A COMPLETE CHARACTERIZATION OF THE ROBUST ISOLATED CALMNESS OF NUCLEAR NORM REGULARIZED CONVEX OPTIMIZATION PROBLEMS*

Ying Cui

The Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, California, USA
Email: yingcui@usc.edu

Defeng Sun

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong
Email: defeng.sun@polyu.edu.hk

Abstract

In this paper, we provide a complete characterization of the robust isolated calmness of the Karush-Kuhn-Tucker (KKT) solution mapping for convex constrained optimization problems regularized by the nuclear norm function. This study is motivated by the recent work in [8], where the authors show that under the Robinson constraint qualification at a local optimal solution, the KKT solution mapping for a wide class of conic programming problems is robustly isolated calm if and only if both the second order sufficient condition (SOSC) and the strict Robinson constraint qualification (SRCQ) are satisfied. Based on the variational properties of the nuclear norm function and its conjugate, we establish the equivalence between the primal/dual SOSC and the dual/primal SRCQ. The derived results lead to several equivalent characterizations of the robust isolated calmness of the KKT solution mapping and add insights to the existing literature on the stability of nuclear norm regularized convex optimization problems.

Mathematics subject classification: 90C25, 90C31, 65K10.
Key words: Robust isolated calmness, Nuclear norm, Second order sufficient condition, Strict Robinson constraint qualification

1. Introduction

Let $\mathbb{X}$ and $\mathbb{Y}$ be two finite dimensional Euclidean spaces. Let $G : \mathbb{X} \rightrightarrows \mathbb{Y}$ be a set-valued mapping. The graph of G is defined as $\operatorname{gph} G := \{(x, y) \in \mathbb{X} \times \mathbb{Y} \mid y \in G(x)\}$. Consider any $(\bar{x}, \bar{y}) \in \operatorname{gph} G$. The mapping G is said to be isolated calm at $\bar{x}$ for $\bar{y}$ if there exist a constant $\kappa > 0$ and open neighborhoods $\mathcal{X}$ of $\bar{x}$ and $\mathcal{Y}$ of $\bar{y}$ such that

$$G(x) \cap \mathcal{Y} \subset \{\bar{y}\} + \kappa \|x - \bar{x}\| \mathbb{B}_{\mathbb{Y}}, \quad \forall \, x \in \mathcal{X}, \tag{1.1}$$

where $\mathbb{B}_{\mathbb{Y}}$ is the unit ball in $\mathbb{Y}$ (cf. e.g., [9, 3.9 (3I)]). The mapping G is said to be robustly isolated calm at $\bar{x}$ for $\bar{y}$ if (1.1) holds and $G(x) \cap \mathcal{Y} \neq \emptyset$ for any $x \in \mathcal{X}$ [8, Definition 2].

In this paper, we are interested in characterizing the robust isolated calmness of the Karush-Kuhn-Tucker (KKT) solution mapping associated with the following nuclear norm regularized

convex optimization problem:

$$\min_{X} \quad h(\mathcal{F}X) + \langle C, X \rangle + \|X\|_*$$
$$\text{s.t.} \quad \mathcal{A}X - b \in \mathcal{Q}, \tag{1.2}$$

where the function $h : \mathbb{R}^d \to \mathbb{R}$ is twice continuously differentiable on $\operatorname{dom} h$, which is assumed to be a non-empty open convex set, and is also essentially strictly convex (i.e., h is strictly convex on every convex subset of $\operatorname{dom} \partial h$), $\mathcal{F} : \mathbb{R}^{m \times n} \to \mathbb{R}^d$ and $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^e$ are linear operators, $C \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^e$ are given data, $\mathcal{Q} \subseteq \mathbb{R}^e$ is a nonempty convex polyhedral cone, $\| \cdot \|_*$ denotes the nuclear norm function in $\mathbb{R}^{m \times n}$, i.e., the sum of all the singular values of a given matrix, and m, n, d, e are non-negative integers. The nuclear norm regularizer has been extensively used in diverse disciplines due to its ability in promoting a low rank solution. See the references [2, 3, 14, 17, 20, 21] for a sample of applications.

The concept of the isolated calmness is of fundamental importance in variational analysis. The monograph [9] by Dontchev and Rockafellar contains a comprehensive study on this subject. Besides its own interest in sensitivity analysis and perturbation theory, the isolated calmness of the KKT solution mapping can be readily applied to provide linear rate convergence analysis for many important primal dual type methods such as the proximal augmented Lagrangian method [16] and the alternating direction method of multipliers [10] that can be employed efficiently to solve large scale matrix optimization problems such as (1.2). With this application in mind, in this paper we aim to derive easy-to-understand conditions to characterize the the isolated calmness of the KKT solution mapping for [10].

Obviously problem (1.2) can be equivalently formulated as the following conic programming problem

$$\min_{X,t} \quad h(\mathcal{F}X) + \langle C, X \rangle + t$$
$$\text{s.t.} \quad \mathcal{A}X - b \in \mathcal{Q}, \quad (X, t) \in \operatorname{epi} \| \cdot \|_*, \tag{1.3}$$

where $\operatorname{epi} \| \cdot \|_*$ denotes the epigraph of the function $\| \cdot \|_*$. Since $\operatorname{epi} \| \cdot \|_*$ is not a polyhedral set, the sensitivity results in the conventional nonlinear programming are not applicable for problem (1.3). Recently, some progress has been achieved in characterizing the isolated calmness of KKT solution mappings for problems involving non-polyhedral functions. For example, Zhang and Zhang [23] show that for the nonlinear semidefinite programming, the second order sufficient condition (SOSC) and the strict Robinson constraint qualification (SRCQ) at a local optimal solution together are sufficient for the KKT solution mapping to be isolated calm. Adding to this result, Han, Sun and Zhang [11] show that the SRCQ is also necessary to ensure the isolated calmness of the KKT solution mapping for such problems. In [12], Liu and Pan extend the aforementioned results to problems constrained by the epigraph of the Ky Fan k-norm function. The most recent work of Ding, Sun and Zhang [8] indicates that under the Robinson constraint qualification (RCQ) at a local optimal solution, the KKT solution mapping for a wide class of conic programming is robustly isolated calm at the origin for a KKT point if and only if both the SOSC and the SRCQ hold at the reference point.

The results developed in [8] can be directly applied to problem (1.3). Thus, by examining the relationships between the SOSCs, the (strict) RCQs as well as the robust isolated calmness of the solution mappings corresponding to problem (1.2) and problem (1.3), we are able to extend the work in [8] to the nuclear norm regularized convex optimization problem (1.2). Additionally, due to the special structure of problem (1.2) and its dual, we could provide more insightful characterizations about the isolated calmness of the KKT solution mapping. Note

that the Lagrangian dual of problem (1.2) is given by

$$
\begin{aligned}
\max_{y,w,S} \quad & -\langle b, y \rangle - \delta_{\mathcal{Q}^\circ}(y) - h^*(w) \\
\text{s.t.} \quad & \mathcal{A}^* y + \mathcal{F}^* w + S + C = 0, \quad \|S\|_2 \le 1,
\end{aligned}
\tag{1.4}
$$

where $\mathcal{A}^*$ and $\mathcal{F}^*$ are the adjoint of $\mathcal{A}$ and $\mathcal{F}$, respectively, $h^*(\cdot)$ is the conjugate function of h and $\| \cdot \|_2$ denotes the spectral norm in $\mathbb{R}^{m \times n}$, i.e., the largest singular value of a given matrix. We shall show that for problem (1.2), its SRCQ is equivalent to the SOSC of problem (1.4), and conversely, its SOSC is equivalent to the SRCQ of problem (1.4). Armed with these results, we are led to a relatively complete understanding of the robust isolated calmness for the KKT solution mapping for the nuclear norm regularized convex optimization problems.

The remaining parts of this paper are organized as follows. In the next section, we provide some preliminary results on variational analysis. In Section 3, we demonstrate how to translate the results of set-constrained problems in [8] into the language of nonsmooth optimization problems. In particular, this translation provides us a characterization of the robust isolated calmness of the KKT solution mapping for the nuclear norm regularized convex optimization problems. Section 4 is devoted to the study of the variational properties of the nuclear norm function. The derived results play an important role in our subsequent analysis. In Section 5, we establish the equivalence between the SOSC for the primal/dual problem and the SRCQ for the dual/primal problem. This establishment enables us to describe the robust isolated calmness of the KKT solution mapping for problem (1.2) via several equivalent conditions.

The following notation will be used throughout our paper.

- For a given positive integer p, we use $\mathbb{S}^p$ to denote the linear space of all $p \times p$ real symmetric matrices, $\mathbb{S}^p_+$ the cone of all $p \times p$ positive semidefinite matrices and $\mathbb{S}^p_-$ the cone of all $p \times p$ negative semidefinite matrices.

- For a given proper closed convex function $\theta : \mathbb{X} \to (-\infty, +\infty]$, we use $\operatorname{dom} \theta$ to denote its effective domain, $\operatorname{epi} \theta$ to denote its epigraph, θ^* to denote its conjugate and $\partial \theta$ to denote its subdifferential, all as in standard convex analysis [15]. We also use $\operatorname{Prox}_\theta$ to denote its proximal mapping with unit step length, i.e. (see, e.g., [18, Definition 1.22]),

$$
\operatorname{Prox}_\theta(x) := \arg\min_{x' \in \mathbb{X}} \left\{ \theta(x') + \frac{1}{2} \|x' - x\|^2 \right\}, \quad x \in \mathbb{X}.
$$

- Let $D \subseteq \mathbb{R}^{m \times n}$ be a non-empty closed convex set. We write $\delta_D(\cdot)$ as the indicator function over D, i.e., $\delta_D(X) = 0$ if $X \in D$, and $\delta_D(X) = \infty$ if $X \notin D$. We write $\Pi_D(\cdot)$ as the metric projection onto D, i.e., $\Pi_D(X) := \arg\min_Y \{ \|Y - X\| \mid Y \in D \}$ for $X \in \mathbb{R}^{m \times n}$.

- For any $z \in \mathbb{R}^m$, we denote $\operatorname{Diag}(z)$ as the $m \times m$ diagonal matrix whose i-th diagonal entry is z_i for $i = 1, \ldots, m$. Let $\alpha \subseteq \{1, ..., m\}$ and $\beta \subseteq \{1, ..., n\}$ be two index sets. For any $Z \in \mathbb{R}^{m \times n}$, we write Z_α as the sub-matrix of Z by removing all the columns of Z not in α, and $Z_{\alpha\beta}$ to be the $|\alpha| \times |\beta|$ sub-matrix of Z obtained by removing all the rows of Z not in α and all the columns of Z not in β.

- Let $\mathcal{O}^n$ be the set of all $n \times n$ orthogonal matrices. For any $X \in \mathbb{R}^{m \times n}$, let $\sigma(X) \in \mathbb{R}^m$ be the vector of all singular values of X with the entries $\sigma_1(X) \ge \sigma_2(X) \ge \ldots \ge \sigma_m(X)$, and let $\mathcal{O}^{m,n}(X)$ be the set of paired orthogonal matrices satisfying the singular value decomposition of X, i.e.,

$$
\mathcal{O}^{m,n}(X) = \{ (U, V) \in \mathcal{O}^m \times \mathcal{O}^n \mid X = U[\operatorname{Diag}(\sigma(X)) \;\; 0] V^T \}.
$$

2. Preliminaries

In this section, we gather some knowledge on variational analysis that will be used in our subsequent developments. One can refer to the monograph [1] of Bonnans and Shapiro for detailed discussions on this subject.

