mathbb{Z} \rightrightarrows \mathbb{W}$ be a given multifunction. Consider an arbitrary point $(\overline{z}, \overline{w}) \in \text{gph}\mathcal{F}$ at which $\mathcal{F}$ is locally closed, where 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 $(\overline{z}, \overline{w}) \in \text{gph}\mathcal{F}$ 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 a 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 $D^*\mathcal{F}(\overline{z}|\overline{w}) : \mathbb{W} \rightrightarrows \mathbb{Z}$ defined by

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

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

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

LEMMA 2.1 (see [24, 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 [17, 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 [33, Example 2] we have

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

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 [21] the coderivative of the subdifferential mapping $\partial\|\cdot\|_*$. For this purpose, in what follows, for two positive integers k_1 and k_2 with $k_2 \ge k_1$, 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 index sets associated to its singular values,

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

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

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

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

$$\big(\Omega_2\big)_{ij} := \begin{cases} \frac{\min(1,\sigma_i(\overline{Z}))+\min(1,\sigma_j(\overline{Z}))}{\sigma_i(\overline{Z})+\sigma_j(\overline{Z})} & \text{if } \sigma_i(\overline{Z})+\sigma_j(\overline{Z}) \ne 0;\\ 0 & \text{otherwise,}\end{cases} \quad i,j \in \{1,2,\ldots,m\}, \tag{11b}$$

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

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

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

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

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

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

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

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

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

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

LEMMA 2.3 (see [21, Theorem 3.2]). *Fix an arbitrary $(X,W) \in \mathrm{gph}\,\partial\|\cdot\|_*$ 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, $(G,H) \in \mathcal{N}_{\mathrm{gph}\,\partial\|\cdot\|_*}(X,W)$ iff the following relations hold:*

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

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

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

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 the linear mappings defined by

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

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

3. Four classes of stationary points and their relations.

To introduce the stationary points of (1), for each $\phi \in \mathscr{L}$ and the associated ψ, write

$$\widehat{\phi}(t) := \phi(|t|) \quad\text{and}\quad \widehat{\psi}(t) := \psi(|t|) \quad\text{for } t \in \mathbb{R}; \tag{18a}$$

$$\widehat{\Phi}(x) := \textstyle\sum_{i=1}^m \widehat{\phi}(x_i) \quad\text{and}\quad \widehat{\Psi}(x) := \sum_{i=1}^m \widehat{\psi}(x_i) \quad\text{for } x \in \mathbb{R}^m. \tag{18b}$$

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:

$$\|X\|_*-\langle X,W\rangle=0,\|W\|\le 1 \Longleftrightarrow W\in \mathop{\arg\max}_{Z\in\mathbb{B}}\langle Z,X\rangle \Longleftrightarrow X\in\mathcal{N}_{\mathbb{B}}(W) \tag{19a}$$

$$\Longleftrightarrow W\in\partial\|\cdot\|_*(X)\Longleftrightarrow (X,W)\in \operatorname{gph}\partial\|\cdot\|_*. \tag{19b}$$

3.1. R-stationary point. Recall that $\overline{X}\in\mathbb{R}^{m\times n}$ is a regular critical point of F if $0\in\widehat{\partial}F(\overline{X})$ and the rank function is regular by [34, Lemma 2.1] and [18, Corollary 7.5]. Together with the assumption on f and [32, Corollary 10.9], it follows that $\widehat{\partial}F(\overline{X})\supseteq \nu\partial f(\overline{X})+\partial\mathrm{rank}(\overline{X})+\mathcal{N}_\Omega(\overline{X})$. 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 an R-stationary point of the problem* (1) *if*

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

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

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

Thus, when $\partial(\mathrm{rank}+\delta_\Omega)(\overline{X})\subset\partial\mathrm{rank}(\overline{X})+\mathcal{N}_\Omega(\overline{X})$, the limiting critical point of F is the same as its regular critical point and also coincides with the R-stationary point of (1), and in this case every local minimizer of (1) is also an R-stationary point of (1).

3.2. M-stationary point. By invoking the relation (19b), clearly, the MPEC (4) can be compactly written as

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

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

$$\partial\widetilde{F}(X,W)\subseteq\big[\nu\partial f(X)+\mathcal{N}_\Omega(X)\big]\times\partial(\widehat{\Phi}\circ\sigma)(W)+\mathcal{N}_{\operatorname{gph}\partial\|\cdot\|_*}(X,W). \tag{20}$$

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

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*

$$0\in\nu\partial f(\overline{X})+\mathcal{N}_\Omega(\overline{X})+D^*\partial\|\cdot\|_*(\overline{X}|\overline{W})(\Delta W). \tag{21}$$

Remark 3.2. (i) Clearly, if $(\overline{X},\overline{W})$ is a limiting critical point of $\widetilde{F}$, then by (20) under a suitable CQ the matrix $\overline{X}$ is necessarily an M-stationary point of the problem (1).

(ii) Definition 3.2 is an extension of [5, Definition 6.1] to mathematical programs with a general matrix equilibrium constraint. Indeed, when $\Omega\subseteq\mathbb{S}^n_+$, the MPEC (4) is precisely the mathematical program with PSD cone complementarity constraints (SDCMPCC)

$$\begin{aligned}&\min_{X,W\in\mathbb{S}^n} \nu f(X)+\textstyle\sum_{i=1}^m \phi(\sigma_i(W))+\delta_\Omega(X)\\ &\quad\text{s.t. } \langle I-W,X\rangle=0,\ I-W\in\mathbb{S}^n_+,\ W\in\mathbb{S}^n_+.\end{aligned} \tag{22}$$

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

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

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

$$\begin{cases} 0\in\nu\partial f(\overline{X})+\overline{\Gamma}_1+\mathcal{N}_\Omega(\overline{X}), & \text{(23a)}\\ 0\in\partial(\widehat{\Phi}\circ\sigma)(\overline{W})+\overline{\Gamma}_2+\mathcal{N}_{\mathbb{S}^n_+}(\overline{W}), & \text{(23b)}\\ \overline{\Gamma}_1\in D^*\mathcal{N}_{\mathbb{S}^n_+}(\overline{X}|\overline{W}-I)(-\overline{\Gamma}_2). & \text{(23c)}\end{cases}$$

That is, $(\overline{X},\overline{W})$ is an M-stationary point of the problem (22) defined by [5, Definition 6.1] with $(\overline{\Gamma}_1,\overline{\Gamma}_2)$ being the multiplier associated to the constraint $(X,W-I)\in \mathrm{gph}\mathcal{N}_{\mathbb{S}^n_+}$.

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

PROPOSITION 3.1. *If $\overline{X}$ is an M-stationary point of the problem* (1) *associated to $\phi\in\mathscr{L}$, 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 $(\overline{U},\overline{V})\in\mathbb{O}^{m,n}(\overline{Z})$,*

$$\Delta\widetilde{\Gamma}:=\overline{U}^{\mathbb{T}}\Delta\Gamma\overline{V}=\begin{bmatrix}0_{\alpha\alpha} & 0_{\alpha\beta} & 0_{\alpha\gamma} & 0_{\alpha c}\\ 0_{\beta\alpha} & (\Delta\widetilde{\Gamma})_{\beta\beta} & (\Delta\widetilde{\Gamma})_{\beta\gamma} & (\Delta\widetilde{\Gamma})_{\beta c}\\ 0_{\gamma\alpha} & (\Delta\widetilde{\Gamma})_{\gamma\beta} & (\Delta\widetilde{\Gamma})_{\gamma\gamma} & (\Delta\widetilde{\Gamma})_{\gamma c}\end{bmatrix}\quad\text{and}\quad \mathcal{S}[(\Delta\widetilde{\Gamma})_{\beta\beta}]=0. \tag{24}$$

