中国科学院大学
University of Chinese Academy of Sciences

博士学位论文

复合矩阵优化问题的扰动分析及其算法应用

作者姓名：王石玮

指导教师：丁超 副研究员

中国科学院数学与系统科学研究院

学位类别：理学博士

学科专业：运筹学与控制论

培养单位：中国科学院数学与系统科学研究院

2023 年 6 月

Perturbation analysis of composite matrix optimization and its applications

A dissertation submitted to

University of Chinese Academy of Sciences

in partial fulfillment of the requirement

for the degree of

Doctor of Philosophy

in Operational Research and Cybernetics

By

WANG Shiwei

Supervisor: Professor Ding Chao

Academy of Mathematics and Systems Science

Chinese Academy of Sciences

June, 2023

Abstract

This thesis mainly focuses on the perturbation properties and applications of a class of nonpolyhedral optimization problem named composite matrix optimization (CMatOP), which includes several important optimization problems arising from practical application, e.g., nonlinear semidefinite programming (NLSDP), largest eigenvalue minimization and so on. We firstly provide the explicit characterizations of the corresponding perturbation properties, e.g., Lipschitzian full stability, isolated calmness and semi-isolated calmness. During this process, the explicit forms of the corresponding coderivates are given. On the other hand, augmented Lagrangian method (ALM) is a well known method for its elegant theory and impressive numerical performance. Thus in the second part of the thesis, we establish new local fast linear convergence results of ALM for CMatOP, in particular for NLSDP, without requiring the uniqueness of multiplier. Our first attempt is to study this topic under semi-isolated calmness. By assuming the second order sufficient condition (SOSC) and the semi-isolated calmness of the KKT solution together with some condition which is hard to be verified, the local (asymptotic Q-superlinear) Q-linear convergence rate of the primal-dual sequences generated by ALM for NLSDP is established. To avoid this difficult condition, our second attempt is to consider the convergence analysis under strong variational sufficient condition. We prove the equivalence between the strong SOSC and the strong variational sufficiency, which is newly proposed by Rockafellar. This property turns out to be of great use in the convergence analysis of multiplier methods. Based on this equivalence, the local asymptotic superlinear convergence rate of ALM for NLSDP can be established under strong SOSC in the absence of constraint qualifications. Moreover, we illustrate that for NLSDP, the strong SOSC seems to be not only sufficient to the local fast linear convergence rate of ALM, but also necessary for the invertibility of generalized Hessian of augmented Lagrangian function for NLSDP, which is crucial for the semi-smooth Newton CG method for solving the ALM subproblem.

Key Words: Composite matrix optimization problem, Lipschitzian full stability, Augmented Lagrangian method, Nonlinear semidefinite programming, Strong variational sufficiency

Contents

List of Figures

List of Tables

Notations

Some frequently occurred notations are listed below.

• We use the blackboard bold letters (e.g., $\mathbb{R}^n, \mathbb{X}$) to denote Euclidean space and capital Latin letters (e.g., A, B) to denote matrices. In particular, E denote the matrix whose all the components are 1. We use decorated capital Latin letters (e.g., $\mathcal{A}, \mathcal{B}$) to denote sets. Specially, let $\mathcal{O}^n$ be the set of all $n \times n$ orthogonal matrices and $\mathcal{P}^n$ be the set of all permutation matrices in $\mathbb{R}^{n\times n}$. Use small Latin letters (e.g., x, b^i) to denote vectors or scalars. Index sets are denoted by Greek letters (e.g. α, ι).

• Given set $\mathcal{S}$, we use $\mathrm{cl}\,\mathcal{S}$ to denote its closure. We use $\mathcal{B}$ to denote the unite closed ball unless specifically mentioned.

• For any $Z \in \mathbb{R}^{m\times n}$ and a given index set $\alpha \subseteq \{1, \ldots, n\}$, we use Z_α to denote the sub-matrix of Z obtained by removing all the columns of Z not in α. In particular, we use Z_j to represent the j-th column of Z, $j = 1, \ldots, n$. Let $\alpha \subseteq \{1, \ldots, m\}$ and $\beta \subseteq \{1, \ldots, n\}$ be two index sets. For any $Z \in \mathbb{R}^{m\times n}$, we use $Z_{\alpha\beta}$ to denote the $|\alpha| \times |\beta|$ sub-matrix of Z obtained by removing all the rows of Z not in α and all the columns of Z not in β.