A cone $\mathcal{Q} \subseteq \mathbb{Y}$ is said to be pointed if $y \in \mathcal{Q}$ and $-y \in \mathcal{Q}$ implies that $y = 0$. Let $\mathcal{Q} \subseteq \mathbb{Y}$ be a pointed convex closed cone. The closed convex set $\mathcal{K} \subseteq \mathbb{X}$ is said to be $\mathcal{C}^2$-cone reducible at $x \in \mathcal{K}$ to the cone $\mathcal{Q}$, if there exist an open neighborhood $\mathcal{W} \subseteq \mathbb{X}$ of x and a twice continuously differentiable mapping $\Xi : \mathcal{W} \to \mathbb{Y}$ such that: (i) $\Xi(x) = 0 \in \mathbb{Y}$; (ii) the derivative mapping $\Xi(x) : \mathbb{X} \to \mathbb{Y}$ is onto; (iii) $\mathcal{K} \cap \mathcal{W} = \{x \in \mathcal{W} \mid \Xi(x) \in \mathcal{Q}\}$. We say that $\mathcal{K}$ is $\mathcal{C}^2$-cone reducible if $\mathcal{K}$ is $\mathcal{C}^2$-cone reducible at every $x \in \mathcal{K}$. A proper closed convex function $\theta : \mathbb{X} \to (-\infty, \infty]$ is said to be $\mathcal{C}^2$-cone reducible at $x \in \operatorname{dom} \theta$ if $\operatorname{epi} \theta$ is $\mathcal{C}^2$-cone reducible at $(x, \theta(x))$. Moreover, θ is said to be $\mathcal{C}^2$-cone reducible if it is $\mathcal{C}^2$-cone reducible at every $x \in \operatorname{dom} \theta$.

Given a subset $\mathcal{K} \subseteq \mathbb{X}$ and $x \in \mathcal{K}$, the inner tangent cone and the contingent cone of $\mathcal{K}$ at x are defined as

$$\mathcal{T}_{\mathcal{K}}^i(x) = \liminf_{t \downarrow 0} \frac{\mathcal{K} - x}{t}$$

and

$$\mathcal{T}_{\mathcal{K}}(x) = \limsup_{t \downarrow 0} \frac{\mathcal{K} - x}{t},$$

respectively, where the limit is taken in the sense of Painleve-Kuratowski set convergence. If $\mathcal{K}$ is convex, $\dfrac{\mathcal{K} - x}{t}$ is a monotone decreasing function of t such that $\mathcal{T}_{\mathcal{K}}(x) = \mathcal{T}_{\mathcal{K}}^i(x)$ for any $x \in \mathcal{K}$ [1, Proposition 2.55]. In this case, both $\mathcal{T}_{\mathcal{K}}(x)$ and $\mathcal{T}_{\mathcal{K}}^i(x)$ are called the tangent cone of $\mathcal{K}$ at x. Given $x \in \mathcal{K}$ and a direction $d \in \mathbb{X}$, define the inner and outer second order tangent sets at x in the direction d as

$$\mathcal{T}_{\mathcal{K}}^{i,2}(x; d) := \liminf_{t \downarrow 0} \frac{\mathcal{K} - x - td}{\frac{1}{2} t^2}$$

and

$$\mathcal{T}_{\mathcal{K}}^2(x; d) := \limsup_{t \downarrow 0} \frac{\mathcal{K} - x - td}{\frac{1}{2} t^2},$$

respectively. The inner and outer second order tangent sets not necessarily coincide in general, even if the set $\mathcal{K}$ is closed and convex. However, if $\mathcal{K}$ is a $\mathcal{C}^2$-cone reducible convex set, then $\mathcal{T}_{\mathcal{K}}^{i,2}(x; d) = \mathcal{T}_{\mathcal{K}}^2(x; d)$ [1, Proposition 3.136].

For a given function $\theta : \mathbb{X} \to (-\infty, +\infty]$, the lower and upper directional epiderivatives of θ at $x \in \operatorname{dom} \theta$ in the direction $h \in \mathbb{X}$ are defined as

$$\theta_-^{\downarrow}(x; h) := \liminf_{\substack{t \downarrow 0 \\ h' \to h}} \frac{\theta(x + th') - \theta(x)}{t}$$

and

$$\theta_+^{\downarrow}(x; h) := \sup_{\{t_n\} \in \Sigma} \left(\liminf_{\substack{n \to \infty \\ h' \to h}} \frac{\theta(x + t_n h') - \theta(x)}{t_n} \right),$$

respectively, where Σ denotes the set of positive real sequences $\{t_n\}$ converging to 0. The contingent and inner tangent cone of $\operatorname{epi} \theta$ are closely related to the lower and upper directional epiderivative of θ [1, Proposition 2.58]. Specifically, for any $x \in \operatorname{dom} \theta$,

$$\mathcal{T}_{\operatorname{epi} \theta}\big(x, \theta(x)\big) = \operatorname{epi} \theta_-^{\downarrow}(x; \cdot), \quad \mathcal{T}_{\operatorname{epi} \theta}^i\big(x, \theta(x)\big) = \operatorname{epi} \theta_+^{\downarrow}(x; \cdot). \tag{2.1}$$

One can observe from the above equations that if θ is a closed convex function, then $\theta_-^\downarrow(x;\cdot) = \theta_+^\downarrow(x;\cdot)$ for any $x \in \mathrm{dom}\,\theta$. In this case we say that θ is directionally epidifferentiable at x and write the common value as $\theta^\downarrow(x;\cdot)$. If $\theta_+^\downarrow(x;d)$ and $\theta_-^\downarrow(x;d)$ are finite for $x \in \mathrm{dom}\,\theta$ and $d \in \mathbb{X}$, we also define the following lower and upper second order epiderivatives for $w \in \mathbb{X}$:

$$\theta_-^{\downarrow\downarrow}(x;d,w) := \liminf_{\substack{t\downarrow 0 \\ w'\to w}} \frac{\theta(x + td + \frac{1}{2}t^2 w') - \theta(x) - t\theta_-^\downarrow(x;d)}{\frac{1}{2}t^2},$$

$$\theta_+^{\downarrow\downarrow}(x;d,w) := \sup_{t_n \in \Sigma}\left(\liminf_{\substack{n\to\infty \\ w'\to w}} \frac{\theta(x + t_n d + \frac{1}{2}t_n^2 w') - \theta(x) - t_n\theta_+^\downarrow(x;d)}{\frac{1}{2}t_n^2}\right).$$

Similarly to (2.1), the inner and outer second order tangent sets of $\mathrm{epi}\,\theta$ are closely related to the lower and upper second order epiderivative of θ [1, Proposition 3.41]. Specifically, for any $x \in \mathrm{dom}\,\theta$ and $d \in \mathbb{X}$, if $\theta_+^\downarrow(x;d)$ and $\theta_-^\downarrow(x;d)$ are finite, then

$$\mathcal{T}_{\mathrm{epi}\,\theta}^{i,2}((x,\theta(x));(d,\theta_+^\downarrow(x;d))) = \mathrm{epi}\,\theta_+^{\downarrow\downarrow}(x;d,\cdot), \quad \mathcal{T}_{\mathrm{epi}\,\theta}^2((x,\theta(x));(d,\theta_-^\downarrow(x;d))) = \mathrm{epi}\,\theta_-^{\downarrow\downarrow}(x;d,\cdot). \tag{2.2}$$

3. Constraint Qualifications, Second Order Sufficient Optimality Conditions and Robust Isolated Calmness

Consider the following canonical perturbation of a general class of nonsmooth optimization problems (not necessarily convex):

$$\min \{f(x) + \theta(x) - \langle \delta_1, x\rangle \mid g(x) + \delta_2 \in \mathcal{P}\}, \tag{3.1}$$

where $f : \mathbb{X} \to \mathbb{R}$ and $g : \mathbb{X} \to \mathbb{Y}$ are twice continuously differentiable functions, $\mathcal{P} \subseteq \mathbb{Y}$ is a closed convex set, $\theta : \mathbb{X} \to (-\infty, +\infty]$ is a closed proper convex function, and $\delta_1 \in \mathbb{X}$ and $\delta_2 \in \mathbb{Y}$ are perturbation parameters. The purpose to perturb both the objective function and the constraint in (3.1) is to study its KKT solution mapping that involves both the primal solution and the corresponding Lagrangian multiplier.

Note that problem (3.1) can be equivalently written as the following optimization problem:

$$\min \{f(x) + t - \langle \delta_1, x\rangle \mid g(x) + \delta_2 \in \mathcal{P}, \ (x,t) \in \mathcal{K}\}, \tag{3.2}$$

where $\mathcal{K} := \mathrm{epi}\,\theta$ is a closed convex set. The constraint qualifications and SOSCs for problem (3.2) have been extensively explored in [1, Section 3], through the study of the (second order) tangent sets of $\mathcal{P}$ and $\mathcal{K}$ at a stationary point. In the following, by employing the equations in (2.1) and (2.2), we reduce these properties to the (second order) directional epiderivatives of θ. This reduction leads to a direct approach to the sensitivity analysis of the nonsmooth optimization problem (3.1).

For notational simplicity, denote $\mathbb{Z} := \mathbb{X} \times \mathbb{R} \times \mathbb{Y} \times \mathbb{X} \times \mathbb{R}$. Let $(\delta_1, \delta_2) \in \mathbb{X} \times \mathbb{Y}$ be given. We say that $(\bar{x}, \bar{t})$ is a feasible solution to problem (3.2) if

$$(\bar{x}, \bar{t}) \in \widehat{F}(\delta_1, \delta_2) := \{(x,t) \in \mathbb{X} \times \mathbb{R} \mid g(x) + \delta_2 \in \mathcal{P}, \ (x,t) \in \mathcal{K}\}.$$

For any $(x, t, y, z, \tau) \in \mathbb{Z}$, the Lagrangian function of (3.2) with $\delta = (\delta_1, \delta_2)$ can be written as

$$\mathcal{L}^\delta(x,t;y,z,\tau) := f(x) + t - \langle \delta_1, x\rangle + \langle y, g(x) + \delta_2\rangle + \langle z, x\rangle + t\tau.$$