In particular, if $t^*=0$, 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 $\phi\in\mathscr{L}$. 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 (21) 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 next argue that $\Delta\Gamma$ satisfies (24). Since $\overline{W}\in\partial\|\cdot\|_*(\overline{X})$, from (19a) we have $\overline{W}=\Pi_{\mathbb{B}}(\overline{Z})$. Together with $(\overline{U},\overline{V})\in\mathbb{O}^{m,n}(\overline{Z})$, it follows that $\overline{W}$ and $\overline{X}$ have the following form:

$$\overline{W}=\overline{U}\begin{bmatrix}I_{\alpha\alpha} & 0_{\alpha\beta} & 0_{\alpha\gamma} & 0_{\alpha c}\\ 0_{\beta\alpha} & I_{\beta\beta} & 0_{\beta\gamma} & 0_{\beta c}\\ 0_{\gamma\alpha} & 0_{\gamma\beta} & \mathrm{Diag}(\sigma_\gamma(\overline{Z})) & 0_{\gamma c}\end{bmatrix}\overline{V}^{\mathbb{T}}, \tag{25a}$$

$$\overline{X}=\overline{U}\begin{bmatrix}\mathrm{Diag}(\sigma_\alpha(\overline{Z})-e) & 0_{\alpha\beta} & 0_{\alpha\gamma} & 0_{\alpha c}\\ 0_{\beta\alpha} & 0_{\beta\beta} & 0_{\beta\gamma} & 0_{\beta c}\\ 0_{\gamma\alpha} & 0_{\gamma\beta} & 0_{\gamma\gamma} & 0_{\gamma c}\end{bmatrix}\overline{V}^{\mathbb{T}}. \tag{25b}$$

Since $\widehat{\Phi}$ is absolutely symmetric and $\Delta W\in\partial(\widehat{\Phi}\circ\sigma)(\overline{W})$, by [18, Theorem 7.1] and (25a) 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

$$\Delta W = \widehat{U}[\mathrm{Diag}(\overline{w}) \;\; 0]\widehat{V}^{\mathbb{T}}. \tag{26}$$

From $\overline{w} \in \partial\widehat{\Phi}(\sigma(\overline{W}))$, the definition of $\widehat{\Phi}$, and (25a), it follows that

$$\overline{w}_i \in \partial\phi(1) \text{ for } i \in \alpha\cup\beta;\;\; \overline{w}_i \in \partial\phi(\sigma_i(\overline{Z})) \text{ for } i \in \gamma_1;\;\; \overline{w}_i \in \partial\widehat{\phi}(0) \text{ for } i \in \gamma_0. \tag{27}$$

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

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

Since $(\widehat{U},\widehat{V}) \in \mathbb{O}^{m,n}(\overline{W})$, from (25a) 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

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

Together with (26) and (27), it is not difficult to obtain that

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

and consequently

$$\Delta\widetilde{W} := \overline{U}^{\mathbb{T}}\Delta W\overline{V} = \begin{bmatrix}\mathrm{Diag}(\overline{w}_{\alpha\cup\beta\cup\gamma_1}) & 0\\ 0 & Q'[\mathrm{Diag}(\overline{w}_{\gamma_0}) \;\; 0](Q'')^{\mathbb{T}}\end{bmatrix}. \tag{28}$$

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

$$\Theta_1 \circ \mathcal{S}(\Delta\widetilde{W}_1) + \Theta_2 \circ \mathcal{S}(\Delta\widetilde{\Gamma}_1) + \Sigma_1 \circ \mathcal{X}(\Delta\widetilde{W}_1) + \Sigma_2 \circ \mathcal{X}(\Delta\widetilde{\Gamma}_1) = 0, \tag{29a}$$

$$(\Delta\widetilde{\Gamma})_{\alpha c} + (\Omega_3)_{\alpha c} \circ [(\Delta\widetilde{W})_{\alpha c} - (\Delta\widetilde{\Gamma})_{\alpha c}] = 0,\; (\Delta\widetilde{W})_{\beta c} = 0,\;\; (\Delta\widetilde{W})_{\gamma c} = 0, \tag{29b}$$

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 $[\Delta\widetilde{W}_1]_{JJ}$ with $J = \alpha\cup\beta\cup\gamma_1$ is a diagonal matrix by (28). Together with (29a)–(29b) and (11a)–(11c), it follows that

$$(\Delta\widetilde{W})_{\alpha c} = 0,\; (\Delta\widetilde{\Gamma})_{\alpha c} = 0,\; (\Delta\widetilde{W})_{\gamma\gamma} = 0, \tag{30a}$$

$$[\mathcal{S}(\Delta\widetilde{\Gamma}_1)]_{\alpha\alpha} + (\Sigma_2)_{\alpha\alpha} \circ [\mathcal{X}(\Delta\widetilde{\Gamma}_1)]_{\alpha\alpha} = 0, \tag{30b}$$

$$(\Theta_2)_{\alpha\beta} \circ [\mathcal{S}(\Delta\widetilde{\Gamma}_1)]_{\alpha\beta} + (\Sigma_2)_{\alpha\beta} \circ [\mathcal{X}(\Delta\widetilde{\Gamma}_1)]_{\alpha\beta} = 0, \tag{30c}$$

$$(\Theta_2)_{\beta\alpha} \circ [\mathcal{S}(\Delta\widetilde{\Gamma}_1)]_{\beta\alpha} + (\Sigma_2)_{\beta\alpha} \circ [\mathcal{X}(\Delta\widetilde{\Gamma}_1)]_{\beta\alpha} = 0, \tag{30d}$$

$$(\Theta_2)_{\alpha\gamma} \circ [\mathcal{S}(\Delta\widetilde{\Gamma}_1)]_{\alpha\gamma} + (\Sigma_2)_{\alpha\gamma} \circ [\mathcal{X}(\Delta\widetilde{\Gamma}_1)]_{\alpha\gamma} = 0, \tag{30e}$$

$$(\Theta_2)_{\gamma\alpha} \circ [\mathcal{S}(\Delta\widetilde{\Gamma}_1)]_{\gamma\alpha} + (\Sigma_2)_{\gamma\alpha} \circ [\mathcal{X}(\Delta\widetilde{\Gamma}_1)]_{\gamma\alpha} = 0. \tag{30f}$$

Notice that (30b) is equivalent to $(E+\Sigma_2)_{\alpha\alpha} \circ (\Delta\widetilde{\Gamma}_1)_{\alpha\alpha} + (E-\Sigma_2)_{\alpha\alpha} \circ (\Delta\widetilde{\Gamma}_1^{\mathbb{T}})_{\alpha\alpha} = 0$ which, by the fact that the entries of Σ_2 belong to $(0,1)$, implies that $(\Delta\widetilde{\Gamma}_1)_{\alpha\alpha} = 0$. Notice that (30c) and (30d) can be equivalently written as

$$(\Theta_2+\Sigma_2)_{\alpha\beta} \circ (\Delta\widetilde{\Gamma}_1)_{\alpha\beta} = (\Sigma_2-\Theta_2)_{\alpha\beta} \circ (\Delta\widetilde{\Gamma}_1^{\mathbb{T}})_{\alpha\beta}, \tag{31a}$$

$$(\Theta_2+\Sigma_2)_{\beta\alpha} \circ (\Delta\widetilde{\Gamma}_1)_{\beta\alpha} = (\Sigma_2-\Theta_2)_{\beta\alpha} \circ (\Delta\widetilde{\Gamma}_1^{\mathbb{T}})_{\beta\alpha}. \tag{31b}$$

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 (31b) we immediately obtain that

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

where "$\oslash$" denotes the entries division operator of two matrices. Substituting this equality into (31a) yields that $(\Delta\widetilde{\Gamma}_1)_{\alpha\beta} = 0$, and then $(\Delta\widetilde{\Gamma}_1)_{\beta\alpha} = 0$. Similarly, from (30e) and (30f), we obtain $(\Delta\widetilde{\Gamma}_1)_{\alpha\gamma} = 0$ and $(\Delta\widetilde{\Gamma}_1)_{\beta\gamma} = 0$. The above arguments show that

$$\Delta\widetilde{\Gamma} = \begin{bmatrix} 0_{\alpha\alpha} & 0_{\alpha\beta} & 0_{\alpha\gamma} & 0_{\alpha c} \\ 0_{\beta\alpha} & (\Delta\widetilde{\Gamma})_{\beta\beta} & (\Delta\widetilde{\Gamma})_{\beta\gamma} & (\Delta\widetilde{\Gamma})_{\beta c} \\ 0_{\gamma\alpha} & (\Delta\widetilde{\Gamma})_{\gamma\beta} & (\Delta\widetilde{\Gamma})_{\gamma\gamma} & (\Delta\widetilde{\Gamma})_{\gamma c} \end{bmatrix}.$$

