THE METRIC SUBREGULARITY OF KKT SOLUTION MAPPINGS OF COMPOSITE CONIC PROGRAMMING

GUO HAN

(B.Sc., WHU, China)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MATHEMATICS

NATIONAL UNIVERSITY OF SINGAPORE

2017

To my parents

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

GUO Han

24 March 2017

Acknowledgements

First of all, I would like to express my extremely grateful to my supervisor Professor SUN Defeng. During my Ph.D study, he guided me not only for my research but also for my life. I benifit a lot from his immense knowledge in mathematics and his wisdom in life. Those are wealth in my future life. I am very thankful for his patience and professionalism in teaching me. It is very fortunate for me to learn from him.

I am also deeply indebted to Professor TOH Kim-Chuan. He widened my knowlege in optimization algorithms. His excellent work on the numerical implementation for algorithms show another important part of optimization learning.

Meanwhile, I would like to say thanks to Professor PANG Chin How, Jeffrey. It is him who gave me a first glance of optimization four years ago. From his courses on mathematical programming, I found my interest in optimization.

My sincere gratitude also go to all the members of our optimization group. During our weekly research seminar, I really learned a lot from presentations and discussions. In particular, I want to thank Dr. DING Chao and Dr. CUI Ying specially. They give me helpful suggestions for this thesis and other research topics. Moreover, I would like to thank my classmates and friends here for them enriching my life in Singapore.

Next, I am thankful to National Unversity of Singapore and Department of Mathematics for offering me an opportunity to study here and providing me the financial support.

And finally, last but by no means least, I would like to say thanks to my parents for their endless and unconditional love. I am grateful for my other family members and friends for their spiritual support and encouragement.

Contents

Summary

In this thesis, we study the stability of a class of composite optimization problems, whose objective functions involve convex composite terms. Many important optimization problems can be reformulated as composite problems, such as nonlinear programming, nonlinear semidefinite programming (SDP), various regularized problems and so on, which frequently arise from various areas such as finance, engineering, applied mathematics, etc.

The study of the stability of composite problems has its own interest in theory. Moreover, the stability has a close relationship with convergent rates of various methods. Due to these facts, there are many studies towards the characterization of the stability, among which the Lagrange multipliers are often required to be unique for different models. However, our study is an extensive work by allowing the Lagrange multiplier set to be non-singleton. To achieve our goals, we conduct our analysis for the composite SDP conic programming and the composite Ky Fan k-norm regularized conic programming. We obtain the metric subregularity of Karush-Kuhn-Tucker (KKT) solution mappings of the aforementioned composite problems under rather weak conditions. Our study is mainly based on the second order analysis of the positive semidefinite cone and the Ky Fan k-norm. Therefore, we explore the variational properties first in each case. The perturbation properties are completely studied for symmetric and nonsymmetric matrices. Under the canonical perturbation, within the assumption of the second order sufficient condition, we obtain an error bound for a locally optimal solution of those underlying composite conic programming. Additionally, if a partial strict complementarity condition holds, an error bound for the corresponding multiplier set is estimated. Since our study of the metric subregularity for composite conic programming is under the nonconvex setting, it can cover the results of convex problems.

Chapter 1

Introduction

1.1 Motivations and backgrounds

In this thesis, we consider the composite optimization problem

$$\begin{aligned} \min \quad & f(x) + \theta(g(x)) \\ \text{s.t.} \quad & h(x) = 0, \end{aligned} \tag{1.1}$$

where $f : \mathcal{X} \to \mathcal{R}$ is a twice continuously differentiable function, $h : \mathcal{X} \to \mathcal{Y}$ and $g : \mathcal{X} \to \mathcal{S}$ are twice continuously differentiable mappings, $\theta : \mathcal{S} \to (-\infty, +\infty]$ is a closed proper convex function (not necessarily smooth), $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{S}$ are finite dimensional real Euclidean spaces. Suppose each of $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{S}$ endowed with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$.

The composite optimization problems in the form of (1.1) are well studied in literatures. It is also an important structured model in optimization. Various optimization problems can be cast in the form of (1.1) by choosing different convex functions θ. For instance, the nonlinear programming (NLP) can be reformulated as a composite problem (e.g., [65]) and so do the nonlinear semidefinite programming (SDP) [90, 100]. Another important application of (1.1) is the convex regularized problem by choosing θ to be vector l_1-norm, vector l_∞-norm, matrix spectral norm, nuclear norm and so on. Such convex regularized problems, especially the low-rank optimization problems, are widely used in many fileds, such as Markov chain problems [9–11], matrix completion problems [40, 47, 78, 79], signal and image processing [17], finance and economics [50, 75] and so on. Moreover, there are also other problems can be modeled in the form of (1.1), such as minmax problems, convex inclusions and penalized constrained problems [12, 35, 55, 83]. Therefore, the study of the composite problem (1.1) is significant.

We consider the Lagrangian function $l : \mathcal{X} \times \mathcal{Y} \times \mathcal{S} :\to \mathcal{R}$ of the problem (1.1) in the form of

$$l(x, y, x) := f(x) + \langle y, h(x)\rangle + \langle S, g(x)\rangle - \theta^*(S), \tag{1.2}$$

where θ^* is the Fenchel conjugate function of the convex function θ.
Let us consider the following canonically perturbed problem of (1.1) with $(u, v, C) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{S}$:

$$\begin{aligned} \min \quad & f(x) + \theta(g(x) + C) - \langle u, x\rangle \\ \text{s.t.} \quad & h(x) + v = 0. \end{aligned} \tag{1.3}$$

For a given perturbation parameter (u, v, C), we consider the following "generalized Karush-Kuhn-Tucker (KKT) conditions" [66] for the problem (1.3):

$$\begin{cases} \nabla f(x) + \nabla h(x)y + \nabla g(x)S = u, \\ h(x) + v = 0, \\ S \in \partial\theta(g(x) + C). \end{cases} \tag{1.4}$$

In general, there has a "gap" between the generalized KKT system (1.4) and the first order optimality conditions of the composite problem (1.3), due to the possibly nonsmooth term $\theta \circ g$ in (1.3). However, under some mild conditions, this gap can be fulfilled (see Section 2.3.3). Therefore, we adpot the KKT system (1.4) for our discussions.

It is well known that a point (x, y, S) solves the KKT system (1.4) if and only if $(u, -v, -C) \in \partial l(x, y, S)$. Thus, we define the multi-valued mapping $\mathcal{T}_l : \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^{m\times n} \rightrightarrows \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^{m\times n}$ associated with the Lagrangian function l at any $(x, y, S) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^{m\times n}$ by

$$\mathcal{T}_l(x, y, S) = \{(u, v, C) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{S} \mid (u, -v, -C) \in \partial l(x, y, S)\}. \tag{1.5}$$

Therefore, the set of all solutions (x, y, S) to the KKT system (1.4) is $\mathcal{T}_l^{-1}(u, v, C)$. We call the corresponding x a stationary point of the problem (1.3) associated with (u, v, C) and (y, S) a Lagrange multiplier associated with (x, u, v, C), if $(x, y, S) \in \mathcal{T}_l^{-1}(u, v, C)$.

Our aim of this thesis is to discuss an important property in perturbation analysis for the problem (1.1) at an optimal solution: the metric subregularity (see Definition 2.11 in Section 2.3.2) of $\mathcal{T}_l$. It is well known that $\mathcal{T}_l$ is metrically subregular at (x, y, S) for (u, v, C) if and only if $\mathcal{T}_l^{-1}$ is calm (see Definition 2.10 in Section 2.3.2) at (u, v, C) for (x, y, S), where $(x, y, S) \in \mathcal{T}_l^{-1}(u, v, C)$. Although these two concepts are equivalent under the inverse operation of mappings, we perfer to investigate the metric subregularity for problems in the form of (1.1). This may because we can deal with the initial program data of the original optimization problem (1.1). Moreover, we also have another equivalence of two strenghthened concepts, that $\mathcal{T}_l$ is strongly metrically subregular (see Definition 2.13 in Section 2.3.2) at (x, y, S) for (u, v, C) if and only if $\mathcal{T}_l^{-1}$ is isolated calm (see Definition 2.12 in Section 2.3.2) at (u, v, C) for (x, y, S)([29, Theorem 5.2] and [30, Theorem 3I.3]), where $(x, y, S) \in \mathcal{T}_l^{-1}(u, v, C)$. In perturbation analysis, we are more interested in $\mathcal{T}_l^{-1}$ possessing the isolated calmness and the locally nonempty-valued (see Definition 2.14 in Section 2.3.2) at the same time, which is the so-called roubust isolated calmness. Over past several decades, there are numerous work have been done towards the characterization of the robust isolated calmness (see Definition 2.15 in Section 2.3.2) of the KKT solution mapping $\mathcal{T}_l^{-1}$, especially for the NLP problem.

When the θ of the problem (1.1) is an indicator function over a polyhedral set, the characterization of the robust isolated calmness is fairly complete, and can even extend to a more general structure, where the smooth term f, g and h in the problem (1.3) are multi-variable functions: $f(x, p)$, $g(x, p)$ and $h(x, p)$ with the parameter $p \in \mathcal{Z}$ and $\mathcal{Z}$ be a finite dimensional Euclidean space [28, 49, 82]. In [28], Dontchev and Rockafellar establish an equivalent relationship for the NLP problem that at a locally optimal solution, the KKT solution mapping $\mathcal{T}_l^{-1}$ prossessing the robust isolated calmness property is equivalent to the strict Mangasarian-Fromovitz constraint qualification (MFCQ) and the second order sufficient condition (SOSC) holding at the same time. Later, Klatte [49] extends the characterization for the $C^{1,1}$ (the class of all differentiable functions having a locally Lipschitzian derivative) nonlinear programming problem. Avoiding to assume the MFCQ, he characterizes the robust isolated calmness at a stationary point of the $C^{1,1}$ nonlinear programming problem by injectivity conditions on the contingent derivative of the Kojima function.

By a powerful tool introduced by King and Rockafellar [48] and Levy [54], we know that the KKT solution mapping $\mathcal{T}_l^{-1}$ possesses the isolated calmness property if and only if its graphical derivative is non-singular, even when the θ in the problem (1.1) is an indicator function over a non-polyhedral set. This nice property enables researchers to investigate the isolated calmness property under the non-polyhedral setting. The following literatures are all under the setting that θ in the problem (1.1) is an indicator function over a non-polyhedral set. In [64] and [63], Mordukhovich et al. obtain an explict formulation of the graphical derivative of $\mathcal{T}_l^{-1}$ at a stationary point, at which they require the constraint non-degeneracy hold; then they characterize the isolated calmness via the non-singularity of the explict formulation obtained. In his analysis, the constraint non-degeneracy condition plays a crucial role. However, in most cases, an explicit formula of the graphical derivative of $\mathcal{T}_l^{-1}$ is not that easy to obtain. Thus, it may be difficult to verify the non-singularity of the graphical derivative of $\mathcal{T}_l^{-1}$. Therefore, people are trying to find sufficient conditions that can gurantee the non-singularity of the graphical derivative of $\mathcal{T}_l^{-1}$ at an interested point. Zhang and Zhang [111] establish that at a locally optimal solution of the nonlinear SDP problem, the so-called strict Robinson consitraint qualification (SRCQ) and the SOSC yield the non-singularity of the graphical derivative of $\mathcal{T}_l^{-1}$. For the contrary implication of the nonlinear SDP problem, Han et al. [41] show that the isolated calmness of $\mathcal{T}_l^{-1}$ at a stationary point yields the SRCQ. Futhermore, they also completely characterize the isolated calmness of $\mathcal{T}_l^{-1}$ for the convex composite quadratic SDP problem. Later, Liu and Pan [56] extend the above results of the nonlinear SDP problem to the nonlinear Ky Fan matrix k-norm problem. Recently, to enlarge the above considered non-polyhedral sets, Ding et al. [26] provide a complete characterization of the robust isolated calmness of $\mathcal{T}_l^{-1}$ by considering a class of C^2-cone reducible sets (see Definition 2.17 in Section 2.4), which contains all the convex polyhedral sets, positive semidefinite (PSD) cone [8, 89], second order cone (SOC) and the epigraph cone of the Ky Fan matrix k-norm [24]. They establish that at a locally optimal solution, $\mathcal{T}_l^{-1}$ is roubustly isolated calm if and only if the SRCQ and the SOSC hold. Milzarek [61], in his thesis, discusses the stability of composite problems involving C^2-fully decomposable functions (see definition in [88]) under the non-degeneracy condition, the strict complementarity condition and the SOSC satisfied.

Over all the aforementioned literatures, we can find that the strict MFCQ for the polyhedral setting and the constraint non-degeneracy condition or SRCQ for the non-polyhedral setting play an essential role in the characterization of the robust isolated calmness of the KKT solution mapping $\mathcal{T}_l^{-1}$. For the NLP problem, Kyparisis [53] proves that the strict MFCQ is equivalent to the uniqueness of Lagrange multipliers at a locally optimal solution. Moreover, under the non-polyhedral setting, although SRCQ is weaker than the constraint non-degeneracy condition, it still implies the uniqueness of Lagrange multipliers. Therefore, we can see that all the analysis conducted for the characterization of the robust isolated calmness of the KKT solution mapping $\mathcal{T}_l^{-1}$, is under the scenario of the uniqueness of Lagrange multipliers. It is very natural to ask what kind of stable property $\mathcal{T}_l^{-1}$ can have when Lagrange multipliers is not unique at a locally optimal solution. In such situation, it is known that $\mathcal{T}_l^{-1}$ no longer possess the robust isolated calmness property by above arguments.

In [82], Robinson shows that if θ of the problem (1.1) is an indicator function over a polyhedral cone, the KKT solution mapping $\mathcal{T}_l^{-1}$ would have upper Lipschitz continuity property under the conditions that the MFCQ and a strong form of the second order sufficient condition hold. The calmness property is the 'localization' of the aforementioned Robinson's upper Lipschitz continuity property. Thus, one can see that the calmness property holds without the requirement of uniqueness of Lagrange multipliers. Therefore, by the equivalence relationship between the calmness and the metric subregularity under the inverse operation, $\mathcal{T}_l$ is metrically subregular under certain aforementioned conditions. Izmailov et al. [46] obtain the metric subregularity of $\mathcal{T}_l$ for the NLP problem with $C^{1,1}$ program data under the existence of a noncritical Lagrange multiplier, which can be implied by the SOSC. The above discussions are conducted under the polyhedral cone setting. How to characterize the metric subregularity or the calmness for non-polyhedral cones is still unknown.

Recently, Cui et al. [22] obtain the metric subregularity of $\mathcal{T}_l$ for the linearly constrained convex SDP problem with Lagrange multipliers non-uniqueness if the SOSC and a partial strict complementarity property hold. Later, Cui et al. [20] extend the study for solution set mappings of the primal and the dual problems to a class of convex matrix optimization problems with the possible nonsmooth terms to be the so-called spectral functions. All their analysis are conducted for convex problems. For the nonconvex composite optimization problem (1.1), by letting θ be a convex piecewise linear function, Mordukhovich et al. [66] explore the equivalence relationship between the metric subregularity of $\mathcal{T}_l$ and the existence of a noncritical Lagrange multiplier. However, the characterization of the metric subregularity for the nonlinear SDP problem is a remaining question to be answered.

In this thesis, we study the metric subregularity for the compositie problem (1.1) with θ be the so-called spectral function [20]. Specifically, we consider the nonlinear SDP problem when $\theta(\cdot) = \delta_{\mathcal{S}^n_+}(\cdot)$ on $\mathcal{S} = \mathcal{S}^n$ the space of all symmetric matrices, and the nonlinear Ky Fan matrix k-norm conic problem when $\theta(\cdot) = \|\cdot\|_{(k)}$ on $\mathcal{S} = \mathcal{R}^{m\times n}$, where $\|\cdot\|_{(k)}$ denotes the Ky Fan matrix k-norm. In [20], Cui et al. show that such θ called the spectral function can be written as a composite matrix function either in the form of $\theta = q \circ \lambda$ with $\lambda(X)$ be the vector of eigenvalues (in nonincreasing order) of $X \in \mathcal{S}^n$ and $q : \mathcal{R}^n \to (-\infty, +\infty]$ be a symmetric function, or in the form of $\theta = q \circ \sigma$ with $\sigma(X)$ be the vector of singular values (in nonincreasing order) of $X \in \mathcal{R}^{m\times n}$ and $q : \mathcal{R}^n \to (-\infty, +\infty]$ be an absolutely symmetric function. They explore that the spectral function θ preserves the nice properties of q; namely, θ is C^2-cone reducible if q is C^2-cone reducible and the subdifferential mapping $\partial\theta$ possesses the metric subregularity if the subdifferential mapping ∂q possesses the metric subregularity. By using these properties, one can check the C^2-cone reducibility and the metric subregularity of θ via checking the C^2-cone reducibility and the metric subregularity of q, which is a function of vectors. Within these knowledge, we will provide rather weakly sufficient conditions to deduce the metric subregularity of $\mathcal{T}_l$ for the nonlinear SDP problem and the nonlinear Ky Fan matrix k-norm conic problem, namely, the metric subregularity of $\mathcal{T}_l$ holds at an optimal solution if a strict SOSC and a partial strict complementarity property are satisfied.

The study of the metric subregularity is not only the interest in the theoretical analysis of the stability, e.g., the metric subregularity is one sufficient condition of the tilt stability for the NLP problem [38], but also because the metric subregularity is a powerful tool for the convergence analysis of various methods. It is well known that "error bounds" are commonly used in the analysis of convergent rates of first order methods [22, 31, 44, 57–59, 71, 95, 97, 98, 108, 114]. Due to the apprealingly simple analysis of convergent rates based only on the error bound property, people also derive error bounds in various measures, e.g., [32,70] and a survey [4]. Another fashion to excavate the error bound property is through detecting its connection with the metric subregularity or the calmness. Under some mild conditions, the metric subregularity can be used to establish the error bound property [31,52,73,99,113].

1.2 Outline of the thesis

We organize the remaining parts of this thesis as follows: In Chapter 2, we give some preliminaries to facilitate the subsequent discussions. Some convex properties and perturbation analysis are briefly reviewed. Chapter 3 focuses on the variational analysis of the positive semidefinite cone and the discussions of the metric subregularity for the nonlinear SDP problem. In Chapter 4, we conduct sensitivity analysis for the nuclear norm function and apply it to obtain the metric subregularity for the nuclear norm regularized problem. An extensive study of the metric subregularity property to the nonlinear Ky Fan matrix k-norm conic problem is conducted in Chapter 5. The perturbation property of the Ky Fan matrix k-norm plays a crucial role in the study. Finally, we draw a conlusion and present some possible topics for future study in Chapter 6.

Chapter 2

Preliminaries

2.1 Notations

- For n be a given integer, $\mathcal{S}^n$ denotes the space of all $n \times n$ symmetric matrices, $\mathcal{S}^n_+$ denotes the cone of all positive semidefinite matrices in $\mathcal{S}^n$ and $\mathcal{S}^n_-$ denotes the cone of all negative semidefinite matrices in $\mathcal{S}^n$. $\mathcal{O}^n$ be the set of all $n \times n$ orthogonal matrices.

- For any $X \in \mathcal{R}^{m\times n}$, we define X_{ij} as the (i,j)-th entry of X.

- For any $X \in \mathcal{R}^{m\times n}$ and any index set $\mathcal{J} \subseteq \{i,\ldots,n\}$, we use $X_{\mathcal{J}}$ to represent the sub-matrix of X obtained by removing all the columns of X not in $\mathcal{J}$. Additionally, for any index set $\mathcal{I} \subseteq \{i,\ldots,m\}$, we use $Z_{\mathcal{I}\mathcal{J}}$ to denote the $|\mathcal{I}| \times |\mathcal{J}|$ sub-matrix of X obtained by removing all the rows of X not in $\mathcal{I}$ and all the columns of X not in $\mathcal{J}$.

- For any $X \in \mathcal{R}^{m\times n}$, $X^\dagger \in \mathcal{R}^{m\times n}$ denotes the Moore-Penrose pseudoinverse of X.

- For any $X \in \mathcal{R}^{m\times n}$, $\|X\|_2$ denotes the spectral norm or the operator norm of X, i.e., the largest singular value of X; and $\|X\|_*$ denotes the nuclear norm of X, i.e., the sum of all the singular values of X.

- For any $X \in \mathcal{R}^{m\times n}$ and any integer $0 < k \leq \min\{m,n\}$, $\|X\|_{(k)}$ denotes the Ky Fan k-norm of X, i.e., the sum of the k-largest singular values of X.

- For any $X \in \mathcal{R}^{m\times n}$, We use $\mathrm{tr}(X)$ to represent the trace of X, i.e., the sum of all the diagonal entries of X.

- We let "$\circ$" denote the Hadamard product between matrices.

- For any $v \in \mathcal{R}^m$, $\mathrm{diag}(v)$ denotes the $m \times m$ diagonal matrix obtained by v with the i-th diagonal entry be v_i, $i = 1, \ldots, m$.

- Given a set C in a finite dimensional real Euclidean space $\mathcal{Z}$ and a point $z \in C$, we denote $\mathrm{ri}(C)$ as its relative interior, $\mathcal{T}_C(z)$ as the tangent cone of C at z and $\mathcal{N}_C(z)$ as the normal cone of C at z. Moreover, for any $z' \in \mathcal{Z}$, we let $\mathrm{dist}(z', C) := \inf_{z\in C} \|z' - z\|$.

- Given a closed convex cone $\mathcal{K} \subseteq \mathcal{Z}$, denote $\mathcal{K}^*$ as the dual cone of $\mathcal{K}$ and $\mathcal{K}^\circ$ as the polar cone of $\mathcal{K}$.

- Given a convex function $\theta : \mathcal{S} \to (-\infty, +\infty]$, we use $\mathrm{dom}\theta$ to denote the effective domain of θ, and $\mathrm{epi}\theta$ to denote the epigraph of θ. Moreover, we let θ^* to denote the Fenchel's conjugate function of θ, and $\partial\theta$ as the subgradient mapping of θ.

2.2 Convex analysis

In this section, we present the following useful concept about the bounded linear regularity of a collection of closed convex sets, which will help us in the subsequent discussions of the metric subregularity or the calmness for the composite problem (1.1). And also, we show some useful characterizations of the subdifferential of convex functions and piecewise linear quadratic functions.

Definition 2.1. *Let $C_1, C_2, ..., C_l \in \mathcal{Z}$ be a sequence of closed convex sets for some positive integer l, where $\mathcal{Z}$ is a finite dimensional real Euclidean space. Assume that*

$C := C_1 \cap C_2 \cap ... \cap C_l$ *is non-empty. The collection* $\{C_1, C_2, ..., C_l\}$ *is said to be boundedly linearly regular if for every bounded set* $\mathcal{D} \in \mathcal{Z}$*, there exists a constant* $\kappa > 0$ *such that*

$$dist(z, C) \leq \kappa \max\{dist(z, C_1), ..., dist(z, C_l)\}, \ \forall z \in \mathcal{D}.$$

One can find the above definition in [5, Definition 5.6]. The next proposition [6, Corollary 3] provides a sufficient condition for the bounded linear regularity.

Proposition 2.1. *Let* $C_1, C_2, ..., C_l \in \mathcal{Z}$ *be a sequence of closed convex sets for some positive integer* l*, where* $\mathcal{W}$ *is a finite dimensional real Euclidean space. Suppose that* $C_1, C_2, ..., C_k$ *are polyhedral for some* $k \in \{0, 1, ..., l\}$*. Then a sufficient condition for* $\{C_1, C_2, ..., C_l\}$ *to be boundedly linearly regular is*

$$\bigcap_{i=1,2,...,k} C_i \ \cap \bigcap_{j=k+1,...,l} ri(C_j) \neq \emptyset.$$

Next, we adopt one useful lemma to characterize the subdifferential of a convex function, one can find in, e.g. [7, Theorem 16.23].

Lemma 2.1. *Suppose that* $\theta : \mathcal{S} \to (-\infty, +\infty]$ *is proper, lower semicontinuous and convex. For any* $s, w \in \mathcal{S}$*. Then the following statements are equivalent:*

(i) $w \in \partial\theta(s)$,

(ii) $(w, -1) \in \mathcal{N}_{\mathrm{epi}\theta}(s, \theta(s))$,

(iii) $\theta(s) + \theta^*(w) = \langle s, w \rangle$,

(iv) $s \in \partial\theta^*(w)$.

Based on this lemma, we can 'decompose' the composite term $\theta \circ g$ in (1.1) by lifting the dimension of the problem (1.1). The followings are the preparations for our later discussions.

Definition 2.2. (cf. [86, 9.57]) *We say a multi-valued mapping* $F : \mathcal{W} \rightrightarrows \mathcal{Z}$ *is piecewise polyhedral if its graph* $\mathrm{gph}(F) := \{(w, z) \in \mathcal{W} \times \mathcal{Z} : z \in F(w)\}$ *is piecewise polyhedral, i.e., expressible as the union of finitely many polyhedral convex sets.*

Definition 2.3. (cf. [86, 10.20]) *A function $\theta : \mathcal{S} \to (-\infty, +\infty]$ is called piecewise linear-quadratic if* $\operatorname{dom}\theta$ *can be represented as the union of finitely many polyhedral sets, relative to ecah of which $\theta(s)$ is given by an expression of the form $\frac{1}{2}\langle s, \mathcal{A}s\rangle + \langle a, s\rangle + \alpha$ for some $\alpha \in \mathcal{R}$, $a \in \mathcal{S}$, and $\mathcal{A}$ is a self-adjoint linear operator on $\mathcal{S}$.*

The next lemma is given by Sun in his PhD thesis [91].

Lemma 2.2. ([86, 11.14, 12.30]) *Suppose that $\theta : \mathcal{S} \to (-\infty, +\infty]$ is proper, lower semicontinuous and convex. Then the following statements are equivalent:*

(a) θ is piecewise linear-quadratic,

(b) θ^* is piecewise linear-quadratic,

(c) the subgradient mapping $\partial\theta$ is piecewise polyhedral,

(d) the subgradient mapping $\partial\theta^*$ is piecewise polyhedral.

2.3 The sensitivity analysis of the optimization problems

In this section, we show some useful concepts corresponding to the variational analysis. The definitions of various Lipschitz-like concepts are given there. Moreover, the first order optimailty conditions are interpreted in different ways.

2.3.1 Directional epidifferentiability and tangent sets

We first provide the definitions of the first and second order tangent sets and the directional epiderivatives. These concepts are introduced by Bonnans and Shapiro in [8, Section 2.2, 3.2].

Definition 2.4. *Let $\mathcal{Z}$ be a finite dimensional real Euclidean space. Given a closed subset $\mathcal{K} \subseteq \mathcal{Z}$ and $z \in \mathcal{K}$, we define the Radial cone of $\mathcal{K}$ at z as*

$$\mathcal{R}_{\mathcal{K}}(z) = \{d \in \mathcal{Z} : \exists \rho^* > 0, s.t.\ z + \rho d \in \mathcal{K},\ \forall \rho \in [0, \rho^*]\}.$$

Furthermore, the contingent (Bouligand) cone of $\mathcal{K}$ *at* z *is defined by*

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

and the inner tangent cone of $\mathcal{K}$ *at* z *is defined by*

$$\mathcal{T}^i_{\mathcal{K}}(z) = \liminf_{\rho \downarrow 0} \frac{\mathcal{K} - z}{\rho}.$$

Moreover, by the above definition, we have the following equivalent interpretations of the contingent and inner tangent cones, respectively.

$$\mathcal{T}_{\mathcal{K}}(z) = \{d \in \mathcal{Z} : \exists \rho_k \downarrow 0, \operatorname{dist}(z + \rho_k d, \mathcal{K}) = o(\rho_k)\},$$
$$\mathcal{T}^i_{\mathcal{K}}(z) = \{d \in \mathcal{Z} : \operatorname{dist}(z + \rho d, \mathcal{K}) = o(\rho), \forall \rho \geq 0\}.$$

By the definitions of the contingent and inner tangent cones, we always have that $\mathcal{T}_{\mathcal{K}}(z) = \mathcal{T}^i_{\mathcal{K}}(z)$ if and only if $\lim_{\rho \downarrow 0} \frac{\mathcal{K} - z}{\rho}$ exists. It is not hard to see that $\frac{\mathcal{K} - z}{\rho}$ is a monotone decreasing function of ρ when $\mathcal{K}$ is convex. Thus, $\mathcal{T}_{\mathcal{K}}(z) = \mathcal{T}^i_{\mathcal{K}}(z)$ for any $z \in \mathcal{K}$, when $\mathcal{K}$ is convex [8, Proposition 2.55]. Meanwhile, we also use $\mathcal{T}_{\mathcal{K}}(z)$ to denote both $\mathcal{T}^i_{\mathcal{K}}(z)$ and $\mathcal{T}_{\mathcal{K}}(z)$ when $\mathcal{T}_{\mathcal{K}}(z) = \mathcal{T}^i_{\mathcal{K}}(z)$, and call it the tangent cone of $\mathcal{K}$ at x in total. Based on this first order tangent sets, we introduce the second order tangent sets.

Similar to the first order scenario, we have the inner and outer second order tangent set.

Definition 2.5. *Let* $\mathcal{Z}$ *be a finite dimensional real Euclidean space and let* $\mathcal{K} \subseteq \mathcal{Z}$ *be given. Given* $z \in \mathcal{K}$ *and a direction* $d \in \mathcal{Z}$*, we define the inner second order tangent set to the set* $\mathcal{K}$ *at the point* z *in the direction* d *as*

$$\mathcal{T}^{i,2}_{\mathcal{K}}(z, d) := \liminf_{\rho \downarrow 0} \frac{\mathcal{K} - z - \rho d}{\frac{1}{2}\rho^2},$$

and the outer second order tangent set to the set $\mathcal{K}$ *at the point* z *in the direction* d *as*

$$\mathcal{T}^{2}_{\mathcal{K}}(z, d) := \limsup_{\rho \downarrow 0} \frac{\mathcal{K} - z - \rho d}{\frac{1}{2}\rho^2}.$$

Analogously, the inner and outer second order tangent sets have the following equivalent interpretations, respectively.

$$\mathcal{T}_{\mathcal{K}}^{i,2}(z,d) = \{w \in \mathcal{Z} : \mathrm{dist}(z + \rho d + \frac{1}{2}\rho^2 w, \mathcal{K}) = o(\rho^2),\ \forall \rho \geq 0\},$$
$$\mathcal{T}_{\mathcal{K}}^{2}(z,d) = \{w \in \mathcal{Z} : \exists \rho_k \downarrow 0,\ \mathrm{dist}(z + \rho_k d + \frac{1}{2}\rho_k^2 w, \mathcal{K}) = o(\rho_k^2)\}.$$

However, different from the first order tangent cones, it is no longer hold that the inner and outer second order tangent sets are identical in general, even under the set $\mathcal{K}$ is closed and convex. When the set $\mathcal{K}$ is convex, we can only have that the inner second order tangent set $\mathcal{T}_{\mathcal{K}}^{i,2}(z,d)$ is convex but the outer second order tangent set $\mathcal{T}_{\mathcal{K}}^{2}(z,d)$ can be nonconvex [8, Section 3.2]. Although within this fact, there are many sets satisfied the identical form. One commonly knwon is the C^2-cone reducible sets [8, Proposition 3.136]. If $\mathcal{T}_{\mathcal{K}}^{i,2}(z;d) = \mathcal{T}_{\mathcal{K}}^{2}(z;d)$, we use $\mathcal{T}_{\mathcal{K}}^{2}(z;d)$ to represent both and call it the second order tangent set to $\mathcal{K}$ at z in the direction d. It is well known that the second order tangent set is closely related to the so called "sigma term" in second order variational analysis of non-polyhedral cone constrained optimization problems.

The definitions of the first order tangent cones are the limits of the difference quotient, which are quite similar to the form of 'derivatives'. In [8], Bonnans and Shapiro introduce the directional epiderivatives to characterize the first order tangent cones of epi θ for a function θ.

Definition 2.6. *Suppose $\theta : \mathcal{Z} \to (-\infty, +\infty]$ be a proper, extended real valued function and $z \in \mathrm{dom}\theta$. The lower and upper directional epiderivatives of θ at z in the direction $h \in \mathcal{Z}$ are defined by, respectively,*

$$\theta_-^{\downarrow}(z;h) := \liminf_{\substack{\rho \downarrow 0 \\ \widetilde{h} \to h}} \frac{\theta(z + \rho\widetilde{h}) - \theta(z)}{\rho},$$

and

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

where Σ *denotes the set of all positive real sequences* $\{\rho_n\}$ *converging to* 0. *Moreover, we say* θ *is directionally epidifferentiable at* z *in the direction* h *if* $\theta^{\downarrow}_{-}(z;h) = \theta^{\downarrow}_{+}(z;h)$, *and denote them by* $\theta^{\downarrow}(z;h)$.

The following proposition shows a 'one to one' relationship between the first order tangent cones and the directional epiderivatives.

Proposition 2.2. *(cf. [8, Proposition 2.58]) Suppose that* $\theta : \mathcal{Z} \to (-\infty, +\infty]$ *is proper and* $z \in \mathrm{dom}\theta$. *Then,*

$$\mathcal{T}_{\mathrm{epi}\theta}\big(z, \theta(z)\big) = \mathrm{epi}\, \theta^{\downarrow}_{-}(z;\cdot), \tag{2.1}$$

and

$$\mathcal{T}^{i}_{\mathrm{epi}\theta}\big(z, \theta(z)\big) = \mathrm{epi}\, \theta^{\downarrow}_{+}(z;\cdot). \tag{2.2}$$

Particularly, when θ is proper and convex, we obtain $\theta^{\downarrow}_{-}(z;\cdot) = \theta^{\downarrow}_{+}(z;\cdot)$ immediately by the convexity of $\mathrm{epi}\,\theta$. Therefore, in this case, θ is directionally epidifferentiable at $z \in \mathrm{dom}\,\theta$ and $\theta^{\downarrow}(z;\cdot)$ is closed, convex and positively homogeneous.

One more thing we want to clarify here is that the directional epiderivative and the conventional directional derivative (denoted as $\theta'(z;\cdot)$) of a function θ may not be identical in general. For a proper convex function θ under the assumptions that both of the directional epiderivative and the conventional directional derivative exist in $\mathrm{dom}\,\theta$, we can only have the following relationship [8, Theorem 2.58 and Theorem 2.60]:

$$\theta^{\downarrow}(z;\cdot) = \mathrm{cl}\,\theta'(z;\cdot), \quad \forall z \in \mathrm{dom}\,\theta.$$

But if the function θ is Lipschitz continuous near z, we have $\theta^{\downarrow}_{-}(z;\cdot) = \theta'_{-}(z;\cdot)$ and $\theta^{\downarrow}_{+}(z;\cdot) = \theta'_{+}(z;\cdot)$. Therefore, we call θ a regular function if θ is Lipschitz continuous and directionally epidifferentiable at every $z \in \mathrm{dom}\theta$. One obviously regular function is a convex and Lipschitz continuous function [8, Theorem 2.126].

Next, we have an analogous terminology for the second order directional epiderivatives.

Definition 2.7. *Suppose $\theta : \mathcal{Z} \to (-\infty, +\infty]$ with $z \in \mathrm{dom}\theta$, and the lower directional epiderivatives $\theta^{\downarrow}_{-}(z;h)$ and upper directional epiderivatives $\theta^{\downarrow}_{+}(z;h)$ are finite for z in the direction $h \in \mathcal{Z}$. Then the lower and upper (parabolic) second order directional epiderivative of θ at z in the direction h are defined as, respectively,*

$$\theta^{\downarrow\downarrow}_{-}(z;h,d) := \liminf_{\substack{\rho\downarrow 0 \\ d'\to d}} \frac{\theta(z+\rho h+\frac{1}{2}\rho^2 d') - \theta(z) - \rho\theta^{\downarrow}_{-}(z;h)}{\frac{1}{2}\rho^2},$$

and

$$\theta^{\downarrow\downarrow}_{+}(z;h,d) := \sup_{\rho_n\in\Sigma}\left(\liminf_{\substack{n\to\infty \\ d'\to d}} \frac{\theta(z+\rho_n h+\frac{1}{2}\rho_n^2 d') - \theta(z) - \rho_n\theta^{\downarrow}_{+}(z;h)}{\frac{1}{2}\rho_n^2}\right),$$

where Σ denotes the set of all positive real sequences $\{\rho_n\}$ converging to 0. Moreover, if θ is directionally epidifferentiable at z in the direction h, and $\theta^{\downarrow\downarrow}_{-}(z;h,d) = \theta^{\downarrow\downarrow}_{+}(z;h,d)$ for all $d \in \mathcal{Z}$, we say θ twice (parabolically) directionally epidifferentiable at z in the direction h and denote it by $\theta^{\downarrow\downarrow}(z;h,d)$.

A 'one to one' relationship between the second order tangent sets and the second order directional epiderivatives is given by (cf. [8, Proposition 3.41]).

Proposition 2.3. *Suppose that $\theta : \mathcal{Z} \to (-\infty, +\infty]$ is proper and $z \in \mathrm{dom}\theta$. For $h \in \mathcal{Z}$, $\theta^{\downarrow}_{-}(z;h)$ and $\theta^{\downarrow}_{+}(z;h)$ are assumed to be finite. Then, we have*

$$\mathcal{T}^2_{\mathrm{epi}\,\theta}((z,\theta(z)),(h,\theta^{\downarrow}_{-}(z;h)) = \mathrm{epi}\,\theta^{\downarrow\downarrow}_{-}(z;h,\cdot). \tag{2.3}$$

and

$$\mathcal{T}^{i,2}_{\mathrm{epi}\theta}((z,\theta(z)),(h,\theta^{\downarrow}_{+}(z;h)) = \mathrm{epi}\,\theta^{\downarrow\downarrow}_{+}(z;h,\cdot), \tag{2.4}$$

Thus, all analysis of the second order tangent sets can be used here to characterize the second order directional epiderivatives. Under the assumptions that θ is convex and $\theta^{\downarrow}(z;h)$ is finite, we can obtain that the upper second order directional epiderivative $\theta^{\downarrow\downarrow}_{+}(z;h,\cdot)$ is convex.

2.3.2 Lipschitz-like properties

In this section, three pairs of Lipschitz-like properties are introduced. Within each pair, they are equivalence under the inverse operation of mappings. These Lipschitz-like properties are the most commonly used in the senstivity analysis of optimization problems. Some literature reviews are showed in our introduction.

Let us consider two finite dimensional real Euclidean spaces $\mathcal{W}$ and $\mathcal{Z}$, and a multi-valued mapping $F : \mathcal{W} \rightrightarrows \mathcal{Z}$. We define the graph of the mapping F as $\mathrm{gph}(F) := \{(w, z) \in \mathcal{W} \times \mathcal{Z} : z \in F(w)\}$ and denote $\mathcal{B}_{\mathcal{Z}} := \{z \in \mathcal{Z} : \|z\| \leq 1\}$. For the multi-valued mapping F, we define the following three pairs terminologies: Aubin property and metric regularity, isolated calmness and strong metric subregularity, and calmness and metric subregularity. These definitions and relationships can be found in, e.g., [30, 62, 74, 86].

Definition 2.8. [**Aubin property**] *A multi-valued mapping* $F : \mathcal{W} \rightrightarrows \mathcal{Z}$ *has the Aubin property at* $\bar{w}$ *for* $\bar{z}$ *with* $(\bar{w}, \bar{z}) \in \mathrm{gph}(F)$ *if there exists a constant* $\kappa \geq 0$ *and open neighborhoods* $\mathcal{U}$ *of* $\bar{w}$ *and* $\mathcal{V}$ *of* $\bar{z}$ *such that*

$$F(w) \cap \mathcal{V} \subset F(w') + \kappa\|w - w'\|\mathcal{B}_{\mathcal{Z}}, \quad \forall w, w' \in \mathcal{U}.$$

Definition 2.9. [**Metric regularity**] *A multi-valued mapping* $G : \mathcal{Z} \rightrightarrows \mathcal{W}$ *is said to be metrically regular at* $\bar{z}$ *for* $\bar{w}$ *with* $(\bar{z}, \bar{w}) \in \mathrm{gph}(G)$ *if there exists a constant* $\kappa \geq 0$ *along with open neighborhoods* $\mathcal{V}$ *of* $\bar{z}$ *and* $\mathcal{U}$ *of* $\bar{w}$ *such that*

$$\mathrm{dist}(z, G^{-1}(w)) \leq \kappa \mathrm{dist}(w, G(z)), \quad \text{for all } z \in \mathcal{V},\ w \in \mathcal{U}.$$

The Aubin property is also called "Lipschitz-like" or "pseudo-Lipschitzian" [3]. A well known criterion to characterize the Aubin property is the so called Mordukhovich criterion, that is, $F : \mathcal{W} \rightrightarrows \mathcal{Z}$ has the Aubin property at $\bar{w}$ for $\bar{z}$ with $(\bar{w}, \bar{z}) \in \mathrm{gph}(F)$ if and only if the limiting coderivative $D^*F(\bar{w}, \bar{z})$ is nonsingular at 0, i.e., $D^*F(\bar{w}, \bar{z})(0) = \{0\}$ [86, Theorem 9.40]. The equivalence of the Aubin property and the inverse metric regularity is established as below.

Proposition 2.4. *A multi-valued mapping $F : \mathcal{W} \rightrightarrows \mathcal{Z}$ has the Aubin property at $\bar{w}$ for $\bar{z}$ with $(\bar{w}, \bar{z}) \in \mathrm{gph}(F)$ if and only if its inverse $F^{-1} : \mathcal{Z} \rightrightarrows \mathcal{W}$ is metrically regular at $\bar{z}$ for $\bar{w}$.*

A relaxed 'one-point' variant of the Aubin property and the metric regularity is the so called calmness and metric subregularity. In this thesis, we investigate the metric subregularity property.

Definition 2.10. **[Calmness]** *We say a multi-valued mapping $F : \mathcal{W} \rightrightarrows \mathcal{Z}$ is calm at $\bar{w}$ for $\bar{z}$ if $(\bar{w}, \bar{z}) \in \mathrm{gph}(F)$ and there exist a constant $\kappa \geq 0$ and neighborhoods $\mathcal{U}$ of $\bar{w}$ and $\mathcal{V}$ of $\bar{z}$ such that*

$$F(w) \cap \mathcal{V} \subseteq F(\bar{w}) + \kappa \|w - \bar{w}\| \mathcal{B}_{\mathcal{Z}}, \quad \forall w \in \mathcal{U}.$$

Or alternatively, we say F is calm at $\bar{w}$ for $\bar{z}$ with modulus $\kappa > 0$ if there exists a neighborhood $\mathcal{V}'$ of $\bar{z}$ such that

$$F(w) \cap \mathcal{V}' \subseteq F(\bar{w}) + \kappa \|w - \bar{w}\| \mathcal{B}_{\mathcal{Z}}, \quad \forall w \in \mathcal{W}.$$

Definition 2.11. **[Metric subregularity]** *We say a multi-valued mapping $G : \mathcal{Z} \rightrightarrows \mathcal{W}$ is metrically subregular at $\bar{z}$ for $\bar{w}$ if $(\bar{z}, \bar{w}) \in \mathrm{gph}(G)$ and there exist a constant $\kappa \geq 0$ along with neighborhoods $\mathcal{V}$ of $\bar{z}$ and $\mathcal{U}$ of $\bar{w}$ such that*

$$\mathrm{dist}(z, G^{-1}(\bar{w})) \leq \kappa \mathrm{dist}(\bar{w}, G(z) \cap \mathcal{U}), \quad \forall z \in \mathcal{V}.$$

Or alternatively, we say G is metrically subregular at $\bar{z}$ for $\bar{w}$ with modulus $\kappa > 0$ if there exists a neighborhood $\mathcal{V}'$ of $\bar{z}$ such that

$$\mathrm{dist}(z, G^{-1}(\bar{w})) \leq \kappa \mathrm{dist}(\bar{w}, G(z)), \quad \forall z \in \mathcal{V}'.$$

The calmness is also called "upper Lipschitzian" by Robinson [80]. Analogously, we have the equivalent relationship between the calmness and the metric subregularity.