For any given $(\delta_1, \delta_2) \in \mathbb{X} \times \mathbb{Y}$, the KKT optimality condition for problem (3.2) is

$$
\begin{cases}
0 = \nabla_x \mathcal{L}^\delta(x,t,y,z,\tau) = \nabla f(x) - \delta_1 + \nabla g(x)y + z, \\
\nabla_t \mathcal{L}^\delta(x,t,y,z,\tau) = 1 + \tau = 0, \qquad\qquad (x,t,y,z,\tau) \in \mathbb{Z}, \qquad (3.3) \\
y \in \mathcal{N}_\mathcal{P}(g(x) + \delta_2), \quad (z,\tau) \in \mathcal{N}_\mathcal{K}(x,t),
\end{cases}
$$

where $\mathcal{N}_C(s)$ denotes the normal cone of a given convex set C at $s \in C$. Throughout the rest of this paper, we shall omit the superscript δ from the Lagrangian function $\mathcal{L}^\delta(\cdot)$ if $\delta = (0,0)$. Let $\widehat{S}_{\mathrm{KKT}} : \mathbb{X} \times \mathbb{Y} \to \mathbb{Z}$ be the following KKT solution mapping:

$$
\widehat{S}_{\mathrm{KKT}}(\delta_1, \delta_2) := \{(x,t,y,z,\tau) \in \mathbb{Z} \mid (x,t,y,z,\tau) \text{ satisfies } (3.3)\}, \quad (\delta_1, \delta_2) \in \mathbb{X} \times \mathbb{Y}. \quad (3.4)
$$

For any given $(\delta_1, \delta_2) \in \mathbb{X} \times \mathbb{Y}$, we call $(\bar{x}, \bar{t})$ a stationary point of problem (3.1) if there exists a Lagrangian multiplier $(\bar{y}, \bar{z}, \bar{\tau}) \in \mathbb{Y} \times \mathbb{X} \times \mathbb{R}$ such that $(\bar{x}, \bar{t}, \bar{y}, \bar{z}, \bar{\tau}) \in \widehat{S}_{\mathrm{KKT}}(\delta_1, \delta_2)$. Denote $\widehat{\mathcal{M}}(\bar{x}, \bar{t}, \delta_1, \delta_2) \subseteq \mathbb{Y} \times \mathbb{X} \times \mathbb{R}$ as the set of all such Lagrange multipliers $(\bar{y}, \bar{z}, \bar{\tau})$ associated with $(\bar{x}, \bar{t})$. Note from (3.3) that $\bar{\tau} \equiv -1$.

Let $(\delta_1, \delta_2) = 0$ in problem (3.2). The Robinson constraint qualification (RCQ) is said to hold at a feasible solution $(\bar{x}, \theta(\bar{x}))$ of problem (3.2) if

$$
\begin{pmatrix} (g'(\bar{x}), 0) \\ (\mathcal{I}_\mathbb{X}, 1) \end{pmatrix} (\mathbb{X} \times \mathbb{R}) + \begin{pmatrix} \mathcal{T}_\mathcal{P}(g(\bar{x})) \\ \mathcal{T}_\mathcal{K}(\bar{x}, \theta(\bar{x})) \end{pmatrix} = \begin{pmatrix} \mathbb{Y} \\ \mathbb{X} \times \mathbb{R} \end{pmatrix}, \quad (3.5)
$$

where $\mathcal{I}_\mathbb{X}$ is the identity mapping from $\mathbb{X}$ to $\mathbb{X}$. It is known that the RCQ (3.5) holds at a local optimal solution $(\bar{x}, \theta(\bar{x}))$ if and only if $\widehat{\mathcal{M}}(\bar{x}, \theta(\bar{x}), 0, 0)$ is a nonempty, convex and compact set (cf., e.g., [1, Theorem 3.9 and Proposition 3.17]). The SRCQ is said to hold at $(\bar{x}, \theta(\bar{x}))$ for $(\bar{y}, \bar{z}, -1) \in \widehat{\mathcal{M}}(\bar{x}, \theta(\bar{x}), 0, 0)$ if

$$
\begin{pmatrix} (g'(\bar{x}), 0) \\ (\mathcal{I}_\mathbb{X}, 1) \end{pmatrix} (\mathbb{X} \times \mathbb{R}) + \begin{pmatrix} \mathcal{T}_\mathcal{P}(g(\bar{x})) \cap \bar{y}^\perp \\ \mathcal{T}_\mathcal{K}(\bar{x}, \theta(\bar{x})) \cap (\bar{z}, -1)^\perp \end{pmatrix} = \begin{pmatrix} \mathbb{Y} \\ \mathbb{X} \times \mathbb{R} \end{pmatrix}. \quad (3.6)
$$

Obviously the SRCQ (3.6) is stronger than the RCQ (3.5). It follows from [1, Proposition 4.50] that $\widehat{M}(\bar{x}, \theta(\bar{x}), 0, 0)$ is a singleton if the SRCQ holds at $(\bar{x}, \theta(\bar{x}))$. The critical cone at a feasible point $(\bar{x}, \theta(\bar{x}))$ for problem (3.2) takes the form of

$$
\widehat{\mathcal{C}}(\bar{x}, \theta(\bar{x})) := \{(d_1, d_2) \in \mathbb{X} \times \mathbb{R} \mid g'(\bar{x})d_1 \in \mathcal{T}_\mathcal{P}(g(\bar{x})), \ (d_1, d_2) \in \mathcal{T}_\mathcal{K}(\bar{x}, \theta(\bar{x})), \ f'(\bar{x})d_1 + d_2 \le 0\}.
$$

Furthermore, if $(\bar{x}, \theta(\bar{x}))$ is a stationary point of problem (3.2) and there exists $(\bar{y}, \bar{z}, -1) \in \widehat{\mathcal{M}}(\bar{x}, \theta(\bar{x}), 0, 0)$, then

$$
\begin{aligned}
\widehat{\mathcal{C}}(\bar{x}, \theta(\bar{x})) &= \{(d_1, d_2) \in \mathbb{X} \times \mathbb{R} \mid g'(\bar{x})d_1 \in \mathcal{T}_\mathcal{P}(g(\bar{x})), \ (d_1, d_2) \in \mathcal{T}_\mathcal{K}(\bar{x}, \theta(\bar{x})), \ f'(\bar{x})d_1 + d_2 = 0\} \\
&= \{(d_1, d_2) \in \mathbb{X} \times \mathbb{R} \mid g'(\bar{x})d_1 \in \mathcal{T}_\mathcal{P}(g(\bar{x})) \cap \bar{y}^\perp, \ (d_1, d_2) \in \mathcal{T}_\mathcal{K}(\bar{x}, \theta(\bar{x})) \cap (\bar{z}, -1)^\perp\}.
\end{aligned}
$$

Assume that epi θ is second order regular at $(\bar{x}, \theta(\bar{x}))$ (see [1, Definition 3.85] for the definition of the second order regularity). This assumption is particularly satisfied when $\theta(\cdot) = \|\cdot\|_*$, since in this case epi θ is a $\mathcal{C}^2$-cone reducible set [5, Proposition 4.3]. Then the SOSC at a stationary point $(\bar{x}, \theta(\bar{x}))$ for problem (3.2) with $(\delta_1, \delta_2) = 0$ is said to hold if for any $d \in \widehat{\mathcal{C}}(\bar{x}, \theta(\bar{x})) \setminus \{0\}$,

$$
\sup_{(\bar{y}, \bar{z}, -1) \in \widehat{\mathcal{M}}(\bar{x}, \theta(\bar{x}), 0, 0)} \langle d, \nabla^2_{(x,t)(x,t)} \mathcal{L}(\bar{x}, \theta(\bar{x}), \bar{y}, \bar{z}, -1)d \rangle - \sigma\big((\bar{z}, -1), \mathcal{T}^2_\mathcal{K}((\bar{x}, \theta(\bar{x})); (g'(\bar{x}), 0)d)\big) > 0,
$$

$$
(3.7)
$$

where $\sigma(s, C) := \sup\{\langle s', s \rangle \mid s' \in C\}$ denotes the support function of a given set C at s. The above SOSC implies the quadratic growth condition at $(\bar{x}, \theta(\bar{x}))$ (cf. e.g., [1, Theorem 3.86]), that is, there exist a constant $\kappa > 0$ and a neighborhood N of $(\bar{x}, \theta(\bar{x}))$ such that

$$f(x) + t \geq f(\bar{x}) + \theta(\bar{x}) + \kappa \|(x, t) - (\bar{x}, \theta(\bar{x}))\|^2, \quad \forall (x, t) \in \widehat{F}(0, 0) \cap N.$$

The following proposition, which is taken from [8, Theorem 24], characterizes the robust isolated calmness of problem (3.2) via the SOSC (3.7) and the SRCQ (3.6).

Proposition 3.1. *Suppose that $(\bar{x}, \theta(\bar{x})) \in \mathbb{X} \times \mathbb{R}$ is a feasible solution of problem (3.2) with $(\delta_1, \delta_2) = 0$. Suppose that the RCQ (3.5) holds at $(\bar{x}, \theta(\bar{x}))$. Assume that $\mathrm{epi}\,\theta$ is $\mathcal{C}^2$-cone reducible at $(\bar{x}, \theta(\bar{x}))$. Let $(\bar{y}, \bar{z}, -1) \in \widehat{\mathcal{M}}(\bar{x}, \theta(\bar{x}), 0, 0)$. Then the following two statements are equivalent to each other:*
(i) The SOSC (3.7) holds at $(\bar{x}, \theta(\bar{x}))$ and the SRCQ (3.6) holds at $(\bar{x}, \theta(\bar{x}))$ for $(\bar{y}, \bar{z}, -1)$.
(ii) The point $(\bar{x}, \theta(\bar{x}))$ is a local optimal solution of problem (3.2) and the KKT solution mapping $\widehat{S}_{KKT}$ is robustly isolated calm at the origin for $(\bar{x}, \theta(\bar{x}), \bar{y}, \bar{z}, -1)$.

Now we return to the nonsmooth optimization problem (3.1). Let $(\delta_1, \delta_2) \in \mathbb{X} \times \mathbb{Y}$ be given. We say that $\bar{x}$ is a feasible solution to problem (3.1) if

$$\bar{x} \in F(\delta_1, \delta_2) := \{x \in \mathrm{dom}\,\theta \mid g(x) + \delta_2 \in \mathcal{P}\}.$$

Denote $l : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$ by

$$l(x, y) := f(x) + \langle g(x), y \rangle, \quad (x, y) \in \mathbb{X} \times \mathbb{Y}.$$

Then the KKT optimality condition takes the form of

$$\begin{cases} \delta_1 \in \nabla_x l(x, y) + \partial\theta(x), \\ y \in \mathcal{N}_{\mathcal{P}}(g(x) + \delta_2), \end{cases} \quad (x, y) \in \mathbb{X} \times \mathbb{Y}. \tag{3.8}$$

Let $S_{\mathrm{KKT}} : \mathbb{X} \times \mathbb{Y} \to \mathbb{X} \times \mathbb{Y}$ be the following KKT solution mapping:

$$S_{\mathrm{KKT}}(\delta_1, \delta_2) := \{(x, y) \in \mathbb{X} \times \mathbb{Y} \mid (x, y) \text{ satisfies (3.8)}\}, \ (\delta_1, \delta_2) \in \mathbb{X} \times \mathbb{Y}. \tag{3.9}$$

For any given $(\delta_1, \delta_2) \in \mathbb{X} \times \mathbb{Y}$, we call $\bar{x}$ a stationary point of problem (3.1) if there exists a Lagrangian multiplier $\bar{y} \in \mathbb{Y}$ such that $(\bar{x}, \bar{y}) \in S_{\mathrm{KKT}}(\delta_1, \delta_2)$. Denote $\mathcal{M}(\bar{x}, \delta_1, \delta_2) \subseteq \mathbb{Y}$ as the set of all such Lagrange multipliers $\bar{y}$ associated with $\bar{x}$.