• We use " $\circ$ " to denote the Hardamard product between matrices, i.e., for any two matrices A and B in $\mathbb{R}^{m\times n}$ the (i, j)-th entry of $Z := A \circ B \in \mathbb{R}^{m\times n}$ is $Z_{ij} = A_{ij}B_{ij}$.

• For $X \in \mathbb{R}^{n\times n}$, $\mathrm{diag}(X)$ denotes the column vector consisting of all the diagonal entries of X being arranged from the first to the last. For any $x \in \mathbb{R}^n$, $\mathrm{Diag}(x)$ denotes the diagonal matrix whose i-th diagonal entry is x_i, $i = 1, \ldots, n$.

Chapter 1 Introduction

1.1 Composite matrix optimization problem

Composite matrix optimization problem (CMatOP) is a classic kind of non-polyhedral optimization problems concerned with matrix variables. It arises from various practical applications in a bunch of fields, including fastest mixing Markov chain problem, the low rank approximations of doubly stochastic matrices and unsupervised learning.

Denote $\mathbb{R}^m$ as the real m-dimensional Euclidean space and $\mathbb{S}^m$ as the $m \times m$ symmetric matrices. The general symmetric composite matrix optimization problem (CMatOP) takes the following form.

$$\begin{array}{ll} \min\limits_{x\in\mathbb{X}} & f(x) + \varpi \circ \lambda(G(x)) \\ \text{s.t.} & h(x) = 0, \end{array} \tag{1-1}$$

where $f : \mathbb{X} \to (-\infty, +\infty]$, $G : \mathbb{X} \to \mathbb{S}^m$, $h : \mathbb{X} \to \mathbb{Y}$ are twice continuously differentiable, $\lambda : \mathbb{S}^m \to \mathbb{R}^m$ is the eigenvalue function of a symmetric matrix with the components being arranged in the non-increasing order. In particular, we can also consider the case where $G : \mathbb{X} \to \mathbb{R}^{m\times n}$ is non-symmetric in the exactly same approach, but we focus on symmetric case here for notation simplicity. Suppose $\varpi : \mathbb{R}^m \to (-\infty, +\infty]$ is a symmetric piecewise affine function with a polyhedron domain, i.e.,

$$\varpi(x) = \max_{1\le i\le p}\{\langle a^i, x\rangle - c_i\}, \tag{1-2}$$

for some $\{(a^i, c_i) \in \mathbb{R}^m \times \mathbb{R}\}_{i=1}^p$ with a positive integer p and

$$\operatorname{dom}\varpi = \widehat{\mathcal{K}} := \{t \in \mathbb{R}^m \mid \psi(t) := \max_{1\le i\le q}\{\langle b^i, t\rangle - d_i\} \le 0\} \tag{1-3}$$

for some $\{(b^i, d_i) \in \mathbb{R}^m \times \mathbb{R}\}_{i=1}^q$ with a positive integer q. It is known from [1, Theorem 2.49] that ϖ can be expressed in the form of

$$\varpi(x) = \underbrace{\max_{1\le i\le p}\left\{\langle a^i, x\rangle - c_i\right\}}_{\text{denoted as } \varpi_1(x)} + \underbrace{\delta_{\operatorname{dom}\varpi}(x)}_{\text{denoted as } \varpi_2(x)}, \quad x \in \mathbb{R}^m, \tag{1-4}$$

where $\delta_{\operatorname{dom}\varpi}(\cdot)$ is the indicator function of $\operatorname{dom}\varpi$. The symmetry here means that for all permutation matrix $U \in \mathbb{R}^{m\times m}$, we have $\varpi(x) = \varpi(Ux)$. It can be checked directly that (1-1) can be transformed into the following form.