Proposition 2.5. *For a a multi-valued mapping $F : \mathcal{W} \rightrightarrows \mathcal{Z}$, let $(\bar{w}, \bar{z}) \in \mathrm{gph}(F)$. Then F is calm at $\bar{w}$ for $\bar{z}$ if and only if F^{-1} is metrically subregular at $\bar{z}$ for $\bar{w}$.*

As mentioned in the introduction, there are many literatures revealing the connection between error bounds and the calmness or the metric subregularity. Under mild conditions, the latter one implies the former one. The study of the calmness property or the metric subregularity is significant, since the error bound is a poweful tool to obtain convergence rates.

In general, we can only have the equivalence between the calmness and the metric subregularity. However, for the subgradient mapping of convex functions, we have another characterization of the metric subregularity.

Theorem 2.1. **([1, Theorem 3.3] and [110, Theorem 4.3])** *Let $\mathcal{H}$ be a real Hilbert space endowed with the inner product $\langle\cdot,\cdot\rangle$ and $\varphi : \mathcal{H} \to (-\infty,+\infty]$ be a proper lower semicontinuous convex function. Let $\bar{z},\bar{w} \in \mathcal{H}$ satisfy $\bar{w} \in \partial\varphi(\bar{z})$. Then $\partial\varphi$ is metric subregular at $\bar{z}$ for $\bar{w}$ if and only if there exists a neighborhood $\mathcal{V}$ of $\bar{z}$ and a positive constant κ such that*

$$\varphi(z) \geq \varphi(\bar{z}) + \langle \bar{w}, z-\bar{z}\rangle + \kappa \mathrm{dist}^2(z,(\partial\varphi)^{-1}(\bar{w})), \quad \forall z \in \mathcal{V}. \tag{2.5}$$

Furthermore, Robinson [81] shows that the calmness property always holds for piecewise polyhedral mappings.

Proposition 2.6. *If a multi-valued mapping $F : \mathcal{W} \rightrightarrows \mathcal{Z}$ is piecewise polyhedral, then F is calm with the same modulus $\kappa \geq 0$ at any w for z whenever $(w,z) \in$ gph(F).*

Therefore, F^{-1} always possesses the metric subregularity when F is a piecewise polyhedral mapping. Moreover, any linear mapping is calm at any point of its graph, so does its inverse mapping. But, the inverse of a linear mapping has the Aubin property at some point if this mapping is surjective, and vice versa.

At the end of this section, we would like to introduce the isolated calmness and the strong metric subregularity, which can be viewed as 'strengthened forms' of the calmness and the metric subregularity.

Definition 2.12. [**Isolated calmness**] *A multi-valued mapping* $F : \mathcal{W} \rightrightarrows \mathcal{Z}$ *is said to be isolated calm at* $\bar{w}$ *for* $\bar{z}$ *if* $(\bar{w}, \bar{z}) \in \mathrm{gph}(F)$ *and there exist a constant* $\kappa \geq 0$ *and neighborhoods* $\mathcal{U}$ *of* $\bar{w}$ *and* $\mathcal{V}$ *of* $\bar{z}$ *such that*

$$F(w) \cap \mathcal{V} \subseteq \{\bar{z}\} + \kappa \|w - \bar{w}\| \mathcal{B}_{\mathcal{Z}}, \quad \forall w \in \mathcal{U}.$$

Definition 2.13. [**Strong metric subregularity**] *A multi-valued mapping* $G : \mathcal{Z} \rightrightarrows \mathcal{W}$ *is said to be strongly metrically subregular at* $\bar{z}$ *for* $\bar{w}$ *if* $(\bar{z}, \bar{w}) \in \mathrm{gph}(G)$ *and there exist a constant* $\kappa \geq 0$ *along with neighborhoods* $\mathcal{V}$ *of* $\bar{z}$ *and* $\mathcal{U}$ *of* $\bar{w}$ *such that*

$$\|z - \bar{z}\| \leq \kappa \mathrm{dist}(\bar{w}, G(z) \cap \mathcal{U}), \quad \forall z \in \mathcal{V}.$$

From the above definitions, it can be regarded as strengthened forms of the calmness and the metric subregularity, since we let $F(\bar{w}) \cap \mathcal{V}$ singleton, i.e., $\bar{z}$ is an isolated point in $F(\bar{w})$. Therefore, we have the next proposition.

Proposition 2.7. [30, **Proposition 3I.1**] *If a multi-valued mapping* $F : \mathcal{W} \rightrightarrows \mathcal{Z}$ *is piecewise polyhedral, then it has the isolated calm at* $\bar{w}$ *for* $\bar{z}$ *if and only if* $\bar{z}$ *is an isolated point of* $F(\bar{w})$*.*

The equivalence of the isolated calmness and the strong metric subregularity of the inverse is given by the following.

Proposition 2.8. *For a a multi-valued mapping* $F : \mathcal{W} \rightrightarrows \mathcal{Z}$*, let* $(\bar{w}, \bar{z}) \in \mathrm{gph}(F)$*. Then* F *is isolated calm at* $\bar{w}$ *for* $\bar{z}$ *if and only if* F^{-1} *is strongly metrically subregular at* $\bar{z}$ *for* $\bar{w}$*.*

Similar with the Aubin property having the Mordukhovich criterion of the coderivative, the isolated calmness has the following criterion of the graphical derivative.

Lemma 2.3. (King and Rockafellar [48]**, Levy** [54]**)** *Let* $(\bar{w}, \bar{z}) \in \mathrm{gph}(F)$*. Then* F *is isolated calm at* $\bar{w}$ *for* $\bar{z}$ *if and only if the graphical derivative* $DF(\bar{w}, \bar{z})$ *is nonsingular at 0, i.e.,*$DF(\bar{w}, \bar{z})(0) = \{0\}$*.*

By using the above lemma, we can obtain a reduced one for continous mapping, which is a natural extension of [41, Lemma 4.4].

Lemma 2.4. *Suppose $F : \mathcal{W} \to \mathcal{Z}$ is a continuous mapping. Let $(\bar{w}, \bar{z}) \in \mathcal{W} \times \mathcal{Z}$ satisfing $F(\bar{w}) = \bar{z}$. Suppose that F is locally Lipschitz continuous around u_0 and directional differentiable at $\bar{w}$. Then F^{-1} is isolated calm at $\bar{z}$ for $\bar{w}$ if and only if*

$$F'(\bar{w}; d) = 0 \Longrightarrow d = 0, \quad \forall d \in \mathcal{W}.$$

Recently, a well studied teminology is the roubust isolated calmness.

Definition 2.14. [Locally nonempty-valued] *We say a multi-valued mapping $F : \mathcal{W} \rightrightarrows \mathcal{Z}$ is locally nonempty-valued at $\bar{w}$ for $\bar{z}$ if $(\bar{w}, \bar{z}) \in \operatorname{gph}(F)$ and there exist neighborhoods $\mathcal{U}$ of $\bar{w}$ and $\mathcal{V}$ of $\bar{z}$ such that*

$$F(w) \cap \mathcal{V} \neq \emptyset, \quad \forall w \in \mathcal{U}.$$

Definition 2.15. [Robust isolated calmness] *We say a multi-valued mapping $F : \mathcal{W} \rightrightarrows \mathcal{Z}$ is robust isolated calm at $\bar{w}$ for $\bar{z}$ with $(\bar{w}, \bar{z}) \in \operatorname{gph}(F)$ if F is both isolated calm and locally nonempty valued at $\bar{w}$ for $\bar{z}$.*

One can find the relationship between the Aubin property and the isolated calmness in [37].

2.3.3 First order optimality conditions

We focus on first order optimality conditions for our composite problem (1.1) in this section. We derive first order optimality conditions of (1.1) in two ways. One way is to decompose the composite term by casting (1.1) in a higher dimensional space, and the other way is to use the conjugate property of convex functions without lifting the dimension of (1.1).

One can see that the composite problem (1.1) can be equivalently written as:

$$\begin{aligned} \min \quad & f(x) + t \\ \text{s.t.} \quad & h(x) = 0, \\ & (g(x), t) \in \mathcal{K}, \end{aligned} \tag{2.6}$$

where $\mathcal{K} := \operatorname{epi}\theta$ is a closed convex set in $\mathcal{S} \times \mathcal{R}$.

Additionally, if $\mathcal{K}$ is a cone, then (2.6) is a conic programming. By transfroming the problem (1.1) to the form (2.6), we can decompose the composite term $\theta \circ g$ with a trade off that the problem (2.6) is one dimensionally higher than (1.1). Before deriving first optimality conditions for the composite problem (1.1), we show the Robinson's constraint qualifications with respect to problems (2.6) and (1.1) first.

The Robinsons's constraint qualification (RCQ) at a feasible solution $(\bar{x}, \theta(g(\bar{x})))$ of the problem (2.6) is defined by

$$0 \in \operatorname{int}\left\{\begin{pmatrix} (h(\bar{x}), 0) \\ (g(\bar{x}), \theta(g(\bar{x}))) \end{pmatrix} + \begin{pmatrix} (h'(\bar{x}), 0) \\ (g'(\bar{x}), 1) \end{pmatrix} (\mathcal{X} \times \mathcal{R}) - \begin{pmatrix} \{0\} \times \{0\} \\ \mathcal{K} \end{pmatrix}\right\}. \tag{2.7}$$

We can obtain the RCQ for our original problem (1.1) by reducing those (2.7) of the problem (2.6). We say that the RCQ holds at a feasible solution $\bar{x}$ of the problem (1.1) if

$$0 \in \operatorname{int}\left\{\begin{pmatrix} h(\bar{x}) \\ g(\bar{x}) \end{pmatrix} + \begin{pmatrix} h'(\bar{x}) \\ g'(\bar{x}) \end{pmatrix} \mathcal{X} - \begin{pmatrix} \{0\} \\ \operatorname{dom}\theta \end{pmatrix}\right\}, \tag{2.8}$$

or equivalently [8, Proposition 2.97],

$$\begin{pmatrix} h'(\bar{x}) \\ g'(\bar{x}) \end{pmatrix} \mathcal{X} + \begin{pmatrix} \{0\} \\ \mathcal{T}_{\operatorname{dom}\theta}(g(\bar{x})) \end{pmatrix} = \begin{pmatrix} \mathcal{Y} \\ \mathcal{S} \end{pmatrix}. \tag{2.9}$$

Remark 2.1. *The Robinson's constraint qualification has the stability property. That is, if (2.8) or (2.9) holds at $\bar{x}$, then it holds at some neighborhood of $\bar{x}$. We call (2.8) or (2.9) the RCQ of the original problem (1.1) in the subsequent discussions.*

Since all program data are 'smooth' in (2.6), we can derive first order optimality conditions with respect to the problem (2.6), and then reduce such first order optimality conditions to those of our original form (1.1) under some mild conditions. This kind of transformation can be found in [8, Section 3.4.1].

For any $(x, t, y, S, \tau) \in \mathcal{X} \times \mathcal{R} \times \mathcal{Y} \times \mathcal{S} \times \mathcal{R}$, we define the Lagrangian function of (2.6) by

$$\mathcal{L}(x, t; y, S, \tau) := f(x) + t + \langle y, h(x) \rangle + \langle S, g(x) \rangle + t\tau. \tag{2.10}$$

We call $(\bar{x}, \bar{t}) \in \mathcal{X} \times \mathcal{R}$ a stationary point of the problem (2.6) and $(\bar{y}, \bar{S}, \bar{\tau})$ a Lagrangian multiplier if $(\bar{x}, \bar{t}, \bar{y}, \bar{S}, \bar{\tau})$ satisfies the following KKT conditions:

$$\begin{cases} \nabla_x \mathcal{L}(\bar{x}, \bar{t}, \bar{y}, \bar{S}, \bar{\tau}) = 0, \\ \nabla_t \mathcal{L}(\bar{x}, \bar{t}, \bar{y}, \bar{S}, \bar{\tau}) = 0, \\ h(\bar{x}) = 0, \\ (\bar{S}, \bar{\tau}) \in \mathcal{N}_{\mathcal{K}}\big((g(\bar{x}), \bar{t})\big). \end{cases} \tag{2.11}$$

Combining them together, we call $(\bar{x}, \bar{t}, \bar{y}, \bar{S}, \bar{\tau})$ a KKT point of the problem (2.6). We use $\widehat{\mathcal{M}}(\bar{x}, \bar{t})$ to denote the set of all the Lagrangian multipliers at a stationary point $(\bar{x}, \bar{t})$ with respect to the problem (2.6).

Especially, let us consider the stationarity at $(\bar{x}, \theta(g(\bar{x})))$. Noting that by the second condition in (2.11), we always have $\bar{\tau} = -1$. Thus, the last inclusion in (2.11) becomes $(\bar{S}, -1) \in \mathcal{N}_{\mathcal{K}}\big((g(\bar{x}), \theta(g(\bar{x}))\big)$, which is equivalent to $\bar{S} \in \partial\theta(g(\bar{x}))$ by Lemma 2.1. Moreover, we obtain the following reduced Lagrangian function by substituting $\bar{\tau} = -1$ into (2.10),

$$L(x, y, S) := f(x) + \langle y, h(x) \rangle + \langle S, g(x) \rangle.$$

Therefore, we can further obtain the following reduced KKT conditions with respect to the problem (1.1) from (2.11):

$$\begin{cases} \nabla_x L(\bar{x}, \bar{y}, \bar{S}) = 0, \\ h(\bar{x}) = 0, \\ \bar{S} \in \partial\theta(g(\bar{x})). \end{cases} \tag{2.12}$$

However, in general, the necessary first order optimality conditions of the problem (1.1) is not coincide with (2.12).

Let $\widetilde{L}: \mathcal{X} \times \mathcal{Y} \to \mathcal{R}$ be the Lagrangian function of the composite problem (1.1) defined by

$$\widetilde{L}(x;y) = f(x) + \theta(g(x)) + \langle y, h(x)\rangle, \ \forall (x,y) \in \mathcal{X} \times \mathcal{Y}.$$

The first order optimality conditions [8, Proposition 3.99] for a stationay point $\bar{x}$ of the problem (1.1) is that there exist $\bar{y} \in \mathcal{Y}$ such that

$$\begin{cases} f'(\bar{x})d + (\theta \circ g)^{\downarrow}_{-}(\bar{x};d) + \langle \bar{y}, h'(\bar{x})d\rangle \geq 0, \ \forall d \in \mathcal{X}, \\ h(\bar{x}) = 0. \end{cases} \tag{2.13}$$

By the arguments in [8, Section 3.4.1] and [61, Section 5.1], we know that there has a gap between the optimality conditions (2.13) and the reduced KKT conditions (2.12). In fact, the latter one is stronger than the former one. Fortunately, we can fulfill this gap if the following reduced RCQ [8, 61] holds at $\bar{x}$, i.e.,

$$0 \in \operatorname{int}\{g(\bar{x}) + g'(\bar{x})\mathcal{X} - \operatorname{dom}\theta\}. \tag{2.14}$$

Therefore, within the reduced RCQ (2.14) satisfied, the first order optimality conditions of the problem (1.1) is the KKT system (2.12).
Similarly, we say $\bar{x}$ a stationary point of the problem (1.1) and $(\bar{y}, \bar{S})$ a corresponding Lagrangian multiplier if $(\bar{x}, \bar{y}, \bar{S})$ satisfies the KKT conditions (2.12). And denote $\mathcal{M}(\bar{x})$ as the set of all the Lagrangian multipliers at a stationary point $\bar{x}$ with respect to the original problem (1.1).

On the one hand, we can obtain first order optimality conditions of the original problem (1.1) via reducing from the optimality conditions imposed in a higher dimensional space with respect to the problem (2.6). On the other hand, for every $(x, y, S) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{S}$, we can consider the Lagrangian function of (1.1) as

$$l(x, y, S) := f(x) + \langle y, h(x)\rangle + \langle S, g(x)\rangle - \theta^*(S), \tag{2.15}$$

where θ^* is the Fenchel conjugate function of the convex function θ. This kind of form for composite problems is introduced by Burke [13, 14]. It is easy to see that we can obtain the first order optimality conditions (2.12) by (2.15) directly. The Lagrangian function l (2.15) contains all data infomation of the problem (1.1), we adopt this form for our later disscussions.

It has been showed in [115] that the RCQ (2.7) holds at $(\bar{x}, \theta(g(\bar{x}))$ if and only if the multiplier set $\widehat{\mathcal{M}}(\bar{x}, \theta(g(\bar{x}))$ associated with the problem (2.6) is nonempty, convex and compact. By the previous reduction, we can transfer the property of $\widehat{\mathcal{M}}(\bar{x}, \theta(g(\bar{x}))$ to the multiplier set $\mathcal{M}(\bar{x})$ with respect to the problem (1.1).

Proposition 2.9. *Let $\bar{x}$ be a local optimal solution of the problem (1.1). Then the set of Lagrangian multipliers $\mathcal{M}(\bar{x})$ of (1.1) is nonempty, convex and compact if and only if the RCQ (2.8) or (2.9) holds.*

Another stronger form of the RCQ is the so called strict Robinsons's constraint qualification (SRCQ), which is frequently used in perturbation analysis. The SRCQ is a sufficient condition for the multiplier set to be a singleton for the problem (2.6). While, the RCQ can only guarantee the boundedness of the multiplier set.
The SRCQ for problem (2.6) at the stationary point $(\bar{x}, \theta(g(\bar{x}))$ with respect to $(\bar{y}, \bar{S}, -1) \in \widehat{\mathcal{M}}(\bar{x}, \theta(g(\bar{x}))$ is defined by:

$$\begin{pmatrix} (h'(\bar{x}), 0) \\ (g'(\bar{x}), 1) \end{pmatrix} (\mathcal{X} \times \mathcal{R}) + \begin{pmatrix} \{(0,0)\} \\ \mathcal{T}_{\mathcal{K}}\big(g(\bar{x}), \theta(g(\bar{x}))\big) \end{pmatrix} \cap (\bar{y}, 0, \bar{S}, -1)^{\perp} = \begin{pmatrix} \mathcal{Y} \times \{0\} \\ \mathcal{S} \times \mathcal{R} \end{pmatrix}. \tag{2.16}$$

Analogously, we can derive the SRCQ for the problem (1.1) by the same reduction. Such reduced SRCQ restricts the uniqueness of the corresponding multipliers. For simplify notations, we define a set-valued mapping $\mathcal{C}_\theta : \operatorname{dom}\theta \times \mathcal{S} \to \mathcal{S}$ associated with a closed proper convex function θ as

$$\mathcal{C}_\theta(w, z) := \{h \in \mathcal{S} : \theta^{\downarrow}(w; h) = \langle h, z \rangle\}, \quad \forall (w, z) \in \operatorname{dom}\theta \times \mathcal{S}. \tag{2.17}$$

Proposition 2.10. *Let $\bar{x}$ be a local optimal solution of the problem (1.1). Assume that $\mathcal{M}(\bar{x})$ is nonempty. Suppose the following condition holds at $\bar{x}$ with respect to $(\bar{y}, \bar{S}) \in \mathcal{M}(\bar{x})$:*

$$\begin{pmatrix} h'(\bar{x}) \\ g'(\bar{x}) \end{pmatrix} \mathcal{X} + \begin{pmatrix} 0 \\ \mathcal{C}_\theta(g(\bar{x}), \bar{S}) \end{pmatrix} = \begin{pmatrix} \mathcal{Y} \\ \mathcal{S} \end{pmatrix}, \tag{2.18}$$

where $\mathcal{C}_\theta(\cdot,\cdot)$ is define as in (2.17). Then $\mathcal{M}(\bar{x}) = \{(\bar{y}, \bar{S})\}$ is a singleton.

Remark 2.2. *One can find that $\mathcal{C}_\theta\big(g(\bar{x}), \bar{S}\big) = \mathcal{N}_{\partial\theta(g(\bar{x}))}(\bar{S})$ as $\bar{S} \in \partial\theta(g(\bar{x}))$. Such $\mathcal{C}_\theta\big(g(\bar{x}), \bar{S}\big)$ can be viewed as a cirtical cone of $\partial\theta(g(\bar{x}))$ at $g(\bar{x}) + \bar{S}$.*

To end this section, we define the critical cones of problems (2.6) and (1.1). The critical cone consists of those directions in which the objective value is nonincreasing and the constraints are satisfied.

Let $(\bar{x}, \bar{t}) \in \mathcal{X} \times \mathcal{R}$ be a feasible point of the problem (2.6), then the critical cone $\widehat{\mathcal{C}}(\bar{x}, \bar{t})$ of the problem (2.6) takes the form of

$$\begin{aligned} \widehat{\mathcal{C}}(\bar{x}, \bar{t}) := \{(d_1, d_2) \in \mathcal{X} \times \mathcal{R} : h'(\bar{x})d_1 = 0, (g'(\bar{x})d_1, d_2) \in \mathcal{T}_{\mathcal{K}}\big(g(\bar{x}), \bar{t}\big), \\ f'(\bar{x})d_1 + d_2 \leq 0\}. \end{aligned}$$

Furthermore, if $(\bar{x}, \theta(g(\bar{x})))$ is a local optimal solution of the problem (2.6) and $\widehat{\mathcal{M}}(\bar{x}, \theta(g(\bar{x})))$ is nonempty, then for any $(\bar{y}, \bar{S}, -1) \in \widehat{\mathcal{M}}(\bar{x}, \theta(g(\bar{x})))$,

$$\begin{aligned} \widehat{\mathcal{C}}(\bar{x}, \theta(g(\bar{x}))) &= \{(d_1, d_2) \in \mathcal{X} \times \mathcal{R} : h'(\bar{x})d_1 = 0, (g'(\bar{x})d_1, d_2) \in \mathcal{T}_{\mathcal{K}}\big(g(\bar{x}), \theta(g(\bar{x}))\big), \\ &\qquad f'(\bar{x})d_1 + d_2 = 0\} \\ &= \{(d_1, d_2) \in \mathcal{X} \times \mathcal{R} : h'(\bar{x})d_1 = 0, (g'(\bar{x})d_1, d_2) \in \mathcal{T}_{\mathcal{K}}\big(g(\bar{x}), \theta(g(\bar{x}))\big), \\ &\qquad (g'(\bar{x})d_1, d_2) \in (\bar{S}, -1)^{\perp}\}. \end{aligned} \tag{2.19}$$

One should note that $(g'(\bar{x})d_1, d_2) \in \mathcal{T}_{\mathcal{K}}\big(g(\bar{x}), \theta(g(\bar{x}))\big)$ is equivalent to $(g'(\bar{x})d_1, d_2) \in \operatorname{epi}\theta^{\downarrow}(g(\bar{x}); \cdot)$, i.e., $(d_1, d_2) \in \operatorname{epi}\theta^{\downarrow}(g(\bar{x}); g'(\bar{x})(\cdot))$, by using $\mathcal{K} = \operatorname{epi}\theta$ and Proposition 2.2. However, we cannot directly obtain the critical cone of the problem (1.1) by reducing that of problem (2.6), since the chain rule for the directional epiderivative of $\theta \circ g$ does not hold. We want to claim that

$$(\theta \circ g)^{\downarrow}_{-}(\bar{x}; d) \geq \theta^{\downarrow}(g(\bar{x}); g'(\bar{x})d), \ \forall d \in \mathcal{X}. \tag{2.20}$$

Let $(d, \mu) \in \operatorname{epi}(\theta \circ g)^{\downarrow}_{-}(\bar{x}; \cdot)$ be arbitrary given. Then, due to Proposition 2.2 and the definition of the tangent cone, there exist sequences $\{\rho_n\}_{n\geq 0}$ with $\rho_n \downarrow 0$, $d_n \to d$, and $\mu_k \to \mu$ such that $(\bar{x} + \rho_n d_n, \theta(\bar{x}) + \rho_n \mu_n) \in \operatorname{epi}(\theta \circ g)$, i.e.,

$$\theta(g(\bar{x} + \rho_n d_n)) - \theta(g(\bar{x})) \leq \rho_n \mu_n.$$

By the Taylor expansion, we have

$$\theta(g(\bar{x}) + \rho_n g'(\bar{x}) d_n + o(\rho_n)) - \theta(g(\bar{x})) \leq \rho_n \mu_n.$$

Dividing by ρ_n on both sides and then taking limit, we obtain

$$\begin{aligned} \mu &\geq \liminf_{n\to\infty} \frac{\theta(g(\bar{x}) + \rho_n g'(\bar{x}) d_n + o(\rho_n)) - \theta(g(\bar{x}))}{\rho_n} \\ &\geq \liminf_{\substack{n\to\infty \\ \widetilde{H}\to g'(\bar{x})d}} \frac{\theta(g(\bar{x}) + \rho_n \widetilde{H}) - \theta(g(\bar{x}))}{\rho_n} \\ &= \theta^{\downarrow}(g(\bar{x}); g'(\bar{x})d). \end{aligned}$$

Therefore, we show (2.20). It is equivalent to $\operatorname{epi}(\theta \circ g)^{\downarrow}_{-}(\bar{x}; \cdot) \subset \operatorname{epi} \theta^{\downarrow}(g(\bar{x}); g'(\bar{x})(\cdot))$. Generally, we do not have the inverse inclusion. Thus, we cannot obtain the critical cone of the problem (1.1) via reduction of (2.19). Fortunately, the inverse inclusion can be obtained under an additional condition — the reduced RCQ (2.14).

Lemma 2.5. (cf. [8, Proposition 2.136]) *Suppose that $\theta : \mathcal{S} \to (-\infty, +\infty]$ is proper, lower semicontinuous and convex and $g : \mathcal{X} \to \mathcal{S}$ is continuously differentiable mappings. If the reduced RCQ (2.14) is satisfied at $\bar{x} \in g^{-1}(\operatorname{dom}\theta)$, then the composite function $\theta \circ g$ is directionally epidifferentiable at $\bar{x}$ and*

$$(\theta \circ g)^{\downarrow}(\bar{x}; d) = \theta^{\downarrow}(g(\bar{x}); g'(\bar{x})d), \ \forall d \in \mathcal{X}. \tag{2.21}$$

Therefore, within the reduced RCQ (2.14), we can obtain the following reduced critical cone of the problem (1.1) by reduction of $\widehat{\mathcal{C}}(\bar{x}, \theta(g(\bar{x})))$ in (2.19).

Proposition 2.11. *Let $\bar{x} \in \mathcal{X}$ be a feasible point of the problem (1.1) with the reduced RCQ (2.14) holds at $\bar{x}$. Then $\widehat{\mathcal{C}}(\bar{x}, \theta(g(\bar{x})))$ can be written as*

$$\widehat{\mathcal{C}}(\bar{x}, \theta(g(\bar{x}))) = \{(d_1, d_2) \in \mathcal{X} \times \mathcal{R} : d_1 \in \mathcal{C}(\bar{x}), \theta^{\downarrow}(g(\bar{x}); g'(\bar{x})d_1) \leq d_2 \leq -f'(\bar{x})d_1\},$$

where the critical cone $\mathcal{C}(\bar{x})$ *of problem (1.1) is defined as*

$$\mathcal{C}(\bar{x}) := \left\{ d \in \mathcal{X} : h'(\bar{x})d = 0,\ f'(\bar{x})d + (\theta \circ g)^{\downarrow}(\bar{x}; d) \leq 0 \right\}. \tag{2.22}$$

Furthermore, if $\bar{x}$ *is a a local optimal solution of the problem (1.1),* $\mathcal{M}(\bar{x})$ *is non-empty and* $\bar{S} \in \mathcal{M}(\bar{x})$*, then* $\mathcal{C}(\bar{x})$ *defined in (2.22) can be written as*

$$\mathcal{C}(\bar{x}) = \left\{ d \in \mathcal{X} : h'(\bar{x})d = 0, g'(\bar{x})d \in \mathcal{C}_{\theta}\big(g(\bar{x}), \bar{S}\big) \right\}, \tag{2.23}$$

where $\mathcal{C}_{\theta}(\cdot, \cdot)$ *is define as in (2.17).*

The critical cone (2.23) plays an essential role in second order conditions.

2.4 The spectral functions

We give an introduction to spectral functions of matrices and employ some nice properties of these functions. By using these properties, the characterization of such matrix functions can be reduced to correspongding underlying functions of vectors.

Definition 2.16. *A function* $q : \mathcal{R}^m \to (-\infty, +\infty]$ *is said to be symmetric if* $q(Qx) = Qq(x)$ *for any* $x \in \mathcal{R}^m$ *and any permutation matrix* $Q \in \mathcal{R}^{m\times m}$*, and is said to be absolutely symmetric if* $q(x) = q(Qx)$ *for any* $x \in \mathcal{R}^m$ *and any signed permutation matrix* Q*, i.e., an* $m \times m$ *matrix with each row or column has one nonzero entry which is* ± 1*.*

Let $\theta : \mathcal{S} \to (-\infty, +\infty]$ be a closed proper convex function with $\mathcal{S}$ be either a rectangular matrix space $\mathcal{R}^{m\times n}$ $(m \leq n)$ or a symmetric matrix space $\mathcal{S}^m$.
We call θ a spectral function if θ can be written as a composite matrix function either in the form of

$$\theta(X) = q \circ \sigma(X), \quad \forall X \in \mathcal{R}^{m\times n}, \tag{2.24}$$

where $\sigma(X) = (\sigma_1(X), \ldots, \sigma_m(X))$ with $\sigma_1(X) \geq \sigma_2(X) \geq \ldots \geq \sigma_m(X) \geq 0$ being the singular values of X arranged in the non-increasing order, and $q : \mathcal{R}^m \to$

$(-\infty, +\infty]$ is absolutely symmetric,

or in the form of

$$\theta(X) = q \circ \lambda(X), \quad \forall X \in \mathcal{S}^m, \tag{2.25}$$

where $\lambda(X) = (\lambda_1(X), \lambda_2(X), \ldots, \lambda_m(X))$ with $\lambda_1(X) \geq \lambda_2(X) \geq \ldots \geq \lambda_m(X)$ being the eigenvelues of X arranged in the non-increasing order, and $q : \mathcal{R}^m \to (-\infty, +\infty]$ is symmetric.

In the following discussions, we also assume that q is closed proper convex. By the above statements of spectral functions, we can find that such spectral functions include $\delta_{\mathcal{S}^m_+}(\cdot)$ when $q(\cdot) = \delta_{\mathcal{R}^m_+}(\cdot)$, and the matrix Ky Fan k-norm $\|\cdot\|_{(k)}$ when q is the vector k-norm, i.e., the sum of k $(1 \leq k \leq m)$ largest absolute components of a vector in $\mathcal{R}^m$.

Before we show more properties of spectral functions. We would like to adopt the follwing definition in [8, Definition 3.135] here.

Definition 2.17. **[C^2-cone reducible set]** *The closed convex set $\mathcal{K}$ is said to be C^2-cone reducible at $\bar{A} \in \mathcal{K}$, if there exist a open neighborhood $\mathcal{W} \subseteq \mathcal{Z}$ of $\bar{A}$, a pointed closed convex cone $\mathcal{Q}$ (a cone is said to be pointed if and only if its lineality space is the origin) in a finite dimensional space $\mathcal{U}$ and a twice continuously differentiable mapping $\Xi : \mathcal{W} \to \mathcal{U}$ such that: (i) $\Xi(\bar{A}) = 0 \in \mathcal{U}$; (ii) the derivative mapping $\Xi'(\bar{A}) : \mathcal{Z} \to \mathcal{U}$ is onto; (iii) $\mathcal{K} \cap \mathcal{W} = \{A \in \mathcal{W} \mid \Xi(A) \in \mathcal{Q}\}$. We say that $\mathcal{K}$ is C^2-cone reducible if $\mathcal{K}$ is C^2-cone reducible at every $\bar{A} \in \mathcal{K}$.*

We say a closed proper convex function θ is C^2-cone reducible if its epigraph epiθ is C^2-cone reducible.

As we have mentioned in section 2.3.1 that, in general, $\mathcal{T}^2_{\mathcal{K}}(x; d) \neq \mathcal{T}^{i,2}_{\mathcal{K}}(x; d)$ even if $\mathcal{K}$ is convex ([8, Section 3.3]). However, it follows from [8, Proposition 3.136] that if $\mathcal{K}$ is a C^2-reducible convex set, then the equality always holds.

Next, we list some useful results for the "sigma term" of the C^2-cone reducible sets, which associated with the second order optimality condition for the problem (1.1). One can reference [8, (3.266) and (3.274)].

Lemma 2.6. *Let $\bar{A} \in \mathcal{K}$ be given. Then, there exist an open neighborhood $\mathcal{W} \subseteq \mathcal{Z}$ of $\bar{A}$, a pointed closed convex cone $\mathcal{Q}$ in a finite dimensional space $\mathcal{U}$ and a twice continuously differentiable mapping $\Xi : \mathcal{W} \to \mathcal{U}$ satisfying conditions (i)-(iii) in Definition 2.17 such that for all $A \in \mathcal{W}$ sufficiently close to $\bar{A}$,*

$$\mathcal{N}_{\mathcal{K}}(A) = \Xi'(A)^* \mathcal{N}_{\mathcal{Q}}(\Xi(A))), \tag{2.26}$$

where $\Xi'(A)^* : \mathcal{U} \to \mathcal{Z}$ is the adjoint of $\Xi'(A)$. In particular, for any $\bar{B} \in \mathcal{N}_{\mathcal{K}}(\bar{A})$, there ia a unique element u in $\mathcal{N}_{\mathcal{Q}}(\Xi(\bar{A}))$ such that $\bar{B} = \Xi'(\bar{A})^* u$, denoted by $(\Xi'(\bar{A})^*)^{-1}\bar{B}$. Furthermore, We have for any $D \in \mathcal{C}_{\mathcal{K}}(\bar{A} + \bar{B})$,

$$\sigma(\bar{B}, \mathcal{T}^2_{\mathcal{K}}(\bar{A}, D)) = -\langle (\Xi'(\bar{A})^*)^{-1}\bar{B}, \Xi''(\bar{A})(D, D)\rangle. \tag{2.27}$$

Then, we go back to discuss spectral functions. The next three properties [20, Proposition 2-3 and 10-14] show the conjugacy of spectral functions, and that the C^2-cone reducibility and the metric subregularity of the spectral function θ can be checked via its underlying function q.

Let $\mathcal{S} = \mathcal{R}^{m \times n}$ or $\mathcal{S}^m$ in below. We always assume that underlying function q is closed proper convex in the following statements.

Proposition 2.12. *Let the function $\theta : \mathcal{S} \to (-\infty, +\infty]$ be a spectral function in form of (2.24) or (2.25), and $q : \mathcal{R}^m \to (-\infty, +\infty]$ be the underlying absolutely symmetric or symmetric function associated with θ. Then,*

(i) the conjugate function q^* is absolutely symmetric and $(q \circ \sigma)^* = q^* \circ \sigma$ if q is absolutely symmetric;

(ii) the conjugate function q^* is symmetric and $(q \circ \lambda)^* = q^* \circ \lambda$ if q is symmetric.

Proposition 2.13. *Let the function $\theta : \mathcal{S} \to (-\infty, +\infty]$ be a spectral function in form of (2.24) or (2.25), and $q : \mathcal{R}^m \to (-\infty, +\infty]$ be the underlying absolutely symmetric or symmetric function associated with θ. Then, for any $X \in \operatorname{dom}\theta$, θ is C^2-cone reducible at X if q is C^2-cone reducible at $\sigma(X)$ (or $\lambda(X)$).*

Proposition 2.14. *Let the function $\theta : \mathcal{S} \to (-\infty, +\infty]$ be a spectral function in form of (2.24) or (2.25), and $q : \mathcal{R}^m \to (-\infty, +\infty]$ be the underlying absolutely symmetric or symmetric function associated with θ. Let $(X, W) \in \operatorname{gph} \partial\theta$. Then the subdifferential mapping $\partial\theta$ is metrically subregular at X for W if the subdifferential mapping ∂q is metrically subregular at $\sigma(X)$ (or $\lambda(X)$) for $\sigma(W)$ (or $\lambda(W)$).*

By these nice properties of spectral functions, we would like to show the C^2-cone reducibility and the metric subregularity of $\delta_{\mathcal{S}^m_+}(\cdot)$ and $\|\cdot\|_{(k)}$. As we know that the underlying functions $q(\cdot)$ of $\delta_{\mathcal{S}^m_+}(\cdot)$ and $\|\cdot\|_{(k)}$ are $\delta_{\mathcal{R}^m_+}(\cdot)$ and the vector k-norm function, respectively, thus we only need to show the C^2-cone reducibility and the metric subregularity of $\delta_{\mathcal{R}^m_+}(\cdot)$ and the vector k-norm function. Since $\delta_{\mathcal{R}^m_+}(\cdot)$ and the vector k-norm function are polyhedral convex functions and so do their conjugate functions [85, Theorem 19.2], thus they are C^2-cone reducible [8]. Moreover, by [86, 12.31], we have the subgradient mapping $\partial\delta_{\mathcal{R}^m_+}(\cdot) = \mathcal{N}_{\mathcal{R}^m_+}(\cdot)$ of $\delta_{\mathcal{R}^m_+}$ is piecewise polyhedral. And it is well studied in [106] that the Fenchel conjugate function of the vector k-norm function is an indicator over a polyhedral convex set. Similarly, by using [86, 12.31] again, we obtain that the subgradient mapping of the conjugate function of the vector k-norm function is piecewise polyhedral. Combining these facts with Lemma 2.2 and Proposition 2.5-2.6, we have the following results.

Proposition 2.15. *The spectral functions $\delta_{\mathcal{S}^m_+}(\cdot)$, $\|\cdot\|_{(k)}$ and their conjugate functions are C^2-cone reducible. Moreover, each of $\partial\delta_{\mathcal{S}^m_+}$, $(\partial\delta_{\mathcal{S}^m_+})^{-1}$, $\partial\|\cdot\|_{(k)}$ and its inverse mapping possesses the metric subregularity property with the same modulus $\kappa \geq 0$ at every point in its graph set.*

For more discussions on the congugate function and the dual norm of the Ky Fan matrix k-norm $\|\cdot\|_{(k)}$, one can refer to [25, 106].

After these preparations, we are ready to present our main results.

Chapter 3

The metric subregularity of the KKT solution mapping for composite semidefinite programming

In this chapter, we consider (1.1) with a special θ chosen as the indicator function over the positive semidefinite cone $\mathcal{S}_+^n$, i.e., $\theta(\cdot) = \delta_{\mathcal{S}_+^n}(\cdot)$.

We can interpret the problem (1.1) as

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & h(x) = 0, \\ & g(x) \in \mathcal{S}_+^n, \end{array} \tag{3.1}$$

where $f : \mathcal{X} \to \mathcal{R}$ is twice continuously differentiable function, $h : \mathcal{X} \to \mathcal{Y}$ and $g : \mathcal{X} \to \mathcal{S}^n$ are twice continuously differentiable mappings, $\mathcal{X}$ and $\mathcal{Y}$ are finite dimensional real Euclidean spaces.

The form (3.1) is the conventional nonlinear semidefinite programming. The nonlinear SDP problem has various applications in diverse areas, such as pooling and blending problems [36, 101], robustness analysis and robust process design [23, 87], quantitative finance and engineering [42,76,77,107,112], etc. Due to the importance of the semidefinite programming in optimization, the study of stable properties of the corresponding optimality systems is significant. It has been shown [26, 41] that at a locally optimal solution, the KKT solution mapping of the problem (3.1) is robustly isolated calm at origin for a corresponding KKT point of the problem (3.1) if and only if the SRCQ and the SOSC hold at that locally optimal solution. The Proposition 2.10 says that the SRCQ implies the singleton of the multiplier set. Therefore, the isolated calmness property requires the uniqueness of the multipliers. This is a restrictive requirement, since it may fail the uniqueness of the multipliers in practice. Recently, Cui et al. [22] remove the SRCQ by establishing the metric subregularity for the KKT solution mapping of linearly constrained convex SDP problem without multipliers to be unique. Such metric subregularity property also guarantees the fast convergent property of some proper first order methods, such as augmented Lagrangian method, etc. Motivated by these facts, we would like to study the metric subregularity for the nonlinear SDP problem (3.1), which is not a convex problem. Our study in this chapter is just a straightforward extension of [22].

3.1 The sensitivity analysis of SDP cone

Some useful results of SDP cone are adapoted here. These are preparations for our later discussions.

Let $A \in \mathcal{S}^n_+$ and $B \in \mathcal{S}^n$ satisfying $B \in \partial\delta_{\mathcal{S}^n_+}(A)$. Then, noting that $\partial\delta_{\mathcal{S}^n_+}(\cdot) = \mathcal{N}_{\mathcal{S}^n_+}(\cdot)$, it is easy to see that $AB = BA = 0$ and $A = \Pi_{\mathcal{S}^n_+}(A+B)$. Set $X := A + B$. Suppose that X has its eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$ being arranged in a non-increasing order. Denote

$$\alpha := \{i \mid \lambda_i > 0, 1 \leq i \leq n\},\ \beta := \{i \mid \lambda_i = 0, 1 \leq i \leq n\}, \gamma := \{i \mid \lambda_i < 0, 1 \leq i \leq n\}.$$

Then there exists an orthogonal matrix $P \in \mathcal{O}^n$ such that

$$M = P\begin{pmatrix} \Lambda_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Lambda_\gamma \end{pmatrix}P^T,\ A = P\begin{pmatrix} \Lambda_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0_\gamma \end{pmatrix}P^T,\ B = P\begin{pmatrix} 0_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Lambda_\gamma \end{pmatrix}P^T, \tag{3.2}$$

where $\Lambda_\alpha = \mathrm{diag}(\lambda_i \mid i \in \alpha) \succ 0$, $\Lambda_\gamma = \mathrm{diag}(\lambda_j \mid j \in \gamma) \prec 0$. Denote $P = [P_\alpha\ P_\beta\ P_\gamma]$ with $P_\alpha \in \mathcal{R}^{n\times|\alpha|}$, $P_\beta \in \mathcal{R}^{n\times|\beta|}$ and $P_\gamma \in \mathcal{R}^{n\times|\gamma|}$. Then by [2], we can characterize the tangent cones and normal cones as

$$\begin{cases} \mathcal{T}_{\mathcal{S}^n_+}(A) & = \{H \in \mathcal{S}^n \mid [P_\beta\ P_\gamma]^T H [P_\beta\ P_\gamma] \succeq 0\}, \\ \mathcal{T}_{\mathcal{S}^n_-}(B) & = \{H \in \mathcal{S}^n \mid [P_\alpha\ P_\beta]^T H [P_\alpha\ P_\beta] \preceq 0\}, \\ \mathcal{N}_{\mathcal{S}^n_+}(A) & = \{H \in \mathcal{S}^n \mid [P_\beta\ P_\gamma]^T H [P_\beta\ P_\gamma] \preceq 0, P_\alpha^T H P = 0\}, \\ \mathcal{N}_{\mathcal{S}^n_-}(B) & = \{H \in \mathcal{S}^n \mid [P_\alpha\ P_\beta]^T H [P_\alpha\ P_\beta] \succeq 0, P_\gamma^T H P = 0\}. \end{cases} \tag{3.3}$$

By the fact that $\partial\delta_{\mathcal{S}^n_+}(\cdot) = \mathcal{N}_{\mathcal{S}^n_+}(\cdot)$, we also obtain the characterization of $\partial\delta_{\mathcal{S}^n_+}(A)$. Define the critical cone of $\mathcal{S}^n_+$ at A associated with B as

$$\mathcal{C}_{\mathcal{S}^n_+}(A, B) := \mathcal{T}_{\mathcal{S}^n_+}(A) \cap B^\perp = \{H \in \mathcal{S}^n \mid P_\gamma^T H [P_\beta\ P_\gamma] = 0, P_\beta^T H P_\beta \succeq 0\},$$

and the critical cone of $\mathcal{S}^n_+$ at B associated with A as

$$\mathcal{C}_{\mathcal{S}^n_-}(B, A) := \mathcal{T}_{\mathcal{S}^n_-}(B) \cap A^\perp = \{H \in \mathcal{S}^n \mid P_\alpha^T H [P_\alpha\ P_\beta] = 0, P_\beta^T H P_\beta \preceq 0\}.$$

From (3.3), it is easy to see the following proposition.