The following proposition establishes the equivalence between the robust isolated calmness of the KKT solution mappings with respect to problem (3.1) and problem (3.2).

Proposition 3.2. *Let $(\bar{x}, \theta(\bar{x})) \in \mathbb{X} \times \mathbb{R}$ be a local optimal solution of problem (3.2) with $\widehat{\mathcal{M}}(\bar{x}, \theta(\bar{x}), 0, 0) \neq \emptyset$. Let $(\bar{y}, \bar{z}, -1) \in \widehat{\mathcal{M}}(\bar{x}, \theta(\bar{x}), 0, 0)$. If the KKT solution mapping $\widehat{S}_{KKT}$ given in (3.4) is robustly isolated calm at the origin for $(\bar{x}, \theta(\bar{x}), \bar{y}, \bar{z}, -1)$, then the KKT solution mapping S_{KKT} given in (3.9) is robustly isolated calm at the origin for $(\bar{x}, \bar{y})$. The reverse implication is true if the function θ is Lipschitz continuous at $\bar{x}$.*

Proof. Note from [4, Corollary 2.4.9] that

$$(z, -1) \in \mathcal{N}_{\mathrm{epi}\,\theta}\big(x, \theta(x)\big) \iff z \in \partial\theta(x), \quad \forall\, x, z \in \mathbb{X}.$$

Then for any $(\delta_1, \delta_2) \in \mathbb{X} \times \mathbb{Y}$ and any $(x, t, y, z, -1) \in \widehat{S}_{\mathrm{KKT}}(\delta_1, \delta_2)$, we know from (3.3) and (3.8) that $(x, y) \in S_{\mathrm{KKT}}(\delta_1, \delta_2)$. Thus, the first part of this proposition follows easily from the definition of the robust isolated calmness.

Conversely, consider any $(\delta_1, \delta_2) \in \mathbb{X} \times \mathbb{Y}$ and $(x, y) \in S_{\mathrm{KKT}}(\delta_1, \delta_2)$. Let $z = \delta_1 - \nabla f(x) - \nabla g(x)y$. By the similar arguments as above, we have $(x, \theta(x), y, z, -1) \in \widehat{S}_{\mathrm{KKT}}(\delta_1, \delta_2)$. Since f and g are assumed to be twice continuously differentiable, $\nabla f(\cdot)$ and $\nabla g(\cdot)$ are locally Lipschitz continuous at $\bar{x}$. Then there exists a constant $\kappa > 0$ (only depending on $\bar{x}$ and $\bar{y}$) such that for any x sufficiently close to $\bar{x}$,

$$\begin{cases} \|\theta(x) - \theta(\bar{x})\| \leq \kappa \|x - \bar{x}\|, \\ \|z - \bar{z}\| \leq \|\delta_1\| + \|\nabla f(x) - \nabla f(\bar{x})\| + \|\nabla g(x)y - \nabla g(\bar{x})\bar{y}\| \leq k(\|\delta_1\| + \|x - \bar{x}\| + \|y - \bar{y}\|). \end{cases}$$

Consequently, the second assertion of this proposition also follows from the definition of the robust isolated calmness.

As mentioned in Section 2, the closed convex function $\theta(\cdot)$ is always directionally epidifferentiable at $x \in \mathrm{dom}\,\theta$. Then by [15, Theorem 23.2], the KKT optimality condition (3.8) is equivalent to

$$\begin{cases} \theta^{\downarrow}(x; d) + \langle \nabla_x l(x, y) - \delta_1, d \rangle \geq 0, \\ y \in \mathcal{N}_{\mathcal{P}}(g(x) + \delta_2), \end{cases} \quad \forall\, d \in \mathbb{X},$$

where $(x, y) \in \mathbb{X} \times \mathbb{Y}$. Define the critical of the function θ by

$$\mathcal{C}_\theta(x, z) := \{d \in \mathcal{T}_{\mathrm{dom}\,\theta}(x) \mid \theta^{\downarrow}(x; d) = \langle d, z \rangle\}, \quad (x, z) \in \mathrm{dom}\,\theta \times \mathbb{X}. \tag{3.10}$$

Let $(\delta_1, \delta_2) = 0$. The RCQ is said to hold at a feasible solution $\bar{x}$ of problem (3.1) if

$$\begin{pmatrix} g'(\bar{x}) \\ \mathcal{I}_{\mathbb{X}} \end{pmatrix} \mathbb{X} + \begin{pmatrix} \mathcal{T}_{\mathcal{P}}(g(\bar{x})) \\ \mathcal{T}_{\mathrm{dom}\,\theta}(\bar{x}) \end{pmatrix} = \begin{pmatrix} \mathbb{Y} \\ \mathbb{X} \end{pmatrix}. \tag{3.11}$$

By the equations in (2.1), the SRCQ is said to hold at a stationary point $\bar{x}$ for $\bar{y} \in \mathcal{M}(\bar{x}, 0, 0)$ if

$$\begin{pmatrix} g'(\bar{x}) \\ \mathcal{I}_{\mathbb{X}} \end{pmatrix} \mathbb{X} + \begin{pmatrix} \mathcal{T}_{\mathcal{P}}(g(\bar{x})) \cap \bar{y}^{\perp} \\ \mathcal{C}_\theta(\bar{x}, -\nabla_x l(\bar{x}, \bar{y})) \end{pmatrix} = \begin{pmatrix} \mathbb{Y} \\ \mathbb{X} \end{pmatrix}. \tag{3.12}$$

The critical cone at a feasible solution $\bar{x}$ of problem (3.1) is given by

$$\mathcal{C}(\bar{x}) := \left\{ d \in \mathbb{X} \mid g'(\bar{x})d \in \mathcal{T}_{\mathcal{P}}(g(\bar{x})),\ d \in \mathcal{T}_{\mathrm{dom}\,\theta}(\bar{x}),\ f'(\bar{x})d + \theta^{\downarrow}(\bar{x}; d) \leq 0 \right\}.$$

If $\bar{x}$ is a stationary point of problem (3.1) and $\bar{y} \in \mathcal{M}(\bar{x}, 0, 0)$, then

$$\begin{aligned} \mathcal{C}(\bar{x}) &= \left\{ d \in \mathbb{X} \mid g'(\bar{x})d \in \mathcal{T}_{\mathcal{P}}(g(\bar{x})),\ d \in \mathcal{T}_{\mathrm{dom}\,\theta}(\bar{x}),\ f'(\bar{x})d + \theta^{\downarrow}(\bar{x}; d) = 0 \right\} \\ &= \left\{ d \in \mathbb{X} \mid g'(\bar{x})d \in \mathcal{T}_{\mathcal{P}}(g(\bar{x})) \cap \bar{y}^{\perp},\ d \in \mathcal{C}_\theta(\bar{x}, -\nabla f(\bar{x})) \right\}. \end{aligned}$$

Based on the equations in (2.2), we know that if θ is $\mathcal{C}^2$-cone reducible at $\bar{x}$, then the SOSC at $\bar{x}$ for problem (3.1) with $(\delta_1, \delta_2) = 0$ takes the form of

$$\sup_{\bar{y} \in \mathcal{M}(\bar{x}, 0, 0)} \left\{ \langle d, \nabla^2_{xx} l(\bar{x}, \bar{y})d \rangle - \psi^*_{(\bar{x}, d)}(-\nabla_x l(\bar{x}, \bar{y})) \right\} > 0, \quad \forall\, d \in \mathcal{C}(\bar{x}) \setminus \{0\}, \tag{3.13}$$

where $\psi^*_{(x, d)}(\cdot)$ is the conjugate function of $\psi_{(x, d)}(\cdot) = \theta^{\downarrow\downarrow}_-(x; d, \cdot)$ for any $x \in \mathrm{dom}\,\theta$ and any $d \in \mathbb{X}$.

Combining Proposition 3.1 and Proposition 3.2, we are ready to state the main result of this section.

Proposition 3.3. *Let $\bar{x} \in \mathbb{X}$ be a local optimal solution of problem (3.1) with $(\delta_1, \delta_2) = 0$. Suppose that the RCQ (3.11) holds at $\bar{x}$. Assume that θ is $\mathcal{C}^2$-cone reducible and Lipschitz continuous at $\bar{x}$. Let $\bar{y} \in \mathcal{M}(\bar{x}, 0, 0)$. Then the following two statements are equivalent to each other:*
(i) The SOSC (3.13) holds at $\bar{x}$ and the SRCQ (3.12) holds at $\bar{x}$ for $\bar{y}$.
(ii) The point $\bar{x}$ is a local optimal solution of problem (3.1) and the KKT solution mapping S_{KKT} is robustly isolated calm at the origin for $(\bar{x}, \bar{y})$.

4. Variational Analysis of the Nuclear Norm Function

Throughout this section, we denote $\theta : \mathbb{R}^{m \times n} \to \mathbb{R}$ as the nuclear norm function. Obviously $\operatorname{dom} \theta = \mathbb{R}^{m \times n}$. Since the nuclear norm function is convex and globally Lipschitz continuous, for any $X \in \mathbb{R}^{m \times n}$, both $\theta_-^{\downarrow}(X; \cdot)$ and $\theta_+^{\downarrow}(X; \cdot)$ defined in Section 2 coincide with $\theta'(X; \cdot)$, the conventional directional derivative of θ at X [1, Section 2.2.3]. In addition, it is known that the singular value function in $\mathbb{R}^{m \times m}$ is second order directionally differentiable [22, Theorem 3.1]. This further implies the second order directional differentiability of the nuclear norm function, i.e., both $\theta_-^{\downarrow\downarrow}(X; \cdot)$ and $\theta_+^{\downarrow\downarrow}(X; \cdot)$ coincide with $\theta''(X; \cdot)$, the conventional second order directional derivative of θ at X.