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 $\Xi_1 \in \mathcal{U}_{|\beta|}$ of the form (14) for some partition $(\beta_+, \beta_0, \beta_-)$ of β such that

$$\Xi_1 \circ Q^{\mathbb{T}}(\Delta\widetilde{W})_{\beta\beta}Q + \Xi_2 \circ \mathcal{S}\big[Q^{\mathbb{T}}(\Delta\widetilde{\Gamma})_{\beta\beta}Q\big] + \Xi_2 \circ \mathcal{X}\big[Q^{\mathbb{T}}(\Delta\widetilde{W})_{\beta\beta}Q\big] = 0, \tag{32}$$

$$Q_{\beta_0}^{\mathbb{T}}(\Delta\widetilde{\Gamma})_{\beta\beta}Q_{\beta_0} \succeq 0,\ Q_{\beta_0}^{\mathbb{T}}(\Delta\widetilde{W})_{\beta\beta}Q_{\beta_0} \preceq 0, \tag{33}$$

where the matrix Ξ_2 associated with Ξ_1 has the form of (15). From (28) and the first equality in (27), $(\Delta\widetilde{W})_{\beta\beta} = \mathrm{Diag}(\overline{w}_\beta)$. Notice that $\partial\phi(1) = [\phi'_-(1), \phi'_+(1)]$ and $t^* \in [0, 1)$ is the unique minimum of ϕ over $[0, 1]$. We have $\overline{w}_\beta \in \mathbb{R}_{++}^{|\beta|}$, which along with the second inequality of (33) implies $\beta_0 = \emptyset$. Since $\mathcal{X}[Q^{\mathbb{T}}(\Delta\widetilde{W})_{\beta\beta}Q] = 0$, (32) 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 $Q^{\mathbb{T}}\mathrm{Diag}(\overline{w}_\beta)Q \in \mathrm{int}(\mathbb{S}_+^{|\beta|})$, by 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. Now we assume that $t^* = 0$. Notice that $\overline{w}_{\gamma_1} \in \mathbb{R}_{++}^{|\gamma_1|}$ by (27) since $t^* = 0$ is the unique minimum of ϕ in $[0, 1]$. Together with $(\Delta\widetilde{W})_{\gamma\gamma} = 0$ and (28), we deduce that $\gamma_1 = \emptyset$. If $0 \notin \partial\widehat{\phi}(0)$, then $\overline{w}_{\gamma_0} \neq 0$ by (28). Together with $(\Delta\widetilde{W})_{\gamma\gamma} = 0$ and (28), we deduce that $\gamma_0 = \emptyset$. The proof is completed. ∎

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 $\phi \in \mathscr{L}$, then it is an R-stationary point. Conversely, if $\overline{X}$ is an 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 $\phi \in \mathscr{L}$. 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 $\Delta\Gamma$ satisfies (24). 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 an arbitrary $(P, P') \in \mathbb{O}^{m-|\alpha|, n-|\alpha|}(\Delta\widetilde{Z})$. Write $\widetilde{U} = [\overline{U}_\alpha \quad \overline{U}_{\beta\cup\gamma}P]$ and $\widetilde{V} = [\overline{V}_\alpha \quad \overline{V}_{\beta\cup\gamma\cup c}P']$. 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 (25b), 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 [16, 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 an R-stationary point.

Now let $\overline{X}$ be an R-stationary point of (1) with $\mathrm{rank}(\overline{X})=\overline{r}$. Suppose that $\overline{r}\ge 1$. 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 [16, Theorem 4], there exists $(\overline{U},\overline{V})\in\mathbb{O}^{m,n}(\overline{X})$ such that

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

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

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

$$\overline{w}_i=\phi'_-(1)\ \text{ for } i\in\alpha\ \text{ and }\ \overline{w}_i=0\in\partial\widehat{\phi}(0)\ \text{ for } i\in\gamma_0. \tag{34}$$

Since ϕ is convex, from [32, Proposition 10.19(a)] it follows that $\overline{w}_i\in\partial\widehat{\phi}(1)$ for $i\in\alpha$. Then $\Delta W=\overline{U}[\mathrm{Diag}(\overline{w})\ \ 0]\overline{V}^{\mathbb{T}}\in\partial(\widehat{\Phi}\circ\sigma)(\overline{W})$. 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_1 and Θ_2 defined as in section 2.3 and $\beta=\emptyset=\gamma_1$, it is easy to verify that $(-\Delta\widetilde{\Gamma},-\Delta\widetilde{W})$ satisfies

$$\Theta_1\circ\mathcal{S}(\Delta\widetilde{W}_1)+\Theta_2\circ\mathcal{S}(\Delta\widetilde{\Gamma}_1)+\Sigma_1\circ\mathcal{X}(\Delta\widetilde{W}_1)+\Sigma_2\circ\mathcal{X}(\Delta\widetilde{\Gamma}_1)=0, \tag{35a}$$

$$(\Delta\widetilde{\Gamma})_{\alpha c}+(\Omega_3)_{\alpha c}\circ[(\Delta\widetilde{W})_{\alpha c}-(\Delta\widetilde{\Gamma})_{\alpha c}]=0,\ (\Delta\widetilde{W})_{\beta c}=0,\ (\Delta\widetilde{W})_{\gamma c}=0. \tag{35b}$$

Since $\beta=\emptyset$, from Lemma 2.3 it follows that $(-\Delta\Gamma,-\Delta W)\in\mathcal{N}_{\mathrm{gph}\,\partial\|\cdot\|_*}(\overline{X},\overline{W})$, 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 $\Delta W=\overline{U}[\mathrm{Diag}(\overline{w})\ \ 0]\overline{V}^{\mathbb{T}}$ with

$$\overline{w}_i=\phi'_-(1)\ \text{ for } i\in\alpha\ \text{ and }\ \overline{w}_i=0\in\partial\phi(t^*)\ \text{ for } i\in\gamma_1. \tag{36}$$

Using the same arguments as those for Case 1, one 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 an M-stationary point associated to ϕ. The proof is then completed. ∎

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 of (1) associated to ϕ. By Proposition A.1 in the appendix, if $(\overline{X},\overline{W})$ is a local optimum of (4) with $\mathrm{rank}(\overline{X})=$

$\sum_{i=1}^{m}\phi(\sigma_i(\overline{W}))$, then $\overline{X}$ is a local minimizer of the problem (1), and conversely, if $\overline{X}$ is a local minimum of (1), then $(\overline{X},\overline{W})$ with $\overline{W}=\overline{U}_1\overline{V}_1^{\mathbb{T}}+t^*\overline{U}_2[\mathrm{Diag}(e)\ 0]\overline{V}_2^{\mathbb{T}}$ for $(\overline{U},\overline{V})\in\mathbb{O}^{m,n}(\overline{X})$ is locally optimal to (4), where $\overline{U}=[\overline{U}_1\ \ \overline{U}_2]$ and $\overline{V}=[\overline{V}_1\ \ \overline{V}_2]$ with $\overline{U}_1$ and $\overline{V}_1$ being the matrix consisting of the first $\overline{r}=\mathrm{rank}(\overline{X})$ columns of $\overline{U}$ and $\overline{V}$, respectively.

PROPOSITION 3.2. *Let $(\overline{X},\overline{W})$ be a local minimum 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, which is also an R-stationary point of the problem* (1), *provided that the following inclusion holds:*

$$\mathcal{N}_{(\Omega\times\mathbb{R}^{m\times n})\cap\mathrm{gph}\partial\|\cdot\|_*}(\overline{X},\overline{W})\subseteq\mathcal{N}_{\mathrm{gph}\partial\|\cdot\|_*}(\overline{X},\overline{W})+\mathcal{N}_{\Omega}(\overline{X})\times\{0\}. \tag{37}$$

Proof. By invoking the relation (19b), (X,W) is a feasible point of (4) iff $(X,W)\in(\Omega\times\mathbb{R}^{m\times n})\cap\mathrm{gph}\partial\|\cdot\|_*$. This implies that (4) can be compactly written as

$$\min_{X,W\in\mathbb{R}^{m\times n}}\Big\{\nu f(X)+\widehat{\Phi}(\sigma(W))+\delta_{(\Omega\times\mathbb{R}^{m\times n})\cap\mathrm{gph}\partial\|\cdot\|_*}(X,W)\Big\}.$$