Proposition 3.1. *Let $A \in \mathcal{S}^n_+$ and $B \in \partial\delta_{\mathcal{S}^n_+}(A)$. Suppose that A and B have the eigenvalue decompositions as in (3.2). Then it holds that:*

(i) $\mathcal{N}_{\mathcal{S}^n_+}(A)$ is a polyhedral set if and only if $|\alpha| \geq n-1$;

(ii) $B \in \mathrm{ri}\,(\mathcal{N}_{\mathcal{S}^n_+}(A))$ if and only if $|\beta| = 0$, i.e., $\mathrm{rank}(A) + \mathrm{rank}(B) = n$.

In Cui's PhD thesis [19, Section 2.5.2], she has proved the metric subregularity of $\partial\delta_{\mathcal{S}^n_+}(\cdot)$ and $\partial\delta_{\mathcal{S}^n_-}(\cdot)$ as follows, which play an important role in the later discussions of the metric subregularity for the problem (3.1). Moreover, these results can be covered by Proposition 2.15.

Proposition 3.2. *Let $A \in \mathcal{S}^n_+$ and $B \in \partial\delta_{\mathcal{S}^n_+}(A)$. Then the subgradient mapping $\partial\delta_{\mathcal{S}^n_+}(\cdot)$ is metrically subregular at A for B and the subgradient mapping $\partial\delta_{\mathcal{S}^n_-}(\cdot)$ is metrically subregular at B for A.*

Another useful result we need for our subsequent discussions is the following perturbation property on the SDP cone (e.g., [22, Proposition 3.4]).

Proposition 3.3. *Let $A \in \mathcal{S}^n_+$ and $B \in \partial\delta_{\mathcal{S}^n_+}(A)$. Suppose that A and B have the eigenvalue decompositions as in (3.2). Then for all $(A', B') \in \mathcal{S}^n \times \mathcal{S}^n$ satisfying $B' \in \partial\delta_{\mathcal{S}^n_+}(A')$ and is sufficiently close to $(A, B) \in \mathcal{S}^n \times \mathcal{S}^n$, we have*

$$\begin{cases} \widetilde{A}'_{\alpha\alpha} = \Lambda_\alpha + O(\|\Delta A\|), \ \widetilde{A}'_{\alpha\beta} = O(\|\Delta A\|), \ \widetilde{A}'_{\alpha\gamma} = O(\min\{\|\Delta A\|, \|\Delta B\|\}), \\ \widetilde{A}'_{\beta\beta} = O(\|\Delta A\|) \in \mathcal{S}^{|\beta|}_+, \ \widetilde{A}'_{\beta\gamma} = O(\|\Delta A\|\|\Delta B\|), \ \widetilde{A}'_{\gamma\gamma} = O(\|\Delta A\|\|\Delta B\|), \\ \widetilde{B}'_{\alpha\alpha} = O(\|\Delta A\|\|\Delta B\|), \ \widetilde{B}'_{\alpha\beta} = O(\|\Delta A\|\|\Delta B\|), \ \widetilde{B}'_{\alpha\gamma} = O(\min\{\|\Delta A\|, \|\Delta B\|\}), \\ \widetilde{B}'_{\beta\beta} = O(\|\Delta B\|) \in \mathcal{S}^{|\beta|}_-, \ \widetilde{B}'_{\beta\gamma} = O(\|\Delta B\|), \ \widetilde{B}'_{\alpha\alpha} = \Lambda_\gamma + O(\|\Delta B\|), \end{cases} \tag{3.4}$$

$$\widetilde{B}'_{\alpha\gamma} + \Lambda_\alpha^{-1}\widetilde{A}'_{\alpha\gamma}\Lambda_\gamma = O(\|\Delta A\|\|\Delta B\|), \tag{3.5}$$

$$\langle \widetilde{A}'_{\beta\beta}, \widetilde{B}'_{\beta\beta} \rangle = \begin{cases} O(\|\Delta A\|\|\Delta B\|)(\|\Delta A\| + \|\Delta B\|) & \text{if } |\alpha| > 0, \\ O(\|\Delta A\|\|\Delta B\|^2) & \text{if } |\alpha| = 0, \end{cases} \tag{3.6}$$

where $\Delta A := A' - A$, $\Delta B := B' - B$, $\widetilde{A}' := P^T A' P$ and $\widetilde{B}' := P^T B' P$.

This perturbation property contains the second order information of $\delta_{\mathcal{S}^n_+}$. One should pay attention to (3.5), which is closely related to the "sigma term" in the SOSC.

3.2 The metric subregularity of the solution mapping for composite SDP problem

In order to conduct our main discussions, we need the following perturbation analysis and notations. Our following perturbation analysis is conducted under the canonically perturbed structure (1.3). The first order optimality conditions in Section 2.3.3 can be adopted here.

For any $(x, y, S) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{S}^n$, the Lagrangian function l associated with problem (3.1) is defined as

$$l(x, y, S) := f(x) + \langle y, h(x) \rangle + \langle S, g(x) \rangle - \delta_{\mathcal{S}^n_-}(S). \tag{3.7}$$

Define the multi-valued mapping $\mathcal{T}_l : \mathcal{X} \times \mathcal{Y} \times \mathcal{S}^n \rightrightarrows \mathcal{X} \times \mathcal{Y} \times \mathcal{S}^n$ associated with the Lagrangian function l at any $(x, y, S) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{S}^n$ by

$$\mathcal{T}_l(x, y, S) = \{(u, v, C) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{S}^n \mid (u, -v, -C) \in \partial l(x, y, S)\}. \tag{3.8}$$

Suppose that the optimal solution set of the problem (3.1) is nonempty and consider an optimal solution $\bar{x} \in \mathcal{X}$ of the problem (3.1). Then, $(\bar{y}, \bar{S}) \in \mathcal{Y} \times \mathcal{S}^n$ is a Lagrangian multiplier corresponding to $\bar{x}$ if and only if $(\bar{x}, \bar{y}, \bar{S})$ satisfies the following KKT system:

$$\begin{cases} \nabla f(\bar{x}) + \nabla h(\bar{x})\bar{y} + \nabla g(\bar{x})\bar{S} = 0, \\ h(\bar{x}) = 0, \\ \bar{S} \in \mathcal{N}_{\mathcal{S}^n_+}(g(\bar{x})). \end{cases} \tag{3.9}$$

Denote $\mathcal{M}(\bar{x})$ as the set of all Lagrangian multipliers corresponding to $\bar{x}$. By our arguments in Section 2.3.3, one should note that the KKT system (3.9) is also the first order optimality conditions of the problem (1.1) if the reduced RCQ holds at $\bar{x}$, i.e.,

$$0 \in \operatorname{int}\{g(\bar{x}) + g'(\bar{x})\mathcal{X} - \mathcal{S}^n_+\}, \tag{3.10}$$

or euivalently,

$$g'(\bar{x})\mathcal{X} + \mathcal{T}_{\mathcal{S}^n_+}(g(\bar{x})) = \mathcal{S}^n. \tag{3.11}$$

By the third inclusion of (3.9), we assume that $g(\bar{x})$ and $\bar{S}$ have the eigenvalue decompositions as in (3.2) with $A = g(\bar{x})$ and $B = \bar{S}$.

For a perturbed point $(u, v, C) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{S}^n$, it is easy to check that $(x, y, S) \in \mathcal{T}_l^{-1}(u, v, C)$ can be equivalently interpreted as the following perturbed KKT system:

$$\begin{cases} \nabla f(x) + \nabla h(x)y + \nabla g(x)S = u, \\ h(x) + v = 0, \\ S \in \mathcal{N}_{\mathcal{S}^n_+}(g(x) + C). \end{cases} \tag{3.12}$$

One can find that $\mathcal{T}_l^{-1}(0,0,0)$ is the set of all the KKT points $(\bar{x},\bar{y},\bar{S})$ of the problem (3.1) satisfying (3.9).

Next, we conduct our discussions about the metric subregularity of $\mathcal{T}_l$ at a KKT point for the origin.

Let $(\bar{x},\bar{y},\bar{S}) \in \mathcal{T}_l^{-1}(0,0,0)$. We adopt the critical cone of the problem (3.1) at $\bar{x}$ from (2.23) as

$$\mathcal{C}(\bar{x}) := \{d \in \mathcal{X} \mid h'(\bar{x})d = 0,\ g'(\bar{x})d \in \mathcal{C}_{\mathcal{S}^n_+}(g(\bar{x}),\bar{S})\}.$$

By Proposition 2.11, one should note that such critical cone is also the critical cone of the problem (1.1) with $\theta = \delta_{\mathcal{S}^n_+}$ at $\bar{x}$ if the reduced RCQ (3.10) or (3.11) holds.

Here, we define a more restrictive second-order sufficient condition for the problem (3.1) at $\bar{x}$ with respect to the multiplier $(\bar{y},\bar{S}) \in \mathcal{M}(\bar{x})$ if

$$\langle d, \nabla^2_{xx} l(\bar{x},\bar{y},\bar{S})d\rangle + 2\langle \bar{S}, g'(\bar{x})d[g(\bar{x})]^{\dagger} g'(\bar{x})d\rangle > 0,\ \forall\ 0 \neq d \in \mathcal{C}(\bar{x}). \tag{3.13}$$

The conventionally used second-order sufficient condition is taking the supreme of the left hand side of (3.13) over the multiplier set $\mathcal{M}(\bar{x})$. Here, we only impose the condition at one fixed multiplier $(\bar{y},\bar{S})$ rather than the whole set. Moreover, the second term in the left hand side is the so called sigma term. And this sigma term can be expressed as the conjugate of the lower (parabolic) second order directional epiderivative of $\delta_{\mathcal{S}^n_+}$, since $\delta_{\mathcal{S}^n_+}$ is C^2-cone reducible.

For the convenience of the later discussions, we define the following joint 'critical cone' associated with the problem (3.1) as

$$\widetilde{\mathcal{C}}(\bar{x},\bar{y},\bar{S}) := \left\{ \begin{array}{c|l} & h'(\bar{x})d_x = 0, \\ (d_x,d_y,d_S) \in & g'(\bar{x})d_x \in \mathcal{C}_{\mathcal{S}^n_+}(g(\bar{x}),\bar{S}), \\ \mathcal{X}\times\mathcal{Y}\times\mathcal{S}^n & d_S \in \mathcal{C}_{\mathcal{S}^n_-}(\bar{S},g(\bar{x})), \\ & (g(\bar{x})^{1/2})^{\dagger}(g'(\bar{x})d_x)\bar{S}^{1/2} + (g(\bar{x})^{1/2}d_S(\bar{S}^{1/2})^{\dagger} = 0 \end{array} \right\}. \tag{3.14}$$

Now, we are ready to present our main result for the nonlinear SDP problem (3.1).

Theorem 3.1. *Let $\bar{x}$ be an optimal solution to the problem (3.1) and $(\bar{y},\bar{S}) \in \mathcal{M}(\bar{x})$ be a Lagrangian multiplier corresponding to $\bar{x}$. Denote $\Phi := \big(\mathcal{C}_{\mathcal{S}^n_+}(g(\bar{x}),\bar{S})\big)^{\circ}$. Let the following assumptions be satisfied:*

(i) *The set* $g'(\bar{x})^T\Phi$ *is closed.*

(ii) $\langle \Pi_\Phi(-g'(\bar{x})d_x),\ \Pi_\Phi(d_S)\rangle = 0$ *for all* $(d_x, d_y, d_S) \in \widetilde{\mathcal{C}}(\bar{x}, \bar{y}, \bar{S})$*, where* $\Pi_\Phi(\cdot)$ *denotes the projection onto the set* Φ *and the set* $\widetilde{\mathcal{C}}(\bar{x}, \bar{y}, \bar{S})$ *is defined as (3.14).*

(iii) *The second-order sufficient condition (3.13) holds at* $\bar{x}$ *with respect to the multiplier* $(\bar{y}, \bar{S}) \in \mathcal{M}(\bar{x})$ *for the problem (3.1).*

Then there exist a constant $\kappa > 0$ *and a neighborhood* $\mathcal{U}$ *of* $(\bar{x}, \bar{y}, \bar{S})$ *such that for any* $(u, v, C) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{S}^n$*,*

$$\|x - \bar{x}\| \leq \kappa\|(u, v, C)\|,\ \forall (x, y, S) \in \mathcal{T}_l^{-1}(u, v, C) \cap \mathcal{U}. \tag{3.15}$$

Moreover, if there exists $(\widehat{y}, \widehat{S}) \in \mathcal{M}(\bar{x})$ *such that* $rank(g(\bar{x})) + rank(\widehat{S}) = n$*, then* $\mathcal{T}_l$ *is metrically subregular at* $(\bar{x}, \bar{y}, \bar{S})$ *for the origin.*

Proof. Firstly, we show that under the assumptions (i)-(iii), there exist a constant $\kappa > 0$ and a neighborhood $\mathcal{U}$ of $(\bar{x}, \bar{y}, \bar{S})$ such that (3.15) holds. We seek a contradiction to settle this.

Suppose that (3.15) does not hold. It means that there exist some sequences $\{(u^k, v^k, C^k)\}_{k\geq 0} \subset \mathcal{X} \times \mathcal{Y} \times \mathcal{S}^n$ and $\{(x^k, y^k, S^k)\}_{k\geq 0} \subset \mathcal{X} \times \mathcal{Y} \times \mathcal{S}^n$ such that $(u^k, v^k, C^k) \to 0,\ (x^k, y^k, S^k) \to (\bar{x}, \bar{y}, \bar{S})$ with every $(x^k, y^k, S^k) \in \mathcal{T}_l^{-1}(u^k, v^k, C^k)$, and

$$\|x^k - \bar{x}\| \geq \delta_k \|(u^k, v^k, C^k)\|$$

with some $0 < \delta_k$ such that $\delta_k \to \infty$. Denote $t_k := \|x^k - \bar{x}\|$, by taking a subsequence if necessary, we can assume that $(x^k - \bar{x})/t_k \to d_{\bar{x}} \in \mathcal{X}$ with $\|d_{\bar{x}}\| = 1$.

From the perturbed KKT system (3.12), we can have that for all $k \geq 0$ large enough,

$$\begin{aligned} 0 &= h(x^k) + v^k - h(\bar{x}) \\ &= h'(\bar{x})(x^k - \bar{x}) + o(t_k) + v^k. \end{aligned} \tag{3.16}$$

Dividing by t_k on both sides of (3.16) and taking limits $k \to \infty$, we get

$$h'(\bar{x})d_{\bar{x}} = 0. \tag{3.17}$$

For simplify notations, we set

$$\Omega := \{W \in \mathcal{S}^n_+ \mid [P_\beta \ P_\gamma]^T W [P_\beta \ P_\gamma] = 0\},$$

$$A := g(\bar{x}), \ B := \bar{S},$$

and for all $k \geq 0$,

$$\begin{cases} A^k := g(x^k) + C^k, \ \widetilde{A}^k := P^T A^k P, \ B^k := S^k, \ \widetilde{B}^k := P^T B^k P, \\ H^k := \Pi_\Omega\big((B^k - B)/t_k\big), \ G^k := (B^k - B)/t_k - H^k \in \Phi. \end{cases} \tag{3.18}$$

Thus $A^k \to A$ and $B^k \to B$ by the assumptions. Moreover, similar to (3.16), we can derive that

$$\frac{1}{t_k}(A^k - A) \to g'(\bar{x})d_{\bar{x}} \text{ as } k \to \infty. \tag{3.19}$$

Since $B \in \partial\delta_{\mathcal{S}^n_+}(A)$ and $B^k \in \partial\delta_{\mathcal{S}^n_+}(A^k)$, we can derive the following estimates by Proposition 3.3 that for all (A^k, B^k) sufficiently close to (A, B),

$$\begin{cases} \widetilde{A}^k_{\beta\gamma} := O(\|A^k - A\|\|B^k - B\|), \ \widetilde{A}^k_{\gamma\gamma} := O(\|A^k - A\|\|B^k - B\|), \\ \widetilde{B}^k_{\alpha\alpha} := O(\|A^k - A\|\|B^k - B\|), \ \widetilde{B}^k_{\alpha\beta} := O(\|A^k - A\|\|B^k - B\|), \\ \widetilde{B}^k_{\alpha\gamma} := -\Lambda_\alpha^{-1}\widetilde{A}^k_{\alpha\gamma}\Lambda_\gamma + O(\|A^k - A\|\|B^k - B\|), \\ \widetilde{A}^k_{\beta\beta} \in \mathcal{S}^{|\beta|}_+, \ \widetilde{B}^k_{\beta\beta} \in \mathcal{S}^{|\beta|}_-. \end{cases} \tag{3.20}$$

Combining the above (3.19) and (3.20), we obtain

$$\begin{cases} g'(\bar{x})d_{\bar{x}} \in \mathcal{C}_{\mathcal{S}^n_+}(g(\bar{x}), \bar{S}), \\ H_1 := \lim\limits_{k\to\infty} H^k = P \begin{pmatrix} 0 & 0 & -\Lambda_\alpha^{-1}(\widetilde{D}_{\bar{x}})_{\alpha\gamma}\Lambda_\gamma \\ 0 & 0 & 0 \\ \big(-\Lambda_\alpha^{-1}(\widetilde{D}_{\bar{x}})_{\alpha\gamma}\Lambda_\gamma\big)^T & 0 & 0 \end{pmatrix} P^T, \end{cases} \tag{3.21}$$

where $\widetilde{D}_{\bar{x}} := P^T g'(\bar{x})d_{\bar{x}} P$.

Again, by the perturbed KKT system (3.12), we can deduce that for $k \geq 0$ large enough,

$$\begin{aligned} u^k &= \nabla f(x^k) + \nabla h(x^k)y^k + \nabla g(x^k)S^k - \big(\nabla f(\bar{x}) + \nabla h(\bar{x})\bar{y} + \nabla g(\bar{x})\bar{S}\big) \\ &= \nabla^2_{xx} f(\bar{x})(x^k - \bar{x}) + \langle y^k, h''(\bar{x})(x^k - \bar{x})\rangle + \langle S^k, g''(\bar{x})(x^k - \bar{x})\rangle \\ &\quad + \nabla h(\bar{x})(y^k - \bar{y}) + \nabla g(\bar{x})(S^k - \bar{S}). \end{aligned} \tag{3.22}$$

Dividing by t_k on both side of (3.22), it gives

$$\frac{u^k}{t_k} - \nabla^2_{xx} f(\bar{x})\frac{(x^k-\bar{x})}{t_k} - \langle y^k, h''(\bar{x})\frac{(x^k-\bar{x})}{t_k}\rangle - \langle S^k, g''(\bar{x})\frac{(x^k-\bar{x})}{t_k}\rangle - \nabla g(\bar{x})H^k$$

$$= \nabla h(\bar{x})\frac{(y^k-\bar{y})}{t_k} + \nabla g(\bar{x})G^k \in \mathrm{Im}\nabla h(\bar{x}) + \nabla g(\bar{x})\Phi,$$

where the set in the right hand side, as a sum of a linear subspace and a closed set, is closed, since $g'(\bar{x})^T\Phi$ is supposed to be closed. Then by taking limit as $k \to \infty$, it yeilds

$$-\nabla^2_{xx} l(\bar{x},\bar{y},\bar{S})d_{\bar{x}} - \nabla g(\bar{x})H_1 \in \mathrm{Im}\nabla h(\bar{x}) + \nabla g(\bar{x})\Phi. \tag{3.23}$$

The inclusion (3.23) means that there exists $(d_{\bar{y}}, H_2) \in \mathcal{Y} \times \Phi$ such that

$$\nabla^2_{xx} l(\bar{x},\bar{y},\bar{S})d_{\bar{x}} + \nabla g(\bar{x})H_1 + \nabla h(\bar{x})d_{\bar{y}} + \nabla g(\bar{x})H_2 = 0. \tag{3.24}$$

Let $d_{\bar{S}} := H_1 + H_2$ and $\widetilde{d}_{\bar{S}} := P^T d_{\bar{S}} P$. Then combining (3.17) and (3.21), we have $(d_{\bar{x}}, d_{\bar{y}}, d_{\bar{S}}) \in \widetilde{\mathcal{C}}(\bar{x},\bar{y},\bar{S})$. This further indicates that $0 \neq d_{\bar{x}} \in \mathcal{C}(\bar{x})$ of problem (3.1). Therefore, by making use of the assumption (ii), we have

$$\begin{aligned} & \langle d_{\bar{x}}, \nabla^2_{xx} l(\bar{x},\bar{y},\bar{S})d_{\bar{x}}\rangle + 2\langle \bar{S}, g'(\bar{x})d_{\bar{x}}[g(\bar{x})]^\dagger g'(\bar{x})d_{\bar{x}}\rangle \\ = \;& -\langle d_{\bar{y}}, h'(\bar{x})d_{\bar{x}}\rangle - \langle d_{\bar{S}}, g'(\bar{x})d_{\bar{x}}\rangle + 2\langle \bar{S}, g'(\bar{x})d_{\bar{x}}[g(\bar{x})]^\dagger g'(\bar{x})d_{\bar{x}}\rangle \\ = \;& -\langle (\widetilde{D}_{\bar{x}})_{\beta\beta}, (\widetilde{d}_{\bar{S}})_{\beta\beta}\rangle = \langle \Pi_\Phi(-g'(\bar{x})d_{\bar{x}}), \Pi_\Phi(d_{\bar{S}})\rangle = 0, \end{aligned}$$

which contradicts the assumption (iii) that the second-order sufficient condition (3.13) holds at $\bar{x}$ with respect to the multiplier $(\bar{y},\bar{S}) \in \mathcal{M}(\bar{x})$. Hence, there exist a constant $\kappa > 0$ and a neighborhood $\mathcal{U}$ of $(\bar{x},\bar{y},\bar{S})$ such that (3.15) holds.

Next, we prove that $\mathcal{T}_l$ is metrically subregular at $(\bar{x},\bar{y},\bar{S})$ for the origin under an additional assumption, which requires that there exist $(\widehat{y},\widehat{S}) \in \mathcal{M}(\bar{x})$ such that $\mathrm{rank}(g(\bar{x})) + \mathrm{rank}(\widehat{S}) = n$. In another word, it is equivalent to show that there exist a constant $\kappa' > 0$ and a neighborhood $\mathcal{U}'$ of $(\bar{x},\bar{y},\bar{S})$ such that for any $(u,v,C) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{S}^n$,

$$\mathrm{dist}((x,y,S), \mathcal{T}_l^{-1}(0)) \leq \kappa' \|((u,v,C)\|, \;\forall (x,y,S) \in \mathcal{T}_l^{-1}((u,v,C) \cap \mathcal{U}'. \tag{3.25}$$

For the convenience, we set

$$\Psi := \{(y,S)|\ (\bar{x},y,S) \in \mathcal{T}_l^{-1}(0,0,0)\},$$

$$\Xi_1 := \{(y,S)|\ \nabla f(\bar{x}) + \nabla h(\bar{x})y + \nabla g(\bar{x})S = 0\},\ \Xi_2 := \{(y,S)|\ S \in \mathcal{N}_{\mathcal{S}^n_+}(g(\bar{x}))\}.$$

One can easily find that $\Psi = \Xi_1 \cap \Xi_2$ and $(\widehat{y}, \widehat{S}) \in \Xi_1 \cap \mathrm{ri}(\Xi_2)$. Thus, by Proposition 2.1, we have that there exists a constant $\kappa_1 > 0$ such that for any $(x,y,S) \in \mathcal{U}'$,

$$\mathrm{dist}((y,S),\Phi) \le \kappa_1\big(\mathrm{dist}((y,S),\Xi_1) + \mathrm{dist}((y,S),\Xi_2)\big). \tag{3.26}$$

For any given point $(x,y,S) \in \mathcal{T}_l^{-1}((u,v,C) \cap \mathcal{U}'$, we assume that $\|(y,S)\| \le \eta$ with some $\eta > 0$ by shrinking $\mathcal{U}'$ if necessary. Fixing that given point, using Hoffman's error bound (e.g., [43] and [33, Lemma 3.2.2])and the twice continuous differentiability of f, h and g, shrinking $\mathcal{U}'$ if necessary, we obtain that there exist constants $\kappa_2 > 0$ and $\kappa_2' > 0$ such that

$$\begin{aligned} \mathrm{dist}((y,S),\Xi_1) &\le \kappa_2\|\nabla f(\bar{x}) + \nabla h(\bar{x})y + \nabla g(\bar{x})S\| \\ &\le \kappa_2(\|\nabla f(x) - \nabla f(\bar{x})\| + \|\nabla h(x) - \nabla h(\bar{x})\|\|y\| \\ &\quad + \|\nabla g(x) - \nabla g(\bar{x})\|\|S\| + \|u\|) \\ &\le \kappa_2'(\|x-\bar{x}\| + \|u\|). \end{aligned} \tag{3.27}$$

By Proposition 3.2, we have $\partial\delta_{\mathcal{S}^n_-}(\cdot) = \mathcal{N}_{\mathcal{S}^n_-}(\cdot)$ is metrically subregular at $\bar{S}$ for $g(\bar{x})$. Together with $g(x) + C \in \mathcal{N}_{\mathcal{S}^n_-}(S)$ and the twice continuous differentiabilty of g , we can duduce, shrinking $\mathcal{U}'$ if necessary, that there exist contants $\kappa_3 > 0$ and $\kappa_3' > 0$ such that

$$\begin{aligned} \mathrm{dist}((y,S),\Xi_2) &= \mathrm{dist}(S, \mathcal{N}_{\mathcal{S}^n_+}(g(\bar{x}))) \\ &\le \kappa_3 \mathrm{dist}(g(\bar{x}), \mathcal{N}_{\mathcal{S}^n_-}(S)) \\ &\le \kappa_3\|g(x) + C - g(\bar{x})\| \\ &\le \kappa_3'(\|x-\bar{x}\| + \|C\|). \end{aligned} \tag{3.28}$$

Therefore, we can find that there exist a constant $\kappa' > 0$ and a neighborhood $\mathcal{U}'$ of $(\bar{x},\bar{y},\bar{S})$ such that (3.25) holds, by using the inequalities (3.15) and (3.26)-(3.28). This implies $\mathcal{T}_l$ is metrically subregular at $(\bar{x},\bar{y},\bar{S})$ for the origin. □

Remark 3.1. *Our proof here follows the pattern of [22, Theorem 3.2] for linearly constrained convex SDP problem with* $C^{1,1}$ *(or* LC^1*, the class of all differentiable functions having a locally Lipschitzian derivative) program data. We also adpot some ideas from [46, 49] for the NLP problem with* $C^{1,1}$ *program data. Therefore, our assumptions of* C^2 *program data can be relaxed to* $C^{1,1}$ *program data in above discussions of the metric subregularity. In this situation, if the gradients* ∇f*,* ∇g *and* ∇h *are directionally differentiable at* $\bar{x}$*, then we only need to change the SOSC in the assumption (iii) of Theorem 3.1 to the following form [22]:*

$$\langle d, (\nabla_x l)'(\bar{x}, \bar{y}, \bar{S}; d)\rangle + 2\langle \bar{S}, g'(\bar{x})d[g(\bar{x})]^{\dagger} g'(\bar{x})d\rangle > 0, \ \forall\, 0 \neq d \in \mathcal{C}(\bar{x}), \tag{3.29}$$

where $(\nabla_x l)'(\bar{x}, \bar{y}, \bar{S}; d)$ *denotes the directional derivative of* $\nabla_x l(\cdot, \bar{y}, \bar{S})$ *at* $\bar{x}$ *in the direction* d*.*

Remark 3.2. *If* Φ *is a polyhedral cone, by [72, Theorem 1.1], the closedness is always satisfied for* Φ *under linear transformations. Thus, if* $|\beta| = 0$ *or* $|\beta| = 1$*, assumption (i) can be omitted here. Moreover, one should note in the proof that* $\langle \Pi_\Phi(-g'(\bar{x})d_{\bar{x}}), \Pi_\Phi(d_{\bar{S}})\rangle = -\langle (\widetilde{D}_{\bar{x}})_{\beta\beta}, (\widetilde{d_{\bar{S}}})_{\beta\beta}\rangle$*. Therefore, if* $\beta = \emptyset$*, assumption (ii) can be omitted here. Additionally, if* $|\alpha| \geq n-1$*, the partial strict complementarity condition can also be omitted, that is, there is no need to exist a* $(\widehat{y}, \widehat{S}) \in \mathcal{M}(\bar{x})$ *such that* $\operatorname{rank}(g(\bar{x})) + \operatorname{rank}(\widehat{S}) = n$*.*

In Theorem 3.1, we obtain the metric subregularity for the KKT solution mapping of the nonlinear SDP problem under rather weak conditions. This result can cover the convex case. One can see that the perturbation property of the symmetric cone $\mathcal{S}^n_+$ is crucial for the first part of proof. Due to the symmetric property of SDP cone, of which the perturbation property is not that complicated. Thus, we want to move forward without the symmetric property. That is what we shall study later.

Chapter 4

The metric subregularity of the KKT solution mapping for composite nuclear norm problem

We move to non-symmetric matrix analysis in this chapter. One particular and useful case is the nuclear norm regularized problem.

Let us consider (1.1) with a special θ chosen as the nuclear norm function on $\mathcal{R}^{m\times n}$ $(m \le n)$. We can restate the problem (1.1) as

$$\begin{aligned} \min \quad & f(x) + \|g(x)\|_* \\ \text{s.t.} \quad & h(x) = 0, \end{aligned} \tag{4.1}$$

where $f : \mathcal{X} \to \mathcal{R}$ is twice continuously differentiable function, $h : \mathcal{X} \to \mathcal{Y}$ and $g : \mathcal{X} \to \mathcal{R}^{m\times n}$ are twice continuously differentiable mappings, $\mathcal{X}$ and $\mathcal{Y}$ are finite dimensional real Euclidean spaces, and $\theta : \mathcal{R}^{m\times n} \to \mathcal{R}$ denotes the nuclear norm function with $\theta(X) = \|X\|_*$ for all $X \in \mathcal{R}^{m\times n}$.

This nuclear norm regularized problem (4.1) arises in various applications such as the matrix norm approximation, low-rank problems and so on [18, 27, 39, 45, 51, 79]. We want to extend our result of the nonlinear SDP problem to the problem (4.1) here. Similar to the SDP case, the robust isolated calmness can be equivalently characterized by the SRCQ and the SOSC at a locally optimal solution [21, 26, 56]. Due to the restrictive requirement of Lagrange multipliers to be unique, many models in practice cannot possess the isolated calmness. To settle down this issue, we study the metric subregularity for the problem (4.1). And this study can be a useful tool for convergence analysis of various methods.

4.1 The sensitivity analysis of the nuclear norm

By using the preliminary results in the Section 2.3.1, it is very natural to obtain some useful properties of the nuclear norm. We list some useful results of the nuclear norm as a preparation of the main result. A perturbation property of the nuclear norm showed at the end of this section.

Let $A, B \in \mathcal{R}^{m\times n}$ satisfying $B \in \partial\theta(A)$ and denote $M := A + B$. A well known equivalent form [67] is given by

$$A = \mathrm{Prox}_{\theta}(M), \quad B = \mathrm{Prox}_{\theta^*}(M). \tag{4.2}$$

Suppose that M admits the following singular-value decomposition (SVD):

$$M = U[\Sigma(M)\ 0]V^T = U[\Sigma(M)\ 0][V_1\ V_2]^T = U\Sigma(M)V_1^T, \tag{4.3}$$

where $U \in \mathcal{O}^m$, $V := [V_1\ V_2] \in \mathcal{O}^n$ with $V_1 \in \mathcal{R}^{n\times m}$ and $V_2 \in \mathcal{R}^{n\times(n-m)}$ are singular vectors of M, and $\Sigma(M) := \mathrm{Diag}(\sigma_1(M), \sigma_2(CM), \ldots, \sigma_m(M))$ is the diagonal matrix of the singular values of M with $\sigma_1(M) \geq \sigma_2(M) \geq \ldots \geq \sigma_m(M) \geq 0$ being arranged in a non-increasing order.

For simplicity, we let $\sigma(M) := (\sigma_1(M), \sigma_2(M), \ldots, \sigma_m(M))$, thus $\Sigma(M) = \mathrm{diag}(\sigma(M))$.

Given the SVD of M as (4.3), it is follows [25] that A and B admit the following SVD:

$$\begin{aligned} A &= U[\Sigma(A)\ 0]V^T = U\Sigma(A)V_1^T, \\ B &= U[\Sigma(B)\ 0]V^T = U\Sigma(B)V_1^T, \end{aligned} \tag{4.4}$$

where $\Sigma(A) := \mathrm{Diag}(\sigma_1(A), \sigma_2(A), \ldots, \sigma_m(A))$, $\Sigma(B) := \mathrm{Diag}(\sigma_1(B), \sigma_2(B), \ldots, \sigma_m(B))$

and

$$\sigma_i(A) = (\sigma_i(M) - 1)_+, \quad \sigma_i(B) = \sigma_i(M) - \sigma_i(A), \quad i = 1, 2, \ldots, m. \tag{4.5}$$

Obviously, $\sigma_1(A) \geq \sigma_2(A) \geq \ldots \geq \sigma_m(A) \geq 0$ and $\sigma_1(B) \geq \sigma_2(B) \geq \ldots \geq \sigma_m(B) \geq 0$. Similarly, we let $\sigma(A) := (\sigma_1(A), \sigma_2(A), \ldots, \sigma_m(A))$ and $\sigma(B) := (\sigma_1(B), \sigma_2(B), \ldots, \sigma_m(B))$.

For simplicity of the subsequent discussions, we define the following three index sets:

$$\alpha := \{1 \leq i \leq m : \sigma_i(A) > 0\}, \ \beta := \{1 \leq i \leq m : \sigma_i(A) = 0\}, \ c := \{m+1, \ldots, n\}. \tag{4.6}$$

Furthermore, let $\nu_1(A) > \nu_2(A) > \ldots > \nu_{r_0}(A) > 0$ with some nonnegative integer r_0 be the distinct nonzero singular values of A. Hence, we can divide α regarding to the distinct nonzero singular values as

$$\alpha = \bigcup_{1 \leq l \leq r_0} a_l, \quad a_l := \{i \in \alpha : \sigma_i(A) = \nu_l(A)\}, \quad l = 1, 2, \ldots, r_0. \tag{4.7}$$

By the relationship (4.5), we can see that $\sigma_\alpha(B) = e_\alpha$ and $0 \leq \sigma_i(B) \leq 1$ for $i \in \beta$. To divide the set β, we define the following three subsets:

$$\beta_1 := \{i \in \beta : \sigma_i(B) = 1\}, \quad \beta_2 := \{i \in \beta : 0 < \sigma_i(B) < 1\}, \quad \beta_3 := \{i \in \beta : \sigma_i(B) = 0\}. \tag{4.8}$$

Actually, we can also interpret (4.6), (4.7) and (4.8) as a classification about the singular values of M. By the relationship (4.5), it is easy to obtain that

$$\begin{aligned}
&\alpha = \{1 \leq i \leq m : \sigma_i(M) > 1\}, \quad \beta = \{1 \leq i \leq m : 0 \leq \sigma_i(M) \leq 1\}, \\
&a_l = \{i \in \alpha : \sigma_i(M) = \nu_l(M)\}, \quad l = 1, 2, \ldots, r_0, \\
&\beta_1 := \{i \in \beta : \sigma_i(M) = 1\}, \quad \beta_2 := \{i \in \beta : 0 < \sigma_i(M) < 1\}, \quad \beta_3 := \{i \in \beta : \sigma_i(M) = 0\},
\end{aligned} \tag{4.9}$$

where $\nu_1(M) > \nu_2(M) > \ldots > \nu_{r_0}(M) > 1$ denotes the distinct singular values of M that are larger than 1.

In fact, the above relationships among the sigular values of A, B and M can be obtained by the following lemma, which is a special case of the characterization in [69, 102].

Lemma 4.1. *Suppose $\sigma(A)$ and $\sigma(B)$ are singular values of A and B respectively. Then $B \in \partial\theta(A)$ if and only if $\sigma(A)$ and $\sigma(B)$ satisfy the following conditions:*

$$\sigma_\alpha(B) = e_\alpha, \quad 0 \le \sigma_\beta(B) \le e_\beta \quad \text{and} \quad \sum_{i\in\beta}\sigma_i(B) \le m - |\alpha|, \tag{4.10}$$

where α and β are defined as (4.6).

Based on this lemma, it is easy to find the following observations.

Proposition 4.1. *Let $A \in \mathcal{R}^{m\times n}$ and $B \in \partial\theta(A)$. Suppose that A and B have the SVD as in (4.4), then it holds that*

(a) $\partial\theta(A)$ is a polyhedral set if and only if $\sigma_m(A) > 0$,

(b) $B \in \mathrm{ri}\,(\partial\theta(A))$ if and only if $0 < \sigma_\beta(B) < e_\beta$.

Remark 4.1. *One can find that if $\sigma_m(A) > 0$ in part (a) holds, then θ is differentiable at A [103]. In this case, problem (4.1) turns to a smooth optimization problem.*

Since the nuclear norm is a norm function on $\mathcal{R}^{m\times n}$, then θ is Lipschitz continuous and convex on $\mathrm{dom}\,\theta = \mathcal{R}^{m\times n}$. By Section 2.3.1, we point out that the nuclear norm θ is a regular function [8, Theorem 2.126]. Hence, $\theta^{\downarrow}(X,\cdot) = \theta'(X;\cdot)$ for any $X \in \mathcal{R}^{m\times n}$. Thus, all the results in the Section 2.3.1 regarding to directional epiderivative of θ can be shifted to its conventional directional derivative here. Moreover, it can be obtained from Watson [104] that the subgradient of θ at A has the form:

$$\partial\theta(A) = \left\{U_\alpha V_\alpha^T + U_\beta W[V_\beta \; V_2]^T : W \in \mathcal{R}^{|\beta|\times(n-|\alpha|)}, \, \|W\|_2 \le 1\right\}. \tag{4.11}$$

Therefore, for any $H \in \mathcal{R}^{m\times n}$, the directional derivative of θ at A along H can be explicitly written as

$$\begin{aligned}\theta'(A;H) &= \sup_{S\in\partial\theta(A)} \langle H, S\rangle \\ &= \mathrm{tr}(U_\alpha^T H V_\alpha) + \|U_\beta^T H[V_\beta \; V_2]\|_*.\end{aligned} \tag{4.12}$$

Let us recall the set valued mapping (2.17) at point (A, B) satisfying $B \in \partial\theta(A)$ with $\theta(\cdot) = \|\cdot\|_*$,

$$\mathcal{C}_\theta(A, B) := \{H \in \mathcal{R}^{m\times n} : \theta'(A; H) = \langle H, B\rangle\}. \tag{4.13}$$

Here, we call $\mathcal{C}_\theta(A, B)$ the critical cone of $\partial\theta(A)$ at $A+B$, associated with $B \in \partial\theta(A)$. Then, we can obtain the following characterization of the critical cone $\mathcal{C}_\theta(A, B)$ from [25, proposition 10].

Lemma 4.2. *Suppose $A, B \in \mathcal{R}^{m\times n}$ satisfy $B \in \partial\theta(A)$ and the index sets $\alpha, \beta, \beta_1, \beta_2$, β_3, and c are defined as (4.6) and (4.8). Given any $H \in \mathcal{R}^{m\times n}$, denote $\widetilde{H} = U^T H V$ for U, V satisfying (4.3). Then $H \in \mathcal{C}_\theta(A, B)$ if and only if $\widetilde{H}$ has the following block structure:*

$$\widetilde{H} = \left(\begin{array}{c:c:c:c:c} \widetilde{H}_{\alpha\alpha} & & \widetilde{H}_{\alpha\beta} & & \widetilde{H}_{\alpha c} \\ \hdashline & \Pi_{\mathcal{S}_+^{|\beta_1|}}(\widetilde{H}_{\beta_1\beta_1}) & 0 & 0 & \\ \widetilde{H}_{\beta\alpha} & 0 & 0_{\beta_2\beta_2} & 0 & 0 \\ & 0 & 0 & 0_{\beta_3\beta_3} & \end{array}\right). \tag{4.14}$$

where $\Pi_{\mathcal{S}_+^p}(\cdot)$ denotes the projection onto the $p\times p$ dimensional positive semidefinite cone.

For the convenience of later discussions, define two linear operators $\mathbb{S} : \mathcal{R}^{p\times p} \to \mathcal{S}^p$ and $\mathbb{T} : \mathcal{R}^{p\times p} \to \mathcal{R}^{p\times p}$ for any positive integer p by

$$\mathbb{S}(X) = \frac{1}{2}(X + X^T), \quad \mathbb{T}(X) = \frac{1}{2}(X - X^T), \quad \forall X \in \mathcal{R}^{p\times p}. \tag{4.15}$$

Next, let us consider the Fenchel conjugate function θ^* of θ. By the equivalence of $B \in \partial\theta(A)$ and $A \in \partial\theta^*(B)$, we can similarly define the critical cone $\mathcal{C}_{\theta^*}(B, A)$ of $\partial\theta^*(B)$ at $A+B$ associated with $A \in \partial\theta^*(B)$. One can find a directly derive in [25]. The critical cone $\mathcal{C}_{\theta^*}(B, A)$ is defined as

$$\mathcal{C}_{\theta^*}(B, A) := \{H \in \mathcal{R}^{m\times n} : \vartheta'(B; H) = \langle H, A\rangle = 0\}, \tag{4.16}$$

where $\vartheta(\cdot)$ denotes the dual norm of the nuclear norm θ, i.e., the spectral norm $\|\cdot\|_2$ on $\mathcal{R}^{m\times n}$.