$$\begin{array}{ll} \min\limits_{x\in\mathbb{X}} & f(x,\bar{p}) + \theta_1(G(x,\bar{p})) \\ \text{s.t.} & h(x,\bar{p}) = 0, \\ & G(x,\bar{p}) \in \mathcal{K}, \end{array} \tag{1-5}$$

where $\theta_1(X) = \varpi_1 \circ \lambda(X)$, $\mathcal{K} = \{X \in \mathbb{S}^m \mid \lambda(X) \in \widehat{\mathcal{K}}\}$ and $\bar{p} \in \mathbb{P}$ is the normal basic parameter. In order to make the discussion more general, we allow $m \neq n$ with two different continuously differentiable functions $g_1 : \mathbb{X} \times \mathbb{P} \to \mathbb{S}^m$ and $g_2 : \mathbb{X} \times \mathbb{P} \to \mathbb{S}^n$ in this problem, i.e.,

$$\begin{array}{ll} \min\limits_{x \in \mathbb{X}} & f(x, \bar{p}) + \theta_1(g_1(x, \bar{p})) \\ \text{s.t.} & h(x, \bar{p}) = 0, \\ & g_2(x, \bar{p}) \in \mathcal{K}. \end{array} \tag{1-6}$$

This model is a generalized form of matrix norm minimization and matrix conic programming, which have now developed into a fruitful subject in optimization, e.g., eigenvalue minimization, low rank matrix completion (see [2], [3]), robust PCA, semidefinite programming and so on. Some concrete examples are listed below.

Example 1.1. Eigenvalue minimization: We consider the following problem

$$\begin{array}{ll} \min\limits_{x \in \mathbb{X}} & f(x, \bar{p}) + \lambda_1\big(G(x, \bar{p})\big), \\ \text{s.t.} & h(x, \bar{p}) = 0, \end{array} \tag{1-7}$$

where $\lambda_1\big(G(x, \bar{p})\big)$ denotes the largest eigenvalue of a symmetric matrix $G(x, \bar{p})$. This is a special case of CMatOP (1-6) where

$$\varpi(x) = \max_{1 \le i \le m} \{\langle e^i, x\rangle\}, \quad x \in \mathbb{R}^m$$

with e^i being the unit vector whose i-th component is 1 and others are zeros.

Example 1.2. Nonlinear positive semidefinite programming: We consider the following problem:

$$\begin{array}{ll} \min\limits_{x \in \mathbb{X}} & f(x, \bar{p}) \\ \text{s.t.} & h(x, \bar{p}) = 0, \\ & G(x, \bar{p}) \in \mathbb{S}^n_+. \end{array} \tag{1-8}$$

where $\mathbb{S}^n$ is the linear space of all $n \times n$ real symmetric matrices equipped with the usual Frobenius inner product and its induced norm, $\mathbb{S}^n_+$ ($\mathbb{S}^n_-$) is the closed convex cone of all $n \times n$ positive (negative) semidefinite matrices in $\mathbb{S}^n$. This is a special case of CMatOP (1-6) with

$$\mathcal{K} = \mathbb{S}^n_+ := \{X \in \mathbb{S}^n \mid \psi \circ \lambda(X) := \max_{1 \le i \le n} \{\langle -e^i, \lambda(X)\rangle\} \le 0\}$$

and e^i is the unit vector whose i-th component is 1 and others are zeros.

Although CMatOP includes many well known and practical example, few people focus on its general model (1-6) for its complicated non-polyhedral structure. Recently, [4] explores some important properties on its variational analysis and perturbation theory. In the following chapters, we will also go along this way to establish more on the perturbation analysis of CMatOP.

1.2 Sensitive analysis of matrix optimization problem

The sensitivity analysis of solutions to matrix optimization problem has aroused great attention in the last half century. A bunch of works have focused on this topic, e.g., [1, 5–7]. Consider general optimization problem

$$\begin{aligned} \min_{x\in\mathbb{X}} \quad & f(x) \\ \text{s.t.} \quad & G(x) \in \mathcal{S}, \end{aligned} \tag{1-9}$$

where $f : \mathbb{X} \to \mathbb{R}, G : \mathbb{X} \to \mathbb{Y}$ are twice continuously differentiable and $\mathcal{S}$ is a closed convex set. For polyhedron $\mathcal{S}$, the perturbation analysis results are quite complete. But when it comes to nonpolyhedral ones, much less has been discovered. In [5, 8, 9], systematical perturbation theory has been established for C^2-cone reducible $\mathcal{S}$. The theory of second order optimality conditions has also been established in [5, 10] as they are closely related with perturbation properties. As for NLSDP, [11, 12] provides explicit characterizations of strong regularity, which is one of the important concepts in sensitivity and perturbation analysis introduced by Robinson [13].