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

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

Together with the inclusion (37), it follows that

$$(0,0)\in\big[\nu\partial f(\overline{X})+\mathcal{N}_{\Omega}(\overline{X})\big]\times\{0\}+\{0\}\times\partial(\widehat{\Phi}\circ\sigma)(\overline{W})+\mathcal{N}_{\mathrm{gph}\partial\|\cdot\|_*}(\overline{X},\overline{W}), \tag{38}$$

which is equivalent to saying that there exists $(\Delta X,-\Delta W)\in\mathcal{N}_{\mathrm{gph}\partial\|\cdot\|_*}(\overline{X},\overline{W})$ 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 $(\Delta X,-\Delta W)\in\mathcal{N}_{\mathrm{gph}\partial\|\cdot\|_*}(\overline{X},\overline{W})$ iff $\Delta X\in D^*\partial\|\cdot\|_*(\overline{X}|\overline{W})(\Delta W)$. So, (38) 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 $\overline{W}\in\partial\|\cdot\|_*(\overline{X})$ by (19b). Thus, by Definition 3.2, $\overline{X}$ is an M-stationary point of the problem (1) associated to ϕ, which is also an R-stationary point of (1) by Theorem 3.1. The proof is completed. ∎

Remark 3.3. (i) If $\Omega=\mathbb{R}^{m\times n}$, the inclusion (37) automatically holds. If $\Omega\subset\mathbb{R}^{m\times n}$, by [13, p. 211] the inclusion (37) is implied by the calmness of the multifunction

$$\mathcal{M}(Y_1,Y_2):=\Big\{(X,W)\in\Omega\times\mathbb{R}^{m\times n}\mid(Y_1,Y_2)\in-(X,W)+\mathrm{gph}\partial\|\cdot\|_*\Big\}$$

at the origin for $(\overline{X},\overline{W})$, where $(\overline{X},\overline{W})$ is a local minimizer of the MPEC (4).

(ii) When $\Omega\subseteq\mathbb{S}_+^n$, to achieve the conclusion of Proposition 3.2 for the SDCMPCC (22) at a local minimizer $(\overline{X},\overline{W})$, we need to replace the inclusion (37) by the following one:

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

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

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

or equivalently, there exist $\delta>0$ and $\gamma>0$ such that for all $(X,W)\in\mathbb{B}_\delta(\overline{X},\overline{W})$,

$$\mathrm{dist}((X,W),C)\le\gamma\big[\mathrm{dist}((X,W),\Omega\times\mathbb{S}_+^n)+\mathrm{dist}((X,W-I),\mathrm{gph}\mathcal{N}_{\mathbb{S}_+^n})\big].$$

It is not hard to check that the calmness of $\mathcal{M}$ at the origin for $(\overline{X},\overline{W})$ is equivalent to the calmness of the following multifunction at the origin for $(\overline{X},\overline{W})$:

$$\widetilde{\mathcal{M}}(Z_1,Z_2,Y_1,Y_2):=\Big\{(X,W)\in\mathbb{S}^n\times\mathbb{S}^n\mid Z_1\in-X+\Omega,Z_2\in-W+\mathbb{S}_+^n,\\ (Y_1,Y_2)\in-(X,W-I)+\mathrm{gph}\mathcal{N}_{\mathbb{S}_+^n}\Big\}.$$

By invoking [11, Corollary 1], the multifunction $\widetilde{M}$ is calm at the origin for $(\overline{X},\overline{W})$ if for any $0\ne(H_1,H_2)\in\mathcal{T}_\Omega(\overline{X})\times\mathcal{T}_{\mathbb{S}_+^n}(\overline{W})$ with $(H_1,H_2)\in\mathcal{T}_{\mathrm{gph}\mathcal{N}_{\mathbb{S}_+^n}}(\overline{X},\overline{W}-I)$,

$$\left.\begin{array}{c}\Gamma_1\in-\mathcal{N}_\Omega(\overline{X};H_1),\ \Gamma_2\in-\mathcal{N}_{\mathbb{S}_+^n}(\overline{W};H_2),\\ (\Gamma_1,\Gamma_2)\in\mathcal{N}_{\mathrm{gph}\mathcal{N}_{\mathbb{S}_+^n}}((\overline{X},\overline{W}-I);(H_1,H_2))\end{array}\right\}\Longrightarrow\begin{pmatrix}\Gamma_1\\ \Gamma_2\end{pmatrix}=0. \tag{39}$$

3.3. EP-stationary points. By the definition of $\widehat{\Psi}$, the exact penalty problem (5) can be compactly written as

$$\min_{X,W\in\mathbb{R}^{m\times n}}\Big\{\nu f(X)+\delta_\Omega(X)+\widehat{\Psi}(\sigma(W))+\rho(\|X\|_*-\langle W,X\rangle)\Big\}. \tag{40}$$

By the given assumption on f, if $(\overline{X},\overline{W})$ is a limiting critical point of (40), it holds that

$$\rho\overline{X}\in\partial(\widehat{\Psi}\circ\sigma)(\overline{W})\ \text{ and }\ 0\in\nu\partial f(\overline{X})+\mathcal{N}_\Omega(\overline{X})+\rho\big[\partial\|\cdot\|_*(\overline{X})-\overline{W}\,\big]. \tag{41}$$

In view of this, we introduce the following EP-stationary point of (1).

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* (41) *hold.*

Clearly, every local optimum of (5) is an EP-stationary point of (1) associated to ϕ. By Proposition A.2 in the appendix, every local optimum of (5) satisfying the constraint $\|X\|_*-\langle X,W\rangle=0$ is necessarily a local optimum of the MPEC (4). The following proposition states an important property of the EP-stationary point, which is the key to disclose the relation between the EP-stationary point and the R-stationary point.

PROPOSITION 3.3. *Let $\mathscr{L}_1$ denote the family of those $\phi\in\mathscr{L}$ that are nondecreasing on $[0,1]$. Suppose that $\overline{X}\in\mathbb{R}^{m\times n}$ is an EP-stationary point of the problem* (1) *associated to $\phi\in\mathscr{L}_1$. Then, there exist $\rho>0,\overline{W}\in\mathbb{B}$, and $(\overline{U},\overline{V})\in\mathbb{O}^{m,n}(\overline{W})\cap\mathbb{O}^{m,n}(\overline{X})$ such that* $\mathrm{rank}(\overline{X})\ge|J_1|+|J_2|$ *with $J_1:=\{i\mid\sigma_i(\overline{W})=1\}$ and $J_2:=\big\{i\mid\sigma_i(\overline{W})\in(0,1)\big\}$, and moreover, there also exists $\Delta\Gamma\in\nu\partial f(\overline{X})+\mathcal{N}_\Omega(\overline{X})$ such that*

$$\Delta\Gamma\in\left\{\overline{U}\begin{bmatrix}0&0\\0&\rho Z\end{bmatrix}\overline{V}^{\mathbb{T}}\;\Big|\;Z\in\mathbb{R}^{(m-|J_1|)\times(n-|J_1|)}\text{ with }\|Z\|\le1\right\}. \tag{42}$$

Proof. Let $\overline{X}$ be an EP-stationary point of the problem (1) associated to $\phi \in \mathscr{L}_1$. Then there exist a constant $\rho > 0$ and a matrix $\overline{W} \in \mathbb{B}$ such that the inclusions in (41) hold. Since ψ is a closed proper convex function that is nondecreasing on $[0,1]$, it follows that $\widehat{\psi}$ is a closed proper convex function. Together with its absolute symmetry, by [19, Corollary 2.6] we know that $\widehat{\Psi}\circ\sigma$ is closed proper and convex. Since $\rho\overline{X} \in \partial(\widehat{\Psi}\circ\sigma)(\overline{W})$, by invoking [19, Corollary 2.5] there exists $(\overline{U},\overline{V}) \in \mathbb{O}^{m,n}(\overline{W})$ such that

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

From (2) we know that $t^* = 0$ is the unique minimum of ψ, which implies that $\partial\psi(t) \subset (0,+\infty]$ for any $t > 0$. So, $\mathrm{rank}(\overline{X}) \geq |J_1| + |J_2|$ and the first part follows.