We adopt the following characterization of the critical cone $\mathcal{C}_{\theta^*}(B,A)$ in [25, proposition 12].

Lemma 4.3. *Suppose that all the assumptions in Lemma 4.2 hold here. Then $H \in \mathcal{C}_{\theta^*}(B,A)$ if and only if $\widetilde{H}$ admits the following block structure:*

$$\widetilde{H} = \left(\begin{array}{c:c:c:c:c} \mathbb{T}(\widetilde{H}_{\alpha\alpha}) & \mathbb{T}(\widetilde{H}_{\alpha\beta_1}) & \widetilde{H}_{\alpha\beta_2} & \widetilde{H}_{\alpha\beta_3} & \\ \hdashline \mathbb{T}(\widetilde{H}_{\beta_1\alpha}) & \Pi_{\mathcal{S}_-^{|\beta_1|}}(\mathbb{S}(\widetilde{H}_{\beta_1\beta_1})) + \mathbb{T}(\widetilde{H}_{\beta_1\beta_1}) & \widetilde{H}_{\beta_1\beta_2} & \widetilde{H}_{\beta_1\beta_3} & \widetilde{H}_2 \\ \hdashline \widetilde{H}_{\beta_2\alpha} & \widetilde{H}_{\beta_2\beta_1} & \widetilde{H}_{\beta_2\beta_2} & \widetilde{H}_{\beta_2\beta_3} & \\ \hdashline \widetilde{H}_{\beta_3\alpha} & \widetilde{H}_{\beta_3\beta_1} & \widetilde{H}_{\beta_3\beta_2} & \widetilde{H}_{\beta_3\beta_3} & \end{array} \right), \qquad (4.17)$$

where $\mathbb{S}(\cdot)$ and $\mathbb{T}(\cdot)$ are defined as (4.15), $\Pi_{\mathcal{S}_-^p}(\cdot)$ denotes the projection onto the $p \times p$ dimensional negative semidefinite cone and $\widetilde{H} = [\widetilde{H}_1 \ \widetilde{H}_2] = [U^T H V_1 \ U^T H V_2]$.

Similar to the SDP case, we also need the perturbation property of the nuclear norm for our subsequent discussions about the metric subregularity for the problem (4.1). Before that, the following observations are useful for our perturbation analysis.

Let $\bar{Z} \in \mathcal{R}^{m\times n}$ be any given matrix. Suppose $\bar{Z}$ admit the SVD: $\bar{Z} = \bar{U}[\Sigma(\bar{Z}) \ 0]\bar{V}^T$ with $\bar{U} \in \mathcal{O}^m$ and $\bar{V} \in \mathcal{O}^n$. Define the two index sets a and b by $a := \{1 \le i \le m : \sigma_i(\bar{Z}) > 0\}$ and $b := \{1 \le i \le m : \sigma_i(\bar{Z}) = 0\}$. Similar to (4.7), a can be divided into q subsets $\{\eta_1, \dots, \eta_q\}$ for some positive integer q such that $a = \bigcup_{1\le l\le q} \eta_l$ with the singular values on each subset are the same and $\sigma_i(\bar{Z}) > \sigma_j(\bar{Z})$ for $\forall i \in \eta_{l_1},\ j \in \eta_{l_2}$ with $1 \le l_1 < l_2 \le q$.

Then we can obtain the following properties from [24, Proposition 2.1].

Proposition 4.2. *For any $\mathcal{R}^{m\times n} \ni H \to 0$, let $Z := [\Sigma(\bar{Z}) \ 0] + H$. Suppose that $\widehat{U} \in \mathcal{O}^m$ and $\widehat{V} \in \mathcal{O}^n$ satisfy*

$$Z := [\Sigma(\bar{Z}) \ 0] + H = \widehat{U}[\Sigma(Z) \ 0]\widehat{V}^T.$$

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

$$\widehat{U} = \begin{bmatrix} Q & 0 \\ 0 & Q' \end{bmatrix} + O(\|H\|) \text{ and } \widehat{V} = \begin{bmatrix} Q & 0 \\ 0 & Q'' \end{bmatrix} + O(\|H\|), \qquad (4.18)$$

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

$$\begin{aligned} \Sigma(Z)_{\eta_k\eta_k} - \Sigma(\bar{Z})_{\eta_k\eta_k} &= Q_k^T \mathbb{S}(H_{\eta_k\eta_k}) Q_k + O(\|H\|^2), \quad k = 1, \ldots, q, \\ [\Sigma(Z)_{bb} - \Sigma(\bar{Z})_{bb}\ 0] &= Q'^T [H_{bb}\ H_{b\gamma}] Q'' + O(\|H\|^2), \end{aligned} \tag{4.19}$$

where γ *is defined in (4.6).*

By using Proposition 4.2, we show the perturbation property of the nuclear norm.

Proposition 4.3. *Let* $A \in \mathcal{R}^{m\times n}$ *and* $B \in \partial\theta(A)$. *Suppose that* A *and* B *have the SVD as in (4.4) and the index sets* $\alpha, \beta, \beta_1, \beta_2$, β_3, *and* c *are defined as in (4.6) and (4.8). Then for all* $(A', B') \in \mathcal{R}^{m\times n} \times \mathcal{R}^{m\times n}$ *satisfying* $B' \in \partial\theta(A')$ *and is sufficiently close to* $(A, B) \in \mathcal{R}^{m\times n} \times \mathcal{R}^{m\times n}$, *we have*

$$\begin{cases} \widetilde{A}'_{\alpha\alpha} = \Sigma(A)_{\alpha\alpha} + O(\|\Delta A\|),\ \widetilde{A}'_{\alpha\beta_1} = O(\|\Delta A\|),\ \widetilde{A}'_{\alpha(\beta_2\cup\beta_3\cup c)} = O(\min\{\|\Delta A\|, \|\Delta B\|\}), \\ \widetilde{A}'_{\beta_1\alpha} = O(\|\Delta A\|),\ \widetilde{A}'_{\beta_1\beta_1} = O(\|\Delta A\|),\ \widetilde{A}'_{\beta_1(\beta_2\cup\beta_3\cup c)} = O(\|\Delta A\|\|\Delta B\|), \\ \widetilde{A}'_{(\beta_2\cup\beta_3)\alpha} = O(\min\{\|\Delta A\|, \|\Delta B\|\}),\ \widetilde{A}'_{(\beta_2\cup\beta_3)(\beta\cup c)} = O(\|\Delta A\|\|\Delta B\|), \end{cases} \tag{4.20}$$

and

$$\begin{cases} \widetilde{B}'_{\alpha\alpha} = I_{|\alpha|} + O(\|\Delta A\|),\ \widetilde{B}'_{\alpha\beta_1} = O(\|\Delta A\|),\ \widetilde{B}'_{\alpha(\beta_2\cup\beta_3\cup c)} = O(\min\{\|\Delta A\|, \|\Delta B\|\}), \\ \widetilde{B}'_{\beta_1\alpha} = O(\|\Delta A\|),\ \widetilde{B}'_{\beta_1\beta_1} = I_{|\beta_1|} + O(\|\Delta A\| + \|\Delta B\|), \\ \widetilde{B}'_{\beta_1(\beta_2\cup\beta_3\cup c)} = O(\|\Delta B\|),\ \widetilde{B}'_{(\beta_2\cup\beta_3)\alpha} = O(\min\{\|\Delta A\|, \|\Delta B\|\}), \\ \widetilde{B}'_{(\beta_2\cup\beta_3)\beta_1} = O(\|\Delta B\|),\ \widetilde{B}'_{\beta_2\beta_2} = \Sigma(B)_{\beta_2\beta_2} + O(\|\Delta B\|),\ \widetilde{B}'_{\beta_3\beta_2} = O(\|\Delta B\|), \\ \widetilde{B}'_{(\beta_2\cup\beta_3)(\beta_3\cup c)} = O(\|\Delta B\|). \end{cases} \tag{4.21}$$

Moreover,

$$\begin{cases} \mathbb{S}(\widetilde{B}'_1)_{\alpha\alpha} = I_{|\alpha|} + O(\|\Delta A\|^2), \ \mathbb{S}(\widetilde{B}'_1)_{\alpha\beta_1} = O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|), \\ \mathbb{T}(\widetilde{A}'_1)_{\alpha\alpha} = \frac{1}{2}\big(\Sigma(A)_{\alpha\alpha}\mathbb{T}(\widetilde{B}'_1)_{\alpha\alpha} + \mathbb{T}(\widetilde{B}'_1)_{\alpha\alpha}\Sigma(A)_{\alpha\alpha}\big) + O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|), \\ \mathbb{T}(\widetilde{A}'_1)_{\alpha\beta_1} = \frac{1}{2}\Sigma(A)_{\alpha\alpha}\mathbb{T}(\widetilde{B}'_1)_{\alpha\beta_1} + O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|), \\ \mathbb{S}(\widetilde{A}'_{\beta_1\beta_1}) + O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|) \in \mathcal{S}_+^{|\beta_1|}, \ \mathbb{T}(\widetilde{A}'_{\beta_1\beta_1}) = O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|), \\ \mathbb{S}(\widetilde{B}'_{\beta_1\beta_1}) \preceq I_{|\beta_1|}, \ \widetilde{B}'_{\alpha\beta_2} = \Sigma(A)^{-1}_{\alpha\alpha}\widetilde{A}'_{\alpha\beta_2} - \Sigma(A)^{-1}_{\alpha\alpha}(\widetilde{A}'_{\beta_2\alpha})^T\Sigma(B)_{\beta_2\beta_2} + O(\|\Delta A\|\|\Delta B\|), \\ \widetilde{B}'_{\beta_2\alpha} = \widetilde{A}'_{\beta_2\alpha}\Sigma(A)^{-1}_{\alpha\alpha} - \Sigma(B)_{\beta_2\beta_2}(\widetilde{A}'_{\alpha\beta_2})^T\Sigma(A)^{-1}_{\alpha\alpha} + O(\|\Delta A\|\|\Delta B\|), \\ \widetilde{B}'_{\alpha(\beta_3\cup c)} = \Sigma(A)^{-1}_{\alpha\alpha}\widetilde{A}'_{\alpha(\beta_3\cup c)} + O(\|\Delta A\|\|\Delta B\|), \\ \widetilde{B}'_{\beta_3\alpha} = \widetilde{A}'_{\beta_3\alpha}\Sigma(A)^{-1}_{\alpha\alpha} + O(\|\Delta A\|\|\Delta B\|). \end{cases} \tag{4.22}$$

and

$$\langle \mathbb{S}(\widetilde{A}'_{\beta_1\beta_1}), \mathbb{S}(\widetilde{B}'_{\beta_1\beta_1}) - I_{|\beta_1|}\rangle = O(\|\Delta A\|\|\Delta B\|)(\|\Delta A\| + \|\Delta B\|) + O(\|\Delta A\|^3), \tag{4.23}$$

In above statement, we denote $\Delta A := A' - A$, $\Delta B := B' - B$, $\widetilde{A}' := U^T A' V = [\widetilde{A}'_1 \ \widetilde{A}'_2] = [U^T A' V_1 \ U^T A' V_2]$, $\widetilde{B}' := U^T B' V = [\widetilde{B}'_1 \ \widetilde{B}'_2] = [U^T B' V_1 \ U^T B' V_2]$ *and* I_p *as the identity* p *by* p *matrix.*

Proof. From the above arguments about the SVD of A and B, it is easy to see that there exists $\widetilde{U} \in \mathcal{O}^m$ and $\widetilde{V} \in \mathcal{O}^n$ such that

$$\widetilde{A}' = \widetilde{U}[\Sigma(A') \ 0]\widetilde{V}^T, \quad \widetilde{B}' = \widetilde{U}[\Sigma(B') \ 0]\widetilde{V}^T.$$

By Proposition 4.2, we can see that for all ΔA and ΔB small enough, there exists $Q_1 \in \mathcal{O}^{|\alpha|}$, $Q'_1 \in \mathcal{O}^{|\beta|}$, $Q''_1 \in \mathcal{O}^{n-|\alpha|}$, $Q_2 \in \mathcal{O}^{m-|\beta_3|}$, $Q'_2 \in \mathcal{O}^{|\beta_3|}$ and $Q''_2 \in \mathcal{O}^{n-m+|\beta_3|}$ such that

$$\widetilde{U} = \begin{pmatrix} Q_1 & 0 \\ 0 & Q'_1 \end{pmatrix} + O(\|\Delta A\|) = \begin{pmatrix} Q_2 & 0 \\ 0 & Q'_2 \end{pmatrix} + O(\|\Delta B\|), \tag{4.24}$$

$$\widetilde{V} = \begin{pmatrix} Q_1 & 0 \\ 0 & Q''_1 \end{pmatrix} + O(\|\Delta A\|) = \begin{pmatrix} Q_2 & 0 \\ 0 & Q''_2 \end{pmatrix} + O(\|\Delta B\|), \tag{4.25}$$

where $Q_1 = \mathrm{Diag}(P_1, P_2, \ldots, P_r)$ and $Q_2 = \mathrm{Diag}(P_1', P_2')$ are block diagonal orthogonal matrices with $P_l \in \mathcal{O}^{|a_l|}$, $l = 1, \ldots, r$, $P_1' \in \mathcal{O}^{|\alpha|+|\beta_1|}$ and $P_2' \in \mathcal{O}^{|\beta_2|}$. Moreover,

$$\begin{aligned} \Sigma(A')_{a_l a_l} - \Sigma(A)_{a_l a_l} &= P_l^T \mathbb{S}(\Delta\widetilde{A}_{a_l a_l}) P_l + O(\|\Delta A\|^2), \quad l = 1, \ldots, r, \\ [\Sigma(A')_{\beta\beta} - \Sigma(A)_{\beta\beta} \ 0] &= Q_1'^T [\Delta\widetilde{A}_{\beta\beta} \ \Delta\widetilde{A}_{\beta c}] Q_1'' + O(\|\Delta A\|^2), \end{aligned} \tag{4.26}$$

and

$$\begin{aligned} \Sigma(B')_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)} - I_{|\alpha|+|\beta_1|} &= P_1'^T \mathbb{S}(\Delta\widetilde{B}_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)}) P_1' + O(\|\Delta B\|^2), \\ \Sigma(B')_{\beta_2\beta_2} - \Sigma(B)_{\beta_2\beta_2} &= P_2'^T \mathbb{S}(\Delta\widetilde{B}_{\beta_2\beta_2}) P_2' + O(\|\Delta B\|^2), \\ [\Sigma(B')_{\beta_3\beta_3} - \Sigma(B)_{\beta_3\beta_3} \ 0] &= Q_2'^T [\Delta\widetilde{B}_{\beta_3\beta_3} \ \Delta\widetilde{B}_{\beta_3 c}] Q_2'' + O(\|\Delta B\|^2), \end{aligned} \tag{4.27}$$

where $\Delta\widetilde{A} := U^T \Delta A V$ and $\Delta\widetilde{B} := U^T \Delta B V$. One should note that

$$Q_1^T \Sigma(A)_{\alpha\alpha} Q_1 = \Sigma(A)_{\alpha\alpha}, \quad P_2'^T \Sigma(B)_{\beta_2\beta_2} P_2' = \Sigma(B)_{\beta_2\beta_2}. \tag{4.28}$$

By Lemma 4.1 and the definition of β_1, β_2 and β_3 in (4.8), we can obtain the following properties of $\Sigma(A')$ and $\Sigma(B')$ that

$$\begin{cases} \Sigma(A')_{(\beta_2\cup\beta_3)(\beta_2\cup\beta_3)} = 0, \quad \Sigma(B')_{\alpha\alpha} = I_{|\alpha|}, \\ \Sigma(B')_{\beta_1\beta_1} \preceq I_{|\beta_1|}, \ \langle \Sigma(A')_{\beta_1\beta_1}, \Sigma(B')_{\beta_1\beta_1} - I_{|\beta_1|} \rangle = 0. \end{cases} \tag{4.29}$$

By using (4.24), (4.25) and the fact that for any $N \in \mathcal{R}^{|\beta_1|\times|\beta_1|}$,

$$NN^T = I_{|\beta_1|} + O(\|\Delta A\| + \|\Delta B\|) \Longrightarrow \exists\, \widehat{N} \in \mathcal{O}^{|\beta_1|} \text{ such that } \widehat{N} = N + O(\|\Delta A\| + \|\Delta B\|),$$

we can deduce that there exists $P_{\beta_1} \in \mathcal{O}^{|\beta_1|}$ such that

$$\widetilde{U} = \begin{pmatrix} Q_1 + R_1 & \widetilde{U}_{\alpha\beta_1} & \widetilde{U}_{\alpha\beta_2} & \widetilde{U}_{\alpha\beta_3} \\ \widetilde{U}_{\beta_1\alpha} & P_{\beta_1} + O(\|\Delta A\| + \|\Delta B\|) & O(\|\Delta B\|) & O(\|\Delta B\|) \\ \widetilde{U}_{\beta_2\alpha} & O(\|\Delta B\|) & P_2' + O(\|\Delta B\|) & O(\|\Delta B\|) \\ \widetilde{U}_{\beta_3\alpha} & O(\|\Delta B\|) & O(\|\Delta B\|) & Q_2' + O(\|\Delta B\|) \end{pmatrix} \tag{4.30}$$

and

$$\widetilde{V} = \begin{pmatrix} Q_1 + R_2 & \widetilde{V}_{\alpha\beta_1} & \widetilde{V}_{\alpha\beta_2} & \widetilde{V}_{\alpha(\beta_3\cup c)} \\ \widetilde{V}_{\beta_1\alpha} & P_{\beta_1} + O(\|\Delta A\| + \|\Delta B\|) & O(\|\Delta B\|) & O(\|\Delta B\|) \\ \widetilde{V}_{\beta_2\alpha} & O(\|\Delta B\|) & P_2' + O(\|\Delta B\|) & O(\|\Delta B\|) \\ \widetilde{V}_{(\beta_3\cup c)\alpha} & O(\|\Delta B\|) & O(\|\Delta B\|) & Q_2'' + O(\|\Delta B\|) \end{pmatrix}, \tag{4.31}$$

where

$$
\begin{gathered}
R_1 = O(\|\Delta A\|),\ \widetilde{U}_{\alpha\beta_1} = O(\|\Delta A\|),\ \widetilde{U}_{\alpha(\beta_2\cup\beta_3)} = O(\min\{\|\Delta A\|,\ \|\Delta B\|\}),\\
\widetilde{U}_{\beta_1\alpha} = O(\|\Delta A\|),\ \widetilde{U}_{(\beta_2\cup\beta_3)\alpha} = O(\min\{\|\Delta A\|,\ \|\Delta B\|\}),\\
R_2 = O(\|\Delta A\|),\ \widetilde{V}_{\alpha\beta_1} = O(\|\Delta A\|),\ \widetilde{V}_{\alpha(\beta_2\cup\beta_3\cup c)} = O(\min\{\|\Delta A\|,\ \|\Delta B\|\}),\\
\widetilde{V}_{\beta_1\alpha} = O(\|\Delta A\|) \text{ and } \widetilde{V}_{(\beta_2\cup\beta_3\cup c)\alpha} = O(\min\{\|\Delta A\|,\ \|\Delta B\|\}).
\end{gathered}
\tag{4.32}
$$

Then, combining (4.30)-(4.32) and the orthogonality of $\widetilde{U}$ and $\widetilde{V}$, we can have that for all ΔA and ΔB sufficiently small,

$$
\begin{cases}
Q_1 R_1^T + R_1 Q_1^T = O(\|\Delta A\|^2),\\
Q_1 \widetilde{U}_{\beta_1\alpha}^T + \widetilde{U}_{\alpha\beta_1} P_{\beta_1}^T = O(\|\Delta A\|^2) + O(\|\Delta A\|\|\Delta B\|),\\
Q_1 \widetilde{U}_{\beta_2\alpha}^T + \widetilde{U}_{\alpha\beta_2} P_2'^T = O(\|\Delta A\|\|\Delta B\|),\\
Q_1 R_2^T + R_2 Q_1^T = O(\|\Delta A\|^2),\\
Q_1 \widetilde{V}_{\beta_1\alpha}^T + \widetilde{V}_{\alpha\beta_1} P_{\beta_1}^T = O(\|\Delta A\|^2) + O(\|\Delta A\|\|\Delta B\|),\\
Q_1 \widetilde{V}_{\beta_2\alpha}^T + \widetilde{V}_{\alpha\beta_2} P_2'^T = O(\|\Delta A\|\|\Delta B\|).
\end{cases}
\tag{4.33}
$$

Next, by using (4.24)-(4.32), we can have the following characterization of $\widetilde{A}'$ and $\widetilde{B}'$ that for all ΔA and ΔB sufficiently small,

$$
\begin{cases}
\widetilde{A}'_{\alpha\alpha} &= \Sigma(A)_{\alpha\alpha} + \Gamma_1 + \Sigma(A)_{\alpha\alpha} Q_1 R_2^T + R_1 Q_1^T \Sigma(A)_{\alpha\alpha} + O(\|\Delta A\|^2)\\
&= \Sigma(A)_{\alpha\alpha} + O(\|\Delta A\|),\\
\widetilde{A}'_{\alpha\beta_1} &= \Sigma(A)_{\alpha\alpha} Q_1 \widetilde{V}_{\beta_1\alpha}^T + O(\|\Delta A\|^2) = O(\|\Delta A\|),\\
\widetilde{A}'_{\alpha(\beta_2\cup\beta_3\cup c)} &= \Sigma(A)_{\alpha\alpha} Q_1 \widetilde{V}_{(\beta_2\cup\beta_3\cup c)\alpha}^T + O(\|\Delta A\|\|\Delta B\|) = O(\min\{\|\Delta A\|, \|\Delta B\|\}),\\
\widetilde{A}'_{\beta_1\alpha} &= \widetilde{U}_{\beta_1\alpha} Q_1^T \Sigma(A)_{\alpha\alpha} + O(\|\Delta A\|^2) = O(\|\Delta A\|),\\
\widetilde{A}'_{\beta_1\beta_1} &= P_{\beta_1} \Sigma(A')_{\beta_1\beta_1} P_{\beta_1}^T + O(\|\Delta A\|^2) + O(\|\Delta A\|\|\Delta B\|) = O(\|\Delta A\|),\\
\widetilde{A}'_{\beta_1(\beta_2\cup\beta_3\cup c)} &= O(\|\Delta A\|\|\Delta B\|),\\
\widetilde{A}'_{(\beta_2\cup\beta_3)\alpha} &= \widetilde{U}_{(\beta_2\cup\beta_3)\alpha} Q_1^T \Sigma(A)_{\alpha\alpha} + O(\|\Delta A\|\|\Delta B\|) = O(\min\{\|\Delta A\|, \|\Delta B\|\}),\\
\widetilde{A}'_{(\beta_2\cup\beta_3)(\beta\cup c)} &= O(\|\Delta A\|\|\Delta B\|),
\end{cases}
\tag{4.34}
$$

and

$$\left\{\begin{array}{lcl}
\widetilde{B}'_{\alpha\alpha} & = & I_{|\alpha|} + Q_1 R_2^T + R_1 Q_1^T + O(\|\Delta A\|^2) \\
 & = & I_{|\alpha|} + O(\|\Delta A\|), \\
\widetilde{B}'_{\alpha\beta_1} & = & Q_1 \widetilde{V}^T_{\beta_1\alpha} + \widetilde{U}_{\alpha\beta_1} P^T_{\beta_1} + O(\|\Delta A\|^2) + O(\|\Delta A\|\|\Delta B\|) = O(\|\Delta A\|), \\
\widetilde{B}'_{\alpha\beta_2} & = & Q_1 \widetilde{V}^T_{\beta_2\alpha} + \widetilde{U}_{\alpha\beta_2} P_2'^T \Sigma(B)_{\beta_2\beta_2} + O(\|\Delta A\|\|\Delta B\|) = O(\min\{\|\Delta A\|, \|\Delta B\|\}), \\
\widetilde{B}'_{\alpha(\beta_3\cup c)} & = & Q_1 \widetilde{V}^T_{(\beta_3\cup c)\alpha} + O(\|\Delta A\|\|\Delta B\|) = O(\min\{\|\Delta A\|, \|\Delta B\|\}), \\
\widetilde{B}'_{\beta_1\alpha} & = & \widetilde{U}_{\beta_1\alpha} Q_1^T + P_{\beta_1} \widetilde{V}^T_{\alpha\beta_1} + O(\|\Delta A\|^2) + O(\|\Delta A\|\|\Delta B\|) = O(\|\Delta A\|), \\
\widetilde{B}'_{\beta_1\beta_1} & = & I_{|\beta_1|} + O(\|\Delta A\| + \|\Delta B\|), \\
\widetilde{B}'_{\beta_1(\beta_2\cup\beta_3\cup c)} & = & O(\|\Delta B\|), \\
\widetilde{B}'_{\beta_2\alpha} & = & \widetilde{U}_{\beta_2\alpha} Q_1^T + \Sigma(B)_{\beta_2\beta_2} P_2' \widetilde{V}^T_{\alpha\beta_2} + O(\|\Delta A\|\|\Delta B\|) = O(\min\{\|\Delta A\|, \|\Delta B\|\}), \\
\widetilde{B}'_{\beta_3\alpha} & = & \widetilde{U}_{\beta_3\alpha} Q_1^T + O(\|\Delta A\|\|\Delta B\|) = O(\min\{\|\Delta A\|, \|\Delta B\|\}), \\
\widetilde{B}'_{(\beta_2\cup\beta_3)\beta_1} & = & O(\|\Delta B\|), \ \widetilde{B}'_{\beta_2\beta_2} = \Sigma(B)_{\beta_2\beta_2} + O(\|\Delta B\|), \ \widetilde{B}'_{\beta_3\beta_2} = O(\|\Delta B\|) \\
\widetilde{B}'_{(\beta_2\cup\beta_3)(\beta_3\cup c)} & = & O(\|\Delta B\|),
\end{array}\right. \tag{4.35}$$

where $\Gamma_1 := \mathrm{Diag}(\mathbb{S}(\Delta\widetilde{A}_{a_1a_1}), \ldots, \mathbb{S}(\Delta\widetilde{A}_{a_ra_r}))$, a $|\alpha| \times |\alpha|$ symmetric matrix. Thus, we have showed (4.20) and (4.21). Next, let us prove (4.22). By the results we obtained in (4.34) and (4.35), using the relationships (4.33), we can derive the following useful relationships via fundamental calculations that for all ΔA and ΔB sufficiently small,

$$\left\{\begin{array}{l}
\mathbb{S}(\widetilde{B}'_1)_{\alpha\alpha} = I_{|\alpha|} + O(\|\Delta A\|^2), \ \mathbb{S}(\widetilde{B}'_1)_{\alpha\beta_1} = O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|), \\
\mathbb{T}(\widetilde{A}'_1)_{\alpha\alpha} = \frac{1}{2}\big(\Sigma(A)_{\alpha\alpha}\mathbb{T}(\widetilde{B}'_1)_{\alpha\alpha} + \mathbb{T}(\widetilde{B}'_1)_{\alpha\alpha}\Sigma(A)_{\alpha\alpha}\big) + O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|), \\
\mathbb{T}(\widetilde{A}'_1)_{\alpha\beta_1} = \frac{1}{2}\Sigma(A)_{\alpha\alpha}\mathbb{T}(\widetilde{B}'_1)_{\alpha\beta_1} + O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|), \\
\widetilde{B}'_{\alpha\beta_2} = \Sigma(A)^{-1}_{\alpha\alpha}\widetilde{A}'_{\alpha\beta_2} - \Sigma(A)^{-1}_{\alpha\alpha}(\widetilde{A}'_{\beta_2\alpha})^T\Sigma(B)_{\beta_2\beta_2} + O(\|\Delta A\|\|\Delta B\|), \\
\widetilde{B}'_{\beta_2\alpha} = \widetilde{A}'_{\beta_2\alpha}\Sigma(A)^{-1}_{\alpha\alpha} - \Sigma(B)_{\beta_2\beta_2}(\widetilde{A}'_{\alpha\beta_2})^T\Sigma(A)^{-1}_{\alpha\alpha} + O(\|\Delta A\|\|\Delta B\|), \\
\widetilde{B}'_{\alpha(\beta_3\cup c)} = \Sigma(A)^{-1}_{\alpha\alpha}\widetilde{A}'_{\alpha(\beta_3\cup c)} + O(\|\Delta A\|\|\Delta B\|), \\
\widetilde{B}'_{\beta_3\alpha} = \widetilde{A}'_{\beta_3\alpha}\Sigma(A)^{-1}_{\alpha\alpha} + O(\|\Delta A\|\|\Delta B\|),
\end{array}\right. \tag{4.36}$$

Moreover, for ΔA and ΔB small enough, we have

$$\begin{aligned} &\mathbb{S}(\widetilde{A}'_{\beta_1\beta_1}) + O(\|\Delta A\|\|\Delta B\|) + O(\|\Delta A\|^2) = P_{\beta_1}\Sigma(A')_{\beta_1\beta_1}P_{\beta_1}^T \in \mathcal{S}_+^{|\beta_1|}, \\ &T(\widetilde{A}'_{\beta_1\beta_1}) = O(\|\Delta A\|^2) + O(\|\Delta A\|\|\Delta B\|). \end{aligned} \tag{4.37}$$

By Lemma 4.1, it always holds that $\Sigma(\widetilde{B}') = \Sigma(B') \preceq I_m$. Thus, we can derive that $\widetilde{B}'_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)}\widetilde{B}'^T_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)} \preceq I_{|\alpha|+|\beta_1|}$ and $\widetilde{B}'^T_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)}\widetilde{B}'_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)} \preceq I_{|\alpha|+|\beta_1|}$. By the definition (4.15) of operators $\mathbb{S}(\cdot)$ and $\mathbb{T}(\cdot)$, one can expand the summary

$$\begin{aligned} I_{|\alpha|+|\beta_1|} &\succeq \tfrac{1}{2}\big(\widetilde{B}'_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)}\widetilde{B}'^T_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)} + \widetilde{B}'^T_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)}\widetilde{B}'_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)}\big) \\ &= \tfrac{1}{2}\big(\mathbb{S}(\widetilde{B}'_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)}) + \mathbb{T}(\widetilde{B}'_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)})\big)\big(\mathbb{S}(\widetilde{B}'_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)}) - \mathbb{T}(\widetilde{B}'_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)})\big) \\ &+ \tfrac{1}{2}\big(\mathbb{S}(\widetilde{B}'_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)}) - \mathbb{T}(\widetilde{B}'_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)})\big)\big(\mathbb{S}(\widetilde{B}'_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)}) + \mathbb{T}(\widetilde{B}'_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)})\big) \\ &= \mathbb{S}(\widetilde{B}'_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)})\mathbb{S}(\widetilde{B}'_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)}) + \mathbb{T}(\widetilde{B}'_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)})\mathbb{T}(\widetilde{B}'_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)})^T. \end{aligned} \tag{4.38}$$

Then,

$$\begin{aligned} \mathbb{S}(\widetilde{B}'_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)})\mathbb{S}(\widetilde{B}'_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)}) \preceq I_{|\alpha|+|\beta_1|} \Longrightarrow\ & \mathbb{S}(\widetilde{B}'_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)}) \preceq I_{|\alpha|+|\beta_1|} \\ & \Downarrow \\ & \mathbb{S}(\widetilde{B}'_{\beta_1\beta_1}) \preceq I_{|\beta_1|}. \end{aligned} \tag{4.39}$$

Finally, let us prove the last equation (4.23). Denote $R_3 = \widetilde{U}_{\beta_1\beta_1} - P_{\beta_1} = O(||\Delta A|| + ||\Delta B||)$ and $R_4 = \widetilde{V}_{\beta_1\beta_1} - P_{\beta_1} = O(||\Delta A|| + ||\Delta B||)$, then by the orthogonality of $\widetilde{U}$ and $\widetilde{V}$, we can easily find that

$$\begin{aligned} P_{\beta_1}R_3^T + R_3P_{\beta_1}^T = O(||\Delta A||^2 + ||\Delta B||^2), \\ P_{\beta_1}R_4^T + R_4P_{\beta_1}^T = O(||\Delta A||^2 + ||\Delta B||^2), \end{aligned} \tag{4.40}$$

Then, we can compute

$$\begin{aligned} \widetilde{B}'_{\beta_1\beta_1} &= (P_{\beta_1} + R_3)\Sigma(B')_{\beta_1\beta_1}(P_{\beta_1}^T + R_4^T) + O(||\Delta A||^2 + ||\Delta B||^2) \\ &= P_{\beta_1}\Sigma(B')_{\beta_1\beta_1}P_{\beta_1}^T + P_{\beta_1}R_4^T + R_3P_{\beta_1}^T + O(||\Delta A||^2 + ||\Delta B||^2). \end{aligned} \tag{4.41}$$

Therefore, combining (4.40) and (4.41), we can conclude that

$$\mathbb{S}(\widetilde{B}'_{\beta_1\beta_1}) = P_{\beta_1}\Sigma(B')_{\beta_1\beta_1}P_{\beta_1}^T + O(||\Delta A||^2 + ||\Delta B||^2).$$

Hence,

$$
\begin{aligned}
& \langle \mathbb{S}(\widetilde{A}'_{\beta_1\beta_1}), \mathbb{S}(\widetilde{B}'_{\beta_1\beta_1}) - I_{|\beta_1|}\rangle \\
= \;& \langle P_{\beta_1}\Sigma(A')_{\beta_1\beta_1}P_{\beta_1}^T + O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|), P_{\beta_1}\Sigma(B')_{\beta_1\beta_1}P_{\beta_1}^T - I_{|\beta_1|} \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad + O(||\Delta A||^2 + ||\Delta B||^2)\rangle \\
= \;& \langle P_{\beta_1}\Sigma(A')_{\beta_1\beta_1}P_{\beta_1}^T, P_{\beta_1}\Sigma(B')_{\beta_1\beta_1}P_{\beta_1}^T - I_{|\beta_1|}\rangle \qquad (4.42)\\
+ \;& \langle P_{\beta_1}\Sigma(A')_{\beta_1\beta_1}P_{\beta_1}^T, O(||\Delta A||^2 + ||\Delta B||^2)\rangle \\
+ \;& \langle O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|), P_{\beta_1}\Sigma(B')_{\beta_1\beta_1}P_{\beta_1}^T - I_{|\beta_1|}\rangle \\
= \;& O(\|\Delta A\|\|\Delta B\|)(\|\Delta A\| + \|\Delta B\|) + O(\|\Delta A\|^3),
\end{aligned}
$$

where the first summand of the second equation equals to 0 due to (4.29), and the last two summands estimated by (4.26) and (4.27).

This completes the proof of the proposition. □

Similar to the SDP cone, this perturbation property contains the second order information of the nuclear norm. Moreover, the proof is complicated due to the nonsymmetric of the underlying matrix.

4.2 The metric subregularity of the solution mapping for composite nuclear norm problem

Similar to the SDP case, we still need the following perturbation analysis and notations.

For any $(x, y, S) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^{m\times n}$, the Lagrangian function l associated with problem (4.1) is defined as

$$l(x, y, S) := f(x) + \langle y, h(x)\rangle + \langle S, g(x)\rangle - \theta^*(S). \qquad (4.43)$$

Define the multi-valued mapping $\mathcal{T}_l : \mathcal{X}\times\mathcal{Y}\times\mathcal{R}^{m\times n} \rightrightarrows \mathcal{X}\times\mathcal{Y}\times\mathcal{R}^{m\times n}$ associated with the Lagrangian function l at any $(x, y, S) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^{m\times n}$ by

$$\mathcal{T}_l(x, y, S) = \{((u, v, C) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^{m\times n} | (u, -v, -C) \in \partial l(x, y, S)\}. \qquad (4.44)$$

Suppose that the optimal solution set of the problem (4.1) is nonempty and consider an optimal solution $\bar{x} \in \mathcal{X}$ of the problem (4.1). As the illustrations in

Section 2.3.3, if the reduced RCQ (2.14) holds at $\bar{x}$, we can impose the following first order optimality conditions for the problem (4.1). Then, $(\bar{y}, \bar{S}) \in \mathcal{Y} \times \mathcal{R}^{m\times n}$ is a Lagrangian multiplier corresponding to $\bar{x}$ if and only if $(\bar{x}, \bar{y}, \bar{S})$ satisfies the following KKT system:

$$\begin{cases} \nabla f(\bar{x}) + \nabla h(\bar{x})\bar{y} + \nabla g(\bar{x})\bar{S} = 0, \\ h(\bar{x}) = 0, \\ \bar{S} \in \partial\theta(g(\bar{x})). \end{cases} \tag{4.45}$$

Denote $\mathcal{M}(\bar{x})$ as the set of all Lagrangian multipliers corresponding to $\bar{x}$.
Moreover, since $\operatorname{dom}\theta = \mathcal{R}^{m\times n}$, the reduced RCQ (2.14) always holds at $\bar{x}$, i.e.,

$$0 \in \operatorname{int}\{g(\bar{x}) + g'(\bar{x})\mathcal{X} - \mathcal{R}^{m\times n}\}. \tag{4.46}$$

Therefore, we can have (4.45) as the optimality conditions of the problem (4.1) without assumptions.
By the third inclusion of (4.45), we assume that $g(\bar{x})$ and $\bar{S}$ have the singular value decompositions as in (4.4) with $A = g(\bar{x})$ and $B = \bar{S}$.

For a perturbed point $(u, v, C) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^{m\times n}$, it is easy to check that $(x, y, S) \in \mathcal{T}_l^{-1}((u, v, C)$ can be equivalently interpreted as the following perturbed KKT system:

$$\begin{cases} \nabla f(x) + \nabla h(x)y + \nabla g(x)S = u, \\ h(x) + v = 0, \\ S \in \partial\theta(g(x) + C). \end{cases} \tag{4.47}$$

One can find that $\mathcal{T}_l^{-1}(0,0,0)$ is the set of all the KKT points $(\bar{x}, \bar{y}, \bar{S})$ of the problem (4.1) satisfying (4.45).

Next, we conduct our discussions about the metric subregularity of $\mathcal{T}_l$ at a KKT point for the origin.

Let $(\bar{x}, \bar{y}, \bar{S}) \in \mathcal{T}_l^{-1}(0,0,0)$. As the reduced RCQ (4.46) always holds, by Proposition 2.11, we can have the critical cone of the problem (4.1) at $\bar{x}$ as

$$\mathcal{C}(\bar{x}) := \{d \in \mathcal{X} \mid h'(\bar{x})d = 0,\ g'(\bar{x})d \in \mathcal{C}_\theta(g(\bar{x}), \bar{S})\},$$

where, $\mathcal{C}_\theta(\cdot,\cdot)$ defined as (4.13).

Here, we define a more restrictive second-order sufficient condition for the problem (4.1) at $\bar{x}$ with respect to the multiplier $(\bar{y},\bar{S})\in\mathcal{M}(\bar{x})$ if

$$\langle d,\nabla^2_{xx}l(\bar{x},\bar{y},\bar{S})d\rangle+\Upsilon_{g(\bar{x})}\big(\bar{S},g'(\bar{x})d\big)>0,\ \forall 0\neq\ d\in\mathcal{C}(\bar{x}), \tag{4.48}$$

where $-\Upsilon_{g(\bar{x})}\big(\bar{S},g'(\bar{x})d\big)$ is the so called sigma term in the second-order sufficient condition (4.48) for the problem (4.1).

Furthermore, by Proposition 2.15 or [24, Proposition 4.3], the epigraph of the nuclear norm θ is $\mathcal{C}^2$-cone reducible at every point $(X,t)\in\operatorname{epi}\theta$, and thus second order regular [8, Proposition 3.136]. Therefore, θ is twice (parabolically) directionally epidifferentiable by Proposition 2.3. Moreover, since θ is Lipschitz continuous and so does its directional derivative, we have $\theta^{\downarrow\downarrow}_{+}(X;H,\cdot)=\theta''_{+}(X;H,\cdot)$ and $\theta^{\downarrow\downarrow}_{-}(X;H,\cdot)=\theta''_{-}(X;H,\cdot)$ for any $X,H\in\mathcal{R}^{m\times n}$. In total, θ is twice (parabolically) directionally differentiable and $\theta^{\downarrow\downarrow}(X;H,\cdot)=\theta''(X;H,\cdot)$ for any $X,H\in\mathcal{R}^{m\times n}$.

In [25], Ding shows that the sigma term for nuclear norm regularized problem is just the conjugate function of the parabolic second order directional derivative of the nuclear norm function θ. Moreover, by adopting the sigma term derived by Bonnans and Shapiro [8, Section 3.4.1] for composite problems, we have

$$-\Upsilon_{g(\bar{x})}\big(\cdot,g'(\bar{x})d\big)=\phi^*(\cdot)\text{ with }\phi(\cdot):=\theta''(g(\bar{x});g'(\bar{x})d,\cdot).$$

By using the expression of the second order directional derivative for the eigenvalues and singular values [96, 109], Ding [25] futher provides the explicit expression of this sigma term as below.