One fundamental perturbation property in the sensitive analysis is the Lipschitzian full stability of local minimizers. This concept is put out by Levy, Poliquin and Rockafellar in 2000 (see [14]) to distinguish the local solutions with some Lipschitzian property and local uniqueness under parameter perturbations. This property reveals the nature of a problem, implies the difficulty of solving the corresponding problem and supports computational work. Up to now, a few work has been down on this topic for different problems. [14] characterized Lipschitzian full stability by second order subdifferential for a class of unconstrained parametrically prox-regular problem. In particular, for polyhedral case, [15] also studied fully stable local minimizers for general classes of constrained optimization problems, including problems of composite optimization, mathematical programs with polyhedral constraints, as well as extended and classical nonlinear programming with twice continuously differentiable data.

For non-polyhedral case, Mordukhovich et. al.(see [16]), developed a simpler approach to explore the characterization of Lipschitizan and Hölderian full stability for general problems recently. This is the first time that full stability for non-polyhedral problems have ever been considered. They also showed Lipschitzian full stability is equivalent to strong SOSC under constraint nondegeneracy condition for NLSDP by applying the explicit form of the corresponding coderivative. This gives us a hint of developing a more checkable characterization of the Lipschitzian full stability for (1-1) by characterizing the corresponding coderivative with the tool of second order variational analysis. As listed in [16], the characterization of the corresponding coderivatives for NLP, mathematical programs with polyhedral constraints and second order cone programming have been established (see [17], [15], [18]). However, few works have been done for (1-1) case, though it contains all the problems mentioned above.

1.3 Augmented Lagrangian method

In this section, we will provide a brief introduction of Augmented Lagrangian method. Consider the following general composite optimization problem

$$\min_{x\in\mathbb{X}} \quad f(x)+\theta(G(x)), \tag{1-10}$$

where $\mathbb{X}$ and $\mathbb{Y}$ are two given Euclidean spaces, $f:\mathbb{X}\to\mathbb{R}$ and $G:\mathbb{X}\to\mathbb{Y}$ are twice continuously differentiable, and $\theta:\mathbb{Y}\to(-\infty,\infty]$ is a closed proper convex function. Its Lagrangian function is defined by

$$L(x,Y):=f(x)+\langle Y,G(x)\rangle, \quad (x,Y)\in\mathbb{X}\times\mathbb{Y}. \tag{1-11}$$

For any $Y\in\mathbb{Y}$, denote the first-order and second-order derivatives of $L(\cdot,Y)$ at $x\in\mathbb{X}$ by $L'_x(x,Y)$ and $L''_{xx}(x,Y)$, respectively. Given $\rho>0$, the augmented Lagrangian function of (1-10) takes the following form (cf. [1, Section 11.K] and [19])

$$\mathcal{L}_\rho(x,Y):=f(x)+e_{1/\rho}\theta(G(x)+\frac{Y}{\rho})-\frac{\|Y\|^2}{2\rho}, \tag{1-12}$$

where $e_\rho\theta(y):=\inf\limits_{x\in\mathbb{X}}\left\{\theta(x)+\frac{1}{2\rho}\|x-y\|^2\right\}$ is the Moreau-Yosida regularization, which will be introduced in detail in Section 2.1.1. For a given initial point $(x^0,Y^0)\in\mathbb{X}\times\mathbb{Y}$ and a constant $\rho^0>0$, the $(k+1)$-th iteration of (extended) augmented Lagrangian method (ALM) for (1-10) proposed by [20] takes the following form

$$\begin{cases} x^{k+1}\approx\arg\min\{\mathcal{L}_{\rho^k}(x,Y^k)\}, \\ Y^{k+1}=Y^k+\widetilde{\rho}^k(\mathcal{L}_{\rho^k})'_Y(x^{k+1},Y^k), \end{cases} \tag{1-13}$$

where ρ^k, $\widetilde{\rho}^k>0$. The (extended) ALM (1-13) reduces to the traditional ALM when $\widetilde{\rho}^k=\rho^k$.