Since $(\overline{X},\overline{W})$ satisfies the second inclusion of (41), 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 $\overline{r} = \mathrm{rank}(\overline{X})$. By the last equation and (9),

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

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 matrix consisting of the last $m-\overline{r}$ and $n-\overline{r}$ columns of $\overline{U}$ and $\overline{V}$, respectively. Together with $\overline{W} = \overline{U}[\mathrm{Diag}(\sigma(\overline{W}))\ \ 0]\overline{V}^{\mathbb{T}}$, it immediately follows that

$$\Delta\Gamma = \overline{U}\begin{bmatrix}0 & 0\\ 0 & \rho Z\end{bmatrix}\overline{V}^{\mathbb{T}} \ \text{ with } Z = [\mathrm{Diag}(\sigma_{J_2\cup J_0}(\overline{W}))\ \ 0] - \begin{bmatrix}\mathrm{Diag}(e_{\{1,\ldots,\overline{r}\}\setminus J_1}) & 0\\ 0 & \Gamma\end{bmatrix},$$

where $J_0 := \big\{i \mid \sigma_i(\overline{W}) = 0\big\}$. Since $\overline{r} \geq |J_1| + |J_2|$, the spectral norm of Z satisfies

$$\|Z\| \leq \max\left(\left\|\begin{bmatrix}\mathrm{Diag}(e_{\{1,\ldots,\overline{r}\}\setminus J_1\cup J_2}) & 0\\ 0 & \Gamma\end{bmatrix}\right\|, \|e - \sigma_{J_2}(\overline{W})\|\right) \leq 1.$$

Thus, we obtain the desired inclusion. The proof is then completed. □

THEOREM 3.2. *If $\overline{X}$ is an EP-stationary point of the problem* (1) *associated to $\phi \in \mathscr{L}_1$ and the associated $\overline{W} \in \mathbb{B}$ is such that $\|\overline{X}\|_* - \langle\overline{X},\overline{W}\rangle = 0$, then $\overline{X}$ is an R-stationary point of* (1). *Conversely, if $\overline{X}$ is an R-stationary point of* (1), *then $\overline{X}$ is an EP-stationary point associated to those $\phi \in \mathscr{L}$ with $0 \in \partial\widehat{\psi}(0)$ and $\partial\widehat{\psi}(1) \subseteq [c,+\infty]$ for some $c > 0$.*

Proof. Let $\overline{X}$ be an EP-stationary point of (1) associated to $\phi \in \mathscr{L}_1$. By the assumption and Proposition 3.3, there exist $\rho > 0$, $\overline{W} \in \mathbb{B}$, and $(\overline{U},\overline{V}) \in \mathbb{O}^{m,n}(\overline{W}) \cap \mathbb{O}^{m,n}(\overline{X})$ such that $\mathrm{rank}(\overline{X}) \geq |J_1| + |J_2|$, and there is $Z \in \mathbb{R}^{(m-|J_1|)\times(n-|J_1|)}$ with $\|Z\| \leq 1$ such that

$$\Delta\Gamma \in \nu\partial f(\overline{X}) + \mathcal{N}_\Omega(\overline{X}) \ \text{ and } \ \Delta\Gamma = \overline{U}\begin{bmatrix}0 & 0\\ 0 & \rho Z\end{bmatrix}\overline{V}^{\mathbb{T}},$$

where J_1 and J_2 are the index sets defined as in Proposition 3.3. Since $\|\overline{X}\|_* - \langle\overline{X},\overline{W}\rangle = 0$, we deduce that $\mathrm{rank}(\overline{X}) = |J_1|$. Take an arbitrary $(P,P') \in \mathbb{O}^{m-|J_1|,n-|J_1|}(\rho Z)$. Write $\widetilde{U} = [\overline{U}_1\ \ \overline{U}_2P]$ and $\widetilde{V} = [\overline{V}_1\ \ \overline{V}_2P']$, where $\overline{U}_1$ and $\overline{V}_1$ are the matrix consisting of the first $|J_1|$ columns of $\overline{U}$ and $\overline{V}$, respectively, and $\overline{U}_2$ and $\overline{V}_2$ are the matrix consisting of other $m-|J_1|$ and $n-|J_1|$ columns of $\overline{U}$ and $\overline{V}$, respectively. It is easy to check that $\Delta\Gamma = \widetilde{U}\left[\begin{smallmatrix}0 & 0\\ 0 & [\mathrm{Diag}(\sigma(\rho Z))\ \ 0]\end{smallmatrix}\right]\widetilde{V}^{\mathbb{T}}$. Since

$\mathrm{rank}(\overline{X}) = |J_1|$, we have $(\widetilde{U}, \widetilde{V}) \in \mathbb{O}^{m,n}(\overline{X})$. By [16, Theorem 4], $-\Delta\Gamma \in \partial\mathrm{rank}(\overline{X})$. Thus, $0 \in \nu\partial f(\overline{X}) + \partial\mathrm{rank}(\overline{X}) + \mathcal{N}_\Omega(\overline{X})$, which implies that $\overline{X}$ is an R-stationary point of the problem (1).

Now let $\overline{X}$ be an R-stationary point of (1). Write $\overline{r} = \mathrm{rank}(\overline{X})$. By Definition 3.1 and [16, Theorem 4], there exist $x^* \in \mathbb{R}^m$ with $x_i^* = 0$ for $i = 1, \ldots, \overline{r}$ and $(\overline{U}, \overline{V}) \in \mathbb{O}^{m,n}(\overline{X})$ such that $Z := \overline{U}[\mathrm{Diag}(x^*) \ \ 0]\overline{V}^{\mathbb{T}} \in \partial\mathrm{rank}(\overline{X})$ and $-Z \in \nu\partial f(\overline{X}) + \mathcal{N}_\Omega(\overline{X})$. Take

$$\rho = \max\left(c/\sigma_{\overline{r}}(\overline{X}), \|x^*\|_\infty\right) \ \text{ and } \ \overline{W} = \overline{U}_1\overline{V}_1^{\mathbb{T}},$$

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. Since $x_i^* = 0$ for $i = 1, \ldots, \overline{r}$, it holds that $\|\rho\overline{W} - \overline{U}[\mathrm{Diag}(x^*) \ \ 0]\overline{V}^{\mathbb{T}}\| \le \rho$, which by the expression of $\rho\partial\|\cdot\|_*(\overline{X})$ implies that $\rho\overline{W} - \overline{U}[\mathrm{Diag}(x^*) \ \ 0]\overline{V}^{\mathbb{T}} \in \rho\partial\|\cdot\|_*(\overline{X})$. Then, $-Z \in \rho[\partial\|\cdot\|_*(\overline{X}) - \overline{W}\,]$. Together with $-Z \in \nu\partial f(\overline{X}) + \mathcal{N}_\Omega(\overline{X})$, we obtain

$$0 \in \nu\partial f(\overline{X}) + \mathcal{N}_\Omega(\overline{X}) + \rho[\partial\|\cdot\|_*(\overline{X}) - \overline{W}\,].$$

Take an arbitrary $\phi \in \mathscr{L}$ with $0 \in \partial\widehat{\psi}(0)$ and $\partial\widehat{\psi}(1) \subseteq [c, +\infty]$. Then, $\partial\widehat{\psi}(\sigma_i(\overline{W})) \subseteq [c, +\infty]$ for $i = 1, \ldots, \overline{r}$ and $0 \in \partial\widehat{\psi}(\sigma_i(\overline{W}))$ for $i = \overline{r}+1, \ldots, m$. From [18, Theorem 7.1] and the choice of ρ, it follows that $\rho\overline{X} \in \partial(\widehat{\Psi} \circ \sigma)(\overline{W})$. Together with the last equation, we conclude that $\overline{X}$ is an EP-stationary point of (1) associated to ϕ. ∎

3.4. DC-stationary point. With the conjugate $\widehat{\Psi}^*$ of $\widehat{\Psi}$, the surrogate problem (6) can be equivalently written as

$$\min_{X \in \mathbb{R}^{m\times n}} \Big\{\nu f(X) + \delta_\Omega(X) + \rho\|X\|_* - \widehat{\Psi}^*(\rho\sigma(X))\Big\}. \tag{43}$$

By [19, Lemma 2.3], we know that $\widehat{\Psi}^*$ is also absolutely symmetric. Along with its lsc and convexity, from [19, 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 (43).