We consider $A,B\in\mathcal{R}^{m\times n}$ satisfying $B\in\partial\theta(A)$ and the index sets $\alpha,\beta,\beta_1,\beta_2,\ \beta_3$ and c defined as (4.6) - (4.8), the sigma term

$$-\Upsilon_A\big(B,H\big)=2\sum_{l=1}^{r_0}\operatorname{tr}\big(\Omega_{a_l}(A,H)\big)+2\langle\operatorname{Diag}(\sigma_\beta(B)),U_\beta^T HA^\dagger HV_\beta\rangle, \tag{4.49}$$

where $\sigma_\beta(B)=(\sigma_i(B))_{i\in\beta}$ and

$$\begin{aligned}\Omega_{a_l}(A,H):=\ &(\mathbb{S}(\widetilde{H}_1))^T_{a_l}(\Sigma(A)-\nu_l(A)\mathcal{I}_m)^\dagger(\mathbb{S}(\widetilde{H}_1))_{a_l}-(2\nu_l(A))^{-1}\widetilde{H}_{a_lc}\widetilde{H}^T_{a_lc}\\ &+(\mathbb{T}(\widetilde{H}_1))^T_{a_l}(-\Sigma(A)-\nu_l(A)\mathcal{I}_m)^\dagger(\mathbb{T}(\widetilde{H}_1))_{a_l},\quad l=1,2,\dots r_0,\end{aligned}$$

with $\widetilde{H} = [\widetilde{H}_1 \ \widetilde{H}_2] = [U^T H V_1 \ U^T H V_2]$. We can futher compute the sigam term as

$$\begin{aligned}
\Upsilon_A(B,H) = & \sum_{1\le l,t\le r_0} \frac{2}{\nu_l(A)+\nu_t(A)}\|(\mathbb{T}(\widetilde{H}_1))_{a_l a_t}\|^2 + \sum_{1\le l\le r_0} \frac{4}{\nu_l(A)}\|(\mathbb{T}(\widetilde{H}_1))_{a_l\beta_1}\|^2 \\
& + \sum_{\substack{1\le l\le r_0 \\ 1\le i-|\alpha|-|\beta_1|\le|\beta_2|}} \left(\frac{2(1-\sigma_i(B))}{\nu_l(A)}\|(\mathbb{S}(\widetilde{H}_1))_{a_l i}\|^2 + \frac{2(\sigma_i(B)+1)}{\nu_l(A)}\|(\mathbb{T}(\widetilde{H}_1))_{a_l i}\|^2\right) \\
& + \sum_{1\le l\le r_0} \left(\frac{2}{\nu_l(A)}\|(\mathbb{S}(\widetilde{H}_1))_{a_l\beta_3}\|^2 + \frac{2}{\nu_l(A)}\|(\mathbb{T}(\widetilde{H}_1))_{a_l\beta_3}\|^2\right) \\
& + \sum_{1\le l\le r_0} \frac{1}{\nu_l(A)}\|(\widetilde{H}_2)_{a_l c}\|^2.
\end{aligned} \tag{4.50}$$

For the convenience of the later discussions, recalling the definition of $\mathcal{C}_{\theta^*}(\cdot,\cdot)$ in (4.16), we define the following joint 'critical cone' associated with the problem (4.1) as

$$\widetilde{\mathcal{C}}(\bar{x},\bar{y},\bar{S}) := \left\{ \begin{array}{c|l} (d_x,d_y,d_S)\in & h'(\bar{x})d_x = 0, \\ \mathcal{X}\times\mathcal{Y}\times\mathcal{R}^{m\times n} & g'(\bar{x})d_x \in \mathcal{C}_\theta(g(\bar{x}),\bar{S}), d_S \in \mathcal{C}_{\theta^*}(\bar{S},g(\bar{x})), \\ & d_S = U[\mathcal{E}_{\mathbb{S}}\circ\mathbb{S}(\widetilde{D}_{x,1}) + \mathcal{E}_{\mathbb{T}}\circ\mathbb{T}(\widetilde{D}_{x,1}) \quad \mathcal{F}\circ(\widetilde{D}_{x,2})]V^T \\ & \quad +U\Theta_{\bar{S}}V^T \end{array} \right\}, \tag{4.51}$$

where $D_x = g'(\bar{x})d_x$, $\widetilde{D}_x = [\widetilde{D}_{x,1} \ \widetilde{D}_{x,2}] = [U^T D_x V_1 \ U^T D_x V_2] = U^T D_x V$, $\widetilde{d}_S = U^T d_S V$ and $\mathcal{E}_{\mathbb{S}} \in \mathcal{S}^m$, $\mathcal{E}_{\mathbb{T}} \in \mathcal{S}^m$, $\mathcal{F} \in \mathcal{R}^{m\times(n-m)}$ are given by

$$(\mathcal{E}_{\mathbb{S}})_{ij} = \begin{cases} \dfrac{\sigma_i(\bar{S})-\sigma_j(\bar{S})}{\sigma_i(g(\bar{x}))-\sigma_j(g(\bar{x}))} & \text{if } \sigma_i(g(\bar{x}))\neq\sigma_j(g(\bar{x})), \\ 0 & \text{otherwise}, \end{cases} \qquad i,j\in\{1,\ldots,m\},$$

$$(\mathcal{E}_{\mathbb{T}})_{ij} = \begin{cases} \dfrac{\sigma_i(\bar{S})+\sigma_j(\bar{S})}{\sigma_i(g(\bar{x}))+\sigma_j(g(\bar{x}))} & \text{if } \sigma_i(g(\bar{x}))+\sigma_j(g(\bar{x}))\neq 0, \\ 0 & \text{otherwise}, \end{cases} \qquad i,j\in\{1,\ldots,m\},$$

and

$$(\mathcal{F})_{ij} = \begin{cases} \dfrac{\sigma_i(\bar{S})}{\sigma_i(g(\bar{x}))} & \text{if } \sigma_i(g(\bar{x}))\neq 0, \\ 0 & \text{otherwise}, \end{cases} \qquad i\in\{1,\ldots,m\},\ j=\{1,\ldots,n-m\};$$

moreover, we define $\Theta_{\bar{S}} := \begin{pmatrix} 0_{\alpha\alpha} & 0 & 0 \\ 0 & (\widetilde{d}_S)_{\beta\beta} & (\widetilde{d}_S)_{\beta c} \end{pmatrix} \in \mathcal{R}^{m\times n}$.

Our main results are as follows.

Theorem 4.1. *Let $\bar{x}$ be an optimal solution to the problem (4.1) and $(\bar{y}, \bar{S}) \in \mathcal{M}(\bar{x})$ be a Lagrangian multiplier corresponding to $\bar{x}$. Denote $\Phi := \big(\mathcal{C}_\theta(g(\bar{x}), \bar{S})\big)^\circ$. Let the following assumptions be satisfied:*

(i) The set $g'(\bar{x})^T\Phi$ is closed.

(ii) $\langle \Pi_\Phi(-g'(\bar{x})d_x),\ \Pi_\Phi(d_S)\rangle = 0$ for all $(d_x, d_y, d_S) \in \widetilde{\mathcal{C}}(\bar{x}, \bar{y}, \bar{S})$, where $\Pi_\Phi(\cdot)$ denotes the projection onto the set Φ and the set $\widetilde{\mathcal{C}}(\bar{x}, \bar{y}, \bar{S})$ is defined as (4.51).

(iii) The second-order sufficient condition (4.48) holds at $\bar{x}$ with respect to the multiplier $(\bar{y}, \bar{S}) \in \mathcal{M}(\bar{x})$ for the problem (4.1).

Then there exist a constant $\kappa > 0$ and a neighborhood $\mathcal{U}$ of $(\bar{x}, \bar{y}, \bar{S})$ such that for any $((u, v, C) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^{m\times n}$,

$$\|x - \bar{x}\| \le \kappa \|((u, v, C)\|,\ \forall (x, y, S) \in \mathcal{T}_l^{-1}((u, v, C) \cap \mathcal{U}. \tag{4.52}$$

Moreover, if there exists $(\widehat{y}, \widehat{S}) \in \mathcal{M}(\bar{x})$ such that $0 < \sigma_\beta(\widehat{S}) < e_\beta$, then $\mathcal{T}_l$ is metrically subregular at $(\bar{x}, \bar{y}, \bar{S})$ for the origin.

Proof. Firstly, we show that under the assumptions (i)-(iii), there exist a constant $\kappa > 0$ and a neighborhood $\mathcal{U}$ of $(\bar{x}, \bar{y}, \bar{S})$ such that (4.52) holds.
Suppose that (4.52) does not hold. It means that there exist some sequences $\{(u^k, v^k, C^k)\}_{k\ge 0} \subset \mathcal{X}\times\mathcal{Y}\times\mathcal{R}^{m\times n}$ and $\{(x^k, y^k, S^k)\}_{k\ge 0} \subset \mathcal{X}\times\mathcal{Y}\times\mathcal{R}^{m\times n}$ such that $(u^k, v^k, C^k) \to 0$, $(x^k, y^k, S^k) \to (\bar{x}, \bar{y}, \bar{S})$ with every $(x^k, y^k, S^k) \in \mathcal{T}_l^{-1}(u^k, v^k, C^k)$, and

$$\|x^k - \bar{x}\| \ge \delta_k \|(u^k, v^k, C^k)\|$$

with some $0 < \delta_k$ such that $\delta_k \to \infty$. Denote $t_k := \|x^k - \bar{x}\|$, by taking a subsequence if necessary, we can assume that $(x^k - \bar{x})/t_k \to d_{\bar{x}} \in \mathcal{X}$ with $\|d_{\bar{x}}\| = 1$.
From the perturbed KKT system (4.47), we can have that for all $k \ge 0$ large enough,

$$\begin{aligned} 0 &= h(x^k) + v^k - h(\bar{x}) \\ &= h'(\bar{x})(x^k - \bar{x}) + o(t_k) + v^k. \end{aligned} \tag{4.53}$$

Dividing by t_k on both sides of (4.53) and taking limits $k \to \infty$, we get

$$h'(\bar{x})d_{\bar{x}} = 0. \tag{4.54}$$

For simplify notations, we set

$$\Omega := \{W \in \mathcal{R}^{m\times n} \mid U_\beta^T W[V_\beta \ V_c] = 0\},$$

$$A := g(\bar{x}), \ B := \bar{S},$$

and for all $k \geq 0$,

$$\begin{cases} A^k := g(x^k) + c^k, \ \widetilde{A}^k := U^T A^k V = [\widetilde{A}_1^k \ \widetilde{A}_2^k] = [U^T A^k V_1 \ U^T A^k V_2], \\ B^k := S^k, \ \widetilde{B}^k := U^T B^k V = [\widetilde{B}_1^k \ \widetilde{B}_2^k] = [U^T B^k V_1 \ U^T B^k V_2], \\ H^k := \Pi_\Omega\big((B^k - B)/t_k\big), \ \Delta A^k := A^k - A, \ \Delta B^k := B^k - B. \end{cases} \tag{4.55}$$

Thus $A^k \to A$ and $B^k \to B$ by the assumptions. Moreover, similar to (4.53), we can derive that

$$\frac{1}{t_k}(A^k - A) \to g'(\bar{x})d_{\bar{x}} \text{ as } k \to \infty. \tag{4.56}$$

Since $B \in \partial\theta(A)$ and $B^k \in \partial\theta(A^k)$, we can derive the following estimates by Proposition 4.3 that for all (A^k, B^k) sufficiently close to (A, B),

$$\begin{cases} \mathbb{S}(\widetilde{A}^k_{\beta_1\beta_1}) + O(\|\Delta A^k\|^2 + \|\Delta A^k\|\|\Delta B^k\|) \in \mathcal{S}_+^{|\beta_1|}, \ \mathbb{T}(\widetilde{A}^k_{\beta_1\beta_1}) = O(\|\Delta A^k\|^2 + \|\Delta A^k\|\|\Delta B^k\|), \\ \widetilde{A}^k_{\beta_1(\beta_2\cup\beta_3\cup c)} = O(\|\Delta A^k\|\|\Delta B^k\|), \ \widetilde{A}^k_{(\beta_2\cup\beta_3)(\beta\cup c)} = O(\|\Delta A^k\|\|\Delta B^k\|), \\ \mathbb{S}(\widetilde{B}_1^k)_{\alpha\alpha} = I_{|\alpha|} + O(\|\Delta A^k\|^2 + \|\Delta A^k\|\|\Delta B^k\|), \ \mathbb{S}(\widetilde{B}_1^k)_{\alpha\beta_1} = O(\|\Delta A^k\|^2 + \|\Delta A^k\|\|\Delta B^k\|), \\ \mathbb{S}(\widetilde{B}^k_{\beta_1\beta_1}) \preceq I_{|\beta_1|}, \end{cases} \tag{4.57}$$

and

$$\begin{cases} \mathbb{T}(\widetilde{A}_1^k)_{\alpha\alpha} = \frac{1}{2}\big(\Sigma(A)_{\alpha\alpha}\mathbb{T}(\widetilde{B}_1^k)_{\alpha\alpha} + \mathbb{T}(\widetilde{B}_1^k)_{\alpha\alpha}\Sigma(A)_{\alpha\alpha}\big) + O(\|\Delta A^k\|^2 + \|\Delta A^k\|\|\Delta B^k\|), \\ \mathbb{T}(\widetilde{A}_1^k)_{\alpha\beta_1} = \frac{1}{2}\Sigma(A)_{\alpha\alpha}\mathbb{T}(\widetilde{B}_1^k)_{\alpha\beta_1} + O(\|\Delta A^k\|^2 + \|\Delta A^k\|\|\Delta B^k\|), \\ \widetilde{B}^k_{\alpha\beta_2} = \Sigma(A)^{-1}_{\alpha\alpha}\widetilde{A}^k_{\alpha\beta_2} - \Sigma(A)^{-1}_{\alpha\alpha}(\widetilde{A}^k_{\beta_2\alpha})^T\Sigma(B)_{\beta_2\beta_2} + O(\|\Delta A^k\|\|\Delta B^k\|), \\ \widetilde{B}^k_{\beta_2\alpha} = \widetilde{A}^k_{\beta_2\alpha}\Sigma(A)^{-1}_{\alpha\alpha} - \Sigma(B)_{\beta_2\beta_2}(\widetilde{A}^k_{\alpha\beta_2})^T\Sigma(A)^{-1}_{\alpha\alpha} + O(\|\Delta A^k\|\|\Delta B^k\|), \\ \widetilde{B}^k_{\alpha(\beta_3\cup c)} = \Sigma(A)^{-1}_{\alpha\alpha}\widetilde{A}^k_{\alpha(\beta_3\cup c)} + O(\|\Delta A^k\|\|\Delta B^k\|), \\ \widetilde{B}^k_{\beta_3\alpha} = \widetilde{A}^k_{\beta_3\alpha}\Sigma(A)^{-1}_{\alpha\alpha} + O(\|\Delta A^k\|\|\Delta B^k\|). \end{cases} \tag{4.58}$$

Combining the above (4.56) - (4.58) with Lemma 4.2, we obtain

$$\begin{cases} g'(\bar{x})d_{\bar{x}} \in \mathcal{C}_\theta(g(\bar{x}), \bar{S}), \;\; G^k := (B^k - B)/t_k - H^k \in \Phi, \\ H := \lim\limits_{k\to\infty} H^k = U[\mathcal{E}_{\mathbb{S}} \circ \mathbb{S}(\widetilde{D}_{\bar{x},1}) + \mathcal{E}_{\mathbb{T}} \circ \mathbb{T}(\widetilde{D}_{\bar{x},1}) \quad \mathcal{F} \circ (\widetilde{D}_{\bar{x},2})]V^T, \end{cases} \tag{4.59}$$

where $\widetilde{D}_{\bar{x}} := U^T g'(\bar{x})d_{\bar{x}}V = [\widetilde{D}_{\bar{x},1} \;\; \widetilde{D}_{\bar{x},2}] = [U^T D_{\bar{x}} V_1 \;\; U^T D_{\bar{x}} V_2]$.

Again, by the perturbed KKT system (4.47), we can deduce that for $k \geq 0$ large enough,

$$\begin{aligned} u^k \; &= \nabla f(x^k) + \nabla h(x^k)y^k + \nabla g(x^k)S^k - \big(\nabla f(\bar{x}) + \nabla h(\bar{x})\bar{y} + \nabla g(\bar{x})\bar{S}\big) \\ &= \nabla^2_{xx} f(\bar{x})(x^k - \bar{x}) + \langle y^k, h''(\bar{x})(x^k - \bar{x})\rangle + \langle S^k, g''(\bar{x})(x^k - \bar{x})\rangle \\ &+ \nabla h(\bar{x})(y^k - \bar{y}) + \nabla g(\bar{x})(S^k - \bar{S}). \end{aligned} \tag{4.60}$$

Dividing by t_k on both side of (4.60), it gives

$$\frac{u^k}{t_k} - \nabla^2_{xx} f(\bar{x})\frac{(x^k - \bar{x})}{t_k} - \langle y^k, h''(\bar{x})\frac{(x^k - \bar{x})}{t_k}\rangle - \langle S^k, g''(\bar{x})\frac{(x^k - \bar{x})}{t_k}\rangle - \nabla g(\bar{x})H^k$$

$$= \nabla h(\bar{x})\frac{(y^k - \bar{y})}{t_k} + \nabla g(\bar{x})G^k \in \mathrm{Im}\nabla h(\bar{x}) + \nabla g(\bar{x})\Phi,$$

where the set in the right hand side, as a sum of a linear subspace and a closed set, is closed, since $g'(\bar{x})^T\Phi$ is supposed to be closed. Then by taking limit as $k \to \infty$, it yeilds

$$-\nabla^2_{xx} l(\bar{x}, \bar{y}, \bar{S})d_{\bar{x}} - \nabla g(\bar{x})H \in \mathrm{Im}\nabla h(\bar{x}) + \nabla g(\bar{x})\Phi. \tag{4.61}$$

The inclusion (4.61) means that there exists $(d_{\bar{y}}, G) \in \mathcal{Y} \times \Phi$ such that

$$\nabla^2_{xx} l(\bar{x}, \bar{y}, \bar{S})d_{\bar{x}} + \nabla g(\bar{x})H + \nabla h(\bar{x})d_{\bar{y}} + \nabla g(\bar{x})G = 0. \tag{4.62}$$

Let $d_{\bar{S}} := H + G$ and $\widetilde{d}_{\bar{S}} := U^T d_{\bar{S}} V$. Then combining (4.54) and (4.59) with Lemma 4.3, we have $(d_{\bar{x}}, d_{\bar{y}}, d_{\bar{S}}) \in \widetilde{\mathcal{C}}(\bar{x}, \bar{y}, \bar{S})$. This further indicates that $0 \neq d_{\bar{x}} \in \mathcal{C}(\bar{x})$ of problem (4.1).

Therefore, by making use of the assumption (ii), we have

$$
\begin{aligned}
& \langle d_{\bar{x}}, \nabla^2_{xx} l(\bar{x},\bar{y},\bar{S}) d_{\bar{x}}\rangle + \Upsilon_{g(\bar{x})}\big(\bar{S}, g'(\bar{x})d_{\bar{x}}\big) \\
= \;& -\langle d_{\bar{y}}, h'(\bar{x})d_{\bar{x}}\rangle - \langle d_{\bar{S}}, g'(\bar{x})d_{\bar{x}}\rangle + \Upsilon_{g(\bar{x})}\big(\bar{S}, g'(\bar{x})d_{\bar{x}}\big) \\
= \;& -\langle G, g'(\bar{x})d_{\bar{x}}\rangle - \langle H, g'(\bar{x})d_{\bar{x}}\rangle + \Upsilon_{g(\bar{x})}\big(\bar{S}, g'(\bar{x})d_{\bar{x}}\big) \\
= \;& -\langle G, g'(\bar{x})d_{\bar{x}}\rangle - \langle [\mathcal{E}_S \circ \mathbb{S}(\widetilde{D}_{\bar{x},1}) + \mathcal{E}_{\mathbb{T}} \circ \mathbb{T}(\widetilde{D}_{\bar{x},1}) \quad \mathcal{F}\circ(\widetilde{D}_{\bar{x},2})], \widetilde{D}_{\bar{x}}\rangle + \Upsilon_{g(\bar{x})}\big(\bar{S}, g'(\bar{x})d_{\bar{x}}\big) \\
= \;& -\langle G, g'(\bar{x})d_{\bar{x}}\rangle - \langle \mathcal{E}_S \circ \mathbb{S}(\widetilde{D}_{\bar{x},1}), \mathbb{S}(\widetilde{D}_{\bar{x},1})\rangle - \langle \mathcal{E}_{\mathbb{T}} \circ \mathbb{T}(\widetilde{D}_{\bar{x},1}), \mathbb{T}(\widetilde{D}_{\bar{x},1})\rangle - \langle \mathcal{F}\circ(\widetilde{D}_{\bar{x},2}), \widetilde{D}_{\bar{x},2}\rangle \\
& + \Upsilon_{g(\bar{x})}\big(\bar{S}, g'(\bar{x})d_{\bar{x}}\big) \\
= \;& -\langle G, g'(\bar{x})d_{\bar{x}}\rangle + \Upsilon_{g(\bar{x})}\big(\bar{S}, g'(\bar{x})d_{\bar{x}}\big) \\
& - \sum_{1\le l,t\le r_0} \frac{2}{\nu_l(g(\bar{x})) + \nu_t(g(\bar{x}))}\|(\mathbb{T}(\widetilde{D}_{\bar{x},1}))_{a_l a_t}\|^2 - \sum_{1\le l\le r_0} \frac{4}{\nu_l(g(\bar{x}))}\|(\mathbb{T}(\widetilde{D}_{\bar{x},1}))_{a_l \beta_1}\|^2 \\
& - \sum_{\substack{1\le l\le r_0 \\ 1\le i-|\alpha|-|\beta_1|\le|\beta_2|}} \left(\frac{2(1-\sigma_i(\bar{S}))}{\nu_l(g(\bar{x}))}\|(\mathbb{S}(\widetilde{D}_{\bar{x},1}))_{a_l i}\|^2 + \frac{2(\sigma_i(\bar{S})+1)}{\nu_l(g(\bar{x}))}\|(\mathbb{T}(\widetilde{D}_{\bar{x},1}))_{a_l i}\|^2\right) \\
& - \sum_{1\le l\le r_0} \left(\frac{2}{\nu_l(g(\bar{x}))}\|(\mathbb{S}(\widetilde{D}_{\bar{x},1}))_{a_l \beta_3}\|^2 + \frac{2}{\nu_l(g(\bar{x}))}\|(\mathbb{T}(\widetilde{D}_{\bar{x},1}))_{a_l \beta_3}\|^2\right) \\
& - \sum_{1\le l\le r_0} \frac{1}{\nu_l(g(\bar{x}))}\|(\widetilde{D}_{\bar{x}})_{a_l c}\|^2 \\
= \;& \langle (\widetilde{D}_{\bar{x}})_{\beta_1\beta_1}, (\widetilde{d}_{\bar{S}})_{\beta_1\beta_1}\rangle + \Upsilon_{g(\bar{x})}\big(\bar{S}, g'(\bar{x})d_{\bar{x}}\big) - \Upsilon_{g(\bar{x})}\big(\bar{S}, g'(\bar{x})d_{\bar{x}}\big) \\
= \;& \langle (\widetilde{D}_{\bar{x}})_{\beta_1\beta_1}, (\widetilde{d}_{\bar{S}})_{\beta_1\beta_1}\rangle = \langle \Pi_\Phi(-g'(\bar{x})d_{\bar{x}}), \Pi_\Phi(d_{\bar{S}})\rangle = 0,
\end{aligned}
$$

which contradicts the assumption (iii) that the second-order sufficient condition (4.48) holds at $\bar{x}$ with respect to the multiplier $(\bar{y},\bar{S}) \in \mathcal{M}(\bar{x})$. Hence, there exist a constant $\kappa > 0$ and a neighborhood $\mathcal{U}$ of $(\bar{x},\bar{y},\bar{S})$ such that (4.52) holds.

The following proof is essentially the same as Theorem 3.1, we still present here for the completeness.

Next, we will prove that $\mathcal{T}_l$ is metrically subregular at $(\bar{x},\bar{y},\bar{S})$ for the origin under an additional assumption, which requires that there exist $(\widehat{y},\widehat{S}) \in \mathcal{M}(\bar{x})$ such that $0 < \sigma_\beta(\widehat{S}) < e_\beta$ and $\sum_{i\in\beta}\sigma_i(\widehat{S}) < m - |\alpha|$. In another word, it is equivalent to show that there exist a constant $\kappa' > 0$ and a neighborhood $\mathcal{U}'$ of $(\bar{x},\bar{y},\bar{S})$ such that for any $((u, v, C) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^{m\times n}$,

$$\mathrm{dist}((x,y,S),\mathcal{T}_l^{-1}(0)) \leq \kappa' \|((u,v,C)\|,\ \forall (x,y,S) \in \mathcal{T}_l^{-1}((u,v,C) \cap \mathcal{U}'. \tag{4.63}$$

For the convenience, we set

$$\Psi := \{(y,S)|\ (\bar{x},y,S) \in \mathcal{T}_l^{-1}(0,0,0)\},$$

$$\Xi_1 := \{(y,S)|\ \nabla f(\bar{x}) + \nabla h(\bar{x})y + \nabla g(\bar{x})S = 0\},\ \Xi_2 := \{(y,S)|\ S \in \partial\theta(g(\bar{x}))\}.$$

One can easily find that $\Psi = \Xi_1 \cap \Xi_2$ and $(\widehat{y}, \widehat{S}) \in \Xi_1 \cap \mathrm{ri}(\Xi_2)$. Thus, by Proposition 2.1, we have that there exists a constant $\kappa_1 > 0$ such that for any $(x,y,S) \in \mathcal{U}'$,

$$\mathrm{dist}((y,S),\Phi) \leq \kappa_1\big(\mathrm{dist}((y,S),\Xi_1) + \mathrm{dist}((y,S),\Xi_2)\big). \tag{4.64}$$

For any given point $(x,y,S) \in \mathcal{T}_l^{-1}((u_1,u_2,C)\cap\mathcal{U}'$, we assume that $\|(y,S)\| \leq \eta$ with some $\eta > 0$ by shrinking $\mathcal{U}'$ if necessary. Fixing that given point, using Hoffman's error bound and the twice continuous differentiability of f, h and g, shrinking $\mathcal{U}'$ if necessary, we obtain that there exist constants $\kappa_2 > 0$ and $\kappa_2' > 0$ such that

$$\begin{aligned}\mathrm{dist}((y,S),\Xi_1) &\leq \kappa_2\|\nabla f(\bar{x}) + \nabla h(\bar{x})y + \nabla g(\bar{x})S\| \\ &\leq \kappa_2(\|\nabla f(x) - \nabla f(\bar{x})\| + \|\nabla h(x) - \nabla h(\bar{x})\|\|y\| \\ &\quad + \|\nabla g(x) - \nabla g(\bar{x})\|\|S\| + \|u\|) \\ &\leq \kappa_2'(\|x - \bar{x}\| + \|u\|).\end{aligned} \tag{4.65}$$

By Proposition 2.15, we have $(\partial\theta)^{-1}(\cdot) = \partial\theta^*(\cdot)$ is metrically subregular at $\bar{S}$ for $g(\bar{x})$. Together with $g(x) + C \in \partial\theta^*(S)$ and the twice continuous differentiabilty of g , we can duduce, shrinking $\mathcal{U}'$ if necessary, that there exist contants $\kappa_3 > 0$ and $\kappa_3' > 0$ such that

$$\begin{aligned}\mathrm{dist}((y,S),\Xi_2) &= \mathrm{dist}(S,\partial\theta(g(\bar{x}))) \\ &\leq \kappa_3\mathrm{dist}(g(\bar{x}),\partial\theta^*(S)) \\ &\leq \kappa_3\|g(x) + C - g(\bar{x})\| \\ &\leq \kappa_3'(\|x - \bar{x}\| + \|C\|).\end{aligned} \tag{4.66}$$

Therefore, we can find that there exist a constant $\kappa' > 0$ and a neighborhood $\mathcal{U}'$ of $(\bar{x}, \bar{y}, \bar{S})$ such that (4.63) holds, by using the inequalities (4.52) and (4.64)-(4.66). This implies $\mathcal{T}_l$ is metrically subregular at $(\bar{x}, \bar{y}, \bar{S})$ for the origin. $\square$

Remark 4.2. *Analogously, by the same reasons stated in Remark 3.1, we can conduct our above analysis under $C^{1,1}$ program data. In this situation, if the gradients ∇f, ∇g and ∇h are directionally differentiable at $\bar{x}$, then the corresponding SOSC (4.48) in the assumption (iii) of Theorem 4.1 changes to the following form:*

$$\langle d, (\nabla_x l)'(\bar{x}, \bar{y}, \bar{S}; d)\rangle + \Upsilon_{g(\bar{x})}\big(\bar{S}, g'(\bar{x})d\big) > 0, \ \forall\, 0 \neq d \in \mathcal{C}(\bar{x}) \tag{4.67}$$

with $\Upsilon_{g(\bar{x})}\big(\bar{S}, g'(\bar{x})d\big)$ defined as (4.49), where $(\nabla_x l)'(\bar{x}, \bar{y}, \bar{S}; d)$ denotes the directional derivative of $\nabla_x l(\cdot, \bar{y}, \bar{S})$ at $\bar{x}$ in the direction d. Thus, by keeping the rest assumptions and following our above discussions, one can easily get the metric subregularity of $\mathcal{T}_l$ at $(\bar{x}, \bar{y}, \bar{S})$ for the origin.

Remark 4.3. *The same reasons as in Remark 3.2, if $|\beta_1| = 0$ or $|\beta_1| = 1$, assumption (i) can be omitted here. Moreover, it can be found in the proof that, by assumption (ii), $0 = \langle \Pi_\Phi(-g'(\bar{x})d_{\bar{x}}), \Pi_\Phi(d_{\bar{S}})\rangle = -\langle (\widetilde{D}_{\bar{x}})_{\beta_1\beta_1}, (\widetilde{d}_{\bar{S}})_{\beta_1\beta_1}\rangle$. Therefore, if $\beta_1 = \emptyset$, assumption (ii) can be omitted here. Additionally, if $\sigma_m(g(\bar{x})) > 0$ holds, the problem (4.1) reduces to a smooth problem.*

We extend the results of the SDP cone to the nuclear norm here without adding an extra condition. The perturbation property of the nuclear norm helps us to obtain the results. One can see that the second order information revealed by the perturbation property is closely related to the sigma term in the SOSC, which is the conjugate of the parabolic second order directional derivative of the nuclear norm. After these results, we want to cover more useful models in optimizaiton by studying the properties of the Ky Fan k-norm. Since the nuclear norm is a particular case of the Ky Fan k-norm, we only need to cover the other two cases in the next chapter.

Chapter 5

The metric subregularity of the KKT solution mapping for composite Ky Fan k-norm problem

In this chapter, we will extend the nuclear norm case in Chapter 4 to the following Ky Fan k-norm case, where consider (1.1) with θ chosen as the Ky Fan k-norm on $\mathcal{R}^{m\times n}$ $(m \le n)$.

We can restate the problem (1.1) as

$$\begin{aligned} \min \quad & f(x) + \|(g(x))\|_{(k)} \\ \text{s.t.} \quad & h(x) = 0, \end{aligned} \tag{5.1}$$

where $f : \mathcal{X} \to \mathcal{R}$ is twice continuously differentiable function, $h : \mathcal{X} \to \mathcal{Y}$ and $g : \mathcal{X} \to \mathcal{R}^{m\times n}$ are twice continuously differentiable mappings, $\mathcal{X}$ and $\mathcal{Y}$ are finite dimensional real Euclidean spaces, and $\theta : \mathcal{R}^{m\times n} \to \mathcal{R}$ denotes the Ky Fan k-norm function with $\theta(X) = \|X\|_{(k)}$ for all $X \in \mathcal{R}^{m\times n}$.

The problem (4.1) is one particular case of the problem (5.1). Thus, the various modifications of the problem (5.1) include all the applications of the nuclear norm regularized problem (4.1). Howerver, not only the aforementioned applications in Chapter 4, there are also other problems can be modified in the form of (5.1), such as Lasso problems, rank minimization, matrix completion, machine learning , etc [15, 16, 47, 60, 92–94, 105]. Similar to the nuclear norm regularized problem, we want to characterize the stability of the problem (5.1) allowing the multipliers to be nonunique.

5.1 The sensitivity analysis of the Ky Fan k-norm

The structure of this section is the same as Section 4.1. Thus, the analysis in this section is an extension of which in Section 4.1.

Let $A, B \in \mathcal{R}^{m\times n}$ satisfying $B \in \partial\theta(A)$ and denote $M := A + B$. A well known equivalent form [67] is given by

$$A = \text{Prox}_{\theta}(M), \quad B = \text{Prox}_{\theta^*}(M). \tag{5.2}$$

Suppose that M admits the following singular-value decomposition (SVD):

$$M = U[\Sigma(M)\ 0]V^T = U[\Sigma(M)\ 0][V_1\ V_2]^T = U\Sigma(M)V_1^T, \tag{5.3}$$

where $U \in \mathcal{O}^m$, $V := [V_1\ \ V_2] \in \mathcal{O}^n$ with $V_1 \in \mathcal{R}^{n\times m}$ and $V_2 \in \mathcal{R}^{n\times(n-m)}$ are the singular vectors of M, and $\Sigma(M) := \text{Diag}(\sigma_1(M), \sigma_2(CM), \ldots, \sigma_m(M))$ are the singular values of M with $\sigma_1(M) \geq \sigma_2(M) \geq \ldots \geq \sigma_m(M) \geq 0$ being arranged in a non-increasing order. For simplicity, we let $\sigma(M) := (\sigma_1(M), \sigma_2(M), \ldots, \sigma_m(M))$.

It is known by [25] that given the SVD of M as (5.3), A and B admit the following SVD:

$$\begin{aligned} A &= U[\Sigma(A)\ 0]V^T = U\Sigma(A)V_1^T, \\ B &= U[\Sigma(B)\ 0]V^T = U\Sigma(B)V_1^T, \end{aligned} \tag{5.4}$$

where $\Sigma(A) := \text{Diag}(\sigma_1(A), \sigma_2(A), \ldots, \sigma_m(A))$, $\Sigma(B) := \text{Diag}(\sigma_1(B), \sigma_2(B), \ldots, \sigma_m(B))$ and

$$\sigma_i(B) = \sigma_i(M) - \sigma_i(A), \quad i = 1, 2, \ldots, m, \tag{5.5}$$

with $\sigma_1(A) \geq \sigma_2(A) \geq \ldots \geq \sigma_m(A) \geq 0$ and $\sigma_1(B) \geq \sigma_2(B) \geq \ldots \geq \sigma_m(B) \geq 0$. Similarly, we set

$$\sigma(A) := (\sigma_1(A), \sigma_2(A), \ldots, \sigma_m(A)),$$

$$\sigma(B) := (\sigma_1(B), \sigma_2(B), \ldots, \sigma_m(B)).$$

For simplicity of the subsequent discussions, we define the following three index sets:

$$a := \{1 \le i \le m : \sigma_i(M) > 0\}, \quad b := \{1 \le i \le m : \sigma_i(M) = 0\}, \quad c := \{m+1, \ldots, n\}. \tag{5.6}$$

To further refine the nonzero singular values of M, we let $\nu_1(M) > \nu_2(M) > \ldots > \nu_r(M) > 0$ with some nonnegative integer r be the distinct nonzero singular values of M. Thus, we can divide the set a as

$$a = \bigcup_{1\le l\le r} a_l, \quad a_l := \{i \in a : \sigma_i(M) = \nu_l(M)\}, \quad l = 1, 2, \ldots, r. \tag{5.7}$$

To obtain the relationships among the sigular values of A, B and M, we shall adopt the following lemma, which can be derived directly from the characterization in [69, 102].

Lemma 5.1. *Suppose $\sigma(A)$ and $\sigma(B)$ are singular values of A and B respectively. Then $B \in \partial\theta(A)$ if and only if $\sigma(A)$ and $\sigma(B)$ satisfy the following conditions:*

(i) If $\sigma_k(A) > 0$, then

$$\sigma_\alpha(B) = e_\alpha, \quad 0 \le \sigma_\beta(B) \le e_\beta \quad \sum_{i\in\beta}\sigma_i(B) = k - k_0 \quad \text{and} \quad \sigma_\gamma(B) = 0, \tag{5.8}$$

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

$$\begin{aligned} &\sigma_1(A) \ge \ldots \ge \sigma_{k_0}(A) > \sigma_{k_0+1}(A) = \ldots = \sigma_k(A) = \ldots = \sigma_{k_1}(A) \\ &> \sigma_{k_1+1}(A) \ge \ldots \ge \sigma_m(A) \ge 0, \end{aligned} \tag{5.9}$$

and

$$\alpha = \{1, \ldots, k_0\}, \quad \beta = \{k_0+1, \ldots, k_1\} \quad \text{and} \quad \gamma = \{k_1+1, \ldots, m\}. \tag{5.10}$$

(ii) If $\sigma_k(A) = 0$, then

$$\sigma_\alpha(B) = e_\alpha, \quad 0 \le \sigma_\beta(B) \le e_\beta \quad \text{and} \quad \sum_{i\in\beta}\sigma_i(B) \le k - k_0, \tag{5.11}$$

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

$$\sigma_1(A) \ge \ldots \ge \sigma_{k_0}(A) > \sigma_{k_0+1}(A) = \ldots = \sigma_k(A) = \ldots = \sigma_m(A) = 0, \quad (5.12)$$

and

$$\alpha = \{1, \ldots, k_0\} \quad \text{and} \quad \beta = \{k_0+1, \ldots, m\}. \quad (5.13)$$

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

$$\begin{gathered}\beta_1 := \{i \in \beta : \sigma_i(B) = 1\}, \quad \beta_2 := \{i \in \beta : 0 < \sigma_i(B) < 1\} \\ \text{and} \quad \beta_3 := \{i \in \beta : \sigma_i(B) = 0\}.\end{gathered} \quad (5.14)$$

For $M = A + B$, let the index sets a, b, c and a_l, $l = 1, \ldots, r$ defined by (5.6) and (5.7) with respect to M. From the above Lemma 5.1 and (5.5), we have the following relationships among index sets a_l, b, α, β_i and γ, $i = 1, 2, 3$, $l = 1, \ldots, r$. For the sake of convenience, we set $a_{r+1} = b$.
If $\sigma_k(A) > 0$, then there exist integers $r_0 \le r_1 \in \{1, \ldots, r+1\}$, $r_0 \le \widetilde{r}_0 \le r_0 + 1$ and $r_1 - 1 \le \widetilde{r}_1 \le r_1$ such that

$$\alpha = \bigcup_{l=1}^{r_0} a_l, \ \beta_1 = \bigcup_{l=r_0+1}^{\widetilde{r}_0} a_l, \ \beta_2 = \bigcup_{l=\widetilde{r}_0+1}^{\widetilde{r}_1} a_l, \ \beta_3 = \bigcup_{l=\widetilde{r}_1+1}^{r_1} a_l \text{ and } \gamma = \bigcup_{l=r_1+1}^{r+1} a_l, \quad (5.15)$$

if $\sigma_k(A) = 0$, then there exist integers $r_0 \in \{0, 1, \ldots, r+1\}$ and $r_0 \le \widetilde{r}_0 \le r_0 + 1$ such that

$$\alpha = \bigcup_{l=1}^{r_0} a_l, \quad \beta_1 = \bigcup_{l=r_0+1}^{\widetilde{r}_0} a_l, \quad \beta_2 = \bigcup_{l=\widetilde{r}_0+1}^{r} a_l \quad \text{and} \quad \beta_3 = b. \quad (5.16)$$

Moreover, we can have the singular values classification of A and B by the above observations.
Namely, if $\sigma_k(A) > 0$, the distinct nonzero singular values of A can be denoted as $\nu_1(A) > \nu_2(A) > \ldots > \nu_{r_0}(A) > \sigma_k(A) > \nu_{r_1+1}(A) > \ldots > \nu_r(A) > 0$ and the distinct nonzero singular values of B can be denoted as $1 > \nu_{\widetilde{r}_0+1}(B) > \ldots > \nu_{\widetilde{r}_1}(B) > 0$, with

$$\begin{aligned} &a_l = \{1 \le i \le m : \ \sigma_i(A) = \nu_l(A)\}, \ l = 1, \ldots, r_0, \ r_1+1, \ldots, r, \\ &\beta = \{1 \le i \le m : \ \sigma_i(A) = \sigma_k(A)\}, \ \alpha \cup \beta_1 = \{1 \le i \le m : \ \sigma_i(B) = 1\}, \\ &a_l = \{1 \le i \le m : \ \sigma_i(B) = \nu_l(B)\}, \ l = \widetilde{r}_0 + 1, \ldots, \widetilde{r}_1. \end{aligned} \quad (5.17)$$

Analogouly, if $\sigma_k(A) = 0$, the distinct nonzero singular values of A can be denoted as $\nu_1(A) > \nu_2(A) > \ldots > \nu_{r_0}(A) > 0$ and the distinct nonzero singular values of B can be denoted as $1 > \nu_{\widetilde{r}_0+1}(B) > \ldots > \nu_r(B) > 0$, with

$$
\begin{aligned}
&a_l = \{1 \le i \le m : \ \sigma_i(A) = \nu_l(A)\}, \ l = 1, \ldots, r_0, \\
&\beta = \{1 \le i \le m : \ \sigma_i(A) = 0\}, \ \alpha \cup \beta_1 = \{1 \le i \le m : \ \sigma_i(B) = 1\}, \\
&a_l = \{1 \le i \le m : \ \sigma_i(B) = \nu_l(B)\}, \ l = \widetilde{r}_0 + 1, \ldots, r.
\end{aligned}
\tag{5.18}
$$

Based on this lemma, it is easy to find the following observations.

Proposition 5.1. *Let $A \in \mathcal{R}^{m\times n}$ and $B \in \partial\theta(A)$. Suppose that A and B have the SVD as in (5.4), then it holds that*

(a) $\partial\theta(A)$ is a polyhedral set if and only if $\sigma_k(A) > \sigma_{k+1}(A)$ (where $\sigma_{m+1}(A)$ is assigned to be 0).

(b) $B \in \mathrm{ri}(\partial\theta(A))$ if and only if

(i) if $\sigma_k(A) > 0$, then $0 < \sigma_\beta(B) < e_\beta$;

(ii) if $\sigma_k(A) = 0$, then $0 < \sigma_\beta(B) < e_\beta$ and $\sum\limits_{i\in\beta}\sigma_i(B) < k - k_0$.

Remark 5.1. *One can find that if $\sigma_k(A) > \sigma_{k+1}(A)$ in part (a) holds, then θ is differentiable at A [103]. In this case, problem (4.1) turns to a smooth optimization problem.*

The same reasons as the nuclear norm, we have $\theta^{\downarrow}(X, \cdot) = \theta'(X; \cdot)$ for any $X \in \mathcal{R}^{m\times n}$ with $\theta(\cdot) = \|\cdot\|_{(k)}$. Thus, all the analysis can be conducted regarding to the conventional directional derivative of the Ky Fan k-norm.

Let us recall the set valued mapping (2.17) at point (A, B) satisfying $B \in \partial\theta(A)$ with $\theta(\cdot) = \|\cdot\|_{(k)}$,

$$
\mathcal{C}_\theta(A, B) := \{H \in \mathcal{R}^{m\times n} : \theta'(A; H) = \langle H, B\rangle\}. \tag{5.19}
$$

Here, we call $\mathcal{C}_\theta(A, B)$ the critical cone of $\partial\theta(A)$ at $A+B$, associated with $B \in \partial\theta(A)$. Then, we can obtain the following characterization of the critical cone $\mathcal{C}_\theta(A, B)$ from [25, proposition 10].