Fix an arbitrary matrix $X \in \mathbb{R}^{m \times n}$. Suppose that X admits the following singular-value decomposition (SVD):

$$X = U[\operatorname{Diag}(\sigma(X))\ 0]V^T = U\operatorname{Diag}(\sigma(X))V_1^T, \tag{4.1}$$

where $U \in \mathcal{O}^m$ and $V = [V_1\ V_2] \in \mathcal{O}^n$ are the left and right singular vectors of X with $V_1 \in \mathbb{R}^{n \times m}$ and $V_2 \in \mathbb{R}^{n \times (n-m)}$. Define the index sets

$$a := \{1 \le i \le m \mid \sigma_i(X) \in (1, +\infty)\}, \quad b := \{1 \le i \le m \mid \sigma_i(X) \in [0,1]\}, \quad c = \{m+1, \ldots, n\}. \tag{4.2}$$

Denote the distinct singular values of X that are greater than 1 by $\nu_1(X) > \ldots > \nu_r(X) > 1$, where r is a non-negative integer. We further divide the sets a and b into the following subsets:

$$
\begin{aligned}
a_l &:= \{i \in a \mid \sigma_i(X) = \nu_l(X)\}, \quad l = 1, \ldots, r, \\
b_1 &:= \{i \in b \mid \sigma_i(X) = 1\}, \quad b_2 := \{i \in b \mid 0 < \sigma_i(X) < 1\}, \quad b_3 := \{i \in b \mid \sigma_i(X) = 0\}.
\end{aligned} \tag{4.3}
$$

Let us first review the concept of Löwner's operator and its differential properties. Suppose that $X \in \mathbb{R}^{m \times n}$ has the SVD (4.1). For any scalar function $g : \mathbb{R} \to \mathbb{R}$, define the corresponding matrix valued function G by

$$G(X) := U[\operatorname{Diag}(g(\sigma_1(X)), g(\sigma_2(X)), \ldots, g(\sigma_m(X)))\ 0]V^T.$$

Such a function is called Löwner's operator associated with the function g, which is first studied by Löwner in the context of symmetric matrices [13]. In particular, let $\phi : \mathbb{R} \to \mathbb{R}$ be the scalar function

$$\phi(x) := \max\{x - 1, 0\}, \quad x \in \mathbb{R}.$$

It is easy to verify that the proximal mapping of θ can be expressed as:

$$\operatorname{Prox}_\theta(X) = U[\operatorname{Diag}(\phi(\sigma_1(X)), \ldots, \phi(\sigma_m(X)))\ 0]V^T, \quad X \in \mathbb{R}^{m \times n}. \tag{4.4}$$

Clearly $\mathrm{Prox}_\theta(\cdot)$ can be taken as Löwner's operator associated with the function ϕ. The directional derivate of $\mathrm{Prox}_\theta(\cdot)$ can thus be obtained via the general formula regarding the directional derivative of Löwner's operator [7]. Obviously ϕ is directionally differentiable with the directional derivative

$$\phi'(x;d) = \begin{cases} d & \text{if } x > 1, \\ \max\{d,0\} & \text{if } x = 1, \\ 0 & \text{if } x < 1, \end{cases} \quad x \in \mathbb{R}, \ d \in \mathbb{R}.$$

For any positive integer p, define linear matrix operators $S : \mathbb{R}^{p \times p} \to \mathbb{S}^p$ and $T : \mathbb{R}^{p \times p} \to \mathbb{R}^{p \times p}$ by

$$S(X) := \frac{1}{2}(X + X^T), \quad T(X) := \frac{1}{2}(X - X^T), \quad X \in \mathbb{R}^{p \times p}. \tag{4.5}$$

Denote $\Xi_{aa}^2 : \mathbb{R}^{m \times n} \to \mathbb{R}^{|a| \times |a|}$, $\Xi_{ab}^1 : \mathbb{R}^{m \times n} \to \mathbb{R}^{|a| \times |b|}$, $\Xi_{ab}^2 : \mathbb{R}^{m \times n} \to \mathbb{R}^{|a| \times |b|}$ and $\Xi_2 : \mathbb{R}^{m \times n} \to \mathbb{R}^{|a| \times |c|}$ as

$$\begin{cases} ((\Xi_{aa}^2)(X))_{ij} := \dfrac{\sigma_i(X) + \sigma_j(X) - 2}{\sigma_i(X) + \sigma_j(X)}, \ i = 1, 2, \ldots, |a|, \ j = 1, 2, \ldots, |a|, \\[2mm] ((\Xi_{ab}^1)(X))_{ij} := \dfrac{\sigma_i(X) - 1}{\sigma_i(X) - \sigma_{j+|a|}(X)}, \ i = 1, 2, \ldots, |a|, \ j = 1, 2, \ldots, |b|, \\[2mm] ((\Xi_{ab}^2)(X))_{ij} := \dfrac{\sigma_i(X) - 1}{\sigma_i(X) + \sigma_{j+|a|}(X)}, \ i = 1, 2, \ldots, |a|, \ j = 1, 2, \ldots, |b|, \\[2mm] ((\Xi_2)(X))_{ij} := \dfrac{\sigma_i(X) - 1}{\sigma_i(X)}, \quad i = 1, 2, \ldots, |a|, \ j = 1, 2, \ldots, n - m, \end{cases} \quad X \in \mathbb{R}^{m \times n}.$$

Denote $\Gamma_1 : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \to \mathbb{R}^{|a| \times |a|}$, $\Gamma_2 : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \to \mathbb{R}^{|a| \times |b|}$, $\Gamma_3 : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \to \mathbb{R}^{|b| \times |a|}$ and $\Gamma_4 : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \to \mathbb{R}^{|a| \times |c|}$ as

$$\begin{cases} \Gamma_1(X,H) &:= \ (S(H_1))_{aa} + \Xi_{aa}^2(X) \circ (T(H_1))_{aa}, \\[1mm] \Gamma_2(X,H) &:= \ \Xi_{ab}^1(X) \circ (S(H_1))_{ab} + \Xi_{ab}^2(X) \circ (T(H_1))_{ab}, \\[1mm] \Gamma_3(X,H) &:= \ (\Xi_{ab}^1(X))^T \circ (S(H_1))_{ba} + (\Xi_{ab}^2(X))^T \circ (T(H_1))_{ba}, \\[1mm] \Gamma_4(X,H) &:= \ \Xi_2(X) \circ H_{ac}, \end{cases} \quad (X,H) \in \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n},$$

where $\circ$ denotes the Hadamard product between two matrices and $H = [H_1 \ \ H_2]$ with $H_1 \in \mathbb{R}^{m \times m}$ and $H_2 \in \mathbb{R}^{m \times (n-m)}$. Then by [7, Theorem 3], the directional derivative of $\mathrm{Prox}_\theta(\cdot)$ at $X \in \mathbb{R}^{m \times n}$ in the direction $H \in \mathbb{R}^{m \times n}$ takes the form of

$$\mathrm{Prox}_\theta'(X;H) = U \begin{pmatrix} \Gamma_1(X,\widetilde{H}) & \Gamma_2(X,\widetilde{H}) & & \Gamma_4(X,\widetilde{H}) \\ & \begin{array}{cc} \Pi_{\mathbb{S}_+^{|b_1|}}(S(\widetilde{H}_{b_1 b_1})) & 0 \ \ \ 0 \end{array} & & \\ \Gamma_3(X,\widetilde{H}) & \begin{array}{cc} 0 & 0_{b_2 \times b_2} \ \ \ 0 \end{array} & 0 \\ & \begin{array}{cc} 0 & 0 \ \ \ 0_{b_3 \times b_3} \end{array} & & \end{pmatrix} V^T, \tag{4.6}$$

where $\widetilde{H} = U^T H V$.

In [19], Watson shows that the subdifferential of the nuclear norm function takes the following form:

$$\partial\theta(X) = \left\{ U_a V_a^T + U_b W [V_b \ V_2]^T \mid W \in \mathbb{R}^{|b| \times (n-|a|)}, \ \|W\|_2 \leq 1 \right\}, \quad X \in \mathbb{R}^{m \times n} \tag{4.7}$$

Therefore, for any $H \in \mathbb{R}^{m \times n}$, the directional derivative of θ at X in the direction H is given by

$$\theta'(X; H) = \sup_{S \in \partial \theta(X)} \langle H, S \rangle = \operatorname{tr}(U_a^T H V_a) + \|U_b^T H [V_b \ V_2]\|_*. \tag{4.8}$$

Let $A = \operatorname{Prox}_\theta(X)$ and $B = \operatorname{Prox}_{\theta^*}(X)$. Define the critical cone of θ at A for B as

$$\mathcal{C}_\theta(A, B) := \{H \in \mathbb{R}^{m \times n} \mid \theta'(A; H) = \langle H, B \rangle\}. \tag{4.9}$$

Similarly, define the critical cone of θ^* at B for A as

$$\mathcal{C}_{\theta^*}(B, A) := \{H \in \mathbb{R}^{m \times n} \mid (\theta^*)'(B; H) = \langle H, A \rangle\}. \tag{4.10}$$

As can be seen from (3.13), in order to analyze the SOSC for problem (1.2), one needs to compute the conjugate of the second order epiderivative of θ. This has already been done in [6]. Firstly, it follows from (4.4) that $\sigma(A) = \max\{\sigma(X) - 1, 0\}$. Specifically,

$$\sigma_i(A) = \begin{cases} \sigma_i(X) - 1, & \text{if } i \in a, \\ 0 & \text{if } i \in b. \end{cases} \tag{4.11}$$

Clearly A has r numbers of nonzero distinct singular values. Denote them by $\nu_1(A) > \ldots > \nu_r(A)$. The index sets $a_1, \ldots, a_r$ that depending on X in (4.3) also provides a partition of $(\sigma_i(A))_{i \in a}$, i.e.,

$$\sigma_i(A) = \nu_l(A), \quad \forall i \in a_l, \ \forall l = 1, \ldots, r.$$

For $l = 1, \ldots, r$, denote $\Omega_{al} : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \to \mathbb{R}^{a_l \times a_l}$ as

$$\begin{aligned} \Omega_{a_l}(A, H) := \quad & (S(\widetilde{H}_1))_{a_l}^T (\operatorname{Diag}(\sigma(A)) - \nu_l(A) I_m)^\dagger (S(\widetilde{H}_1))_{a_l} + (2\nu_l(A))^{-1} \widetilde{H}_{a_l c} \widetilde{H}_{a_l c}^T \\ & + (T(\widetilde{H}_1))_{a_l}^T (-\operatorname{Diag}(\sigma(A)) - \nu_l(A) I_m)^\dagger (T(\widetilde{H}_1))_{a_l}, \quad (A, H) \in \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n}, \end{aligned}$$

where I_m is the $m \times m$ identity matrix and $Z^\dagger$ denotes the Moore-Penrose pseudoinverse of a given matrix Z. Then for any $H \in \mathbb{R}^{m \times n}$, the conjugate of $\theta''(A; H, \cdot)$ is

$$\psi_{(A,H)}^*(B) := (\theta''(A; H, \cdot))^*(B) = 2 \sum_{l=1}^r \operatorname{tr}(\Omega_{a_l}(A, H)) + 2\langle \operatorname{Diag}(\sigma_b(B)), U_b^T H A^\dagger H V_b \rangle, \tag{4.12}$$

where $\sigma_b(B) = (\sigma_i(B))_{i \in b}$ and $\widetilde{H} = [\widetilde{H}_1 \ \widetilde{H}_2] = [U^T H V_1 \ U^T H V_2]$.

In the following, we present several properties regarding the critical cone of θ and the directional derivative of $\operatorname{Prox}_\theta(\cdot)$.