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*

$$0 \in \nu\partial f(\overline{X}) + \mathcal{N}_\Omega(\overline{X}) + \rho\partial\|\cdot\|_*(\overline{X}) - \rho\partial(\widehat{\Psi}^* \circ \sigma)(\rho\overline{X}). \tag{44}$$

When f is convex, the problem (43) is a DC program, and now $\overline{X}$ is a DC-stationary point of (1) iff it is a critical point of the objective function of (43) defined by Pang, Razaviyayn, and Alvarado [28]. It is worthwhile to point out that the limiting critical point of the objective function of (43) is a DC-stationary point of (1), but the converse does not hold. Indeed, by letting $\overline{X}$ be a limiting critical point of the objective function of (43), we have

$$\begin{aligned}
0 &\in \partial\big[\nu f(\cdot) + \delta_\Omega(\cdot) + \rho\|\cdot\|_* - \widehat{\Psi}^*(\rho\sigma(\cdot))\big](\overline{X}) \\
&\subseteq \partial\big(\nu f(\cdot) + \rho\|\cdot\|_* - \widehat{\Psi}^*(\rho\sigma(\cdot))\big)(\overline{X}) + \mathcal{N}_\Omega(\overline{X}) \\
&\subseteq \nu\partial f(\overline{X}) + \mathcal{N}_\Omega(\overline{X}) + \rho\partial\|\overline{X}\|_* - \partial^C\widehat{\Psi}^*\big(\rho\sigma(\cdot)\big)(\overline{X}) \\
&= \nu\partial f(\overline{X}) + \mathcal{N}_\Omega(\overline{X}) + \rho\partial\|\overline{X}\|_* - \partial\widehat{\Psi}^*\big(\rho\sigma(\overline{X})\big),
\end{aligned}$$

where $\partial^C\widehat{\Psi}^*(\rho\sigma(\cdot))(\overline{X})$ denotes the Clarke subdifferential of $Z \mapsto \widehat{\Psi}^*(\rho\sigma(Z))$ at $\overline{X}$, and the first inclusion is due to the Lipschitiz continuity of $Z \mapsto \nu f(Z) + \rho\|Z\|_* -$

$\widehat{\Psi}^*(\rho\sigma(Z))$ and the Lipschitz continuity of $\widehat{\Psi}^*(\rho\sigma(\cdot))$ is implied by its finite convexity. The above inclusions also show that each local optimum of (6) is a DC-stationary point of (1). Next we focus on the relation between the DC-stationary point and the EP-stationary point.

THEOREM 3.3. *Choose an arbitrary $\phi \in \mathscr{L}_1$. Then, $\overline{X}$ is a DC-stationary point associated to ϕ iff $\overline{X}$ is an EP-stationary point associated to ϕ.*

Proof. $\Longrightarrow$. Since $\overline{X}$ is a DC-stationary point of the problem (1), by Definition 3.4 there exists a constant $\rho > 0$ such that (44) holds. So, there is $\overline{W} \in \partial(\widehat{\Psi}^* \circ \sigma)(\rho\overline{X})$ such that

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

Since ψ is a closed proper convex function, we have range $\partial\widehat{\psi}^* \subseteq \mathrm{dom}\widehat{\psi} = [-1, 1]$ by [31, section 23], which along with $\sigma_i(\overline{W}) \in \partial\widehat{\psi}^*(\sigma_i(\rho\overline{X}))$ for $i = 1, 2, \ldots, m$ implies that $\|\overline{W}\| \le 1$. Let $F(X) := (\widehat{\Psi} \circ \sigma)(X)$ for $X \in \mathbb{R}^{m\times n}$. By using the von Neumann's trace inequality, it is easy to verify that $F^*(Z) = \sup_{X\in\mathbb{R}^{m\times n}} \big\{\langle Z, X\rangle - F(X)\big\} = (\widehat{\Psi}^* \circ \sigma)(Z)$ for $Z \in \mathbb{R}^{m\times n}$. This shows that $\overline{W} \in \partial(\widehat{\Psi}^* \circ \sigma)(\rho\overline{X})$ iff $\overline{W} \in \partial F^*(\rho\overline{X})$. Notice that ψ is nondecreasing, lsc, and convex. Hence, $\widehat{\psi}$ is lsc and convex, which along with its absolute symmetry implies that $\widehat{\Psi}$ is absolutely symmetric, lsc, and convex. From [19, Corollaries 2.5 and 2.6] it follows that $\widehat{\Psi}^*$ is absolutely symmetric and $F^* = \widehat{\Psi}^* \circ \sigma$ is convex and lsc. By using [31, Corollary 23.5.1] and $\overline{W} \in \partial F^*(\rho\overline{X})$, we have $\rho\overline{X} \in \partial F(\overline{W}) = \partial(\widehat{\Psi} \circ \sigma)(\overline{W})$. Along with the last equation and $\|\overline{W}\| \le 1$, $\overline{X}$ is an EP-stationary point of (1).

$\Longleftarrow$. Suppose that $\overline{X}$ is an EP-stationary point associated to ϕ. Then, there exist $\rho > 0$ and $\overline{W} \in \mathbb{B}$ such that the inclusions in (41) hold. From the above arguments, we know that $\widehat{\Psi} \circ \sigma$ is lsc and convex over $\mathbb{R}^{m\times n}$. Together with $\rho\overline{X} \in \partial(\widehat{\Psi} \circ \sigma)(\overline{W})$, it follows that $\overline{W} \in \partial(\widehat{\Psi} \circ \sigma)^*(\rho\overline{X}) = (\widehat{\Psi}^* \circ \sigma)(\rho\overline{X})$. Combining this with the second inclusion in (41) and Definition 3.4, we conclude that $\overline{X}$ is a DC-stationary point of (1). ∎

Now let $\mathscr{L}_2$ denote the family of those $\phi \in \mathscr{L}$ with $0 \in \partial\widehat{\psi}(0)$ and $\partial\widehat{\psi}(1) \subseteq [c, +\infty]$ for some $c > 0$, and let $\mathscr{L}_3$ be the family of those $\phi \in \mathscr{L}$ with $0 \in \partial\widehat{\phi}(0)$. Since $\widehat{\psi}(0) = \partial\widehat{\phi}(0)$, clearly, $\mathscr{L}_2 \subseteq \mathscr{L}_3$. To sum up the previous discussions, we obtain the relations as shown in Figure 1, where the relation among the local optimum of (1) and (4)–(6) is obtained by Propositions A.1–A.3 in the appendix. We see that the set of R-stationary points is almost the same as that of M-stationary points and includes those EP-stationary points associated to $\phi \in \mathscr{L}_1$ and satisfying the constraint $\|X\|_* - \langle W, X\rangle = 0$, while the set of EP-stationary points associated to $\phi \in \mathscr{L}_1$ coincides with that of DC-stationary points.

The following examples provide several $\phi \in \mathscr{L}$ satisfying the conditions in Figure 1.

Example 3.1. Let $\phi(t) := t$ for $t \in \mathbb{R}$. Clearly, $\phi \in \mathscr{L}_1 \cap \mathscr{L}_2$ with $t^* = 0$. After a simple calculation, $\widehat{\psi}^*(s) = |s| - 1$ if $|s| > 1$; otherwise $\widehat{\psi}^*(s) = 0$.

Example 3.2. Let $\phi(t) := \frac{\varphi(t)}{\varphi(1)}$ with $\varphi(t) = \frac{a^2}{4} + (a - \frac{a^2}{2})t$ for $t \in \mathbb{R}$, where $a \in (0, 2]$ is a constant. Clearly, $\phi \in \mathscr{L}_1 \cap \mathscr{L}_2$ with $t^* = 0$. After an elementary calculation,

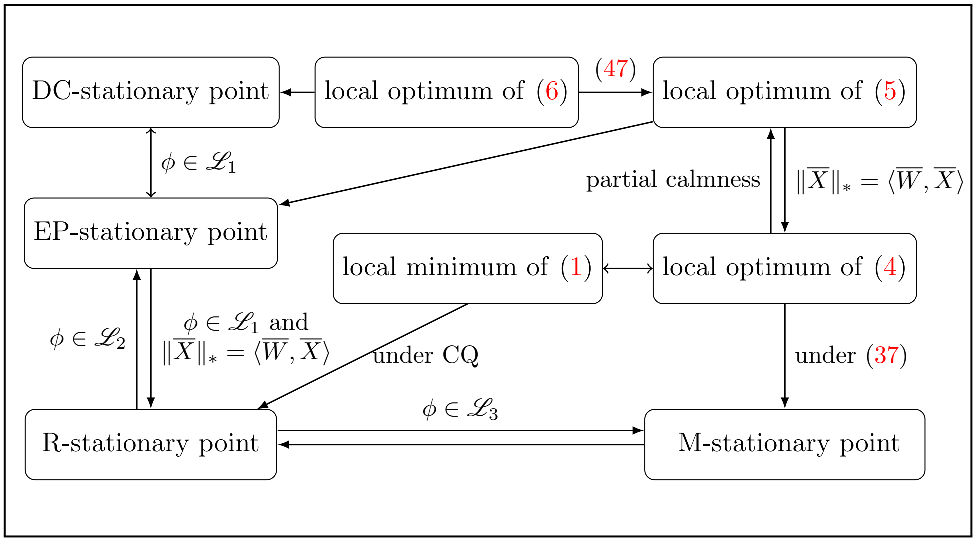

FIG. 1. *Relations among the stationary points and the local optima.*

$$\widehat{\psi}^*(s) = \begin{cases} 0 & \text{if } |s| \le \frac{a-a^2/2}{\varphi(1)}; \\ \frac{1}{a^2\varphi(1)}\left(\frac{a^2-2a}{2} + |s|\varphi(1)\right)^2 & \text{if } \frac{a-a^2/2}{\varphi(1)} < |s| \le \frac{a}{\varphi(1)}; \\ |s| - 1 & |s| > \frac{a}{\varphi(1)}. \end{cases}$$