The augmented Lagrangian method was firstly proposed by Hestenes [21] and Powell [22] for solving the equality constrained problem and was generalized by Rockafellar [23] to NLP. It rapidly grew into popularity for its mathematical elegance and impressive numerical performance in various areas, like statistical optimization (e.g., Lasso [24]), machine learning and game theory. It has also been implemented in many powerful large scale solvers like SDPNAL+ [25, 26], QSDPNAL [27], SuitedLasso [28] and so on.

There are also many works focused on the theory of this algorithm. For convergence analysis of ALM, tremendous work has been established since it was proposed. Powell [22] demonstrated that for the equality constrained problem, if the second-order sufficient condition (SOSC) and linear independence constraint qualification (LICQ) were satisfied, the algorithm should converge locally at a linear rate, without the need for having $\rho\to\infty$. This implies that ALM may provide numerical stability, which the usual penalty methods do not possess. In 1973, Rockafellar [29] and Tretykov [30] proved the global convergence of the augmented Lagrangian method for convex optimization problem with inequality constraints for any $\rho > 0$ based on the saddle point theorem established in [23].

For the convex nonlinear programming (NLP) problem, the local convergence rate of the ALM can be derived through its deep connection with the dual proximal point algorithm (PPA) as studied by Rockafellar in [31]. As stated in [31, Proposition 3, Theorem 2], one can obtain the Q-linear convergence rate of the dual sequence generated by the ALM under the upper Lipschitz continuity of the dual solution mapping at the origin, the boundedness of dual sequence and certain stopping criteria on the inexact computations of the augmented Lagrangian subproblems. For more details about PPA and monotone operators, please see [31–33].

Following this way, the convergence rate of ALM for general convex optimization problems can also be attained under very mild conditions with implementable stopping criteria for the ALM subproblems. In 1984, Luque relaxed the upper Lipschitz continuity of the dual solution mapping used in [31], which required the uniqueness of the optimal solution, by an error bound type condition [34, (2.1)] that is known to be satisfied for polyhedron [35] but difficult to be verified for non-polyhedron. In 2019, Cui et al. [36] established the asymptotic R-superlinear convergence of the KKT residuals and asymptotic Q-superlinear convergence of the dual sequence generated by the ALM for solving convex NLSDP, under a quadratic growth condition on the dual problem that neither local solution nor the multiplier is required to be unique. Their remarkable work improved [31] in giving a practical stopping criterion for ALM subproblem under the Robinson constraint qualification (RCQ) (for the improvement of implementable stopping criteria, see also [37]) and obtaining the convergence of the KKT residuals with the application of KKT residual information. Also, they relaxed Luque's condition by the calmness of the dual solution mapping at the origin.

When it comes to non-convex optimization problems, fruitful results have been established for the polyhedral case. In 1982, Bertsekas [38] established that the generated dual sequence converges Q-linearly and the corresponding primal sequence converges R-linearly under SOSC, LICQ and the strict complementarity for NLP. His result shows that the ratio constant is proportional to $1/\rho$, which implies the convergence can be accelerated by increasing ρ. Efforts are made to weaken the above conditions. Firstly, successful attempts are made to remove the strict complementarity condition, e.g., Conn et al. [39], Contesse-Becker [40], and Ito and Kunisch [41] derived linear convergence rate for the ALM of general NLP. Secondly, it is also crucial to weaken LICQ, which implies the uniqueness of multipliers. As in real-world, multipliers are usually non-unique, e.g., considering Lasso as a dual problem. In 2012, Fernandez and Solodov

[42] firstly studied this topic for NLP without requiring the multiplier to be unique. This work is a milestone to establishing the convergence by removing the uniqueness of the Lagrangian multiplier and strict complementarity. Recently Hang and Sarabi [43] established the local convergence for piecewise linear quadratic composite optimization problems under merely SOSC. Their success relies on the validity of upper Lipschitz continuous of KKT solution mapping/Hoffman error bound when SOSC is satisfied, see [43–47]. However, this does not hold for non-polyhedral case as mentioned in [36] by using [5, Example 4.54]. For comprehensive surveys about the augmented Lagrangian method for nonlinear programming, see [38, 48, 49].