Lemma 5.2. *Suppose $A, B \in \mathcal{R}^{m\times n}$ satisfy $B \in \partial\theta(A)$ and the index sets $\alpha, \beta, \beta_1, \beta_2$, β_3, γ and c are defined as in Lemma 5.1 and (5.14). Given any $H \in \mathcal{R}^{m\times n}$, denote $\widetilde{H} = U^T HV = [U^T HV_1 \; U^T HV_2] = [\widetilde{H}_1 \; \widetilde{H}_2]$ for U, V satisfying (5.3). Then $H \in \mathcal{C}_\theta(A, B)$ if and only if*

(a) If $\sigma_k(A) > 0$, then there exists some $\tau \in \mathcal{R}$ such that

$$\lambda_{|\beta_1|}(\mathbb{S}(\widetilde{H}_1)_{\beta_1\beta_1}) \geq \tau \geq \lambda_1(\mathbb{S}(\widetilde{H}_1)_{\beta_3\beta_3}) \tag{5.20}$$

and $\widetilde{H}$ has the following block structure

$$\widetilde{H} = \left(\begin{array}{c|ccc|c} \widetilde{H}_{\alpha\alpha} & & \widetilde{H}_{\alpha\beta} & & \widetilde{H}_{\alpha(\gamma\cup c)} \\ \hline & \widetilde{H}_{\beta_1\beta_1} & \mathbb{T}(\widetilde{H}_1)_{\beta_1\beta_2} & \mathbb{T}(\widetilde{H}_1)_{\beta_1\beta_3} & \\ \widetilde{H}_{\beta\alpha} & \mathbb{T}(\widetilde{H}_1)_{\beta_2\beta_1} & \tau I_{|\beta_2|} + \mathbb{T}(\widetilde{H}_1)_{\beta_2\beta_2} & \mathbb{T}(\widetilde{H}_1)_{\beta_2\beta_3} & \widetilde{H}_{\beta(\gamma\cup c)} \\ & \mathbb{T}(\widetilde{H}_1)_{\beta_3\beta_1} & \mathbb{T}(\widetilde{H}_1)_{\beta_3\beta_2} & \widetilde{H}_{\beta_3\beta_3} & \\ \hline \widetilde{H}_{\gamma\alpha} & & \widetilde{H}_{\gamma\beta} & & \widetilde{H}_{\gamma(\gamma\cup c)} \end{array}\right). \tag{5.21}$$

(b) If $\sigma_k(A) = 0$ and $||B||_* = k$, then there exists some $\tau \geq 0$ such that

$$\lambda_{|\beta_1|}(\mathbb{S}(\widetilde{H}_1)_{\beta_1\beta_1}) \geq \tau \geq \sigma_1\big([\widetilde{H}_{bb} \; \widetilde{H}_{bc}]\big) \tag{5.22}$$

and $\widetilde{H}$ has the following block structure

$$\widetilde{H} = \left(\begin{array}{c|ccc|c} \widetilde{H}_{\alpha\alpha} & & \widetilde{H}_{\alpha\beta} & & \widetilde{H}_{\alpha c} \\ \hline & \mathbb{S}(\widetilde{H}_1)_{\beta_1\beta_1} & 0 & 0 & 0 \\ \widetilde{H}_{\beta\alpha} & 0 & \tau I_{|\beta_2|} & 0 & 0 \\ & 0 & 0 & \widetilde{H}_{bb} & \widetilde{H}_{bc} \end{array}\right). \tag{5.23}$$

(c) If $\sigma_k(A) = 0$ and $||B||_* < k$, then $\mathbb{S}(\widetilde{H}_1)_{\beta_1\beta_1} \succeq 0$ and $\widetilde{H}$ has the following block structure

$$\widetilde{H} = \left(\begin{array}{c|ccc|c} \widetilde{H}_{\alpha\alpha} & & \widetilde{H}_{\alpha\beta} & & \widetilde{H}_{\alpha c} \\ \hline & \mathbb{S}(\widetilde{H}_1)_{\beta_1\beta_1} & 0 & 0 & \\ \widetilde{H}_{\beta\alpha} & 0 & 0_{\beta_2\beta_2} & 0 & 0 \\ & 0 & 0 & 0_{\beta_3\beta_3} & \end{array}\right). \tag{5.24}$$

In above, $\mathbb{S}(\cdot)$ *and* $\mathbb{T}(\cdot)$ *are defined as (4.15).*

Next, let us consider the Fenchel conjugate function θ^* of θ. By the equivalence of $B \in \partial\theta(A)$ and $A \in \partial\theta^*(B)$, we can similarly define the critical cone $\mathcal{C}_{\theta^*}(B, A)$ of $\partial\theta^*(B)$ at $A+B$ associated with $A \in \partial\theta^*(B)$. One can find a directly derive in [25]. The critical cone $\mathcal{C}_{\theta^*}(B, A)$ is defined as

$$\mathcal{C}_{\theta^*}(B, A) := \{H \in \mathcal{R}^{m\times n} : \vartheta'(B; H) = \langle H, A\rangle = 0\}, \tag{5.25}$$

where $\vartheta(\cdot)$ denotes the dual norm of the Ky Fan k-norm θ. The relationships between the conjugate and the dual norm of the Ky Fan k-norm are well studied in [25]. One can find the following characterization of the critical cone $\mathcal{C}_{\theta^*}(B, A)$ in [25, proposition 12].

Lemma 5.3. *Suppose that all the assumptions in Lemma 5.2 hold here. Then* $H \in \mathcal{C}_{\theta^*}(B, A)$ *if and only if*

(a) *If* $\sigma_k(A) > 0$, *then*

$$\operatorname{tr}(\widetilde{H}_{\beta\beta}) = 0, \quad \mathbb{S}(\widetilde{H}_1)_{\beta_1\beta_1} \preceq 0, \quad \mathbb{S}(\widetilde{H}_1)_{\beta_3\beta_3} \succeq 0 \tag{5.26}$$

and $\widetilde{H}$ *admits the following block structure:*

$$\widetilde{H} = \begin{pmatrix} \mathbb{T}(\widetilde{H}_{\alpha\alpha}) & \mathbb{T}(\widetilde{H}_{\alpha\beta_1}) & \widetilde{H}_{\alpha\beta_2} & \widetilde{H}_{\alpha\beta_3} & \widetilde{H}_{\alpha(\gamma\cup c)} \\ \mathbb{T}(\widetilde{H}_{\beta_1\alpha}) & \widetilde{H}_{\beta_1\beta_1} & \widetilde{H}_{\beta_1\beta_2} & \widetilde{H}_{\beta_1\beta_3} & \widetilde{H}_{\beta_1(\gamma\cup c)} \\ \widetilde{H}_{\beta_2\alpha} & \widetilde{H}_{\beta_2\beta_1} & \widetilde{H}_{\beta_2\beta_2} & \widetilde{H}_{\beta_2\beta_3} & \widetilde{H}_{\beta_2(\gamma\cup c)} \\ \widetilde{H}_{\beta_3\alpha} & \widetilde{H}_{\beta_3\beta_1} & \widetilde{H}_{\beta_3\beta_2} & \mathbb{S}(\widetilde{H}_1)_{\beta_3\beta_3} & 0 \\ \widetilde{H}_{\gamma\alpha} & \widetilde{H}_{\gamma\beta_1} & \widetilde{H}_{\gamma\beta_2} & 0 & 0 \end{pmatrix}. \tag{5.27}$$

(b) *If* $\sigma_k(A) = 0$ *and* $||B||_* = k$, *then*

$$\operatorname{tr}(\widetilde{H}_{(\beta_1\cup\beta_2)(\beta_1\cup\beta_2)}) + ||[\widetilde{H}_{bb}\ \widetilde{H}_{bc}]||_* \leq 0, \quad \mathbb{S}(\widetilde{H}_1)_{\beta_1\beta_1} \preceq 0 \tag{5.28}$$

and $\widetilde{H}$ *admits the following block structure:*

$$\widetilde{H} = \left(\begin{array}{cccc|c} \mathbb{T}(\widetilde{H}_{\alpha\alpha}) & \mathbb{T}(\widetilde{H}_{\alpha\beta_1}) & \widetilde{H}_{\alpha\beta_2} & \widetilde{H}_{\alpha\beta_3} & \\ \mathbb{T}(\widetilde{H}_{\beta_1\alpha}) & \widetilde{H}_{\beta_1\beta_1} & \widetilde{H}_{\beta_1\beta_2} & \widetilde{H}_{\beta_1\beta_3} & \widetilde{H}_2 \\ \widetilde{H}_{\beta_2\alpha} & \widetilde{H}_{\beta_2\beta_1} & \widetilde{H}_{\beta_2\beta_2} & \widetilde{H}_{\beta_2\beta_3} & \\ \widetilde{H}_{\beta_3\alpha} & \widetilde{H}_{\beta_3\beta_1} & \widetilde{H}_{\beta_3\beta_2} & \widetilde{H}_{\beta_3\beta_3} & \end{array}\right). \tag{5.29}$$

(c) If $\sigma_k(A) = 0$ and $||B||_* < k$, then $\mathbb{S}(\widetilde{H}_1)_{\beta_1\beta_1} \preceq 0$ and $\widetilde{H}$ admits the following block structure:

$$\widetilde{H} = \left(\begin{array}{c:c:c:c:c} \mathbb{T}(\widetilde{H}_{\alpha\alpha}) & \mathbb{T}(\widetilde{H}_{\alpha\beta_1}) & \widetilde{H}_{\alpha\beta_2} & \widetilde{H}_{\alpha\beta_3} & \\ \hdashline \mathbb{T}(\widetilde{H}_{\beta_1\alpha}) & \widetilde{H}_{\beta_1\beta_1} & \widetilde{H}_{\beta_1\beta_2} & \widetilde{H}_{\beta_1\beta_3} & \\ \hdashline \widetilde{H}_{\beta_2\alpha} & \widetilde{H}_{\beta_2\beta_1} & \widetilde{H}_{\beta_2\beta_2} & \widetilde{H}_{\beta_2\beta_3} & \widetilde{H}_2 \\ \hdashline \widetilde{H}_{\beta_3\alpha} & \widetilde{H}_{\beta_3\beta_1} & \widetilde{H}_{\beta_3\beta_2} & \widetilde{H}_{\beta_3\beta_3} & \end{array}\right). \tag{5.30}$$

In above, $\mathbb{S}(\cdot)$ and $\mathbb{T}(\cdot)$ are defined as (4.15).

An extensive study of the perturbation properties of the Ky Fan k-norm is conducted here. For the sake of convenience for reading, we divided them into two parts regarding to $\sigma_k(A) > 0$ and $\sigma_k(A) = 0$.

Proposition 5.2. *Let $A \in \mathcal{R}^{m\times n}$ and $B \in \partial\theta(A)$ with $\sigma_k(A) > 0$. Suppose that A and B have the SVD as in (5.4) and the index sets $\alpha, \beta, \beta_1, \beta_2, \beta_3, \gamma$ and c are defined as in Lemma 5.1 and (5.14). Then for all $(A', B') \in \mathcal{R}^{m\times n} \times \mathcal{R}^{m\times n}$ satisfying $B' \in \partial\theta(A')$ and is sufficiently close to $(A, B) \in \mathcal{R}^{m\times n} \times \mathcal{R}^{m\times n}$, we have*

$$\begin{cases} \widetilde{A}'_{\alpha\alpha} = \Sigma(A)_{\alpha\alpha} + O(\|\Delta A\|),\ \widetilde{A}'_{\alpha\beta_1} = O(\|\Delta A\|),\ \widetilde{A}'_{\beta_1\alpha} = O(\|\Delta A\|), \\ \widetilde{A}'_{\alpha\beta_2} = O(\min\{\|\Delta A\|, \|\Delta B\|\}),\ \widetilde{A}'_{\beta_2\alpha} = O(\min\{\|\Delta A\|, \|\Delta B\|\}), \\ \widetilde{A}'_{\alpha\beta_3} = O(\min\{\|\Delta A\|, \|\Delta B\|\} + \|\Delta A\|^2),\ \widetilde{A}'_{\beta_3\alpha} = O(\min\{\|\Delta A\|, \|\Delta B\|\} + \|\Delta A\|^2), \\ \widetilde{A}'_{\alpha(\gamma\cup c)} = O(\min\{\|\Delta A\|, \|\Delta B\|\}),\ \widetilde{A}'_{\gamma\alpha} = O(\min\{\|\Delta A\|, \|\Delta B\|\}), \\ \widetilde{A}'_{\beta_1\beta_1} = \sigma_k(A)I_{|\beta_1|} + O(\|\Delta A\| + \|\Delta B\|),\ \widetilde{A}'_{\beta_1(\beta_2\cup\beta_3)} = O(\|\Delta B\|), \\ \widetilde{A}'_{\beta_2\beta_1} = O(\|\Delta B\|),\ \widetilde{A}'_{\beta_2\beta_2} = \sigma_k(A)I_{|\beta_2|} + O(\|\Delta B\|),\ \widetilde{A}'_{\beta_2\beta_3} = O(\|\Delta B\|), \\ \widetilde{A}'_{\beta_3(\beta_1\cup\beta_2)} = O(\|\Delta B\|),\ \widetilde{A}'_{\beta_3\beta_3} = \sigma_k(A)I_{|\beta_3|} + O(\|\Delta A\| + \|\Delta B\|), \\ \widetilde{A}'_{\beta_1(\gamma\cup c)} = O(\min\{\|\Delta A\|, \|\Delta B\|\} + \|\Delta A\|^2),\ \widetilde{A}'_{\gamma\beta_1} = O(\min\{\|\Delta A\|, \|\Delta B\|\} + \|\Delta A\|^2), \\ \widetilde{A}'_{\beta_2(\gamma\cup c)} = O(\min\{\|\Delta A\|, \|\Delta B\|\}),\ \widetilde{A}'_{\gamma\beta_2} = O(\min\{\|\Delta A\|, \|\Delta B\|\}), \\ \widetilde{A}'_{\beta_3(\gamma\cup c)} = O(\|\Delta A\|),\ \widetilde{A}'_{\gamma\beta_3} = O(\|\Delta A\|), \\ \widetilde{A}'_{\gamma\gamma} = \Sigma(A)_{\gamma\gamma} + O(\|\Delta A\|),\ \widetilde{A}'_{\gamma c} = O(\|\Delta A\|), \end{cases} \tag{5.31}$$

and

$$\left\{\begin{array}{l}
\widetilde{B}'_{\alpha\alpha} = I_{|\alpha|} + O(\|\Delta A\|),\ \widetilde{B}'_{\alpha\beta_1} = O(\|\Delta A\|),\ \widetilde{B}'_{\beta_1\alpha} = O(\|\Delta A\|),\\
\widetilde{B}'_{\alpha\beta_2} = O(\min\{\|\Delta A\|, \|\Delta B\|\}),\ \widetilde{B}'_{\beta_2\alpha} = O(\min\{\|\Delta A\|, \|\Delta B\|\})\\
\widetilde{B}'_{\alpha\beta_3} = O(\min\{\|\Delta A\|, \|\Delta B\|\} + \|\Delta A\|^2),\ \widetilde{B}'_{\beta_3\alpha} = O(\min\{\|\Delta A\|, \|\Delta B\|\} + \|\Delta A\|^2),\\
\widetilde{B}'_{\alpha(\gamma\cup c)} = O(\min\{\|\Delta A\|, \|\Delta B\|\}),\ \widetilde{B}'_{\gamma\alpha} = O(\min\{\|\Delta A\|, \|\Delta B\|\})\\
\widetilde{B}'_{\beta_1\beta_1} = I_{|\beta_1|} + O(\|\Delta A\| + \|\Delta B\|),\ \widetilde{B}'_{\beta_1(\beta_2\cup\beta_3)} = O(\|\Delta B\|),\\
\widetilde{B}'_{\beta_2\beta_1} = O(\|\Delta B\|),\ \widetilde{B}'_{\beta_2\beta_2} = \Sigma(B)_{\beta_2\beta_2} + O(\|\Delta B\|),\ \widetilde{B}'_{\beta_2\beta_3} = O(\|\Delta B\|),\\
\widetilde{B}'_{\beta_3(\beta_1\cup\beta_2)} = O(\|\Delta B\|),\ \widetilde{B}'_{\beta_3\beta_3} = O(\|\Delta B\|),\\
\widetilde{B}'_{\beta_1(\gamma\cup c)} = O(\min\{\|\Delta A\|, \|\Delta B\|\} + \|\Delta A\|^2),\ \widetilde{B}'_{\gamma\beta_1} = O(\min\{\|\Delta A\|, \|\Delta B\|\} + \|\Delta A\|^2),\\
\widetilde{B}'_{\beta_2(\gamma\cup c)} = O(\min\{\|\Delta A\|, \|\Delta B\|\}),\ \widetilde{B}'_{\gamma\beta_2} = O(\min\{\|\Delta A\|, \|\Delta B\|\}),\\
\widetilde{B}'_{\beta_3(\gamma\cup c)} = O(\|\Delta A\|\|\Delta B\|),\ \widetilde{B}'_{\gamma\beta_3} = O(\|\Delta A\|\|\Delta B\|),\\
\widetilde{B}'_{\gamma(\gamma\cup c)} = O(\|\Delta A\|\|\Delta B\|).
\end{array}\right. \tag{5.32}$$

Moreover,

$$\left\{\begin{array}{l}
\mathbb{S}(\widetilde{B}'_1)_{\alpha\alpha} = I_{|\alpha|} + O(\|\Delta A\|^2),\ \mathbb{S}(\widetilde{B}'_1)_{\alpha\beta_1} = O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|),\\
\mathbb{T}(\widetilde{A}'_1)_{\alpha\alpha} = \frac{1}{2}\big(\Sigma(A)_{\alpha\alpha}\mathbb{T}(\widetilde{B}'_1)_{\alpha\alpha} + \mathbb{T}(\widetilde{B}'_1)_{\alpha\alpha}\Sigma(A)_{\alpha\alpha}\big) + O(\|\Delta A\|^2),\\
\mathbb{T}(\widetilde{A}'_1)_{\alpha\beta_1} = \frac{1}{2}\big(\Sigma(A)_{\alpha\alpha} + \sigma_k(A)I_{|\alpha|}\big)\mathbb{T}(\widetilde{B}'_1)_{\alpha\beta_1} + O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|),\\
\widetilde{A}'_{\alpha\beta_2} - (\widetilde{A}'_{\beta_2\alpha})^T\Sigma(B)_{\beta_2\beta_2} = \Sigma(A)_{\alpha\alpha}\widetilde{B}'_{\alpha\beta_2} - \sigma_k(A)(\widetilde{B}'_{\beta_2\alpha})^T + O(\|\Delta A\|\|\Delta B\|)\\
\widetilde{A}'_{\alpha\beta_3} = \Sigma(A)_{\alpha\alpha}\widetilde{B}'_{\alpha\beta_3} - \sigma_k(A)(\widetilde{B}'_{\beta_3\alpha})^T + O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|),\\
\widetilde{A}'_{\alpha\gamma} = \Sigma(A)_{\alpha\alpha}\widetilde{B}'_{\alpha\gamma} - (\widetilde{B}'_{\gamma\alpha})^T\Sigma(A)_{\gamma\gamma} + O(\|\Delta A\|\|\Delta B\|),\\
\widetilde{A}'_{\alpha c} = \Sigma(A)_{\alpha\alpha}\widetilde{B}'_{\alpha c} + O(\|\Delta A\|\|\Delta B\|),
\end{array}\right. \tag{5.33}$$

$$
\begin{cases}
\mathbb{S}(\widetilde{B}'_1)_{\beta_1\beta_1} \preceq I_{|\beta_1|},\ \mathbb{T}(\widetilde{B}'_1)_{\beta_1\beta_1} = \dfrac{1}{\sigma_k(A)}\mathbb{T}(\widetilde{A}'_1)_{\beta_1\beta_1} + O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|),\\
\mathbb{S}(\widetilde{A}'_1)_{\beta_1(\beta_2\cup\beta_3)} = O(\|\Delta A\|\|\Delta B\|),\ \mathbb{S}(\widetilde{A}'_1)_{\beta_2\beta_3} = O(\|\Delta A\|\|\Delta B\|),\\
\mathbb{T}(\widetilde{B}'_1)_{\beta_1\beta_2} = \dfrac{1}{2\sigma_k(A)}\mathbb{T}(\widetilde{A}'_1)_{\beta_1\beta_2}(I_{|\beta_2|} + \Sigma(B)_{\beta_2\beta_2}) + O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|),\\
\mathbb{T}(\widetilde{B}'_1)_{\beta_1\beta_3} = \dfrac{1}{2\sigma_k(A)}\mathbb{T}(\widetilde{A}'_1)_{\beta_1\beta_3} + O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|),\\
\mathbb{T}(\widetilde{B}'_1)_{\beta_2\beta_2} = \dfrac{1}{2\sigma_k(A)}(\Sigma(B)_{\beta_2\beta_2}\mathbb{T}(\widetilde{A}'_1)_{\beta_2\beta_2} + \mathbb{T}(\widetilde{A}'_1)_{\beta_2\beta_2}\Sigma(B)_{\beta_2\beta_2}) + O(\|\Delta A\|\|\Delta B\|),\\
\mathbb{T}(\widetilde{B}'_1)_{\beta_2\beta_3} = \dfrac{1}{2\sigma_k(A)}\Sigma(B)_{\beta_2\beta_2}\mathbb{T}(\widetilde{A}'_1)_{\beta_2\beta_3} + O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|),\\
\mathbb{T}(\widetilde{B}'_1)_{\beta_3\beta_3} = O(\|\Delta A\|\|\Delta B\|),\ (\widetilde{B}'_1)_{\beta_3\beta_3} + O(\|\Delta A\|\|\Delta B\|) \in \mathcal{S}_+^{|\beta_3|},\\
\widetilde{A}'_{\beta_1\gamma} = \sigma_k(A)\widetilde{B}'_{\beta_1\gamma} - (\widetilde{B}'_{\gamma\beta_1})^T\Sigma(A)_{\gamma\gamma} + O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|),\\
\widetilde{A}'_{\beta_1 c} = \sigma_k(A)\widetilde{B}'_{\beta_1 c} + O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|),\\
\widetilde{A}'_{\beta_2\gamma} = \Sigma(B)^{-1}_{\beta_2\beta_2}(\sigma_k(A)\widetilde{B}'_{\beta_2\gamma} - (\widetilde{B}'_{\gamma\beta_2})^T\Sigma(A)_{\gamma\gamma}) + O(\|\Delta A\|\|\Delta B\|),\\
\widetilde{A}'_{\beta_2 c} = \sigma_k(A)\Sigma(B)^{-1}_{\beta_2\beta_2}\widetilde{B}'_{\beta_2 c} + O(\|\Delta A\|\|\Delta B\|),
\end{cases}
\tag{5.34}
$$

and $\sigma_k(A') > 0$ *that*

$$
\begin{aligned}
\mathbb{S}(\widetilde{A}'_1)_{\beta_1\beta_1} + O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|) &\succeq \sigma_k(A')I_{|\beta_1|},\\
\mathbb{S}(\widetilde{A}'_1)_{\beta_2\beta_2} &= \sigma_k(A')I_{|\beta_2|} + O(\|\Delta A\|\|\Delta B\|),\\
\mathbb{S}(\widetilde{A}'_1)_{\beta_3\beta_3} + O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|) &\preceq \sigma_k(A')I_{|\beta_3|},\\
\mathrm{tr}(\widetilde{B}'_{\beta\beta}) &= \mathrm{tr}(\Sigma(B)_{\beta\beta}) + O(\|\Delta A\|^2).
\end{aligned}
\tag{5.35}
$$

In above statement, we denote $\Delta A := A' - A$, $\Delta B := B' - B$, $\widetilde{A}' := U^T A' V = [\widetilde{A}'_1\ \widetilde{A}'_2] = [U^T A' V_1\ U^T A' V_2]$, $\widetilde{B}' := U^T B' V = [\widetilde{B}'_1\ \widetilde{B}'_2] = [U^T B' V_1\ U^T B' V_2]$ *and* I_p *as the identity* p *by* p *matrix.*

Proof. From the above arguments about the SVD of A and B, it is easy to see that there exists $\widetilde{U} \in \mathcal{O}^m$ and $\widetilde{V} \in \mathcal{O}^n$ such that

$$\widetilde{A}' = \widetilde{U}[\Sigma(A')\ 0]\widetilde{V}^T, \quad \widetilde{B}' = \widetilde{U}[\Sigma(B')\ 0]\widetilde{V}^T.$$

Without loss of generality, we assume $\beta_2 \neq \emptyset$. By Proposition 4.2, we can see that for all ΔA and ΔB small enough, there exists $Q_1 \in \mathcal{O}^{|a|}$, $Q'_1 \in \mathcal{O}^{|b|}$, $Q''_1 \in \mathcal{O}^{n-|a|}$,

$Q_2 \in \mathcal{O}^{m-|\beta_3|-|\gamma|}$, $Q_2' \in \mathcal{O}^{|\beta_3|+|\gamma|}$ and $Q_2'' \in \mathcal{O}^{n-m+|\beta_3|+|\gamma|}$ such that

$$\widetilde{U} = \begin{pmatrix} Q_1 & 0 \\ 0 & Q_1' \end{pmatrix} + O(\|\Delta A\|) = \begin{pmatrix} Q_2 & 0 \\ 0 & Q_2' \end{pmatrix} + O(\|\Delta B\|), \tag{5.36}$$

$$\widetilde{V} = \begin{pmatrix} Q_1 & 0 \\ 0 & Q_1'' \end{pmatrix} + O(\|\Delta A\|) = \begin{pmatrix} Q_2 & 0 \\ 0 & Q_2'' \end{pmatrix} + O(\|\Delta B\|), \tag{5.37}$$

where $Q_1 = \mathrm{Diag}(P_1, \ldots, P_{r_0}, P_\beta, P_{r_1+1}, \ldots, P_r)$ and $Q_2 = \mathrm{Diag}(P_1', P_{\widetilde{r}_0+1}', \ldots, P_{\widetilde{r}_1}')$ are block diagonal orthogonal matrices with $P_l \in \mathcal{O}^{|a_l|}$, $l = 1, \ldots, r_0, r_1+1, \ldots, r$, $P_\beta \in \mathcal{O}^{|\beta|}$, $P_1' \in \mathcal{O}^{|\alpha|+|\beta_1|}$ and $P_t' \in \mathcal{O}^{|a_t|}$, $t = \widetilde{r}_0+1, \ldots, \widetilde{r}_1$. Moreover,

$$\begin{aligned}
\Sigma(A')_{a_l a_l} - \Sigma(A)_{a_l a_l} &= P_l^T \mathbb{S}(\Delta\widetilde{A}_{a_l a_l}) P_l + O(\|\Delta A\|^2), \quad l = 1, \ldots, r_0, r_1+1, \ldots r, \\
\Sigma(A')_{\beta\beta} - \sigma_k(A) I_{|\beta|} &= P_\beta^T \mathbb{S}(\Delta\widetilde{A}_{\beta\beta}) P_\beta + O(\|\Delta A\|^2), \\
[\Sigma(A')_{bb} - \Sigma(A)_{bb}\ 0] &= Q_1'^T [\Delta\widetilde{A}_{bb}\ \Delta\widetilde{A}_{bc}] Q_1'' + O(\|\Delta A\|^2),
\end{aligned} \tag{5.38}$$

and

$$\begin{aligned}
&\Sigma(B')_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)} - I_{|\alpha|+|\beta_1|} = P_1'^T \mathbb{S}(\Delta\widetilde{B}_{(\alpha\cup\beta_1)(\alpha\cup\beta_1)}) P_1' + O(\|\Delta B\|^2), \\
&\Sigma(B')_{a_t a_t} - \Sigma(B)_{a_t a_t} = P_t'^T \mathbb{S}(\Delta\widetilde{B}_{a_t a_t}) P_t' + O(\|\Delta B\|^2), \quad t = \widetilde{r}_0+1, \ldots, \widetilde{r}_1, \\
&[\Sigma(B')_{(\beta_3\cup\gamma)(\beta_3\cup\gamma)} - \Sigma(B)_{(\beta_3\cup\gamma)(\beta_3\cup\gamma)}\ 0] = Q_2'^T [\Delta\widetilde{B}_{(\beta_3\cup\gamma)(\beta_3\cup\gamma)}\ \Delta\widetilde{B}_{(\beta_3\cup\gamma)c}] Q_2'' + O(\|\Delta B\|^2),
\end{aligned} \tag{5.39}$$

where $\Delta\widetilde{A} := U^T \Delta A V$ and $\Delta\widetilde{B} := U^T \Delta B V$. One should note that

$$\begin{aligned}
&P_\alpha^T \Sigma(A)_{\alpha\alpha} P_\alpha = \Sigma(A)_{\alpha\alpha}, \quad P_\gamma^T [\Sigma(A)_{\gamma\gamma}\ 0] P_\gamma' = [\Sigma(A)_{\gamma\gamma}\ 0], \\
&\text{and} \quad P_2'^T \Sigma(B)_{\beta_2\beta_2} P_2' = \Sigma(B)_{\beta_2\beta_2}
\end{aligned} \tag{5.40}$$

with $P_\alpha = \mathrm{Diag}(P_1, \ldots, P_{r_0})$, $P_\gamma = \mathrm{Diag}(P_{r_1+1}, \ldots, P_r, Q_1')$, $P_\gamma' = \mathrm{Diag}(P_{r_1+1}, \ldots, P_r, Q_1'')$ and $P_2' = \mathrm{Diag}(P_{\widetilde{r}_0+1}', \ldots, P_{\widetilde{r}_1}')$.

By Lemma 5.1 and the definition of β_1, β_2 and β_3 in (5.14), we can obtain the following properties of $\Sigma(A')$ and $\Sigma(B')$ that

$$\begin{cases} \Sigma(A')_{\beta_2\beta_2} = \sigma_k(A') I_{|\beta_2|} \quad \text{with} \quad \sigma_k(A') > 0, \\ \Sigma(B')_{\alpha\alpha} = I_{|\alpha|}, \quad \mathrm{tr}(\Sigma(B')_{\beta\beta}) = \mathrm{tr}(\Sigma(B)_{\beta\beta}) \quad \text{and} \quad \Sigma(B')_{\gamma\gamma} = 0. \end{cases} \tag{5.41}$$

By using (5.36), (5.37) and the fact that for any $N \in \mathcal{R}^{p\times p}$ with some integer $p > 0$,

$$NN^T = I_p + O(\|\Delta A\| + \|\Delta B\|) \Longrightarrow \exists\, \widehat{N} \in \mathcal{O}^p \text{ such that } \widehat{N} = N + O(\|\Delta A\| + \|\Delta B\|),$$

we can deduce that there exists $P_{\beta_1} \in \mathcal{O}^{|\beta_1|}$ and $P_{\beta_3} \in \mathcal{O}^{|\beta_3|}$ such that

$$\widetilde{U} = \begin{pmatrix} P_\alpha + R_1 & R_3 & \widetilde{U}_{\alpha\beta_2} & \widetilde{U}_{\alpha\beta_3} & \widetilde{U}_{\alpha\gamma} \\ R_2 & P_{\beta_1} + G_1 & K_3 & K_6 & \widetilde{U}_{\beta_1\gamma} \\ \widetilde{U}_{\beta_2\alpha} & K_1 & P_2' + K_4 & K_7 & \widetilde{U}_{\beta_2\gamma} \\ \widetilde{U}_{\beta_3\alpha} & K_2 & K_5 & P_{\beta_3} + G_2 & R_5 \\ \widetilde{U}_{\gamma\alpha} & \widetilde{U}_{\gamma\beta_1} & \widetilde{U}_{\gamma\beta_2} & R_4 & P_\gamma + R_6 \end{pmatrix} \tag{5.42}$$

and

$$\widetilde{V} = \begin{pmatrix} P_\alpha + R_1' & R_3' & \widetilde{V}_{\alpha\beta_2} & \widetilde{V}_{\alpha\beta_3} & \widetilde{V}_{\alpha(\gamma\cup c)} \\ R_2' & P_{\beta_1} + G_1' & K_3' & K_6' & \widetilde{V}_{\beta_1(\gamma\cup c)} \\ \widetilde{V}_{\beta_2\alpha} & K_1' & P_2' + K_4' & K_7' & \widetilde{V}_{\beta_2(\gamma\cup c)} \\ \widetilde{V}_{\beta_3\alpha} & K_2' & K_5' & P_{\beta_3} + G_2' & R_5' \\ \widetilde{V}_{(\gamma\cup c)\alpha} & \widetilde{V}_{(\gamma\cup c)\beta_1} & \widetilde{V}_{(\gamma\cup c)\beta_2} & R_4' & P_\gamma' + R_6' \end{pmatrix}, \tag{5.43}$$

where

$$\begin{aligned} &R_i = O(\|\Delta A\|),\ R_i' = O(\|\Delta A\|),\ i = 1,\dots,6,\ G_l = O(\|\Delta A\| + \|\Delta B\|),\ l = 1,2,\\ &K_j = O(\|\Delta B\|),\ K_j' = O(\|\Delta B\|),\ j = 1,\dots,7,\ G_l' = O(\|\Delta A\| + \|\Delta B\|),\ l = 1,2,\\ &\widetilde{U}_{\alpha(\beta_2\cup\beta_3\cup\gamma)} = O(\min\{\|\Delta A\|,\ \|\Delta B\|\}),\ \widetilde{U}_{(\beta_2\cup\beta_3\cup\gamma)\alpha} = O(\min\{\|\Delta A\|,\ \|\Delta B\|\}),\\ &\widetilde{U}_{(\beta_1\cup\beta_2)\gamma} = O(\min\{\|\Delta A\|,\ \|\Delta B\|\}),\ \widetilde{U}_{\gamma(\beta_1\cup\beta_2)} = O(\min\{\|\Delta A\|,\ \|\Delta B\|\}),\\ &\widetilde{V}_{\alpha(\beta_2\cup\beta_3\cup\gamma\cup c)} = O(\min\{\|\Delta A\|,\ \|\Delta B\|\}),\ \widetilde{V}_{(\beta_2\cup\beta_3\cup\gamma\cup c)\alpha} = O(\min\{\|\Delta A\|,\ \|\Delta B\|\}),\\ &\widetilde{V}_{(\beta_1\cup\beta_2)(\gamma\cup c)} = O(\min\{\|\Delta A\|,\ \|\Delta B\|\}),\ \widetilde{V}_{(\gamma\cup c)(\beta_1\cup\beta_2)} = O(\min\{\|\Delta A\|,\ \|\Delta B\|\}). \end{aligned} \tag{5.44}$$

Futhermore, using (5.36) and (5.37), we can deduce that $\widetilde{V}_{\beta\beta} = \widetilde{U}_{\beta\beta} + O(\|\Delta A\|)$, i.e.,

$$K_j - K_j' = O(\|\Delta A\|),\ j = 1,\dots,7 \text{ and } G_l - G_l' = O(\|\Delta A\|),\ l = 1,2. \tag{5.45}$$

Next, by using (5.38)-(5.44), we can have the characterization of $\widetilde{A}'_{\alpha\alpha}$ and $\widetilde{B}'_{\alpha\alpha}$ that for all ΔA and ΔB sufficiently small,

$$\begin{aligned} \widetilde{A}'_{\alpha\alpha} &= P_\alpha \Sigma(A')_{\alpha\alpha} P_\alpha^T + \Sigma(A)_{\alpha\alpha} P_\alpha R_1'^T + R_1 P_\alpha^T \Sigma(A)_{\alpha\alpha} + O(\|\Delta A\|^2) \\ &= \Sigma(A)_{\alpha\alpha} + O(\|\Delta A\|), \\ \widetilde{B}'_{\alpha\alpha} &= I_{|\alpha|} + P_\alpha R_1'^T + R_1 P_\alpha^T + O(\|\Delta A\|^2) \\ &= I_{|\alpha|} + O(\|\Delta A\|). \end{aligned} \tag{5.46}$$

Then, combining (5.42)-(5.44) and the orthogonality of $\widetilde{U}$ and $\widetilde{V}$, we can have that for all ΔA and ΔB sufficiently small,

$$\begin{cases} P_\alpha R_1^T + R_1 P_\alpha^T = O(\|\Delta A\|^2), \\ P'_\alpha R_1^T + R_1 P_\alpha'^T = O(\|\Delta A\|^2), \end{cases} \tag{5.47}$$

From (5.46) and (5.47), one can derive that

$$\begin{aligned} &\mathbb{S}(\widetilde{B}'_1)_{\alpha\alpha} = I_{|\alpha|} + O(\|\Delta A\|^2), \\ &\mathbb{T}(\widetilde{A}'_1)_{\alpha\alpha} = \tfrac{1}{2}\big(\Sigma(A)_{\alpha\alpha}\mathbb{T}(\widetilde{B}'_1)_{\alpha\alpha} + \mathbb{T}(\widetilde{B}'_1)_{\alpha\alpha}\Sigma(A)_{\alpha\alpha}\big) + O(\|\Delta A\|^2). \end{aligned} \tag{5.48}$$

In a similar way, by using (5.38)-(5.45), for all ΔA and ΔB sufficiently small we can get the rest part of (5.31)-(5.34) except $\mathbb{S}(\widetilde{B}'_1)_{\beta_1\beta_1} \preceq I_{|\beta_1|}$ and $(\widetilde{B}'_1)_{\beta_3\beta_3} + O(\|\Delta A\|\|\Delta B\|) \in \mathcal{S}_+^{|\beta_3|}$. While, $\mathbb{S}(\widetilde{B}'_1)_{\beta_1\beta_1} \preceq I_{|\beta_1|}$ can be obtained from (4.38) and (4.39) in the proof of Proposition 4.2; and $(\widetilde{B}'_1)_{\beta_3\beta_3} + O(\|\Delta A\|\|\Delta B\|) \in \mathcal{S}_+^{|\beta_3|}$ can be derived by direct calculation of

$$\begin{aligned} (\widetilde{B}'_1)_{\beta_3\beta_3} = \; & K_2 \Sigma(B')_{\beta_1\beta_1} K_2^T + K_5 \Sigma(B')_{\beta_2\beta_2} K_5^T + (P_{\beta_3} + G_2)\Sigma(B')_{\beta_3\beta_3}(P_{\beta_3}^T + G_2^T) \\ & + O(\|\Delta A\|\|\Delta B\|). \end{aligned}$$

Thus, we have showed (5.31)-(5.34). Next, let us prove the first three relationships in (5.35). By noting (5.41), it is not difficult for us to derive the characterization of $\mathbb{S}(\widetilde{A}'_1)_{\beta_1\beta_1}$, $\mathbb{S}(\widetilde{A}'_1)_{\beta_2\beta_2}$ and $\mathbb{S}(\widetilde{A}'_1)_{\beta_3\beta_3}$ that for all ΔA and ΔB sufficiently small,

$$\begin{aligned} \mathbb{S}(\widetilde{A}'_1)_{\beta_1\beta_1} &= P_{\beta_1}\Sigma(A')_{\beta_1\beta_1} P_{\beta_1}^T + O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|), \\ \mathbb{S}(\widetilde{A}'_1)_{\beta_2\beta_2} &= P'_2 \Sigma(A')_{\beta_2\beta_2} P_2'^T + O(\|\Delta A\|\|\Delta B\|) \\ &= \sigma_k(A') I_{|\beta_2|} + O(\|\Delta A\|\|\Delta B\|), \\ \mathbb{S}(\widetilde{A}'_1)_{\beta_3\beta_3} &= P_{\beta_3}\Sigma(A')_{\beta_3\beta_3} P_{\beta_3}^T + O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|), \end{aligned} \tag{5.49}$$

which directly yields the first three relationships in (5.35).

Finally, let us prove the last trace equation in (5.35). By (5.36) and (5.37), we know that $\widetilde{U}_{\beta\beta} = P_\beta + R_U$ and $\widetilde{V}_{\beta\beta} = P_\beta + R_V$ with $R_U = O(\|\Delta A\|)$ and $R_V = O(\|\Delta A\|)$. Then by the orthogonality of $\widetilde{U}$ and $\widetilde{V}$, it is easy to see that

$$\begin{aligned} P_\beta^T R_U + R_U^T P_\beta &= O(\|\Delta A\|^2), \\ P_\beta^T R_V + R_V^T P_\beta &= O(\|\Delta A\|^2). \end{aligned} \tag{5.50}$$

Moreover,

$$\begin{aligned} \widetilde{B}'_{\beta\beta} &= \widetilde{U}_{\beta\alpha}\widetilde{V}_{\beta\alpha}^T + \widetilde{U}_{\beta\beta}\Sigma(B')_{\beta\beta}\widetilde{V}_{\beta\beta}^T \\ &= P_\beta\Sigma(B')_{\beta\beta}P_\beta^T + P_\beta\Sigma(B')_{\beta\beta}R_V^T + R_U\Sigma(B')_{\beta\beta}P_\beta^T + O(\|\Delta A\|^2). \end{aligned} \tag{5.51}$$

Since $\mathrm{tr}(P_\beta\Sigma(B')_{\beta\beta}R_V^T) = \mathrm{tr}(R_V^T P_\beta\Sigma(B')_{\beta\beta}) = \mathrm{tr}(P_\beta^T R_V\Sigma(B')_{\beta\beta})$ and $\mathrm{tr}(R_U\Sigma(B')_{\beta\beta}P_\beta^T)$ $= \mathrm{tr}(P_\beta^T R_U\Sigma(B')_{\beta\beta}) = \mathrm{tr}(R_U^T P_\beta\Sigma(B')_{\beta\beta})$. Noting (5.50), we have $\mathrm{tr}(P_\beta\Sigma(B')_{\beta\beta}R_V^T) =$ $O(\|\Delta A\|^2)$ and $\mathrm{tr}(R_U\Sigma(B')_{\beta\beta}P_\beta^T) = O(\|\Delta A\|^2)$. Together these observations and (5.41), we have

$$\mathrm{tr}(\widetilde{B}'_{\beta\beta}) = \mathrm{tr}(\Sigma(B')_{\beta\beta}) + O(\|\Delta A\|^2) = \mathrm{tr}(\Sigma(B)_{\beta\beta}) + O(\|\Delta A\|^2).$$

This completes the proof of the proposition. □

The second part is quite similar to the nuclear norm case. Before stating the perturbation properties, we adopt the following well known von Neumann's trace inequality [68] for our later discussions.

Lemma 5.4. *Let X and Y be two matrices in $\mathcal{R}^{m\times n}$. Then*

$$\langle X, Y\rangle \le \sigma(X)^T\sigma(Y),$$

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

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

Proposition 5.3. *Let $A \in \mathcal{R}^{m\times n}$ and $B \in \partial\theta(A)$ with $\sigma_k(A) = 0$. Suppose that A and B have the SVD as in (5.4) and the index sets $\alpha, \beta, \beta_1, \beta_2, \beta_3$ and c are defined* *as in Lemma 5.1 and (5.14). Then for all* $(A', B') \in \mathcal{R}^{m\times n} \times \mathcal{R}^{m\times n}$ *satisfying* $B' \in \partial\theta(A')$ *and is sufficiently close to* $(A, B) \in \mathcal{R}^{m\times n} \times \mathcal{R}^{m\times n}$*, we have*

(i) If $\|B\|_* = k$*, then*

$$\begin{cases} \widetilde{A}'_{\alpha\alpha} = \Sigma(A)_{\alpha\alpha} + O(\|\Delta A\|),\ \widetilde{A}'_{\alpha\beta_1} = O(\|\Delta A\|),\ \widetilde{A}'_{\alpha(\beta_2\cup\beta_3\cup c)} = O(\min\{\|\Delta A\|, \|\Delta B\|\}), \\ \widetilde{A}'_{\beta_1\alpha} = O(\|\Delta A\|),\ \widetilde{A}'_{\beta_1\beta_1} = O(\|\Delta A\|),\ \widetilde{A}'_{\beta_1(\beta_2\cup\beta_3\cup c)} = O(\|\Delta A\|\|\Delta B\|), \\ \widetilde{A}'_{(\beta_2\cup\beta_3)\alpha} = O(\min\{\|\Delta A\|, \|\Delta B\|\}),\ \widetilde{A}'_{(\beta_2\cup\beta_3)\beta_1} = O(\|\Delta A\|\|\Delta B\|), \\ \widetilde{A}'_{\beta_2\beta_2} = O(\|\Delta A\|),\ \widetilde{A}'_{\beta_2(\beta_3\cup c)} = O(\|\Delta A\|\|\Delta B\|), \\ \widetilde{A}'_{\beta_3\beta_2} = O(\|\Delta A\|\|\Delta B\|),\ \widetilde{A}'_{\beta_3(\beta_3\cup c)} = O(\|\Delta A\|), \end{cases} \tag{5.52}$$

and (4.21) and (4.22) hold here.