Proposition 4.1. *Suppose that $X \in \mathbb{R}^{m \times n}$ has the singular value decomposition (4.1). Let the index sets $a, b, a_1, \ldots, a_l, b_1, b_2, b_3$ be given by (4.2) and (4.3). Given any $H \in \mathbb{R}^{m \times n}$, denote $\widetilde{H} = U^T H V$ for $(U, V) \in \mathcal{O}^{m \times n}(X)$. Denote $A = \operatorname{Prox}_\theta(X)$ and $B = \operatorname{Prox}_{\theta^*}(X)$. Then the following conclusions hold:*

(i) $H \in \mathcal{C}_\theta(A, B)$ if and only if H satisfies

(ii) For any $D \in \mathbb{R}^{m \times n}$, $H = \operatorname{Prox}_\theta'(X; H + D)$ if and only if $H \in \mathcal{C}_\theta(A, B)$ and $\langle H, D \rangle = -\psi_{(A,H)}^*(B)$, where the function $\psi_{(A,H)}^*(\cdot)$ is given in (??).

$$
\widetilde{H} = \left(
\begin{array}{c|ccc|c}
\widetilde{H}_{aa} & & \widetilde{H}_{ab} & & \widetilde{H}_{ac} \\
\hline
& \Pi_{\mathcal{S}_+^{|b_1|}}(\widetilde{H}_{b_1 b_1}) & 0 & 0 & \\
\widetilde{H}_{ba} & 0 & 0_{b_2 \times b_2} & 0 & 0 \\
& 0 & 0 & 0_{b_3 \times b_3} &
\end{array}
\right). \tag{4.13}
$$

Proof. The result of part (i) can be obtained from [6, Proposition 10]. Now we derive (ii). Suppose $H = \mathrm{Prox}_\theta'(C; H + D)$. Denote $\widetilde{H} = [\widetilde{H}_1, \widetilde{H}_2]$ with $\widetilde{H}_1 \in \mathbb{R}^{m \times m}$ and $\widetilde{H}_2 \in \mathbb{R}^{m \times (n-m)}$. Direct computations of $\psi^*_{(A,H)}(B)$ given in (**??**) show that

$$
\begin{aligned}
\psi^*_{(A,H)}(B) =\ & \sum_{1 \le l,t \le r} \frac{2}{-\nu_t(A) - \nu_l(A)} \|(T(\widetilde{H}_1))_{a_l a_t}\|^2 + \sum_{1 \le l \le r} \frac{4}{-\nu_l(A)} \|(T(\widetilde{H}_1))_{a_l b_1}\|^2 \\
& + \sum_{\substack{1 \le l \le r \\ 1 \le i - |a| - |b_1| \le |b_2|}} \left(\frac{2(1 - \sigma_i(B))}{-\nu_l(A)} \|(S(\widetilde{H}_1))_{a_l i}\|^2 + \frac{2(\sigma_i(B) + 1)}{-\nu_l(A)} \|(T(\widetilde{H}_1))_{a_l i}\|^2 \right) \\
& + \sum_{\substack{1 \le l \le r \\ 1 \le i - |a| - |b_1| - |b_2| \le |b_3|}} \left(\frac{2}{-\nu_l(A)} \|(S(\widetilde{H}_1))_{a_l i}\|^2 + \frac{2}{-\nu_l(A)} \|(T(\widetilde{H}_1))_{a_l i}\|^2 \right) \\
& + \sum_{1 \le l \le r} \frac{1}{-\nu_l(A)} \|(\widetilde{H}_c)_{a_l c}\|^2.
\end{aligned}
\tag{4.14}
$$

Recall the formula of $\mathrm{Prox}_\theta'(X; \cdot)$ given in (**??**). We deduce that

$$
\widetilde{H} = \left(
\begin{array}{c|ccc|c}
\widetilde{H}_{aa} & & \widetilde{H}_{ab} & & \widetilde{H}_{ac} \\
\hline
& \widetilde{H}_{b_1 b_1} & 0 & 0 & \\
\widetilde{H}_{ba} & 0 & 0_{b_2 \times b_2} & 0 & 0 \\
& 0 & 0 & 0_{b_3 \times b_3} &
\end{array}
\right)
$$

and

$$
\begin{cases}
\widetilde{D}_{a_l a_t} = \dfrac{2}{\nu_l(X) + \nu_t(X) - 2}(T(\widetilde{H}_1))_{a_l a_t}, \ 1 \le l, t \le r, \\[2mm]
(\widetilde{D}_{ab})_{ij} = \dfrac{1}{\sigma_i(X) - 1}(\widetilde{H}_{ab})_{ij} - \dfrac{\sigma_{j+|a|}(X)}{\sigma_i(X) - 1}(\widetilde{H}_{ab})_{ji}, \ i = 1, 2, \dots |a|, \ j = 1, 2, \dots, |b|, \\[2mm]
(\widetilde{D}_{ba})_{ji} = \dfrac{1}{\sigma_i(X) - 1}(\widetilde{H}_{ab})_{ji} - \dfrac{\sigma_{j+|a|}(X)}{\sigma_i(X) - 1}(\widetilde{H}_{ab})_{ij}, \ i = 1, 2, \dots |a|, \ j = 1, 2, \dots, |b|, \\[2mm]
(\widetilde{D}_{ac})_{ij} = \dfrac{1}{\sigma_i(X) - 1}(\widetilde{H}_{ac})_{ij}, \ i = 1, 2, \dots, |a|, \ j = 1, 2, \dots, n - m, \\[2mm]
\mathbb{S}_+^{|b_1|} \ni \widetilde{H}_{b_1 b_1} = S(\widetilde{H}_{b_1 b_1}) \perp S(\widetilde{D}_{b_1 b_1}) \in \mathbb{S}_-^{|b_1|},
\end{cases}
$$

where $\widetilde{D} = U^T D V$. Consequently, it follows from part (i) of this proposition that $H \in \mathcal{C}_\theta(A, B)$.

Moreover, we have

$$
\begin{aligned}
\langle D, H\rangle \;=\;& \langle \tilde{d}_{aa}, \widetilde{H}_{aa}\rangle + \langle \tilde{d}_{ab}, \widetilde{H}_{ab}\rangle + \langle \tilde{d}_{ba}, \widetilde{H}_{ba}\rangle + \langle \tilde{d}_{a2}, \widetilde{H}_{a2}\rangle \\[2mm]
=\;& \sum_{1\le t,l\le r} \frac{2}{\nu_l(X)+\nu_t(X)-2}\|(T(\widetilde{H}_1))_{a_l a_t}\|^2 + \sum_{1\le l\le r}\frac{4}{\nu_l(X)-1}\|(T(\widetilde{H}_1))_{a_l b_1}\|^2 \\[2mm]
& + \sum_{\substack{1\le l\le r \\ 1\le i-|a|-|\bar b_1|\le |b_2|}} \left(\frac{2(1-\sigma_i(X))}{\nu_l(X)-1}\|(S(\widetilde{H}_1))_{a_l i}\|^2 + \frac{2(\sigma_i(X)+1)}{\nu_l(X)-1}\|(T(\widetilde{H}_1))_{a_l i}\|^2 \right) \\[2mm]
& + \sum_{\substack{1\le l\le r \\ 1\le i-|a|-|\bar b_1|-|b_2|\le |b_3|}} \left(\frac{2}{\nu_l(X)-1}\|(S(\widetilde{H}_1))_{a_l i}\|^2 + \frac{2}{\nu_l(X)-1}\|(T(\widetilde{H}_1))_{a_l i}\|^2 \right) \\[2mm]
& + \sum_{1\le l\le r}\frac{1}{\nu_l(X)-1}\|(\widetilde{H}_2)_{a_l c}\|^2 .
\end{aligned}
$$

Note from (??) that $\nu_l(A) = \nu_l(X)-1$ for any $l = 1,\dots,r$. Hence, $\langle D, H\rangle = -\psi^*_{(A,H)}(B)$ by (??) and the above equation. The converse of this statement can be established by reversing the above arguments.

Proposition 4.2. *Suppose that $X \in \mathbb{R}^{m\times n}$ has the singular value decomposition (4.1). Let the index sets $a, b, a_1,\dots,a_l, b_1, b_2, b_3$ be given by (4.2) and (4.3). Given any $H \in \mathbb{R}^{m\times n}$, denote $\widetilde{H} = U^T H V$ for $(U,V) \in \mathcal{O}^{m\times n}(X)$. Denote $A = \mathrm{Prox}_\theta(X)$ and $B = \mathrm{Prox}_{\theta^*}(X)$. Then the following conclusions hold:*

(i) $H \in \mathcal{C}_{\theta^*}(B, A)$ if and only if H satisfies $S(\widetilde{H}_{b_1 b_1}) \in \mathbb{S}^{|b_1|}_-$ and

$$
\widetilde{H} =
\left(
\begin{array}{c|c|c|c|c}
T(\widetilde{H}_{aa}) & \frac{1}{2}(\widetilde{H}_{ab_1}-\widetilde{H}^T_{b_1 a}) & \widetilde{H}_{ab_2} & \widetilde{H}_{ab_3} & \\
\hline
\frac{1}{2}(\widetilde{H}_{b_1 a}-\widetilde{H}^T_{ab_1}) & \widetilde{H}_{b_1 b_1} & \widetilde{H}_{b_1 b_2} & \widetilde{H}_{b_1 b_3} & \\
\hline
\widetilde{H}_{b_2 a} & \widetilde{H}_{b_2 b_1} & \widetilde{H}_{b_2 b_2} & \widetilde{H}_{b_2 b_3} & \widetilde{H}_c \\
\hline
\widetilde{H}_{b_3 a} & \widetilde{H}_{b_3 b_1} & \widetilde{H}_{b_3 b_2} & \widetilde{H}_{b_3 b_3} &
\end{array}
\right),
\qquad (4.15)
$$

where the linear operators $S(\cdot)$ and $T(\cdot)$ are defined in (4.5).

(ii) $H \in (\mathcal{C}_\theta(A,B))^\circ$ if and only if $\phi^*_{(B,H)}(A) = 0$ and $H \in \mathcal{C}_{\theta^*}(B,A)$, where $\phi^*_{(B,H)}(\cdot)$ is the conjugate function of $\phi_{(B,H)}(\cdot) = (\theta^*)''(B; H, \cdot)$.

(iii) $H \in (\mathcal{C}_{\theta^*}(B,A))^\circ$ if and only if $\psi^*_{(A,H)}(B) = 0$ and $H \in \mathcal{C}_\theta(A,B)$.

Proof. Part (i) follows from [6, Proposition 12]. To prove part (ii), we use a result from [6, Proposition 16] stating that $\phi^*_{(B,H)}(A) = 0$ if and only if $\psi^*_{(A,H)}(B) = 0$ for any $H \in \mathbb{R}^{m\times n}$, which is further equivalent to

$$
\widetilde{H}_{aa} \in \mathbb{S}^{|a|}, \quad \widetilde{H}_{ab_1} = \widetilde{H}^T_{b_1 a}, \quad \widetilde{H}_{ab_2} = \widetilde{H}^T_{b_2 a} = 0, \quad \widetilde{H}_{ab_3} = \widetilde{H}^T_{b_3 a} = 0, \quad \widetilde{H}_{ac} = 0.
$$

Then by part (i) of this proposition, one can see that $\phi^*_{(B,H)}(A) = 0$ and $H \in \mathcal{C}_{\theta^*}(B, A)$ imply that

$$\widetilde{H}_{aa} = 0, \quad \widetilde{H}_{ab} = 0, \quad \widetilde{H}_{ac} = 0, \quad \widetilde{H}_{ba} = 0, \quad S(\widetilde{H}_{b_1 b_1}) \in \mathbb{S}^{|b_1|}_-.$$

In view of Proposition 4.1, this is equivalent to have $H \in (\mathcal{C}_\theta(A, B))^\circ$. To prove part (iii), it suffices to note that either $H \in (\mathcal{C}_{\theta^*}(B, A))^\circ$ or $\psi^*_{(A,H)}(B) = 0$ with $H \in \mathcal{C}_\theta(A, B)$ is equivalent to

$$\widetilde{H} = \left(\begin{array}{ccc:c|c} S(\widetilde{H}_{aa}) & & \frac{1}{2}(\widetilde{H}_{ab_1} + \widetilde{H}^T_{b_1 a}) & 0 & 0 \\ \hdashline \frac{1}{2}(\widetilde{H}_{b_1 a} + \widetilde{H}^T_{ab_1}) & & \Pi_{\mathcal{S}^{|b_1|}_+}(\widetilde{H}_{b_1 b_1}) & 0 & 0 \\ \hdashline 0 & & 0 & 0 & 0 \\ \hline 0 & & 0 & 0 & 0 \end{array} \right. \left. \begin{array}{c} \\ \\ 0 \\ \\ \end{array} \right).$$

The proof of this proposition is completed.