For non-convex non-polyhedral problem, Sun et al. [50] proved the convergence rate of NLSDP under strongly SOSC [11] together with nondegeneracy (cf. [50] or [5]). In 2019, Kanzow and Steck [51] justified the primal-dual linear convergence of ALM under SOSC and strict Robinson constraint qualification (SRCQ) for C^2-cone reducible constrained problems, which include NLSDP and nonlinear second-order cone programming (NLSOC). Recently, the primal-dual linear convergence rate of ALM for NLSOC is also studied under SOSC and the semi-isolated calmness of the KKT solution mapping (see Definition 3.32) in [52] with the multiple uniqueness assumption.

Going through all the papers above, it is not hard to see that existing works usually suppose either the problem is convex (or polyhedral), or the Lagrangian multiplier is (locally) unique. However, few results on the local convergence rate of ALM have been established for non-convex non-polyhedral problems without the multiple uniqueness. It is worth noting that in [38], the author revealed a kind of local duality based on sufficient conditions for local optimality, which turns out to be the key to understanding the convergence of ALM for nonconvex nonlinear programming. It follows that there may be a local reduction from nonconvex optimization to convex optimization. This gives a hint on extending this local duality approach to broader nonconvex problems. Recently, the newly published work [20] opens new doors for the convergence rate analysis of ALM for general nonconvex composite problems through its connection with PPA for the dual problem and the strong variational sufficient condition without any constraint qualifications. He successfully obtained the ALM primal R-linear convergence from the ALM dual Q-linear convergence for generalized NLP by assuming strongly variational sufficiency. As for [20], what strong variational sufficiency really means for general problems remains a great challenging. Thus results on the local convergence rate of ALM for non-convex non-polyhedral problems without the multiple uniqueness is demanding. Recently, [53] proposed a procedure called progressive decoupling algorithm, which reforms PPA from a more general angle to solve the linkage problem. The practical application of this algorithm is illustrated in [53, Section 3] and its equivalence with ALM can be seen in [54, Section 4].

1.4 Variational sufficiency

In this section, we will mainly introduce some important properties proposed by Rockafellar in his recent works [20, 55]. The definition of (strong) variational sufficient condition is officiall given in [55] for general composite optimization problem (1-10). We can recast (1-10) in the form

$$\min\ \phi(x,u) \text{ subject to } u=0, \quad \text{where } \phi(x,u)=f(x)+\theta(G(x)+u). \tag{1-14}$$

The first order local optimality condition for (1-10) of $\bar{x}$ is the existence of $\overline{Y}$ such that

$$L'_x(\bar{x},\overline{Y})=0 \quad \text{with} \quad \overline{Y}\in\partial\,\theta(G(\bar{x})).$$

Define

$$\phi_r := \phi(x,u)+\frac{r}{2}|u|^2. \tag{1-15}$$

The variational (strong) convexity, which is firstly proposed in [54], refers to the existence of open convex neighborhoods $\mathcal{W}$ of $(\bar{x},0)$ and $\mathcal{Z}$ of $(0,\overline{Y})$ such that there exists a proper closed (strongly) convex function $\psi\le\phi_r$ on $\mathcal{W}$ such that

$$(\mathcal{W}\times\mathcal{Z})\cap\operatorname{gph}\partial\psi=(\mathcal{W}\times\mathcal{Z})\cap\operatorname{gph}\partial\phi_r$$

and for $(x,u;v,y)$ belonging to this common set, $\psi(x,u)=\phi_r(x,u)$. As shown in [56], (strongly) variational convexity can be fully characterized via the corresponding Moreau-Yosida regularization and has deep connection with tilt stability under certain constraint qualification.

Definition 1.1. *[55] The (strong) variational sufficient condition for local optimality in* (1-14) *holds with respect to $\bar{x}$ and $\overline{Y}$ satisfying the first order condition if there exists $r>0$ such that $\phi_r(x,u)$ is variationally (strongly) convex with respect to the pair $\big((\bar{x},0),(0,\overline{Y})\big)$ in* $\operatorname{gph}\partial\phi_r$.

The strong variational sufficient condition is firstly introduced to deal with the local convexity of primal augmented Lagrangian function and the augmented tilt stability. It has gained more and more attention for its mathematical elegance and wide applications in the convergence analysis of multiplier methods. Several characterizations of this abstract property are also given in [55]. For instance, it is equivalent to the positive definiteness of the Hessian bundle of the augmented Lagrangian function [55, Theorem 3] or the criterion [55, Theorem 5] involving quadratic bundle (Definition 5.3).