Example 3.3. Let $\phi(t) = \frac{a-1}{a+1}t^2 + \frac{2}{a+1}t$ $(a > 1)$ for $t \in \mathbb{R}$. Clearly, $\phi \in \mathscr{L}_1 \cap \mathscr{L}_2$. After an elementary calculation, the conjugate $\widehat{\psi}^*$ of $\widehat{\psi}$ takes the following form:

$$\widehat{\psi}^*(s) = \begin{cases} 0 & \text{if } |s| \le \frac{2}{a+1}; \\ \frac{((a+1)|s|-2)^2}{4(a^2-1)} & \text{if } \frac{2}{a+1} < |s| \le \frac{2a}{a+1}; \\ |s| - 1 & \text{if } |s| > \frac{2a}{a+1}. \end{cases}$$

Appendix A.

PROPOSITION A.1. *If $\overline{X}$ is a local minimum of rank $\overline{r}$ to the problem* (1), *then* $(\overline{X}, \overline{W})$ *with* $\overline{W} := \overline{U}_1\overline{V}_1^{\mathbb{T}} + t^*\overline{U}_2[\mathrm{Diag}(e)\ 0]\overline{V}_2^{\mathbb{T}}$ *for* $(\overline{U}, \overline{V}) \in \mathbb{O}^{m,n}(\overline{X})$ *is locally optimal to* (4), *where* $\overline{U} = [\overline{U}_1 \quad \overline{U}_2]$ *and* $\overline{V} = [\overline{V}_1 \quad \overline{V}_2]$ *with* $\overline{U}_1$ *and* $\overline{V}_1$ *being the matrix consisting of the first* $\overline{r}$ *columns of* $\overline{U}$ *and* $\overline{V}$, *respectively. Conversely, if* $(\overline{Z}, \overline{Y})$ *is a local optimum of* (4) *with* $\mathrm{rank}(\overline{Z}) = \sum_{i=1}^m \phi(\sigma_i(\overline{Y}))$, *then* $\overline{Z}$ *is a local minimum of the problem* (1).

Proof. Let $\overline{X}$ be a local minimum of rank $\overline{r}$ to (1). Then, there exists $\varepsilon > 0$ such that

$$\nu f(Z) + \mathrm{rank}(Z) \ge \nu f(\overline{X}) + \mathrm{rank}(\overline{X}) \quad \forall Z \in \Omega \cap \mathbb{B}_\varepsilon(\overline{X}). \tag{45}$$

Let $\mathcal{F}$ denote the feasible set of the MPEC (4). Clearly, $(\overline{X}, \overline{W}) \in \mathcal{F}$. To prove that $(\overline{X}, \overline{W})$ is locally optimal to (4), we take an arbitrary $(X, W) \in \mathbb{B}_\varepsilon(\overline{X}, \overline{W}) \cap \mathcal{F}$. Then,

$$\nu f(X) + \textstyle\sum_{i=1}^m \phi(\sigma_i(W)) + \delta_\Omega(X) = \nu f(X) + \sum_{i=1}^m \phi(\sigma_i(W)) \ge \nu f(X) + \mathrm{rank}(X),$$

where the inequality is since $\mathrm{rank}(X)=\min_{\|Y\|\le 1,\|X\|_*-\langle X,Y\rangle=0}\sum_{i=1}^m\phi(\sigma_i(Y))$. Notice that $X\in\Omega\cap\mathbb{B}_\varepsilon(\overline{X})$. From (45), we have $\nu f(X)+\mathrm{rank}(X)\ge \nu f(\overline{X})+\mathrm{rank}(\overline{X})$. Thus,

$$\nu f(X)+\textstyle\sum_{i=1}^m\phi(\sigma_i(W))+\delta_\Omega(X)\ge \nu f(\overline{X})+\mathrm{rank}(\overline{X})+\delta_\Omega(\overline{X}).$$

Notice that $\mathrm{rank}(\overline{X})=\sum_{i=1}^m\phi(\sigma_i(\overline{W}))$ since $\phi(1)=1$ and $\phi(t^*)=0$. By the arbitrariness of $(X,W)\in\mathbb{B}_\varepsilon(\overline{X},\overline{W})\cap\mathcal{F}$, we conclude that $(\overline{X},\overline{W})$ is locally optimal to (4).

Conversely, let $(\overline{Z},\overline{Y})$ be a local optimum of (4). Then there exists $\varepsilon>0$ such that

$$\nu f(X)+\textstyle\sum_{i=1}^m\phi(\sigma_i(W))\ge \nu f(\overline{Z})+\sum_{i=1}^m\phi(\sigma_i(\overline{Y}))\quad \forall (X,W)\in\mathbb{B}_\varepsilon(\overline{Z},\overline{Y})\cap\mathcal{F}. \tag{46}$$

Since $\|\overline{Z}\|_*-\langle\overline{Z},\overline{Y}\rangle=0$ and $\|\overline{Y}\|\le 1$, we deduce that $\overline{Z}$ and $\overline{Y}$ have a simultaneous ordered SVD and $\sigma_i(\overline{Y})=1$ for $i=1,\ldots,\overline{r}=\mathrm{rank}(\overline{Z})$, which along with $\phi\in\mathscr{L}$ and $\overline{r}=\sum_{i=1}^m\phi(\sigma_i(\overline{Y}))$ implies that $\overline{Y}=\overline{U}_1\overline{V}_1^{\mathbb{T}}+t^*\overline{U}_2[\mathrm{Diag}(e)\ 0]\overline{V}_2^{\mathbb{T}}$ for $(\overline{U},\overline{V})\in\mathbb{O}^{m,n}(\overline{Z})$, where $\overline{U}_1,\overline{V}_1$ and $\overline{U}_2,\overline{V}_2$ are defined as above. Choose a $\delta\in(0,\varepsilon/2)$. Fix an arbitrary $Z\in\mathbb{B}_\delta(\overline{Z})$. Let $Y=U_1V_1^{\mathbb{T}}+t^*U_2[\mathrm{Diag}(e)\ \ 0]V_2^{\mathbb{T}}$ for $(U,V)\in\mathbb{O}^{m,n}(Z)$, where U_1 and V_1 are the matrix consisting of the first $r=\mathrm{rank}(Z)$ columns of U and V, respectively, and U_2 and V_2 are the matrix consisting of the rest of the $m-r$ and $n-r$ columns of U and V, respectively. Clearly, $(Z,Y)\in\mathcal{F}$. Moreover [4, Proposition 8], we know that $\|Y-\overline{Y}\|_F\le\varepsilon/2$ by shrinking δ if necessary. Thus, $\|(Z,Y)-(\overline{Z},\overline{Y})\|_F\le\varepsilon$. By using the inequality (46) and noting that $\sum_{i=1}^m\phi(\sigma_i(Y))=\mathrm{rank}(Z)$, it follows that

$$\nu f(Z)+\textstyle\sum_{i=1}^m\phi(\sigma_i(Z))=\nu f(Z)+\mathrm{rank}(Z)\ge \nu f(\overline{Z})+\sum_{i=1}^m\phi(\sigma_i(\overline{Y}))=\nu f(\overline{Z})+\mathrm{rank}(\overline{X}).$$

This, by the arbitrariness of $Z\in\mathbb{B}_\delta(\overline{Z})$, shows that $\overline{Z}$ is a local minimum of (1). □

PROPOSITION A.2. *If $(\overline{X},\overline{W})$ is a local optimum of the problem (4) which is partially calm at $(\overline{X},\overline{W})$, then $(\overline{X},\overline{W})$ is locally optimal to the problem (5) associated to all sufficiently large ρ. Conversely, if $(\overline{X},\overline{W})$ is a local optimum of (5) with $\|\overline{X}\|_*-\langle\overline{X},\overline{W}\rangle=0$, then $(\overline{X},\overline{W})$ is locally optimal to (4).*

Proof. The first part follows by [35, Proposition 2.2], and the second part is immediate by noting that $\|\overline{X}\|_*\ge\langle\overline{X},\overline{W}\rangle$ for all $(X,W)\in\mathbb{R}^{m\times n}\times\mathbb{B}$. □

PROPOSITION A.3. *If $\overline{X}$ is a local optimum of the problem (6), then for each $\overline{W}$ satisfying*

$$\textstyle\sum_{i=1}^m\psi(\sigma_i(\overline{W}))-\rho\langle\overline{W},\overline{X}\rangle=\sum_{i=1}^m\psi^*(\sigma_i(\overline{W})), \tag{47}$$

we have that $(\overline{X},\overline{W})$ is a local optimum of the problem (5).