5. The Robust Isolated Calmness of the KKT Solution Mapping

The aim of this section is to show that the SOSC for the primal problem (1.2) (the dual problem (1.4)) is in fact equivalent to the SRCQ for the dual problem (1.4) (the primal problem (1.2)). Before proceeding, we mention that a variation of this result regarding the linear semidefinite programming problem has been studied in [24].

Following the notation in the previous section, we use θ to denote the nuclear norm function in $\mathbb{R}^{m \times n}$. Let $\Omega_P \subseteq \mathbb{X}$ and $\Omega_D \subseteq \mathbb{R}^e \times \mathbb{R}^d \times \mathbb{R}^{m \times n}$ be the optimal solution sets of the primal problem (1.2) and the dual problem (1.4), respectively, both being assumed to be non-empty. It follows from (3.8) that the KKT optimality condition of problem (1.2) is given by

$$\begin{cases} 0 \in \mathcal{F}^* \nabla h(\mathcal{F}X) + C + \partial\theta(X) + \mathcal{A}^* y, \\ y \in \mathcal{N}_\mathcal{Q}(\mathcal{A}X - b), \end{cases} \quad (X, y) \in \mathbb{R}^{m \times n} \times \mathbb{R}^e. \quad (5.1)$$

We write $\mathcal{M}_P(\overline{X}) \subseteq \mathbb{R}^e$ as the set of Lagrange multipliers $\bar{y}$ associated with $\overline{X} \in \Omega_P$, i.e., $\bar{y} \in \mathcal{M}_P(\overline{X})$ if and only if $(\overline{X}, \bar{y})$ satisfies (5.1). Let $\mathcal{M}_D(\bar{y}, \bar{w}, \overline{S}) \subseteq \mathbb{R}^{m \times n}$ be the set of Lagrange multipliers associated with $(\bar{y}, \bar{w}, \overline{S}) \in \Omega_D$ for problem (1.4), i.e., $\overline{X} \in \mathcal{M}_D(\bar{y}, \bar{w}, \overline{S})$ if and only if $(\overline{X}, \bar{y}, \bar{w}, \overline{S})$ solves the following KKT system:

$$\begin{cases} \mathcal{A}X - b \in \mathcal{N}_{\mathcal{Q}^\circ}(y), \quad \mathcal{F}X \in \partial h^*(w), \\ X \in \partial\theta^*(S), \quad 0 = \mathcal{A}^* y + \mathcal{F}^* w + S + C, \end{cases} \quad (y, w, S, X) \in \mathbb{R}^e \times \mathbb{R}^d \times \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n}. \quad (5.2)$$

Since h is assumed to be essentially strictly convex, h^* is essentially smooth [15, Theorem 26.3]. Thus, ∇h^* is locally Lipschitz continuous and directionally differentiable on $\text{int}\,(\text{dom}\,h^*)$. Moreover, $\partial h^*(w) = \emptyset$ whenever $w \notin \text{int}\,(\text{dom}\,h^*)$ [15, Theorem 26.1]. Therefore, if (5.2) admits at least one solution, this KKT optimality condition can be equivalently written as

$$\begin{cases} \mathcal{A}X - b \in \mathcal{N}_{\mathcal{Q}^\circ}(y), \quad \mathcal{F}X \in \nabla h^*(w), \\ X \in \partial\theta^*(S), \quad 0 = \mathcal{A}^* y + \mathcal{F}^* w + S + C, \end{cases} \quad (y, w, S, X) \in \mathbb{R}^e \times \mathbb{R}^d \times \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n}.$$

As in Section 3, we consider the canonical perturbation of problem (1.2) for the sake of subsequent sensitivity analysis:

$$\min_X \quad h(\mathcal{F}X) + \langle C, X \rangle + \|X\|_* - \langle X, \delta_1 \rangle$$
$$\text{s.t.} \quad \mathcal{A}X - b + \delta_2 \in \mathcal{Q},$$

where $\delta_1 \in \mathbb{R}^{m \times n}$ and $\delta_2 \in \mathbb{R}^e$ are perturbation parameters. For any given $(\delta_1, \delta_2) \in \mathbb{R}^{m \times n} \times \mathbb{R}^e$, the KKT optimality condition then takes the form of

$$\begin{cases} \delta_1 \in \mathcal{F}^* \nabla h(\mathcal{F}X) + C + \partial\theta(X) + \mathcal{A}^* y, \\ y \in \mathcal{N}_{\mathcal{Q}}(\mathcal{A}X - b + \delta_2), \end{cases} \quad (X, y) \in \mathbb{R}^{m \times n} \times \mathbb{R}^e. \tag{5.3}$$

Let $S_{\text{KKT}} : \mathbb{R}^{m \times n} \times \mathbb{R}^e \to \mathbb{R}^{m \times n} \times \mathbb{R}^e$ be the following KKT solution mapping:

$$S_{\text{KKT}}(\delta_1, \delta_2) := \{(x, y) \in \mathbb{R}^{m \times n} \times \mathbb{R}^e \mid (x, y) \text{ satisfies } (5.3)\}, \ (\delta_1, \delta_2) \in \mathbb{R}^{m \times n} \times \mathbb{R}^e. \tag{5.4}$$

The RCQ of problem (1.2) at a feasible solution $\overline{X} \in \mathbb{R}^{m \times n}$ is given by

$$\mathcal{A}\mathbb{R}^{m \times n} + \mathcal{T}_{\mathcal{Q}}(\mathcal{A}\overline{X} - b) = \mathbb{R}^e. \tag{5.5}$$

Let $(y, q) \in \mathbb{R}^e \times \mathbb{R}^e$ satisfy $y \in \mathcal{N}_{\mathcal{Q}}(q)$. We denote the critical cone of $\mathcal{Q}$ at q for y and the critical cone of $\mathcal{Q}^\circ$ at y for q as

$$\mathcal{C}_{\mathcal{Q}}(q, y) := \mathcal{T}_{\mathcal{Q}}(q) \cap y^\perp, \qquad \mathcal{C}_{\mathcal{Q}^\circ}(y, q) := \mathcal{T}_{\mathcal{Q}^\circ}(y) \cap q^\perp.$$

It is easy to verify that

$$(\mathcal{C}_{\mathcal{Q}}(q, y))^\circ = \mathcal{C}_{\mathcal{Q}^\circ}(y, q). \tag{5.6}$$

The following theorem, which is the main result of our paper, demonstrates the equivalence between the primal SOSC and the dual SRCQ.

Theorem 5.1. *Let $\overline{X} \in \mathbb{R}^{m \times n}$ be an optimal solution of problem (1.2) with $\mathcal{M}_P(\overline{X}) \neq \emptyset$. Let $\bar{y} \in \mathcal{M}_P(\overline{X})$. Denote $\overline{S} := -\mathcal{A}^* \bar{y} - \mathcal{F}^* \nabla h(\mathcal{F}\overline{X}) - C$. Then the following two statements are equivalent to each other:*

(i) The SOSC holds at $\overline{X}$ for $\bar{y}$ with respect to the primal problem (1.2), i.e.,

$$\langle \mathcal{F}H, \nabla^2 h(\mathcal{F}\overline{X})\mathcal{F}H \rangle - \psi^*_{(\overline{X}, H)}(\overline{S}) > 0, \quad \forall H \in \mathcal{C}(\bar{x}) \setminus \{0\}, \tag{5.7}$$

where $\mathcal{C}(\overline{X}) := \mathcal{C}_{\mathcal{Q}}(\mathcal{A}\overline{X} - b, \bar{y}) \cap \mathcal{C}_\theta(\overline{X}, \overline{S})$.

(ii) The SRCQ holds at $\bar{y}$ for $\overline{X}$ with respect to the dual problem (1.4), i.e.,

$$\mathcal{F}^* \mathbb{R}^d + \mathcal{A}^* \mathcal{C}_{\mathcal{Q}^\circ}(\bar{y}, \mathcal{A}\overline{X} - b) + \mathcal{C}_{\theta^*}(\overline{S}, \overline{X}) = \mathbb{R}^{m \times n} \tag{5.8}$$

Proof. Firstly, let us assume that the statement (i) holds. Denote

$$\mathbb{E} := \mathcal{F}^* \mathbb{R}^d + \mathcal{A}^* \mathcal{C}_{\mathcal{Q}^\circ}(\bar{y}, \mathcal{A}\overline{X} - b) + \mathcal{C}_{\theta^*}(\overline{S}, \overline{X}).$$

Suppose on the contrary that $\mathbb{E} \neq \mathbb{R}^{m \times n}$. Then $\text{cl}(\mathbb{E}) \neq \mathbb{R}^{m \times n}$ [15, Theorem 6.3]. Hence, there exists $D \in \mathbb{R}^{m \times n}$ but $D \notin \text{cl}(\mathbb{E})$. Note that $\text{cl}(\mathbb{E})$ is a closed convex cone. Let

$$\overline{D} := D - \Pi_{\text{cl}(\mathbb{E})}(D) = \Pi_{(\text{cl}(\mathbb{E}))^\circ}(D) \neq 0.$$

Obviously, $\langle H, \overline{D} \rangle \leq 0$ for any $H \in \mathrm{cl}(\mathbb{E})$. This implies that

$$\mathcal{F}\overline{D} = 0, \quad \mathcal{A}\overline{D} \in (\mathcal{C}_{\mathcal{Q}^\circ}(\bar{y}, \mathcal{A}\overline{X} - b))^\circ, \quad \overline{D} \in (\mathcal{C}_{\theta^*}(\overline{S}, \overline{X}))^\circ.$$

Thus, it follows from (5.6) that $\mathcal{A}\overline{D} \in \mathcal{C}_{\mathcal{Q}}(\mathcal{A}\overline{X} - b, \bar{y})$. From Proposition 4.2, we also have that

$$\psi^*_{(\overline{X}, \overline{D})}(\overline{S}) = 0, \quad \overline{D} \in \mathcal{C}_\theta(\overline{X}, \overline{S}).$$

Therefore, $\overline{D} \in \mathcal{C}(\overline{X}) \setminus \{0\}$ and $\langle \mathcal{F}\overline{D}, \nabla^2 h(\mathcal{F}\overline{X})\mathcal{F}\overline{D} \rangle - \psi^*_{(\overline{X}, \overline{D})}(\overline{S}) = 0$, which contradicts the assumed SOSC (5.7) at $\overline{X}$.