Moreover, $\sigma_k(A') \geq 0$ *and*

$$\begin{aligned} \widetilde{A}'_{\beta_1\beta_1} - \sigma_k(A')I_{|\beta_1|} + O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|) \in \mathcal{S}_+^{|\beta_1|}, \\ \widetilde{A}'_{\beta_2\beta_2} = \sigma_k(A')I_{|\beta_2|} + O(\|\Delta A\|\|\Delta B\|), \\ \sigma_1(\widetilde{A}'_{\beta_3(\beta_3\cup c)}) + O(\|\Delta A\|\|\Delta B\|) \leq \sigma_k(A'), \\ \mathrm{tr}(\widetilde{B}'_{(\beta_1\cup\beta_2)(\beta_1\cup\beta_2)}) + \|\widetilde{B}'_{\beta_3(\beta_3\cup c)}\|_* + O(\|\Delta A\|^2) \leq \mathrm{tr}(\Sigma(B)_{(\beta_1\cup\beta_2)(\beta_1\cup\beta_2)}). \end{aligned} \tag{5.53}$$

(ii) If $\|B\|_* < k$*, then the same conculsion as in Proposition 4.3.*

In above statement, we denote $\Delta A := A' - A$*,* $\Delta B := B' - B$*,* $\widetilde{A}' := U^TA'V = [\widetilde{A}'_1\ \widetilde{A}'_2] = [U^TA'V_1\ U^TA'V_2]$*,* $\widetilde{B}' := U^TB'V = [\widetilde{B}'_1\ \widetilde{B}'_2] = [U^TB'V_1\ U^TB'V_2]$ *and* I_p *as the identity* p *by* p *matrix.*

Proof. The proof of this proposition is largely similar to the proof of Proposition 4.3. We have $\widetilde{U}$ and $\widetilde{V}$ in the form of (4.30) and (4.31) respectively, such that

$$\widetilde{A}' = \widetilde{U}[\Sigma(A')\ 0]\widetilde{V}^T, \quad \widetilde{B}' = \widetilde{U}[\Sigma(B')\ 0]\widetilde{V}^T.$$

with $\Sigma(A)$, $\Sigma(A')$, $\Sigma(B)$ and $\Sigma(B')$ satisfying (4.26) and (4.27).

Case (i) $\|B\|_* = k$. By Lemma 5.1 and the definition of β_1, β_2 and β_3 in (5.14), we can obtain the following properties of $\Sigma(A')$ and $\Sigma(B')$ that

$$\begin{cases} \Sigma(A')_{\beta_2\beta_2} = \sigma_k(A')I_{|\beta_2|} \quad \text{with} \quad \sigma_k(A') \geq 0, \\ \Sigma(B')_{\alpha\alpha} = I_{|\alpha|}, \quad \operatorname{tr}(\Sigma(B')) \leq \operatorname{tr}(\Sigma(B)). \end{cases} \tag{5.54}$$

Then, we have the characterization of $(\widetilde{A}'_1)_{\beta_1\beta_1}$,$(\widetilde{A}'_1)_{\beta_2\beta_2}$ and $\widetilde{A}'_{\beta_3(\beta_3\cup c)}$ that for all ΔA and ΔB sufficiently small,

$$\begin{aligned} (\widetilde{A}'_1)_{\beta_1\beta_1} &= P_{\beta_1}\Sigma(A')_{\beta_1\beta_1}P_{\beta_1}^T + O(\|\Delta A\|^2 + \|\Delta A\|\|\Delta B\|) \\ &= O(\|\Delta A\|), \\ (\widetilde{A}'_1)_{\beta_2\beta_2} &= \sigma_k(A')I_{|\beta_2|} + O(\|\Delta A\|\|\Delta B\|) \\ &= O(\|\Delta A\|), \\ \widetilde{A}'_{\beta_3(\beta_3\cup c)} &= Q'_2[\Sigma(A')_{\beta_3\beta_3}\ 0]Q''^T_2 + O(\|\Delta A\|\|\Delta B\|) \\ &= O(\|\Delta A\|), \end{aligned} \tag{5.55}$$

where P_{β_1}, Q'_2 and Q''_2 are denoted in (4.26) and (4.27). These showed part of (5.52) and the first three relationships in (5.53). While, the rest part of (5.52), (4.21) and (4.22) can be proved similarly to Propostion 4.3, and we omit here. To close this case, we only need to show the last inequality of (5.53). As we have obtained from (4.22) that $\mathbb{S}(\widetilde{B}'_1)_{\alpha\alpha} = I_{|\alpha|} + O(\|\Delta A\|^2)$, then $\operatorname{tr}(\widetilde{B}'_{\alpha\alpha}) = |\alpha| + O(\|\Delta A\|^2)$.
Suppose that $\widetilde{B}'_{\beta_3(\beta_3\cup c)}$ admits the SVD as $\widetilde{B}'_{\beta_3(\beta_3\cup c)} = \widehat{U}[\widehat{\Sigma}\ 0]\widehat{V}^T$ with $\widehat{U} \in \mathcal{O}^{|\beta_3|}$, $\widehat{V} \in \mathcal{O}^{|\beta_3\cup c|}$ and $\widehat{\Sigma}$ be the diagonal matrix of all the singular values of $\widetilde{B}'_{\beta_3(\beta_3\cup c)}$. Then, we construct a matrix $\Gamma = \begin{pmatrix} I_{|\alpha|+|\beta_1|+|\beta_2|} & 0 \\ 0 & \widehat{U}[I_{|\beta_3|}\ 0]\widehat{V}^T \end{pmatrix} \in \mathcal{R}^{m\times n}$ of which the sigular values are all 1's. Moreover,

$$\begin{aligned} \operatorname{tr}(\Sigma(B')) &= \langle[\Sigma(B')\ 0],\ [I_m\ 0]\rangle \\ &\geq \langle \widetilde{B}', \Gamma\rangle \\ &= \operatorname{tr}(\widetilde{B}'_{(\alpha\cup\beta_1\cup\beta_2)(\alpha\cup\beta_1\cup\beta_2)}) + \langle \widetilde{B}'_{\beta_3(\beta_3\cup c)},\ \widehat{U}[I_{|\beta_3|}\ 0]\widehat{V}^T\rangle \\ &= \operatorname{tr}(\widetilde{B}'_{(\alpha\cup\beta_1\cup\beta_2)(\alpha\cup\beta_1\cup\beta_2)}) + \langle[\widehat{\Sigma}\ 0],\ [I_{|\beta_3|}\ 0]\rangle \\ &= \operatorname{tr}(\widetilde{B}'_{\alpha\alpha}) + \operatorname{tr}(\widetilde{B}'_{(\beta_1\cup\beta_2)(\beta_1\cup\beta_2)}) + ||\widetilde{B}'_{\beta_3(\beta_3\cup c)}||_* \end{aligned} \tag{5.56}$$

$$
\begin{aligned}
&= |\alpha| + \mathrm{tr}(\widetilde{B}'_{(\beta_1\cup\beta_2)(\beta_1\cup\beta_2)}) + ||\widetilde{B}'_{\beta_3(\beta_3\cup c)}||_* \\
&\quad +O(\|\Delta A\|^2),
\end{aligned}
$$

where the inequality comes from von Neumann's trace inequality Lemma 5.4. Combining (5.54), (5.56) and $\mathrm{tr}(\Sigma(B)) = |\alpha| + \mathrm{tr}(\Sigma(B)_{\beta\beta})$, we can get the last inequality of (5.53).

Case (ii) $\|B\|_* < k$. Same as Proposition 4.3.

Therefore, this completes the proof of the proposition. □

5.2 The metric subregularity of the solution mapping for composite Ky Fan k-norm problem

The following first order optimality conditions of the problem (5.1) are similar to those of the nuclear norm regularized problem (4.1).

For any $(x, y, S) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^{m\times n}$, the Lagrangian function l associated with the problem (5.1) is defined as

$$
l(x, y, S) := f(x) + \langle y, h(x)\rangle + \langle S, g(x)\rangle - \theta^*(S). \tag{5.57}
$$

Define the multi-valued mapping $\mathcal{T}_l : \mathcal{X}\times\mathcal{Y}\times\mathcal{R}^{m\times n} \rightrightarrows \mathcal{X}\times\mathcal{Y}\times\mathcal{R}^{m\times n}$ associated with the Lagrangian function l at any $(x, y, S) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^{m\times n}$ by

$$
\mathcal{T}_l(x, y, S) = \{(u, v, C) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^{m\times n} | (u, -v, -C) \in \partial l(x, y, S)\}. \tag{5.58}
$$

Suppose that the optimal solution set of the problem (5.1) is nonempty and consider an optimal solution $\bar{x} \in \mathcal{X}$ of the problem (5.1). Since $\mathrm{dom}\,\theta = \mathcal{R}^{m\times n}$, the reduced RCQ (2.14) always holds at $\bar{x}$, therefore we can impose the following first order optimality conditions for the problem (5.1). Then, $(\bar{y}, \bar{S}) \in \mathcal{Y} \times \mathcal{R}^{m\times n}$ is a Lagrangian multiplier corresponding to $\bar{x}$ if and only if $(\bar{x}, \bar{y}, \bar{S})$ satisfies the following KKT system:

$$\begin{cases} \nabla f(\bar{x}) + \nabla h(\bar{x})\bar{y} + \nabla g(\bar{x})\bar{S} = 0, \\ h(\bar{x}) = 0, \\ \bar{S} \in \partial\theta(g(\bar{x})). \end{cases} \tag{5.59}$$

Denote $\mathcal{M}(\bar{x})$ as the set of all Lagrangian multipliers corresponding to $\bar{x}$.

By the third inclusion of (5.59), we assume that $g(\bar{x})$ and $\bar{S}$ have the singular value decompositions as in (5.4) with $A = g(\bar{x})$ and $B = \bar{S}$.

For a perturbed point $(u, v, C) \in \mathcal{X}\times\mathcal{Y}\times\mathcal{R}^{m\times n}$, it is easy to check that $(x, y, S) \in \mathcal{T}_l^{-1}(u, v, C)$ can be equivalently interpreted as the following perturbed KKT system:

$$\begin{cases} \nabla f(x) + \nabla h(x)y + \nabla g(x)S = u, \\ h(x) + v = 0, \\ S \in \partial\theta(g(x) + C). \end{cases} \tag{5.60}$$

One can find that $\mathcal{T}_l^{-1}(0, 0, 0)$ is the set of all the KKT points $(\bar{x}, \bar{y}, \bar{S})$ of the problem (5.1) satisfying (5.59).

Next, we conduct our discussions about the metric subregularity of $\mathcal{T}_l$ at a KKT point for the origin.

Let $(\bar{x}, \bar{y}, \bar{S}) \in \mathcal{T}_l^{-1}(0, 0, 0)$. Since $\operatorname{dom}\theta = \mathcal{R}^{m\times n}$ with $\theta(\cdot) = \|\cdot\|_{(k)}$, we can define the critical cone of the problem (5.1) at $\bar{x}$ by

$$\mathcal{C}(\bar{x}) := \{d \in \mathcal{X} \mid h'(\bar{x})d = 0,\ g'(\bar{x})d \in \mathcal{C}_\theta(g(\bar{x}), \bar{S})\},$$

where, $\mathcal{C}_\theta(\cdot, \cdot)$ defined as (5.19).

Again, we define a more restrictive second-order sufficient condition for problem (5.1) at $\bar{x}$ with respect to the multiplier $(\bar{y}, \bar{S}) \in \mathcal{M}(\bar{x})$ if

$$\langle d, \nabla^2_{xx} l(\bar{x}, \bar{y}, \bar{S})d\rangle + \Upsilon_{g(\bar{x})}\big(\bar{S}, g'(\bar{x})d\big) > 0,\ \forall 0 \neq\ d \in \mathcal{C}(\bar{x}), \tag{5.61}$$

where $-\Upsilon_{g(\bar{x})}\big(\bar{S}, g'(\bar{x})d\big)$ is the so called sigma term in the second-order sufficient condition (5.61) for the problem (5.1).

By the C^2-cone reducibility of $\|\cdot\|_{(k)}$ and the Lipschitz continuity of $\|\cdot\|_{(k)}$ and its directional derivative, we obtain $\theta^{\downarrow\downarrow}(X;H,\cdot) = \theta''(X;H,\cdot)$ for any $X, H \in \mathcal{R}^{m\times n}$. In [25], Ding shows that the sigma term for Ky Fan k-norm regularized problem is just the conjugate function of the parabolic second order directional derivative of the nuclear norm function θ. Moreover, by adopting the sigma term derived by Bonnans and Shapiro [8, Section 3.4.1] for composite problems, we have

$$-\Upsilon_{g(\bar{x})}\big(\cdot, g'(\bar{x})d\big) = \phi^*(\cdot) \text{ with } \phi(\cdot) := \theta''(g(\bar{x}); g'(\bar{x})d, \cdot).$$

By using the expression of the second order directional derivative for the eigenvalues and singular values [96,109], Ding [25] futher provides the explicit expression of this sigma term as below.

We consider $A, B \in \mathcal{R}^{m\times n}$ satisfying $B \in \partial\theta(A)$ and the index sets $\alpha, \beta, \beta_1, \beta_2, \beta_3$, γ and c defined as in Lemma 5.1 and (5.14), then the sigma term is given as below.

(i) If $\sigma_k(A) = 0$, then

$$-\Upsilon_A\big(B,H\big) = 2\sum_{l=1}^{r_0} \mathrm{tr}\big(\Omega_{a_l}(A,H)\big) + 2\langle \Sigma(B)_{\beta\beta},\ U_\beta^T H A^\dagger H V_\beta\rangle, \qquad (5.62)$$

(ii) If $\sigma_k(A) > 0$, then

$$-\Upsilon_A\big(B,H\big) = 2\sum_{l=1}^{r_0} \mathrm{tr}\big(\Omega_{a_l}(A,H)\big) + 2\langle \Sigma(B)_{\beta\beta},\ \Omega_\beta(A,H)\rangle, \qquad (5.63)$$

where

$$\begin{aligned}\Omega_{a_l}(A,H) := \ & (\mathbb{S}(\widetilde{H}_1))_{a_l}^T(\Sigma(A) - \nu_l(A)\mathcal{I}_m)^\dagger(\mathbb{S}(\widetilde{H}_1))_{a_l} - (2\nu_l(A))^{-1}\widetilde{H}_{a_lc}\widetilde{H}_{a_lc}^T \\ & +(\mathbb{T}(\widetilde{H}_1))_{a_l}^T(-\Sigma(A) - \nu_l(A)\mathcal{I}_m)^\dagger(\mathbb{T}(\widetilde{H}_1))_{a_l}, \quad l = 1, 2, \ldots r_0,\end{aligned}$$

and

$$\begin{aligned}\Omega_\beta(A,H) := \ & (\mathbb{S}(\widetilde{H}_1))_\beta^T(\Sigma(A) - \sigma_k(A)\mathcal{I}_m)^\dagger(\mathbb{S}(\widetilde{H}_1))_\beta - (2\sigma_k(A))^{-1}\widetilde{H}_{\beta c}\widetilde{H}_{\beta c}^T \\ & +(\mathbb{T}(\widetilde{H}_1))_\beta^T(-\Sigma(A) - \sigma_k(A)\mathcal{I}_m)^\dagger(\mathbb{T}(\widetilde{H}_1))_\beta,\end{aligned}$$

with $\widetilde{H} = [\widetilde{H}_1\ \ \widetilde{H}_2] = [U^T H V_1\ \ U^T H V_2]$. We can futher compute $\Upsilon_A\big(B,H\big)$ as follows,

(i) if $\sigma_k(A) = 0$,

$$
\begin{aligned}
\Upsilon_A(B,H) = \; & \sum_{1\le l,t\le r_0} \frac{2}{\nu_l(A)+\nu_t(A)}\|(\mathbb{T}(\widetilde{H}_1))_{a_la_t}\|^2 + \sum_{1\le l\le r_0}\frac{4}{\nu_l(A)}\|(\mathbb{T}(\widetilde{H}_1))_{a_l\beta_1}\|^2 \\
& + \sum_{\substack{1\le l\le r_0\\ 1\le i-|\alpha|-|\beta_1|\le|\beta_2|}} \left(\frac{2(1-\sigma_i(B))}{\nu_l(A)}\|(\mathbb{S}(\widetilde{H}_1))_{a_li}\|^2 + \frac{2(1+\sigma_i(B))}{\nu_l(A)}\|(\mathbb{T}(\widetilde{H}_1))_{a_li}\|^2\right) \\
& + \sum_{1\le l\le r_0}\left(\frac{2}{\nu_l(A)}\|(\mathbb{S}(\widetilde{H}_1))_{a_l\beta_3}\|^2 + \frac{2}{\nu_l(A)}\|(\mathbb{T}(\widetilde{H}_1))_{a_l\beta_3}\|^2\right) \\
& + \sum_{1\le l\le r_0}\frac{1}{\nu_l(A)}\|\widetilde{H}_{a_lc}\|^2;
\end{aligned}
\tag{5.64}
$$

(ii) if $\sigma_k(A) > 0$,

$$
\begin{aligned}
\Upsilon_A(B,H) = \; & \sum_{1\le l,t\le r_0} \frac{2}{\nu_l(A)+\nu_t(A)}\|(\mathbb{T}(\widetilde{H}_1))_{a_la_t}\|^2 + \sum_{1\le l\le r_0}\frac{4}{\nu_l(A)+\sigma_k(A)}\|(\mathbb{T}(\widetilde{H}_1))_{a_l\beta_1}\|^2 \\
& + \sum_{\substack{1\le l\le r_0\\ 1\le i-|\alpha|-|\beta_1|\le|\beta_2|}} \left(\frac{2(1-\sigma_i(B))}{\nu_l(A)-\sigma_k(A)}\|(\mathbb{S}(\widetilde{H}_1))_{a_li}\|^2 + \frac{2(1+\sigma_i(B))}{\nu_l(A)+\sigma_k(A)}\|(\mathbb{T}(\widetilde{H}_1))_{a_li}\|^2\right) \\
& + \sum_{1\le l\le r_0}\left(\frac{2}{\nu_l(A)-\sigma_k(A)}\|(\mathbb{S}(\widetilde{H}_1))_{a_l\beta_3}\|^2 + \frac{2}{\nu_l(A)+\sigma_k(A)}\|(\mathbb{T}(\widetilde{H}_1))_{a_l\beta_3}\|^2\right) \\
& + \sum_{\substack{1\le l\le r_0\\ r_1+1\le t\le r+1}}\left(\frac{2}{\nu_l(A)-\nu_t(A)}\|(\mathbb{S}(\widetilde{H}_1))_{a_la_t}\|^2 + \frac{2}{\nu_l(A)+\nu_t(A)}\|(\mathbb{T}(\widetilde{H}_1))_{a_la_t}\|^2\right) \\
& + \sum_{1\le l\le r_0}\frac{1}{\nu_l(A)}\|\widetilde{H}_{a_lc}\|^2 \\
& + \frac{1}{\sigma_k(A)}\|(\mathbb{T}(\widetilde{H}_1))_{\beta_1\beta_1}\|^2 + \sum_{1\le i-|\alpha|-|\beta_1|\le|\beta_2|}\frac{1+\sigma_i(B)}{\sigma_k(A)}\|(\mathbb{T}(\widetilde{H}_1))_{\beta_1 i}\|^2 \\
& + \frac{1}{\sigma_k(A)}\|(\mathbb{T}(\widetilde{H}_1))_{\beta_1\beta_3}\|^2 + \sum_{\substack{1\le i-|\alpha|-|\beta_1|\le|\beta_2|\\ 1\le j-|\alpha|-|\beta_1|\le|\beta_2|}}\frac{\sigma_i(B)+\sigma_j(B)}{2\sigma_k(A)}\|(\mathbb{T}(\widetilde{H}_1))_{ij}\|^2 \\
& + \sum_{1\le i-|\alpha|-|\beta_1|\le|\beta_2|}\frac{\sigma_i(B)}{\sigma_k(A)}\|(\mathbb{T}(\widetilde{H}_1))_{i\beta_3}\|^2 \\
& + \sum_{r_1+1\le l\le r+1}\left(\frac{2}{\sigma_k(A)-\nu_l(A)}\|(\mathbb{S}(\widetilde{H}_1))_{\beta_1a_l}\|^2 + \frac{2}{\sigma_k(A)+\nu_l(A)}\|(\mathbb{T}(\widetilde{H}_1))_{\beta_1a_l}\|^2\right) \\
& + \sum_{\substack{1\le i-|\alpha|-|\beta_1|\le|\beta_2|\\ r_1+1\le l\le r+1}}\left(\frac{2\sigma_i(B)}{\sigma_k(A)-\nu_l(A)}\|(\mathbb{S}(\widetilde{H}_1))_{ia_l}\|^2 + \frac{\sigma_i(B)}{\sigma_k(A)+\nu_l(A)}\|(\mathbb{T}(\widetilde{H}_1))_{ia_l}\|^2\right) \\
& + \frac{1}{\sigma_k(A)}\|\widetilde{H}_{\beta_1c}\|^2 + \sum_{1\le i-|\alpha|-|\beta_1|\le|\beta_2|}\frac{\sigma_i(B)}{\sigma_k(A)}\|\widetilde{H}_{ic}\|^2,
\end{aligned}
\tag{5.65}
$$

where we denote $\nu_{r+1}(A) = 0$.

For the convenience of the later discussions, recalling the definition of $\mathcal{C}_{\theta^*}(\cdot,\cdot)$ in

(5.25), we define the following joint 'critical cone' associated with the problem (5.1) as

$$\widetilde{\mathcal{C}}(\bar{x},\bar{y},\bar{S}) := \left\{ \begin{array}{c|l} & h'(\bar{x})d_x = 0, \\ (d_x, d_y, d_S) \in & g'(\bar{x})d_x \in \mathcal{C}_\theta(g(\bar{x}),\bar{S}), d_S \in \mathcal{C}_{\theta^*}(\bar{S}, g(\bar{x})), \\ \mathcal{X}\times\mathcal{Y}\times\mathcal{R}^{m\times n} & d_S = U[\mathcal{E}_{\mathbb{S}}\circ\mathbb{S}(\widetilde{D}_{x,1}) + \mathcal{E}_{\mathbb{T}}\circ\mathbb{T}(\widetilde{D}_{x,1}) \quad \mathcal{F}\circ(\widetilde{D}_{x,2})]V^T \\ & \qquad +U\Theta_{\bar{S}}V^T \end{array} \right\}, \tag{5.66}$$

where $D_x = g'(\bar{x})d_x$, $\widetilde{D}_x = [\widetilde{D}_{x,1}\ \widetilde{D}_{x,2}] = [U^T D_x V_1\ U^T D_x V_2] = U^T D_x V$, $\widetilde{d}_S = [\widetilde{d}_{S,1}\ \widetilde{d}_{S,2}] = [U^T d_S V_1\ U^T d_S V_2] = U^T d_S V$ and $\mathcal{E}_{\mathbb{S}} \in \mathcal{S}^m$, $\mathcal{E}_{\mathbb{T}} \in \mathcal{S}^m$, $\mathcal{F} \in \mathcal{R}^{m\times(n-m)}$ are given by

$$(\mathcal{E}_{\mathbb{S}})_{ij} = \begin{cases} \dfrac{\sigma_i(\bar{S}) - \sigma_j(\bar{S})}{\sigma_i(g(\bar{x})) - \sigma_j(g(\bar{x}))} & \text{if } \sigma_i(g(\bar{x})) \neq \sigma_j(g(\bar{x})), \\ 0 & \text{otherwise,} \end{cases} \qquad i,j \in \{1,\ldots,m\},$$

$$(\mathcal{E}_{\mathbb{T}})_{ij} = \begin{cases} \dfrac{\sigma_i(\bar{S}) + \sigma_j(\bar{S})}{\sigma_i(g(\bar{x})) + \sigma_j(g(\bar{x}))} & \text{if } \sigma_i(g(\bar{x})) + \sigma_j(g(\bar{x})) \neq 0, \\ 0 & \text{otherwise,} \end{cases} \qquad i,j \in \{1,\ldots,m\},$$

and

$$(\mathcal{F})_{ij} = \begin{cases} \dfrac{\sigma_i(\bar{S})}{\sigma_i(g(\bar{x}))} & \text{if } \sigma_i(g(\bar{x})) \neq 0, \\ 0 & \text{otherwise,} \end{cases} \qquad i \in \{1,\ldots,m\},\ j = \{1,\ldots,n-m\};$$

moreover,

(i) if $\sigma_k(g(\bar{x})) = 0$, we define $\Theta_{\bar{S}} := \begin{pmatrix} 0_{\alpha\alpha} & 0 & 0 \\ 0 & (\widetilde{d}_S)_{\beta\beta} & (\widetilde{d}_S)_{\beta c} \end{pmatrix} \in \mathcal{R}^{m\times n}$;

(ii) if $\sigma_k(g(\bar{x})) > 0$, we define $\Theta_{\bar{S}} := \begin{pmatrix} 0_{\alpha\alpha} & 0 & 0 & 0 \\ 0 & S(\widetilde{d}_{S,1})_{\beta\beta} & 0 & 0 \\ 0 & 0 & 0_{\gamma\gamma} & 0_{\gamma c} \end{pmatrix} \in \mathcal{R}^{m\times n}$.

Theorem 5.1. *Let $\bar{x}$ be an optimal solution to the problem (5.1) and $(\bar{y},\bar{S}) \in \mathcal{M}(\bar{x})$ be a Lagrangian multiplier corresponding to $\bar{x}$. Denote $\Phi := \big(\mathcal{C}_\theta(g(\bar{x}),\bar{S})\big)^\circ$. Let the following assumptions be satisfied:*

(i) *The set* $g'(\bar{x})^T\Phi$ *is closed.*

(ii) $\langle g'(\bar{x})d_x,\ \Pi_\Phi(d_S)\rangle = 0$ *for all* $(d_x, d_y, d_S) \in \widetilde{\mathcal{C}}(\bar{x}, \bar{y}, \bar{S})$*, where* $\Pi_\Phi(\cdot)$ *denotes the projection onto the set* Φ *and the set* $\widetilde{\mathcal{C}}(\bar{x}, \bar{y}, \bar{S})$ *is defined as (5.66).*

(iii) *The second-order sufficient condition (5.61) holds at* $\bar{x}$ *with respect to the multiplier* $(\bar{y}, \bar{S}) \in \mathcal{M}(\bar{x})$ *for the problem (5.1).*

Then there exist a constant $\kappa > 0$ *and a neighborhood* $\mathcal{U}$ *of* $(\bar{x}, \bar{y}, \bar{S})$ *such that for any* $(u, v, C) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^{m\times n}$*,*

$$\|x - \bar{x}\| \le \kappa\|(u, v, C)\|,\ \forall (x, y, S) \in \mathcal{T}_l^{-1}(u, v, C) \cap \mathcal{U}. \tag{5.67}$$

Moreover, if there exists $(\widehat{y}, \widehat{S}) \in \mathcal{M}(\bar{x})$ *such that*

(a) $0 < \sigma_\beta(\widehat{S}) < e_\beta$ *and* $\sum_{i\in\beta}\sigma_i(\widehat{S}) < k - k_0$ *if* $\sigma_k(g(\bar{x})) = 0$*;*

(b) $0 < \sigma_\beta(\widehat{S}) < e_\beta$ *if* $\sigma_k(g(\bar{x})) > 0$.

Then $\mathcal{T}_l$ *is metrically subregular at* $(\bar{x}, \bar{y}, \bar{S})$ *for the origin.*

Proof. Firstly, we show that under the assumptions (i)-(iii), there exist a constant $\kappa > 0$ and a neighborhood $\mathcal{U}$ of $(\bar{x}, \bar{y}, \bar{S})$ such that (5.67) holds.
Suppose that (5.67) does not hold. It means that there exist some sequences $\{(u^p, v^p, C^p)\}_{p\ge 0} \subset \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^{m\times n}$ and $\{(x^p, y^p, S^p)\}_{p\ge 0} \subset \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^{m\times n}$ such that $(u^p, v^p, C^p) \to 0$, $(x^p, y^p, S^p) \to (\bar{x}, \bar{y}, \bar{S})$ with every $(x^p, y^p, S^p) \in \mathcal{T}_l^{-1}(u^p, v^p, C^p)$, and

$$\|x^p - \bar{x}\| \ge \delta_p\|(u^p, v^p, C^p)\|$$

with some $0 < \delta_p$ such that $\delta_p \to \infty$. Denote $t_p := \|x^p - \bar{x}\|$, by taking a subsequence if necessary, we can assume that $(x^p - \bar{x})/t_p \to d_{\bar{x}} \in \mathcal{X}$ with $\|d_{\bar{x}}\| = 1$.
From the perturbed KKT system (5.60), we can have that for all $k \ge 0$ large enough,

$$\begin{aligned} 0 &= h(x^p) + v^p - h(\bar{x}) \\ &= h'(\bar{x})(x^p - \bar{x}) + o(t_p) + v^p. \end{aligned} \tag{5.68}$$

Dividing by t_p on both sides of (5.68) and taking limits $p \to \infty$, we get

$$h'(\bar{x})d_{\bar{x}} = 0. \tag{5.69}$$

Denote $A := g(\bar{x})$ and $B := \bar{S}$ in later discussions.

Case (i) $\sigma_k(g(\bar{x})) = 0$ and $\|\bar{S}\|_* = k - k_0$. For simplify notations, we set

$$\Omega := \{W \in \mathcal{R}^{m\times n} \mid U_\beta^T W[V_\beta \ V_2] = 0\},$$

and for all $p \geq 0$,

$$\begin{cases} A^p := g(x^p) + C^p, \ \widetilde{A}^p := U^T A^p V = [\widetilde{A}_1^p \ \widetilde{A}_2^p] = [U^T A^p V_1 \ U^T A^p V_2], \\ B^p := S^p, \ \widetilde{B}^p := U^T B^p V = [\widetilde{B}_1^p \ \widetilde{B}_2^p] = [U^T B^p V_1 \ U^T B^p V_2], \\ \Delta A^p = A^p - A, \ \Delta B^p = B^p - B, \\ H^p := \Pi_\Omega\big((B^p - B)/t_p\big), \ G^p := (B^p - B)/t_p - H^p. \end{cases} \tag{5.70}$$

Thus, $A^p \to A$ and $B^p \to B$ by the assumptions. Moreover, similar to (5.68), we can derive that

$$\frac{1}{t_p}(A^p - A) \to g'(\bar{x})d_{\bar{x}} \text{ as } p \to \infty. \tag{5.71}$$

Since $B \in \partial\theta(A)$ and $B^p \in \partial\theta(A^p)$, we can derive the following estimates by Proposition 5.3 that for all (A^p, B^p) sufficiently close to (A, B), $\sigma_k(A^p) \geq 0$,

$$\begin{cases} \widetilde{A}^p_{\beta_1\beta_1} - \sigma_k(A^p) I_{|\beta_1|} + O(\|\Delta A^p\|^2 + \|\Delta A^p\|\|\Delta B^p\|) \in \mathcal{S}_+^{|\beta_1|}, \\ \widetilde{A}^p_{\beta_2\beta_2} = \sigma_k(A^p) I_{|\beta_2|} + O(\|\Delta A^p\|\|\Delta B^p\|), \\ \sigma_1(\widetilde{A}^p_{\beta_3(\beta_3\cup c)}) + O(\|\Delta A^p\|\|\Delta B^p\|) \leq \sigma_k(A^p), \ \widetilde{A}^p_{\beta_1(\beta_2\cup\beta_3\cup c)} = O(\|\Delta A^p\|\|\Delta B^p\|), \\ \widetilde{A}^p_{(\beta_2\cup\beta_3)\beta_1} = O(\|\Delta A^p\|\|\Delta B^p\|), \ \widetilde{A}^p_{\beta_2(\beta_3\cup c)} = O(\|\Delta A^p\|\|\Delta B^p\|), \ \widetilde{A}^p_{\beta_3\beta_2} = O(\|\Delta A^p\|\|\Delta B^p\|), \\ \mathbb{S}(\widetilde{B}_1^p)_{\alpha\alpha} = I_{|\alpha|} + O(\|\Delta A^p\|^2), \ \mathbb{S}(\widetilde{B}_1^p)_{\alpha\beta_1} = O(\|\Delta A^p\|^2 + \|\Delta A^p\|\|\Delta B^p\|), \\ \mathbb{S}(\widetilde{B}_1^p)_{\beta_1\beta_1} \preceq I_{|\beta_1|}, \end{cases} \tag{5.72}$$

and

$$
\begin{cases}
\mathbb{T}(\widetilde{A}^p_1)_{\alpha\alpha} = \frac{1}{2}\big(\Sigma(A)_{\alpha\alpha}\mathbb{T}(\widetilde{B}^p_1)_{\alpha\alpha} + \mathbb{T}(\widetilde{B}^p_1)_{\alpha\alpha}\Sigma(A)_{\alpha\alpha}\big) + O(\|\Delta A^p\|^2),\\
\mathbb{T}(\widetilde{A}^p_1)_{\alpha\beta_1} = \frac{1}{2}\Sigma(A)_{\alpha\alpha}\mathbb{T}(\widetilde{B}^p_1)_{\alpha\beta_1} + O(\|\Delta A^p\|^2 + \|\Delta A^p\|\|\Delta B^p\|),\\
\widetilde{B}^p_{\alpha\beta_2} = \Sigma(A)^{-1}_{\alpha\alpha}\widetilde{A}^p_{\alpha\beta_2} - \Sigma(A)^{-1}_{\alpha\alpha}(\widetilde{A}^p_{\beta_2\alpha})^T\Sigma(B)_{\beta_2\beta_2} + O(\|\Delta A^p\|\|\Delta B^p\|),\\
\widetilde{B}^p_{\beta_2\alpha} = \widetilde{A}^p_{\beta_2\alpha}\Sigma(A)^{-1}_{\alpha\alpha} - \Sigma(B)_{\beta_2\beta_2}(\widetilde{A}^p_{\alpha\beta_2})^T\Sigma(A)^{-1}_{\alpha\alpha} + O(\|\Delta A^p\|\|\Delta B^p\|),\\
\widetilde{B}^p_{\alpha(\beta_3\cup\gamma)} = \Sigma(A)^{-1}_{\alpha\alpha}\widetilde{A}^p_{\alpha(\beta_3\cup\gamma)} + O(\|\Delta A^p\|\|\Delta B^p\|),\\
\widetilde{B}^p_{\beta_3\alpha} = \widetilde{A}^p_{\beta_3\alpha}\Sigma(A)^{-1}_{\alpha\alpha} + O(\|\Delta A^p\|\|\Delta B^p\|),\\
\mathrm{tr}(\widetilde{B}^p_{(\beta_1\cup\beta_2)(\beta_1\cup\beta_2)}) + \|\widetilde{B}^p_{\beta_3(\beta_3\cup c)}\|_* + O(\|\Delta A^p\|^2) \le \mathrm{tr}(\Sigma(B)_{(\beta_1\cup\beta_2)(\beta_1\cup\beta_2)}).
\end{cases}
\tag{5.73}
$$

Combining the above (5.71) - (5.73) with Lemma 5.2, we obtain

$$
\begin{cases}
g'(\bar{x})d_{\bar{x}} \in \mathcal{C}_\theta(g(\bar{x}),\bar{S}),\ G^p + \dfrac{1}{t_p}O(\|\Delta A^p\|^2) \in \Phi,\\
H := \lim\limits_{k\to\infty} H^k = U[\mathcal{E}_{\mathbb{S}}\circ\mathbb{S}(\widetilde{D}_{\bar{x},1}) + \mathcal{E}_{\mathbb{T}}\circ\mathbb{T}(\widetilde{D}_{\bar{x},1}) \quad \mathcal{F}\circ(\widetilde{D}_{\bar{x},2})]V^T,
\end{cases}
\tag{5.74}
$$

where $\widetilde{D}_{\bar{x}} := U^T g'(\bar{x})d_{\bar{x}}V = [\widetilde{D}_{\bar{x},1}\ \widetilde{D}_{\bar{x},2}] = [U^T D_{\bar{x}}V_1\ U^T D_{\bar{x}}V_2]$.
Thus, there exist $E^p = \dfrac{1}{t_p}O(\|\Delta A^p\|^2) \in \mathcal{R}^{m\times n}$ such that $G^p + E^p \in \Phi$.
Again, by the perturbed KKT system (5.60), we can deduce that for $p \ge 0$ large enough,

$$
\begin{aligned}
u^p &= \nabla f(x^p) + \nabla h(x^p)y^p + \nabla g(x^p)S^p - \big(\nabla f(\bar{x}) + \nabla h(\bar{x})\bar{y} + \nabla g(\bar{x})\bar{S}\big)\\
&= \nabla^2_{xx}f(\bar{x})(x^p - \bar{x}) + \langle y^p, h''(\bar{x})(x^p-\bar{x})\rangle + \langle S^p, g''(\bar{x})(x^p-\bar{x})\rangle\\
&+ \nabla h(\bar{x})(y^p - \bar{y}) + \nabla g(\bar{x})(S^p - \bar{S}).
\end{aligned}
\tag{5.75}
$$

Dividing by t_p on both sides of (5.75), and then adding $\nabla g(\bar{x})E^p$ on both sides, it gives

$$
\begin{aligned}
& \frac{u^p}{t_p} - \nabla^2_{xx}f(\bar{x})\frac{(x^p-\bar{x})}{t_p} - \langle y^p, h''(\bar{x})\frac{(x^p-\bar{x})}{t_p}\rangle - \langle S^p, g''(\bar{x})\frac{(x^p-\bar{x})}{t_p}\rangle\\
- \quad & \nabla g(\bar{x})H^p + \nabla g(\bar{x})E^p\\
= \quad & \nabla h(\bar{x})\frac{(y^p-\bar{y})}{t_p} + \nabla g(\bar{x})(G^p + E^p) \in \mathrm{Im}\nabla h(\bar{x}) + \nabla g(\bar{x})\Phi,
\end{aligned}
$$

where the set in the right hand side, as a sum of a linear subspace and a closed set, is closed, since $g'(\bar{x})^T\Phi$ is supposed to be closed. Then by taking limit as $p \to \infty$, it yeilds

$$-\nabla^2_{xx} l(\bar x,\bar y,\bar S)d_{\bar x} - \nabla g(\bar x)H \in \mathrm{Im}\nabla h(\bar x) + \nabla g(\bar x)\Phi. \tag{5.76}$$

The inclusion (5.76) means that there exists $(d_{\bar y}, G)\in \mathcal{Y}\times\Phi$ such that

$$\nabla^2_{xx} l(\bar x,\bar y,\bar S)d_{\bar x} + \nabla g(\bar x)H + \nabla h(\bar x)d_{\bar y} + \nabla g(\bar x)G = 0. \tag{5.77}$$

Let $d_{\bar S} := H+G$ and $\widetilde{d}_{\bar S} := U^T d_{\bar S} V$. Then combining (5.69) and (5.74) with Lemma 5.3, we have $(d_{\bar x}, d_{\bar y}, d_{\bar S}) \in \widetilde{\mathcal{C}}(\bar x,\bar y,\bar S)$. This further indicates that $0\neq d_{\bar x}\in\mathcal{C}(\bar x)$ of problem (5.1).

Therefore, by making use of the assumption (ii), we have

$$
\begin{aligned}
& \langle d_{\bar x}, \nabla^2_{xx} l(\bar x,\bar y,\bar S)d_{\bar x}\rangle + \Upsilon_{g(\bar x)}\big(\bar S, g'(\bar x)d_{\bar x}\big)\\
= \;& -\langle d_{\bar y}, h'(\bar x)d_{\bar x}\rangle - \langle d_{\bar S}, g'(\bar x)d_{\bar x}\rangle + \Upsilon_{g(\bar x)}\big(\bar S, g'(\bar x)d_{\bar x}\big)\\
= \;& -\langle G, g'(\bar x)d_{\bar x}\rangle - \langle H, g'(\bar x)d_{\bar x}\rangle + \Upsilon_{g(\bar x)}\big(\bar S, g'(\bar x)d_{\bar x}\big)\\
= \;& -\langle G, g'(\bar x)d_{\bar x}\rangle - \langle [\mathcal{E}_{\mathbb{S}}\circ\mathbb{S}(\widetilde D_{\bar x,1}) + \mathcal{E}_{\mathbb{T}}\circ\mathbb{T}(\widetilde D_{\bar x,1}) \quad \mathcal{F}\circ(\widetilde D_{\bar x,2})], \widetilde D_{\bar x}\rangle + \Upsilon_{g(\bar x)}\big(\bar S, g'(\bar x)d_{\bar x}\big)\\
= \;& -\langle G, g'(\bar x)d_{\bar x}\rangle - \langle \mathcal{E}_{\mathbb{S}}\circ\mathbb{S}(\widetilde D_{\bar x,1}), \mathbb{S}(\widetilde D_{\bar x,1})\rangle - \langle \mathcal{E}_{\mathbb{T}}\circ\mathbb{T}(\widetilde D_{\bar x,1}), \mathbb{T}(\widetilde D_{\bar x,1})\rangle - \langle \mathcal{F}\circ(\widetilde D_{\bar x,2}), \widetilde D_{\bar x,2}\rangle\\
& +\Upsilon_{g(\bar x)}\big(\bar S, g'(\bar x)d_{\bar x}\big)\\
= \;& -\langle G, g'(\bar x)d_{\bar x}\rangle + \Upsilon_{g(\bar x)}\big(\bar S, g'(\bar x)d_{\bar x}\big)\\
& -\sum_{1\le l,t\le r_0}\frac{2}{\nu_l(g(\bar x))+\nu_t(g(\bar x))}\|(\mathbb{T}(\widetilde D_{\bar x,1}))_{a_l a_t}\|^2 - \sum_{1\le l\le r_0}\frac{4}{\nu_l(g(\bar x))}\|(\mathbb{T}(\widetilde D_{\bar x,1}))_{a_l\beta_1}\|^2\\
& -\sum_{\substack{1\le l\le r_0\\ 1\le i-|\alpha|-|\beta_1|\le|\beta_2|}}\left(\frac{2(1-\sigma_i(\bar S))}{\nu_l(g(\bar x))}\|(\mathbb{S}(\widetilde D_{\bar x,1}))_{a_l i}\|^2 + \frac{2(\sigma_i(\bar S)+1)}{\nu_l(g(\bar x))}\|(\mathbb{T}(\widetilde D_{\bar x,1}))_{a_l i}\|^2\right)\\
& -\sum_{1\le l\le r_0}\left(\frac{2}{\nu_l(g(\bar x))}\|(\mathbb{S}(\widetilde D_{\bar x,1}))_{a_l\beta_3}\|^2 + \frac{2}{\nu_l(g(\bar x))}\|(\mathbb{T}(\widetilde D_{\bar x,1}))_{a_l\beta_3}\|^2\right)\\
& -\sum_{1\le l\le r_0}\frac{1}{\nu_l(g(\bar x))}\|(\widetilde D_{\bar x})_{a_l c}\|^2\\
= \;& -\langle G, g'(\bar x)d_{\bar x}\rangle + \Upsilon_{g(\bar x)}\big(\bar S, g'(\bar x)d_{\bar x}\big) - \Upsilon_{g(\bar x)}\big(\bar S, g'(\bar x)d_{\bar x}\big)\\
= \;& -\langle G, g'(\bar x)d_{\bar x}\rangle = -\langle g'(\bar x)d_{\bar x}, \Pi_{\Phi}(d_{\bar S})\rangle = 0,
\end{aligned}
$$

which contradicts the assumption (iii) that the second-order sufficient condition (5.61) holds at $\bar x$ with respect to the multiplier $(\bar y,\bar S)\in\mathcal{M}(\bar x)$. Hence, there exist a constant $\kappa > 0$ and a neighborhood $\mathcal{U}$ of $(\bar{x}, \bar{y}, \bar{S})$ such that (5.67) holds.