In terms of the connection with traditional optimality conditions, Rockafellar [55, Theorem 4] shows that the strong variational sufficiency is equivalent to the well-known strong second order sufficient condition (SOSC), which is expressed entirely via the program data, when the function θ in (1-10) is polyhedral convex (i.e., the epigraph

$\operatorname{epi}\theta$ of θ is a polyhedral convex set). For non-polyhedral problems, the equivalence is still valid if θ in (1-10) is the indicator function of the second order cone with $G(\bar{x}) \neq 0$, where $\bar{x}$ is a local optimal solution (see [55, Example 3] for details). However, an explicit and verifiable characterization of strong variational sufficient condition for general non-polyhedral problems remains unknown. In this thesis, without loss of generality, we mainly focus on the characterization of strong variational sufficiency for the following nonlinear semidefinite programming (NLSDP):

$$\begin{aligned} \min_{x\in\mathbb{X}} \quad & f(x) \\ \text{s.t.} \quad & G(x) \in \mathbb{S}^n_+. \end{aligned} \tag{1-16}$$

One of our contributions lies in uncovering the equivalence between strong variational sufficient condition and strong SOSC (see Definition 3.20 for details) for NLSDP problems without requiring any other constraint qualifications.

1.5 Innovations and difficulties

The main contributions of this thesis are listed below.

- We provide explicit characterization of three important perturbation properties of CMatOP: Lipschitzian full stability, isolated calmness and semi-isolated calmness. During this process, the explicit form of corresponding coderivatives are given. This result includes corresponding NLSDP result as a particular case and can be applied to prove the equivalence between strong regularity and strong SOSC with nondegeneracy. This result is of great theoretical importance as existing works can only obtain strong regularity from strong SOSC and nondegeneracy [16]. This part provide a uniform approach for problems like NLSDP and eigenvalue minimization.
- For NLSDP, we obtain the characterization of strong variational sufficiency and prove its equivalence with strong SOSC.
- We prove new local convergence results of ALM for NLSDP without requiring the uniqueness of multiplier. We also illustrate that strong SOSC seems to be not only sufficient to the local fast linear convergence rate of ALM, but also necessary for the invertibility of generalized Hessian of augmented Lagrangian function for NLSDP.

The main difficulties are discussed below.

- The problem we mainly discussed in this thesis is CMatOP, which includes several nonpolyhedral problems like NLSDP, eigenvalue minimization and so on. Compared to polyhedral problems, nonpolyhedral ones are more complicated as they would lose many good properties like Hoffman error bound. Because of the complicated structure, few work has been done on CMatOP.
- Although strongly variational sufficiency and the coderivative of set-value mapping are of great use in the characterization of perturbation properties and analysis

of multiplier method, it is very hard to characterize them. Most existing works only consider the polyhedral case. While the explicit form of coderivative for SDP has been established, the formula for more general case is still unknown.

- The convergence analysis of ALM for nonconvex nonpolyhedral problems is a long-term concerning problem. Because of the absence of Hoffman error bound, we can not apply the approach for NLP here. Most existing works require the uniqueness of multiplier, which is usually not true in practice, e.g. Lasso. Thus, there still exist some gaps between the theory of ALM and its practical use.

1.6 Outline of this thesis

The thesis is organized as follows: to facilitate later discussions, we give some preliminaries on the eigenvalue decomposition properties and the variational properties of CMatOP, especially for NLSDP in Chapter 2. In Chapter 3, we study the characterization of Lipschitz full stability, isolated calmness and semi-isolated calmness for CMatOP. During this process, the characterization of its corresponding coderivative is also given. Chapter 4 focuses on the convergence analysis of ALM for NLSDP under semi-isolated calmness, which relaxed the commonly used uniqueness of multiplier condition. However, the assumption that is difficult to be verified is still needed. To overcome the complicated assumption used in Chapter 4, Chapter 5 is devoted to the similar topic under strong variational sufficiency. Results in this chapter is more of practical use. Moreover, how to solve the ALM subproblem is discussed and some numerical experiments are conducted to verify our theory. We present conclusions and some possible topic for future research in Chapter 6.