The reverse implication can be proved similarly. Suppose that the SOSC (5.7) fails to hold at $\overline{X}$ for $\bar{y}$. Then there exists $\overline{H} \in \mathcal{C}(\overline{X}) \setminus \{0\}$ such that

$$\langle \mathcal{F}\overline{H}, \nabla^2 h(\mathcal{F}\overline{X})\mathcal{F}\overline{H} \rangle - \psi^*_{(\overline{X}, \overline{H})}(\overline{S}) = 0.$$

Since h is assumed to be essentially strictly convex, $\langle \mathcal{F}H, \nabla^2 h(\mathcal{F}\overline{X})\mathcal{F}H \rangle > 0$ for any $H \in \mathbb{R}^{m \times n}$ such that $\mathcal{F}H \neq 0$. It also follows from [6, Proposition 16] that $\psi^*_{(\overline{X}, H)}(\overline{S}) \leq 0$ for any $H \in \mathbb{R}^{m \times n}$. Consequently,

$$\mathcal{F}\overline{H} = 0, \quad \psi^*_{(\overline{X}, \overline{H})}(\overline{S}) = 0.$$

We have from $\overline{H} \in \mathcal{C}(\overline{X}) \setminus \{0\}$ that

$$\mathcal{A}\overline{H} \in \mathcal{C}_{\mathcal{Q}}(\mathcal{A}\overline{X} - b, \bar{y}), \quad \overline{H} \in \mathcal{C}_\theta(\overline{X}, \overline{S}).$$

Hence, we deduce from (5.6) and Proposition 4.2 that

$$\overline{H} \in (\mathcal{A}^*\mathcal{C}_{\mathcal{Q}^\circ}(\bar{y}, \mathcal{A}\overline{X} - b))^\circ \cap (\mathcal{C}_{\theta^*}(\overline{S}, \overline{X}))^\circ = (\mathcal{A}^*\mathcal{C}_{\mathcal{Q}^\circ}(\bar{y}, \mathcal{A}\overline{X} - b) + \mathcal{C}_{\theta^*}(\overline{S}, \overline{X}))^\circ.$$

By the assumed SRCQ (5.8) at $\bar{y}$ for $\overline{X}$, there exist $\tilde{w} \in \mathbb{R}^d$ and $\widetilde{H} \in \mathcal{A}^*\mathcal{C}_{\mathcal{Q}^\circ}(\bar{y}, \mathcal{A}\overline{X} - b) + \mathcal{C}_{\theta^*}(\overline{S}, \overline{X})$ such that $\overline{H} = \mathcal{F}^*\tilde{w} + \widetilde{H}$. Then

$$\langle \overline{H}, \overline{H} \rangle = \langle \overline{H}, \mathcal{F}^*\tilde{w} + \widetilde{H} \rangle = \langle \overline{H}, \widetilde{H} \rangle \leq 0,$$

which implies $\overline{H} = 0$. This contradicts the previous assumption that $\overline{H} \neq 0$. The proof is thus completed.

One can also establish an analogous result by swapping the roles of the primal and dual problems in Theorem 5.1.

Theorem 5.2. *Let $(\bar{y}, \bar{w}, \overline{S}) \in \mathbb{R}^e \times \mathbb{R}^d \times \mathbb{R}^{m \times n}$ be an optimal solution of problem (1.4) with $\mathcal{M}_D(\bar{y}, \bar{w}, \overline{S}) \neq \emptyset$. Let $\overline{X} \in \mathcal{M}_D(\bar{y}, \bar{w}, \overline{S})$. Then the following two statements are equivalent to each other:*

(i) The SOSC holds at $(\bar{y}, \bar{w}, \overline{S})$ for $\overline{X}$ with respect to the dual problem (1.4), i.e.,

$$\langle H_w, (\nabla h^*)'(\bar{w}; H_w) \rangle - \phi^*_{(\overline{S}, H_S)}(\overline{X}) > 0, \quad \forall (H_y, H_w, H_s) \in \mathcal{C}(\bar{y}, \bar{w}, \overline{S}) \setminus \{0\}, \tag{5.9}$$

where the critical cone $\mathcal{C}(\bar{y}, \bar{w}, \overline{S})$ is defined as

$$\mathcal{C}(\bar{y}, \bar{w}, \overline{S}) := \left\{ (H_y, H_w, H_S) \in \mathbb{R}^e \times \mathbb{R}^d \times \mathbb{R}^{m \times n} \left| \begin{array}{c} \mathcal{A}^*H_y + \mathcal{F}^*H_w + H_S = 0, \\ H_y \in \mathcal{C}_{\mathcal{Q}^\circ}(\bar{y}, \mathcal{A}\overline{X} - b), \ H_S \in \mathcal{C}_{\theta^*}(\overline{S}, \overline{X}) \end{array} \right. \right\}.$$

(ii) The SRCQ holds at $\overline{X}$ for $(\bar{y}, \bar{w}, \overline{S})$ with respect to the primal problem (1.2), i.e.,

$$\left(\begin{array}{c} \mathcal{A} \\ \mathcal{I}_{\mathbb{R}^{m \times n}} \end{array} \right) \mathbb{R}^{m \times n} + \left(\begin{array}{c} \mathcal{C}_{\mathcal{Q}}(\mathcal{A}\overline{X} - b, \bar{y}) \\ \mathcal{C}_\theta(\overline{X}, \overline{S}) \end{array} \right) = \left(\begin{array}{c} \mathbb{R}^e \\ \mathbb{R}^{m \times n} \end{array} \right). \tag{5.10}$$

Proof. With the help of Proposition 4.2, one can establish the assertion in the same fashion as in Theorem 5.1. We omit the details here.

Finally, note that the epigraph of the nuclear norm function is $\mathcal{C}^2$-cone reducible [5, Proposition 4.3]. This fact allows us to establish the following equivalent characterizations of the robust isolated calmness of the KKT solution mapping of problem (1.2) by combining Proposition 3.3, Theorem 5.1 and Theorem 5.2.

Theorem 5.3. *Let $\overline{X} \in \mathbb{R}^{m \times n}$ be an optimal solution of problem (1.2) and $(\bar{y}, \bar{w}, \overline{S}) \in \mathbb{R}^e \times \mathbb{R}^d \times \mathbb{R}^{m \times n}$ be an optimal solution of problem (1.4). Assume that the RCQ (5.5) holds at $\overline{X}$. Then the following statements are equivalent to each other:*

(i) The KKT solution mapping S_{KKT} in (5.4) is robustly isolated calm at the origin for $(\overline{X}, \bar{y})$.

(ii) The SOSC (5.7) holds at $\overline{X}$ for $\bar{y}$ with respect to the primal problem (1.2) and the SRCQ (5.10) holds at $\overline{X}$ for $(\bar{y}, \bar{w}, \overline{S})$ with respect to the primal problem (1.2).

(iii) The SOSC (5.7) holds at $\overline{X}$ for $\bar{y}$ with respect to the primal problem (1.2) and the SOSC (5.9) holds at $(\bar{y}, \bar{w}, \overline{S})$ for $\overline{X}$ with respect to the dual problem (1.4).

(iv) The SRCQ (5.8) holds at $\bar{y}$ for $\overline{X}$ with respect to the dual problem (1.4) and the SRCQ (5.10) holds at $\overline{X}$ for $(\bar{y}, \bar{w}, \overline{S})$ with respect to the primal problem (1.2).

(v) The SOSC (5.9) holds at $(\bar{y}, \bar{w}, \overline{S})$ for $\overline{X}$ with respect to the dual problem (1.4) and the SRCQ (5.8) holds at $\bar{y}$ for $\overline{X}$ with respect to the dual problem (1.4).

It is worth mentioning that in this paper, we focus on the characterizations of the robust isolated calmness of the KKT solution mapping for problem (1.2) when it admits a unique KKT point. It would be certainly interesting to know to what extent our results can be extended to the case when problem (1.2) admits non-unique solutions. We shall leave this as a future research topic.

References

[1] J.F. Bonnans and A. Shapiro, Perturbation Analysis of Optimization Problems, Springer, New York, 2000.

[2] E.J. Candès and B. Recht, Exact matrix completion via convex optimization, *Found. Comput. Math.*, **9** (2008), 717–772.

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

[4] F.H. Clarke, Optimization and Nonsmooth Analysis, Second Edition, Classics in Applied Mathematics, 5, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1990.

[5] C. Ding, *An Introduction to a Class of Matrix Optimization Problems*, PhD thesis, National University of Singapore, 2012.

[6] C. Ding, Variational analysis of the Ky Fan k-norm, *Set-Valued Var. Anal.*, **25** (2017), 265–296.

[7] C. Ding, D.F. Sun, J. Sun, and K.-C. Toh, Spectral operators of matrices, *Math. Program.*, in press. DOI: 10.1007/s10107-017-1162-3.

[8] C. Ding, D.F. Sun, and L. W. Zhang, Characterization of the robust isolated calmness for a class of conic programming problems, *SIAM J. Optimiz.*, **27** (2017), 67–90.

[9] A.L. Dontchev and R.T. Rockafellar, Implicit Functions and Solution Mappings, Springer, New York, 2009.

[10] D.R. Han, D.F. Sun, and L.W. Zhang, Linear rate convergence of the alternating direction method of multipliers for convex composite programming, *Math. Oper. Res.*, in press.

[11] D.R. Han, D.F. Sun, and L.W. Zhang, Linear rate convergence of the alternating direction method of multipliers for convex composite quadratic and semi-definite programming, arXiv:1508.02134 (2015).

[12] Y.L. Liu and S.H. Pan, Locally upper Lipschitz of the perturbed KKT system of Ky Fan k-norm matrix conic optimization problems, arXiv:1509.00681, 2015.

[13] K. Löwner, Über monotone matrixfunktionen, *Math. Z.*, **38** (1934), 177–216.

[14] B. Recht, M. Fazel, and P.A. Parrilo, Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization, *SIAM Rev.*, **52** (2010), 471–501.

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

[16] R.T. Rockafellar, Augmented Lagrangians and applications of the proximal point algorithm in convex programming, *Math. Oper. Res.*, **1** (1976), 97–116.

[17] T.K. Pong, P. Tseng, S.W. Ji, and J.P. Ye, Trace norm regularization: reformulations, algorithms, and multi-task learning, *SIAM J. Optimiz.*, **20** (2010), 3465–3489.

[18] R.T. Rockafellar and R.J.-B. Wets, Variational Analysis, Springer, New York, 1998.

[19] G.A. Watson, Characterization of the subdifferential of some matrix norms, *Linear Algebra Appl.*, **170** (1992), 33–45.

[20] G.A. Watson, On matrix approximation problems with Ky Fan k norms, *Numer. Algorithms*, **5** (1993), 263–272.

[21] M. Yuan, A. Ekici, Z.S. Lu, and R.D.C. Monteiro, Dimension reduction and coefficient estimation in multivariate linear regression, *J. Roy. Stat. Soc. B.*, **69** (2007), 329–346.

[22] L.W. Zhang, N. Zhang, and X.T. Xiao, On the second-order directional derivatives of singular values of matrices and symmetric matrix-valued functions, *Set-Valued Var. Anal.*, **21** (2013), 557–586.

[23] Y.L. Zhang and L.W. Zhang, On the upper Lipschitz property of the KKT mapping for nonlinear semidefinite optimization, *Oper. Res. Lett.*, **44** (2016), 474–478.

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