Proof. Let $\overline{X}$ be a local optimum of the problem (6). Then, there exists $\varepsilon>0$ such that

$$\nu f(X)+\rho\|X\|_*-\textstyle\sum_{i=1}^m\psi^*(\rho\sigma_i(X))\ge \nu f(\overline{X})+\rho\|\overline{X}\|_*-\sum_{i=1}^m\psi^*(\rho\sigma_i(\overline{X}))\quad\forall X\in\mathbb{B}_\varepsilon(\overline{X}).$$

Let $\overline{W}$ be such that $\sum_{i=1}^m\psi(\sigma_i(\overline{W}))-\rho\langle\overline{W},\overline{X}\rangle=\sum_{i=1}^m\psi^*(\sigma_i(\overline{W}))$. Fix an arbitrary $(Z,W)\in\mathbb{B}_\varepsilon(\overline{X},\overline{W})\cap(\Omega\times\mathbb{B})$. Clearly, $Z\in\mathbb{B}_\varepsilon(\overline{X})$. Together with the last equation,

$$\nu f(Z)+\rho\|Z\|_*-\textstyle\sum_{i=1}^m\psi^*(\rho\sigma_i(Z))\ge \nu f(\overline{X})+\rho\|\overline{X}\|_*-\sum_{i=1}^m\psi^*(\rho\sigma_i(\overline{X})).$$

Notice that $\sum_{i=1}^m \psi(\sigma_i(Z)) - \rho\langle W, Z\rangle \geq -\sum_{i=1}^m \psi^*(\rho\sigma_i(Z))$. Therefore, it holds that

$$\begin{aligned}
&\nu f(Z) + \rho(\|Z\|_* - \langle W, Z\rangle) + \textstyle\sum_{i=1}^m \psi(\sigma_i(Z))\\
&\geq \nu f(\overline{X}) + \rho\|\overline{X}\|_* - \textstyle\sum_{i=1}^m \psi^*(\rho\sigma_i(\overline{X}))\\
&= \nu f(\overline{X}) + \rho(\|\overline{X}\|_* - \langle \overline{W}, \overline{X}\rangle) + \textstyle\sum_{i=1}^m \psi(\sigma_i(\overline{X})).
\end{aligned}$$

This, by the arbitrariness of $(Z, W) \in \mathbb{B}_\varepsilon(\overline{X}, \overline{W}) \cap (\Omega \times \mathbb{B})$, implies the result. ∎

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 J. Control Optim., 55 (2017), pp. 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 J. Optim., 26 (2016), pp. 397–425.

[3] M. A. Davenport and J. Romberg, *An overview of low-rank matrix recovery from incomplete observations*, IEEE J. Selected Topics Signal Process., 10 (2016), pp. 608–622.

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

[5] C. Ding, D. F. Sun, and J. J. Ye, *First order optimality conditions for mathematical programs with semidefinite cone complementarity constraints*, Math. Program., 144 (2014), pp. 141–179.

[6] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings*, Springer Monogr. Math., Springer, New York, 2009.

[7] M. Fazel, *Matrix Rank Minimization with Applications*, Ph.D. thesis, Stanford University, 2002.

[8] M. Fazel, H. Hindi, and S. Boyd, *Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices*, in Proceedings of the American Control Conference, 2003, pp. 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 J. Matrix Anal., 34 (2013), pp. 946–977.

[10] M. L. Flegel and C. Kanzow, *On M-stationary points for mathematical programs with equilibrium constraints*, J. Math. Anal. Appl., 310 (2005), pp. 286–302.

[11] H. Gfrerer and D. Klatte, *Lipschitz and Hölder stability of optimization problems and generalized equations*, Math. Program., 158 (2016), pp. 35–75.

[12] D. Gross, *Recovering low-rank matrices from few coefficients in any basis*, IEEE Trans. Inform. Theory, 57 (2011), pp. 1548–1566.

[13] A. D. Ioffe and J. V. Outrata, *On metric and calmness qualification conditions in subdifferential calculus*, Set-Valued Anal., 16 (2008), pp. 199–227.

[14] A. J. King and R. T. Rockafellar, *Sensitivity analysis for nonsmooth generalized equations*, Math. Program., 55 (1992), pp. 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 J. Numer. Anal., 51 (2013), pp. 927–957.

[16] H. Y. Le, *Generalized subdifferentials of the rank function*, Optim. Lett., 7 (2013), pp. 731–743.

[17] A. B. Levy, *Implicit multifunction theorems for the sensitivity analysis of variational conditions*, Math. Program., 74 (1996), pp. 333–350.

[18] A. S. Lewis and H. Sendov, *Nonsmooth analysis of singular values. Part* I: *Theory*, Set-Valued Anal., 13 (2005), pp. 213–241.

[19] A. S. Lewis, *The convex analysis of unitarily invariant matrix functions*, J. Convex Anal., 2 (1995), pp. 173–183.

[20] Y. L. Liu, S. J. Bi, and S. H. Pan, *Equivalent Lipschitz surrogates for zero-norm and rank optimization problems*, J. Global Optim., 72 (2018), pp. 679–704.

[21] 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 Var. Anal., 27 (2019), pp. 71–85.

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

[23] K. Mohan and M. Fazel, *Iterative reweighted algorithm for matrix rank minimization*, J. Mach. Learn. Res., 13 (2012), pp. 3441–3473.

[24] B. S. Mordukhovich, *Complete characterization of openess, metric regularity, and Lipschitzian properties of multifunctions*, Trans. Amer. Math. Soc., 340 (1993), pp. 1–35.
[25] S. Negahban and M. J. Wainwright, *Estimation of (near) low-rank matrices with noise and high-dimensional scaling*, Ann. Statist., 39 (2011), pp. 1068–1097.
[26] F. Nie, H. Huang, and C. Ding, *Low-rank matrix recovery via efficient schatten p-norm minimization*, in Proceedings of the 26th AAAI Conference on Artificial Intelligence, 2012, pp. 655–661.
[27] L. L. Pan, Z. Y. Luo, and N. H. Xiu, *Restricted robinson constraint qualification and optimality for cardinality-constrained cone Programming*, J. Optim. Theory Appl., 175 (2017), pp. 104–118.
[28] J. S. Pang, M. Razaviyayn, and A. Alvarado, *Computing B-stationary points of nonsmooth DC programs*, Math. Oper. Res., 42 (2017), pp. 95–118.
[29] R. Pietersz and P. J. F. Groenen, *Rank reduction of correlation matrices by majorization*, Quant. Finance, 4 (2004), pp. 649–662.
[30] J. Rennie and N. Srebro, *Fast maximum margin matrix factorization for collaborative prediction*, in Proceedings of the 22nd International Conference on Machine Learning, 2005, pp. 713–719.
[31] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[32] R. T. Rockafellar and R. J-B. Wets, *Variational Analysis*, Springer, New York, 1998.
[33] G. A. Watson, *Characterization of the subdifferential of some matrix norms*, Linear Algebra Appl., 170 (1992), pp. 33–45.
[34] Y. Q. Wu, S. H. Pan, and S. J. Bi, *KL Property of Exponent* 1/2 *for Zero-Norm Composite Quadratic Functions*, https://arxiv.org/abs/1811.04371v3, 2019.
[35] J. J. Ye, D. L. Zhu, and Q. J. Zhu, *Exact penalization and necessary optimality conditions for generalized bilevel programming problems*, SIAM J. Optim., 7 (1997), pp. 481–507.