Case (ii) $\sigma_k(g(\bar{x})) = 0$ and $\|\bar{S}\|_* < k - k_0$. Same proof as Theorem 4.1 and we omit here.

Case (iii) $\sigma_k(g(\bar{x})) > 0$. Similar to case (i), we set

$$\Omega := \{W \in \mathcal{R}^{m\times n} \mid \mathbb{S}(U_\beta^T W V_\beta) = 0\},$$

and for all $p \geq 0$,

$$\begin{cases} A^p := g(x^p) + C^p, \ \widetilde{A}^p := U^T A^p V = [\widetilde{A}_1^p \ \widetilde{A}_2^p] = [U^T A^p V_1 \ U^T A^p V_2], \\ B^p := S^p, \ \widetilde{B}^p := U^T B^p V = [\widetilde{B}_1^p \ \widetilde{B}_2^p] = [U^T B^p V_1 \ U^T B^p V_2], \\ \Delta A^p = A^p - A, \ \Delta B^p = B^p - B, \\ H^p := \Pi_\Omega\big((B^p - B)/t_p\big), \ G^p := (B^p - B)/t_p - H^p. \end{cases} \tag{5.78}$$

Thus, $A^p \to A$ and $B^p \to B$ by the assumptions. Moreover, similar to (5.68), we can derive that

$$\frac{1}{t_p}(A^p - A) \to g'(\bar{x})d_{\bar{x}} \text{ as } p \to \infty. \tag{5.79}$$

Since $B \in \partial\theta(A)$ and $B^p \in \partial\theta(A^p)$, we can derive the following estimates by Proposition 5.2 that for all (A^p, B^p) sufficiently close to (A, B), $\sigma_k(A^p) > 0$,

$$\begin{cases} \mathbb{S}(\widetilde{A}_1^p)_{\beta_1\beta_1} + O(\|\Delta A^p\|^2 + \|\Delta A^p\|\|\Delta B^p\|) \succeq \sigma_k(A^p) I_{|\beta_1|}, \\ \mathbb{S}(\widetilde{A}_1^p)_{\beta_2\beta_2} = \sigma_k(A^p) I_{|\beta_2|} + O(\|\Delta A^p\|\|\Delta B^p\|), \\ \mathbb{S}(\widetilde{A}_1^p)_{\beta_3\beta_3} + O(\|\Delta A^p\|^2 + \|\Delta A^p\|\|\Delta B^p\|) \preceq \sigma_k(A^p) I_{|\beta_3|}, \\ \mathbb{S}(\widetilde{A}_1^p)_{\beta_1(\beta_2\cup\beta_3)} = O(\|\Delta A^p\|\|\Delta B^p\|), \ \mathbb{S}(\widetilde{A}_1^p)_{\beta_2\beta_3} = O(\|\Delta A^p\|\|\Delta B^p\|), \\ \mathbb{S}(\widetilde{B}_1^p)_{\alpha\alpha} = I_{|\alpha|} + O(\|\Delta A^p\|^2), \ \mathbb{S}(\widetilde{B}_1^p)_{\alpha\beta_1} = O(\|\Delta A^p\|^2 + \|\Delta A^p\|\|\Delta B^p\|), \\ \mathbb{S}(\widetilde{B}_1^p)_{\beta_1\beta_1} \preceq I_{|\beta_1|}, \ (\widetilde{B}_1^p)_{\beta_3\beta_3} + O(\|\Delta A^p\|\|\Delta B^p\|) \in \mathcal{S}_+^{|\beta_3|}, \\ \widetilde{B}^p_{\beta_3(\gamma\cup c)} = O(\|\Delta A^p\|\|\Delta B^p\|), \ \widetilde{B}^p_{\gamma(\beta_3\cup\gamma\cup c)} = O(\|\Delta A^p\|\|\Delta B^p\|), \\ \mathrm{tr}(\widetilde{B}^p_{\beta\beta}) = \mathrm{tr}(\Sigma(B)_{\beta\beta}) + O(\|\Delta A^p\|^2), \end{cases} \tag{5.80}$$

and

$$\begin{cases} \mathbb{T}(\widetilde{A}_1^p)_{\alpha\alpha} = \frac{1}{2}\big(\Sigma(A)_{\alpha\alpha}\mathbb{T}(\widetilde{B}_1^p)_{\alpha\alpha} + \mathbb{T}(\widetilde{B}_1^p)_{\alpha\alpha}\Sigma(A)_{\alpha\alpha}\big) + O(\|\Delta A^p\|^2), \\ \mathbb{T}(\widetilde{A}_1^p)_{\alpha\beta_1} = \frac{1}{2}\big(\Sigma(A)_{\alpha\alpha} + \sigma_k(A)I_{|\alpha|}\big)\mathbb{T}(\widetilde{B}_1^p)_{\alpha\beta_1} + O(\|\Delta A^p\|^2 + \|\Delta A^p\|\|\Delta B^p\|), \\ \widetilde{A}_{\alpha\beta_2}^p - (\widetilde{A}_{\beta_2\alpha}^p)^T\Sigma(B)_{\beta_2\beta_2} = \Sigma(A)_{\alpha\alpha}\widetilde{B}_{\alpha\beta_2}^p - \sigma_k(A)(\widetilde{B}_{\beta_2\alpha}^p)^T + O(\|\Delta A^p\|\|\Delta B^p\|), \\ \widetilde{A}_{\alpha\beta_3}^p = \Sigma(A)_{\alpha\alpha}\widetilde{B}_{\alpha\beta_3}^p - \sigma_k(A)(\widetilde{B}_{\beta_3\alpha}^p)^T + O(\|\Delta A^p\|^2 + \|\Delta A^p\|\|\Delta B^p\|), \\ \widetilde{A}_{\alpha\gamma}^p = \Sigma(A)_{\alpha\alpha}\widetilde{B}_{\alpha\gamma}^p - (\widetilde{B}_{\gamma\alpha}^p)^T\Sigma(A)_{\gamma\gamma} + O(\|\Delta A^p\|\|\Delta B^p\|), \\ \widetilde{A}_{\alpha c}^p = \Sigma(A)_{\alpha\alpha}\widetilde{B}_{\alpha c}^p + O(\|\Delta A^p\|\|\Delta B^p\|). \end{cases} \tag{5.81}$$

Moreover,

$$\begin{cases} \mathbb{T}(\widetilde{B}_1^p)_{\beta_1\beta_1} = \dfrac{1}{\sigma_k(A)}\mathbb{T}(\widetilde{A}_1^p)_{\beta_1\beta_1} + O(\|\Delta A^p\|^2 + \|\Delta A^p\|\|\Delta B^p\|), \\ \mathbb{T}(\widetilde{B}_1^p)_{\beta_1\beta_2} = \dfrac{1}{2\sigma_k(A)}\mathbb{T}(\widetilde{A}_1^p)_{\beta_1\beta_2}(I_{|\beta_2|} + \Sigma(B)_{\beta_2\beta_2}) + O(\|\Delta A^p\|^2 + \|\Delta A^p\|\|\Delta B^p\|), \\ \mathbb{T}(\widetilde{B}_1^p)_{\beta_1\beta_3} = \dfrac{1}{2\sigma_k(A)}\mathbb{T}(\widetilde{A}_1^p)_{\beta_1\beta_3} + O(\|\Delta A^p\|^2 + \|\Delta A^p\|\|\Delta B^p\|), \\ \mathbb{T}(\widetilde{B}_1^p)_{\beta_2\beta_2} = \dfrac{1}{2\sigma_k(A)}(\Sigma(B)_{\beta_2\beta_2}\mathbb{T}(\widetilde{A}_1^p)_{\beta_2\beta_2} + \mathbb{T}(\widetilde{A}_1^p)_{\beta_2\beta_2}\Sigma(B)_{\beta_2\beta_2}) + O(\|\Delta A^p\|\|\Delta B^p\|), \\ \mathbb{T}(\widetilde{B}_1^p)_{\beta_2\beta_3} = \dfrac{1}{2\sigma_k(A)}\Sigma(B)_{\beta_2\beta_2}\mathbb{T}(\widetilde{A}_1^p)_{\beta_2\beta_3} + O(\|\Delta A^p\|^2 + \|\Delta A^p\|\|\Delta B^p\|), \\ \widetilde{A}_{\beta_1\gamma}^p = \sigma_k(A)\widetilde{B}_{\beta_1\gamma}^o - (\widetilde{B}_{\gamma\beta_1}^p)^T\Sigma(A)_{\gamma\gamma} + O(\|\Delta A^p\|^2 + \|\Delta A^p\|\|\Delta B^p\|), \\ \widetilde{A}_{\beta_1 c}^p = \sigma_k(A)\widetilde{B}_{\beta_1 c}^p + O(\|\Delta A^p\|^2 + \|\Delta A^p\|\|\Delta B^p\|), \\ \widetilde{A}_{\beta_2\gamma}^p = \Sigma(B)_{\beta_2\beta_2}^{-1}(\sigma_k(A)\widetilde{B}_{\beta_2\gamma}^p - (\widetilde{B}_{\gamma\beta_2}^p)^T\Sigma(A)_{\gamma\gamma}) + O(\|\Delta A^p\|\|\Delta B^p\|), \\ \widetilde{A}_{\beta_2 c}^p = \sigma_k(A)\Sigma(B)_{\beta_2\beta_2}^{-1}\widetilde{B}_{\beta_2 c}^p + O(\|\Delta A^p\|\|\Delta B^p\|). \end{cases} \tag{5.82}$$

Combining the above (5.79) - (5.82) with Lemma 5.2, we obtain

$$\begin{cases} g'(\bar{x})d_{\bar{x}} \in \mathcal{C}_\theta(g(\bar{x}), \bar{S}), \;\; G^p + \dfrac{1}{t_p}O(\|\Delta A^p\|^2 + \|\Delta A^p\|\|\Delta B^p\|) \in \Phi, \\ H := \lim\limits_{k\to\infty} H^k = U[\mathcal{E}_{\mathbb{S}} \circ \mathbb{S}(\widetilde{D}_{\bar{x},1}) + \mathcal{E}_{\mathbb{T}} \circ \mathbb{T}(\widetilde{D}_{\bar{x},1}) \quad \mathcal{F} \circ (\widetilde{D}_{\bar{x},2})]V^T, \end{cases} \tag{5.83}$$

where $\widetilde{D}_{\bar{x}} := U^T g'(\bar{x})d_{\bar{x}}V = [\widetilde{D}_{\bar{x},1} \; \widetilde{D}_{\bar{x},2}] = [U^T D_{\bar{x}}V_1 \; U^T D_{\bar{x}}V_2]$.

Thus, there exist $E^p = \dfrac{1}{t_p}O(\|\Delta A^p\|^2 + \|\Delta A^p\|\|\Delta B^p\|) \in \mathcal{R}^{m\times n}$ such that $G^p + E^p \in \Phi$.

Again, by the perturbed KKT system (5.60), we can deduce that for $p \geq 0$ large enough,

$$\begin{aligned} u^p &= \nabla f(x^p) + \nabla h(x^p)y^p + \nabla g(x^p)S^p - \big(\nabla f(\bar{x}) + \nabla h(\bar{x})\bar{y} + \nabla g(\bar{x})\bar{S}\big) \\ &= \nabla^2_{xx} f(\bar{x})(x^p - \bar{x}) + \langle y^p, h''(\bar{x})(x^p - \bar{x})\rangle + \langle S^p, g''(\bar{x})(x^p - \bar{x})\rangle \\ &+ \nabla h(\bar{x})(y^p - \bar{y}) + \nabla g(\bar{x})(S^p - \bar{S}). \end{aligned} \tag{5.84}$$

Dividing by t_p on both sides of (5.84), and then adding $\nabla g(\bar{x})E^p$ on both sides, it gives

$$\begin{aligned} & \frac{u^p}{t_p} - \nabla^2_{xx} f(\bar{x})\frac{(x^p - \bar{x})}{t_p} - \langle y^p, h''(\bar{x})\frac{(x^p - \bar{x})}{t_p}\rangle - \langle S^p, g''(\bar{x})\frac{(x^p - \bar{x})}{t_p}\rangle \\ - & \nabla g(\bar{x})H^p + \nabla g(\bar{x})E^p \\ = & \nabla h(\bar{x})\frac{(y^p - \bar{y})}{t_p} + \nabla g(\bar{x})(G^p + E^p) \in \mathrm{Im}\nabla h(\bar{x}) + \nabla g(\bar{x})\Phi, \end{aligned}$$

where the set in the right hand side, as a sum of a linear subspace and a closed set, is closed, since $g'(\bar{x})^T\Phi$ is supposed to be closed. Then by taking limit as $p \to \infty$, it yeilds

$$-\nabla^2_{xx} l(\bar{x}, \bar{y}, \bar{S})d_{\bar{x}} - \nabla g(\bar{x})H \in \mathrm{Im}\nabla h(\bar{x}) + \nabla g(\bar{x})\Phi. \tag{5.85}$$

The inclusion (5.85) means that there exists $(d_{\bar{y}}, G) \in \mathcal{Y} \times \Phi$ such that

$$\nabla^2_{xx} l(\bar{x}, \bar{y}, \bar{S})d_{\bar{x}} + \nabla g(\bar{x})H + \nabla h(\bar{x})d_{\bar{y}} + \nabla g(\bar{x})G = 0. \tag{5.86}$$

Let $d_{\bar{S}} := H + G$ and $\widetilde{d}_{\bar{S}} := U^T d_{\bar{S}} V$. Then combining (5.69) and (5.74) with Lemma 5.3, we have $(d_{\bar{x}}, d_{\bar{y}}, d_{\bar{S}}) \in \widetilde{\mathcal{C}}(\bar{x}, \bar{y}, \bar{S})$. Similar to case (i), it is easy for us to find that $\langle H, g'(\bar{x})d_{\bar{x}}\rangle = \Upsilon_{g(\bar{x})}\big(\bar{S}, g'(\bar{x})d_{\bar{x}}\big)$ via the explicit expression of H in (5.83). This further indicates that $0 \neq d_{\bar{x}} \in \mathcal{C}(\bar{x})$ of problem (5.1).

Therefore, by making use of the assumption (ii), we have

$$\begin{aligned} & \langle d_{\bar{x}}, \nabla^2_{xx} l(\bar{x}, \bar{y}, \bar{S})d_{\bar{x}}\rangle + \Upsilon_{g(\bar{x})}\big(\bar{S}, g'(\bar{x})d_{\bar{x}}\big) \\ = & -\langle d_{\bar{y}}, h'(\bar{x})d_{\bar{x}}\rangle - \langle d_{\bar{S}}, g'(\bar{x})d_{\bar{x}}\rangle + \Upsilon_{g(\bar{x})}\big(\bar{S}, g'(\bar{x})d_{\bar{x}}\big) \\ = & -\langle G, g'(\bar{x})d_{\bar{x}}\rangle - \langle H, g'(\bar{x})d_{\bar{x}}\rangle + \Upsilon_{g(\bar{x})}\big(\bar{S}, g'(\bar{x})d_{\bar{x}}\big) \\ = & -\langle G, g'(\bar{x})d_{\bar{x}}\rangle + \Upsilon_{g(\bar{x})}\big(\bar{S}, g'(\bar{x})d_{\bar{x}}\big) - \Upsilon_{g(\bar{x})}\big(\bar{S}, g'(\bar{x})d_{\bar{x}}\big) \\ = & -\langle G, g'(\bar{x})d_{\bar{x}}\rangle = -\langle g'(\bar{x})d_{\bar{x}}, \Pi_\Phi(d_{\bar{S}})\rangle = 0, \end{aligned}$$

which contradicts the assumption (iii) that the second-order sufficient condition (5.61) holds at $\bar{x}$ with respect to the multiplier $(\bar{y}, \bar{S}) \in \mathcal{M}(\bar{x})$.
Hence, there exist a constant $\kappa > 0$ and a neighborhood $\mathcal{U}$ of $(\bar{x}, \bar{y}, \bar{S})$ such that (5.67) holds for all the cases.

Next, we will prove that $\mathcal{T}_l$ is metrically subregular at $(\bar{x}, \bar{y}, \bar{S})$ for the origin under an additional assumption, which requires that there exist $(\widehat{y}, \widehat{S}) \in \mathcal{M}(\bar{x})$ such that

(a) $0 < \sigma(\widehat{S})_\beta < e_\beta$ and $\sum_{i\in\beta} < k - k_0$ if $\sigma_k(g(\bar{x})) = 0$;

(b) $0 < \sigma(\widehat{S})_\beta < e_\beta$ if $\sigma_k(g(\bar{x})) > 0$.

In another word, it is equivalent to show that there exist a constant $\kappa' > 0$ and a neighborhood $\mathcal{U}'$ of $(\bar{x}, \bar{y}, \bar{S})$ such that for any $(u, v, C) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^{m\times n}$,

$$\operatorname{dist}((x, y, S), \mathcal{T}_l^{-1}(0)) \leq \kappa' \|(u, v, C)\|, \ \forall (x, y, S) \in \mathcal{T}_l^{-1}(u, v, C) \cap \mathcal{U}'. \tag{5.87}$$

For the convenience, we set

$$\Psi := \{(y, S) |\ (\bar{x}, y, S) \in \mathcal{T}_l^{-1}(0, 0, 0)\},$$

$$\Xi_1 := \{(y, S) |\ \nabla f(\bar{x}) + \nabla h(\bar{x}) y + \nabla g(\bar{x}) S = 0\}, \ \Xi_2 := \{(y, S) |\ S \in \partial\theta(g(\bar{x}))\}.$$

One can easily find that $\Psi = \Xi_1 \cap \Xi_2$ and $(\widehat{y}, \widehat{S}) \in \Xi_1 \cap \operatorname{ri}(\Xi_2)$. Thus, by Proposition 2.1, we have that there exists a constant $\kappa_1 > 0$ such that for any $(x, y, S) \in \mathcal{U}'$,

$$\operatorname{dist}((y, S), \Phi) \leq \kappa_1 \big(\operatorname{dist}((y, S), \Xi_1) + \operatorname{dist}((y, S), \Xi_2)\big). \tag{5.88}$$

For any given point $(x, y, S) \in \mathcal{T}_l^{-1}(u, v, C) \cap \mathcal{U}'$, we assume that $\|(y, S)\| \leq \eta$ with some $\eta > 0$ by shrinking $\mathcal{U}'$ if necessary. Fixing that given point, using Hoffman's error bound and the twice continuous differentiability of f, h and g, shrinking $\mathcal{U}'$ if necessary, we obtain that there exist constants $\kappa_2 > 0$ and $\kappa_2' > 0$ such that

$$\begin{aligned} \operatorname{dist}((y, S), \Xi_1) &\leq \kappa_2 \|\nabla f(\bar{x}) + \nabla h(\bar{x}) y + \nabla g(\bar{x}) S\| \\ &\leq \kappa_2 (\|\nabla f(x) - \nabla f(\bar{x})\| + \|\nabla h(x) - \nabla h(\bar{x})\| \|y\| \\ &\quad + \|\nabla g(x) - \nabla g(\bar{x})\| \|S\| + \|u\|) \\ &\leq \kappa_2' (\|x - \bar{x}\| + \|u\|). \end{aligned} \tag{5.89}$$

By Proposition 2.15, we have $(\partial\theta)^{-1}(\cdot) = \partial\theta^*(\cdot)$ is metrically subregular at $\bar{S}$ for $g(\bar{x})$. Together with $g(x) + C \in \partial\theta^*(S)$ and the twice continuous differentiabilty of g , we can duduce, shrinking $\mathcal{U}'$ if necessary, that there exist contants $\kappa_3 > 0$ and $\kappa_3' > 0$ such that

$$\begin{aligned} \mathrm{dist}((y,S),\Xi_2) &= \mathrm{dist}(S, \partial\theta(g(\bar{x}))) \\ &\leq \kappa_3 \mathrm{dist}(g(\bar{x}), \partial\theta^*(S)) \\ &\leq \kappa_3 \|g(x) + C - g(\bar{x})\| \\ &\leq \kappa_3'(\|x-\bar{x}\| + \|C\|). \end{aligned} \tag{5.90}$$

Therefore, we can find that there exist a constant $\kappa' > 0$ and a neighborhood $\mathcal{U}'$ of $(\bar{x}, \bar{y}, \bar{S})$ such that (5.87) holds, by using the inequalities (5.67) and (5.88)-(5.90). This implies $\mathcal{T}_l$ is metrically subregular at $(\bar{x}, \bar{y}, \bar{S})$ for the origin. □

Remark 5.2. *One can obtain a similar result as Remark 4.2. Especially, when the program data of the problem (5.1) is relaxed to* $C^{1,1}$*, if the gradients* ∇f*,* ∇g *and* ∇h *are directionally differentiable at* $\bar{x}$*, then we can obtain the metric subregularity of* $\mathcal{T}_l$ *at* $(\bar{x}, \bar{y}, \bar{S})$ *for the origin by changing the corresponding SOSC (5.61) in the assumption (iii) of Theorem 5.1 to the following form:*

$$\langle d, (\nabla_x l)'(\bar{x}, \bar{y}, \bar{S}; d)\rangle + \Upsilon_{g(\bar{x})}\big(\bar{S}, g'(\bar{x})d\big) > 0, \ \forall\, 0 \neq d \in \mathcal{C}(\bar{x}) \tag{5.91}$$

with $\Upsilon_{g(\bar{x})}\big(\bar{S}, g'(\bar{x})d\big)$ *defined as (5.62) or (5.63), where* $(\nabla_x l)'(\bar{x}, \bar{y}, \bar{S}; d)$ *denotes the directional derivative of* $\nabla_x l(\cdot, \bar{y}, \bar{S})$ *at* $\bar{x}$ *in the direction* d*.*

Remark 5.3. *The assumption (ii) in the above theorem is*

$$0 = \langle g'(\bar{x})d_x,\ \Pi_\Phi(d_S)\rangle,$$

whereas the assumption (ii) in Theorem 4.1 of the nuclear norm case is

$$0 = \langle \Pi_\Phi(-g'(\bar{x})d_{\bar{x}}), \Pi_\Phi(d_{\bar{S}})\rangle.$$

Here we briefly explain why these two assumptions coincide in the nuclear norm case. In Theorem 4.1, we have

$$0 = \langle \Pi_\Phi(-g'(\bar{x})d_{\bar{x}}), \Pi_\Phi(d_{\bar{S}}) \rangle = -\langle (\widetilde{D}_{\bar{x}})_{\beta_1\beta_1}, (\widetilde{d}_{\bar{S}})_{\beta_1\beta_1} \rangle.$$

And, by the assumption (ii) in Theorem 5.1 restricted to the nuclear norm case, we also have

$$0 = \langle g'(\bar{x})d_x,\ \Pi_\Phi(d_S) \rangle = \langle (\widetilde{D}_{\bar{x}})_{\beta_1\beta_1}, (\widetilde{d}_{\bar{S}})_{\beta_1\beta_1} \rangle.$$

Thus, both of these two assumptions are equivalent to $\langle (\widetilde{D}_{\bar{x}})_{\beta_1\beta_1}, (\widetilde{d}_{\bar{S}})_{\beta_1\beta_1} \rangle = 0$.

Chapter 6

Conclusions

In this thesis, we study the stability of composite optimization problems, whose objective functions involve convex composite terms. Many important optimization problems arising from various areas such as finance, engineering, applied mathematics and so on, can be reformulated as composite problems. Due to the interest in theory and practice, we study the stability of the composite SDP conic programming and the composite Ky Fan k-norm regularized conic programming. Different from previous studies of the stability with the requirement of the Lagrange multipliers to be unique, our discussions allow the multiplier set of the aforementioned composite problems to be non-singleton.

Within the multiplier set to be non-singleton, motivated by recent studies for nonlinear programming [46] and convex SDP problems [20, 22], we investigate the metric subregularity for the KKT solution mappings of the composite problems, which may not be convex. The study of the metric subregularity is mainly based on the second order sensitivity analysis of the SDP cone and the Ky Fan k-norm. To explore sufficient conditions for the metric subregularity, we extend the perturbation analysis of symmetric matrices to nonsymmetric matrices. Such perturbation properties reveal the curvature information of the SDP cone and the Ky Fan k-norm. Meanwhile, the curvature of the SDP cone and the Ky Fan k-norm are also taken into account by the second order sufficient condition. Therefore, under the canonical perturbation of composite problems, within the assumption of the second order sufficient condition, we obtain an error bound for a locally optimal solution of those underlying composite conic programming. Additionally, if a partial strict complementarity condition holds, an error bound for the corresponding multiplier set is estimated. Our study plays a transition role from a convex problem to a nonconvex problem and from a symmetric conic programming to a nonsymmetric conic programming. Compared to the study of the NLP, our discussions of the metric subregularity are conducted under a more complicated situation and can cover those for the NLP.

Those error bound results can be applied to obtain fast convergent rates of primal-dual methods, e.g., the alternating direction method of multipliers [34, 41] and proximal augmented Lagrange methods [84] of convex problems. This application is one direct extension of our work. There are also many other intereting topics for our future study. First of all, it is interesting to explore the stability when the partial strict complementarity condition fails but the multiplier set is non-singleton. Moreover, in this thesis, we only discuss two types of composite programming here — the SDP cone and the Ky Fan k-norm, both of which can be cast in the class of spectral functions. An extensive study of composite programming involving spectral functions under general settings is one attractive topic. Or even the characterization of the curvature of spectral functions via perturbation analysis is a challenge. Moreover, in our research, we discover the sufficient conditions for the metric subregularity. One can provide weaker sufficient conditions or sufficient and necessary conditions to characterize the metric subregularity in future. Finally, a further study of the stability beyond the composite programming is another research direction.

Bibliography

[1] F. J. Aragón Artacho and M. H. Geoffroy, *Characterization of metric regularity of subdifferentials*, Journal of Convex Analysis, 15 (2008), pp. 365–380.

[2] V. I. Arnold, *On matrices depending on parameters*, Russian Mathematical Surveys, 26 (1971), pp. 29–43.

[3] J.-P. Aubin, *Lipschitz behavior of solutions to convex minimization problems*, Mathematics of Operations Research, 9 (1984), pp. 87–111.

[4] D. Azé, *A survey on error bounds for lower semicontinuous functions*, in ESAIM: Proceedings, vol. 13, EDP Sciences, 2003, pp. 1–17.

[5] H. H. Bauschke and J. M. Borwein, *On projection algorithms for solving convex feasibility problems*, SIAM Review, 38 (1996), pp. 367–426.

[6] H. H. Bauschke, J. M. Borwein, and W. Li, *Strong conical hull intersection property, bounded linear regularity, Jameson's property (G), and error bounds in convex optimization*, Mathematical Programming, 86 (1999), pp. 135–160.

[7] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer Science & Business Media, 2011.

[8] J. F. Bonnans and A. Shapiro, *Perturbation analysis of optimization problems*, Springer Science & Business Media, 2000.

[9] S. Boyd, P. Diaconis, P. Parrilo, and L. Xiao, *Fastest mixing Markov chain on graphs with symmetries*, SIAM Journal on Optimization, 20 (2009), pp. 792–819.

[10] S. Boyd, P. Diaconis, J. Sun, and L. Xiao, *Fastest mixing Markov chain on a path*, The American Mathematical Monthly, 113 (2006), pp. 70–74.

[11] S. Boyd, P. Diaconis, and L. Xiao, *Fastest mixing Markov chain on a graph*, SIAM Review, 46 (2004), pp. 667–689.

[12] J. V. Burke, *Descent methods for composite nondifferentiable optimization problems*, Mathematical Programming, 33 (1985), pp. 260–279.

[13] ——, *Second order necessary and sufficient conditions for convex composite ndo*, Mathematical Programming, 38 (1987), pp. 287–302.

[14] J. V. Burke and R. Poliquin, *Optimality conditions for non-finite valued convex composite functions*, Mathematical Programming, 57 (1992), pp. 103–120.

[15] E. Candes and B. Recht, *Exact matrix completion via convex optimization*, Communications of the ACM, 55 (2012), pp. 111–119.

[16] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, *Rank-sparsity incoherence for matrix decomposition*, SIAM Journal on Optimization, 21 (2011), pp. 572–596.

[17] Y. Chen and Y. Chi, *Spectral compressed sensing via structured matrix completion.*, in ICML (3), 2013, pp. 414–422.

[18] M. T. Chu, R. E. Funderlic, and R. J. Plemmons, *Structured low rank approximation*, Linear Algebra and Its Applications, 366 (2003), pp. 157–172.

[19] Y. Cui, *Large scale composite optimization problems with coupled objective functions: theory, algorithms and applications*, PhD thesis, National University of Singapore, 2016.

[20] Y. Cui, C. Ding, and X. Zhao, *Quadratic growth conditions for convex matrix optimization problems associated with spectral functions*, arXiv preprint arXiv:1702.03262, (2017).

[21] Y. Cui and D. Sun, *A complete characterization on the robust isolated calmness of the nuclear norm regularized convex optimization problems*, arXiv preprint arXiv:1702.05914, (2017).

[22] Y. Cui, D. Sun, and K.-C. Toh, *On the asymptotic superlinear convergence of the augmented Lagrangian method for semidefinite programming with multiple solutions*, arXiv preprint arXiv:1610.00875, (2016).

[23] R. R. De Gaston and M. G. Safonov, *Exact calculation of the multiloop stability margin*, IEEE Transactions on Automatic Control, 33 (1988), pp. 156–171.

[24] C. Ding, *An introduction to a class of matrix optimization problems*, PhD thesis, National University of Singapore, 2012.

[25] C. Ding, *Variational analysis of the Ky Fan k-norm*, Set-Valued and Variational Analysis, (2016), pp. 1–32.

[26] C. Ding, D. Sun, and L. Zhang, *Characterization of the robust isolated calmness for a class of conic programming problems*, SIAM Journal on Optimization, 27 (2017), pp. 67–90.

[27] V. Dobrynin, *On the rank of a matrix associated with a graph*, Discrete Mathematics, 276 (2004), pp. 169–175.

[28] A. L. Dontchev and R. Rockafellar, *Characterizations of Lipschitzian stability in nonlinear programming*, in Mathematical Programming with Data Perturbations, Marcel Dekker, Inc, 1998, pp. 65–82.

[29] A. L. Dontchev and R. Rockafellar, *Regularity and conditioning of solution mappings in variational analysis*, Set-Valued Analysis, 12 (2004), pp. 79–109.

[30] A. L. Dontchev and R. T. Rockafellar, *Implicit functions and solution mappings*, Springer Monogr. Math., (2014).

[31] D. Drusvyatskiy and A. S. Lewis, *Error bounds, quadratic growth, and linear convergence of proximal methods*, arXiv preprint arXiv:1602.06661, (2016).

[32] M. J. Fabian, R. Henrion, A. Y. Kruger, and J. V. Outrata, *Error bounds: necessary and sufficient conditions*, Set-Valued and Variational Analysis, 18 (2010), pp. 121–149.

[33] F. Facchinei and J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems*, Springer Science & Business Media, 2007.

[34] M. Fazel, T. K. Pong, D. Sun, and P. Tseng, *Hankel matrix rank minimization with applications to system identification and realization*, SIAM Journal on Matrix Analysis and Applications, 34 (2013), pp. 946–977.

[35] R. Fletcher, *Practical methods of optimization*, John Wiley & Sons, 2013.

[36] C. A. Floudas and V. Visweswaran, *A global optimization algorithm (GOP) for certain classes of nonconvex NLPs – I. Theory*, Computers & Chemical Engineering, 14 (1990), pp. 1397–1417.

[37] P. Fusek, *Isolated zeros of Lipschitzian metrically regular-functions*, Optimization, 49 (2001), pp. 425–446.

[38] H. Gfrerer and B. S. Mordukhovich, *Complete characterizations of tilt stability in nonlinear programming under weakest qualification conditions*, SIAM Journal on Optimization, 25 (2015), pp. 2081–2119.

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

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

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

[42] N. J. Higham, *Computing the nearest correlation matrix—a problem from finance*, (2002).

[43] A. J. Hoffman, *On approximate solutions of systems of linear inequalities*, Journal of Research of the National Bureau of Standards, 49 (1952), pp. 263–265.

[44] M. Hong and Z.-Q. Luo, *On the linear convergence of the alternating direction method of multipliers*, Mathematical Programming, (2012), pp. 1–35.

[45] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, *Fast and accurate matrix completion via truncated nuclear norm regularization*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 35 (2013), pp. 2117–2130.

[46] A. F. Izmailov, A. S. Kurennoy, and M. V. Solodov, *A note on upper Lipschitz stability, error bounds, and critical multipliers for Lipschitz-continuous KKT systems*, Mathematical Programming, 142 (2013), pp. 591–604.

[47] R. H. Keshavan, A. Montanari, and S. Oh, *Matrix completion from a few entries*, IEEE Transactions on Information Theory, 56 (2010), pp. 2980–2998.

[48] A. J. King and R. T. Rockafellar, *Sensitivity analysis for nonsmooth generalized equations*, Mathematical Programming, 55 (1992), pp. 193–212.

[49] D. Klatte, *Upper Lipschitz behavior of solutions to perturbed $C^{1,1}$ programs*, Mathematical Programming, 88 (2000), pp. 285–311.

[50] F. Kleibergen and R. Paap, *Generalized reduced rank tests using the singular value decomposition*, Journal of Econometrics, 133 (2006), pp. 97–126.

[51] V. Koltchinskii, K. Lounici, and A. B. Tsybakov, *Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion*, The Annals of Statistics, (2011), pp. 2302–2329.

[52] A. Y. Kruger, *Error bounds and metric subregularity*, Optimization, 64 (2015), pp. 49–79.

[53] J. Kyparisis, *On uniqueness of Kuhn-Tucker multipliers in nonlinear programming*, Mathematical Programming, 32 (1985), pp. 242–246.

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

[55] A. S. Lewis and S. J. Wright, *A proximal method for composite minimization*, Mathematical Programming, 158 (2016), pp. 501–546.

[56] Y. Liu and S. Pan, *Locally upper Lipschitz of the perturbed KKT system of Ky Fan k-norm matrix conic optimization problems*, arXiv preprint arXiv:1509.00681, (2015).

[57] Z.-Q. Luo and P. Tseng, *Error bound and convergence analysis of matrix splitting algorithms for the affine variational inequality problem*, SIAM Journal on Optimization, 2 (1992), pp. 43–54.

[58] ——, *On the linear convergence of descent methods for convex essentially smooth minimization*, SIAM Journal on Control and Optimization, 30 (1992), pp. 408–425.

[59] ——, *Error bounds and convergence analysis of feasible descent methods: a general approach*, Annals of Operations Research, 46 (1993), pp. 157–178.

[60] W. Miao, S. Pan, and D. Sun, *A rank-corrected procedure for matrix completion with fixed basis coefficients*, Mathematical Programming, 159 (2016), pp. 289–338.

[61] A. Milzarek, *Numerical methods and second order theory for nonsmooth problems*, PhD thesis, Dissertation, München, Technische Universität München, 2016.

[62] B. S. Mordukhovich, *Variational analysis and generalized differentiation I: Basic theory*, vol. 330, Springer Science & Business Media, 2006.

[63] B. S. Mordukhovich, J. V. Outrata, and H. Ramírez, *Graphical derivatives and stability analysis for parameterized equilibria with conic constraints*, Set-Valued and Variational Analysis, 23 (2015), pp. 687–704.

[64] B. S. Mordukhovich, J. V. Outrata, and H. Ramirez C, *Second-order variational analysis in conic programming with applications to optimality and stability*, SIAM Journal on Optimization, 25 (2015), pp. 76–101.

[65] B. S. Mordukhovich, R. T. Rockafellar, and M. E. Sarabi, *Characterizations of full stability in constrained optimization*, SIAM Journal on Optimization, 23 (2013), pp. 1810–1849.

[66] B. S. Mordukhovich and M. E. Sarabi, *Critical multipliers in variational systems via second-order generalized differentiation*, arXiv preprint arXiv:1609.07199, (2016).

[67] J.-J. Moreau, *Proximité et dualité dans un espace Hilbertien*, Bulletin de la Société Mathématique de France, 93 (1965), pp. 273–299.

[68] J. v. Neumann, *Some matrix inequalities and metrization of matrix space*, Tomsk Univ. Rev, 1 (1937), pp. 286–300.

[69] M. Overton and R. Womersley, *Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices*, Mathematical Programming, 62 (1993), pp. 321–357.

[70] J.-S. Pang, *A posteriori error bounds for the linearly-constrained variational inequality problem*, Mathematics of Operations Research, 12 (1987), pp. 474–484.

[71] ——, *Error bounds in mathematical programming*, Mathematical Programming, 79 (1997), pp. 299–332.

[72] G. Pataki, *On the closedness of the linear image of a closed convex cone*, Mathematics of Operations Research, 32 (2007), pp. 395–412.

[73] J.-P. Penôt, *Error bounds, calmness and their applications in nonsmooth analysis*, Contemporary Math., 514 (2010), pp. 225–247.

[74] J.-P. Penot, *Calculus without derivatives*, vol. 266, Springer Science & Business Media, 2012.

[75] R. Pietersz 4 and P. J. Groenen, *Rank reduction of correlation matrices by majorization*, Quantitative Finance, 4 (2004), pp. 649–662.

[76] H. Qi and D. Sun, *A quadratically convergent newton method for computing the nearest correlation matrix*, SIAM Journal on Matrix Analysis and Applications, 28 (2006), pp. 360–385.

[77] R. Rebonato and P. Jäckel, *The most general methodology to create a valid correlation matrix for risk management and option pricing purposes*, (2011).

[78] B. Recht, *A simpler approach to matrix completion*, Journal of Machine Learning Research, 12 (2011), pp. 3413–3430.

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

[80] S. M. Robinson, *Generalized equations and their solutions, Part I: Basic theory*, in Point-to-Set Maps and Mathematical Programming, Springer, 1979, pp. 128–141.

[81] ——, *Some continuity properties of polyhedral multifunctions*, in Mathematical Programming at Oberwolfach, Springer, 1981, pp. 206–214.

[82] ——, *Generalized equations and their solutions, Part II: Applications to nonlinear programming*, in Optimality and Stability in Mathematical Programming, Springer, 1982, pp. 200–221.

[83] R. Rockafellar, *First-and second-order epi-differentiability in nonlinear programming*, Transactions of the American Mathematical Society, 307 (1988), pp. 75–108.

[84] R. T. Rockafellar, *Augmented lagrangians and applications of the proximal point algorithm in convex programming*, Mathematics of Operations Research, 1 (1976), pp. 97–116.

[85] ——, *Convex analysis*, Princeton university press, 1976.

[86] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, vol. 317, Springer Science & Business Media, 2009.

[87] H. S. Ryoo and N. V. Sahinidis, *Global optimization of nonconvex NLPs and MINLPs with applications in process design*, Computers & Chemical Engineering, 19 (1995), pp. 551–566.

[88] A. Shapiro, *On a class of nonsmooth composite functions*, Mathematics of Operations Research, 28 (2003), pp. 677–692.

[89] ——, *Sensitivity analysis of generalized equations*, Journal of Mathematical Sciences, 115 (2003), pp. 2554–2565.

[90] D. Sun, *The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications*, Mathematics of Operations Research, 31 (2006), pp. 761–776.

[91] J. Sun, *On monotropic piecewise quadratic programming*, PhD thesis, University of Washington, Seattle, 1986.

[92] R. J. Tibshirani et al., *The lasso problem and uniqueness*, Electronic Journal of Statistics, 7 (2013), pp. 1456–1490.

[93] K.-C. Toh, *GMRES vs. ideal GMRES*, SIAM Journal on Matrix Analysis and Applications, 18 (1997), pp. 30–36.

[94] K.-C. Toh and L. N. Trefethen, *The Chebyshev polynomials of a matrix*, SIAM Journal on Matrix Analysis and Applications, 20 (1998), pp. 400–419.

[95] P. Tseng, *Error bounds and superlinear convergence analysis of some Newton-type methods in optimization*, in Nonlinear Optimization and Related Topics, Springer, 2000, pp. 445–462.

[96] ——, *Convergence of a block coordinate descent method for nondifferentiable minimization*, Journal of Optimization Theory and Applications, 109 (2001), pp. 475–494.

[97] ——, *Approximation accuracy, gradient methods, and error bound for structured convex optimization*, Mathematical Programming, 125 (2010), pp. 263–295.

[98] P. Tseng and S. Yun, *A coordinate gradient descent method for nonsmooth separable minimization*, Mathematical Programming, 117 (2009), pp. 387–423.

[99] H. Van Ngai and M. Théra, *Error bounds in metric spaces and application to the perturbation stability of metric regularity*, SIAM Journal on Optimization, 19 (2008), pp. 1–20.

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

[101] V. Visweswaran and C. Floudast, *A global optimization algorithm (GOP) for certain classes of nonconvex NLPs – II. Application of theory and test problems*, Computers & Chemical Engineering, 14 (1990), pp. 1419–1434.

[102] G. Watson, *Linear best approximation using a class of polyhedral norms*, Numerical Algorithms 2, (1992), pp. 321–336.

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

[104] G. A. Watson, *Characterization of the subdifferential of some matrix norms*, Linear Algebra and its Applications, 170 (1992), pp. 33–45.

[105] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, *Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization*, in Advances in Neural Information Processing Systems, 2009, pp. 2080–2088.

[106] B. Wu, C. Ding, D. Sun, and K.-C. Toh, *On the Moreau–Yosida regularization of the vector k-norm related functions*, SIAM Journal on Optimization, 24 (2014), pp. 766–794.

[107] L. Wu, *Fast at-the-money calibration of libor market model through lagrange multipliers*, Journal of Computational Finance, 6 (2002), pp. 39–77.

[108] H. Zhang, J. Jiang, and Z.-Q. Luo, *On the linear convergence of a proximal gradient method for a class of nonsmooth convex minimization problems*, Journal of the Operations Research Society of China, 1 (2013), pp. 163–186.

[109] L. Zhang, N. Zhang, and X. Xiao, *On the second-order directional derivatives of singular values of matrices and symmetric matrix-valued functions*, Set-Valued and Variational Analysis, 21 (2013), pp. 557–586.

[110] R. Zhang and J. Treiman, *Upper-Lipschitz multifunctions and inverse subdifferentials*, Nonlinear Analysis: Theory, Methods & Applications, 24 (1995), pp. 273–286.

[111] Y. Zhang and L. Zhang, *On the upper Lipschitz property of the KKT mapping for nonlinear semidefinite optimization*, Operations Research Letters, 44 (2016), pp. 474–478.

[112] Z. Zhang and L. Wu, *Optimal low-rank approximation to a correlation matrix*, Linear Algebra and Its Applications, 364 (2003), pp. 161–187.

[113] Z. Zhou and A. M.-C. So, *A unified approach to error bounds for structured convex optimization problems*, Mathematical Programming, (2017), pp. 1–40.

[114] Z. Zhou, Q. Zhang, and A. M.-C. So, *$l_{1,p}$-norm regularization: Error bounds and convergence rate analysis of first-order methods.*, in ICML, 2015, pp. 1501–1510.

[115] J. Zowe and S. Kurcyusz, *Regularity and stability for the mathematical programming problem in Banach spaces*, Applied mathematics and Optimization, 5 (1979), pp. 49–62.