Chapter 2 Preliminaries

2.1 Basic variational properties

In the section, we list some basic definitions and properties in variational analysis. Details on these concepts can be found from [1, 57, 58] and others. Let $\mathbb{X}$ and $\mathbb{P}$ be two Euclidean spaces. The following definitions of the proximal, regular and limiting normal cones of sets are taken from [1, Page 213], [57, Definition 1.1 (ii)] and [58, Page 62 and Theorem 6.1(b)], respectively.

Definition 2.1. *Let $\mathcal{S}$ be a nonempty subset of $\mathbb{X}$ and $\bar{x} \in \operatorname{cl} \mathcal{S}$ be given. We call*

$$\mathcal{N}_{\mathcal{S}}^{\pi}(\bar{x}) := \{d \in \mathbb{X} \mid \exists\, c > 0 \text{ such that } \langle d, x - \bar{x}\rangle \le c\|x - \bar{x}\|^2 \quad \forall\, x \in \mathcal{S}\} \tag{2-1}$$

the proximal normal cone to set $\mathcal{S}$ at point $\bar{x}$, The regular/Fréchet normal cone of $\mathcal{S}$ at x is defined by

$$\widehat{N}_{\mathcal{S}}(\bar{x}) := \left\{ d \in \mathbb{X} \mid \limsup_{x \xrightarrow{\mathcal{S}} \bar{x}} \frac{\langle d, x - \bar{x}\rangle}{\|x - \bar{x}\|} \le 0 \right\},$$

where $x \xrightarrow{\mathcal{S}} \bar{x}$ means $x \to \bar{x}$ and $x \in \mathcal{S}$. The limiting normal cone (also known as Mordukhovich normal cone or basic normal cone) to set $\mathcal{S}$ at point $\bar{x}$ is defined as

$$\mathcal{N}_{\mathcal{S}}(\bar{x}) := \{\lim_{i\to\infty} d_k \mid d_k \in \mathcal{N}_{\mathcal{S}}^{\pi}(x_k), \;\; x_k \to \bar{x}, \;\; x_k \in \mathcal{S}\} \tag{2-2}$$

It is well-known that in an Euclidean space the limiting normal cone can also be defined by the regular normal cone (see [1, Page 234] or [57, Section 2.5.2(D)] for details). When $\mathcal{S}$ is convex, the proximal normal cone and limiting normal cone coincide with the normal cone in the sense of convex analysis [59], i.e., $\mathcal{N}_{\mathcal{S}}(\bar{x}) := \{d \in \mathbb{X} \mid \langle d, x - \bar{x}\rangle \le 0 \;\forall\, x \in \mathcal{S}\}$. For a set-valued mapping $\mathcal{F} : \mathbb{X} \rightrightarrows \mathbb{X}$, its $\limsup$ means

$$\limsup_{x\to\bar{x}} \mathcal{F}(x) := \left\{ d \in \mathbb{X} \mid \exists\, x_k \to \bar{x}, \; y_k \to d \text{ such that } y_k \in \mathcal{F}(x_k), \; \forall\, k \right\}.$$

Definition 2.2. *[1, Definition 6.1, Proposition 6.2] The tangent cone of a set $\mathcal{S}$ to $\bar{x}$ is a closed cone expressed as*

$$T_{\mathcal{S}}(\bar{x}) = \limsup_{t\to 0} \frac{\mathcal{S} - \bar{x}}{t}.$$

The following definitions of proximal and (general) subdifferentials of functions are adopted from [1, Definition 8.45 and 8.3].

Definition 2.3. *Consider a function $f : \mathbb{X} \to (-\infty, +\infty]$ and a point $\bar{x}$ with $f(\bar{x})$ finite. The proximal subdifferential of f at $\bar{x}$ is defined as*

$$\partial^{\pi} f(\bar{x}) := \{d \mid \exists\, l > 0, c > 0 \text{ such that } f(x) \geq f(\bar{x}) + \langle d, x - \bar{x}\rangle - l\|x - \bar{x}\|^2,$$

$$\forall\, x \in \mathcal{B}_c(\bar{x})\},$$

where $\mathcal{B}_c(\bar{x})$ is the ball centered at $\bar{x}$ with radius c. The (general) subdifferential of f at $\bar{x}$ is $\partial f(\bar{x}) := \limsup_{x \xrightarrow{f} \bar{x}} \partial^{\pi} f(x)$, where $x \xrightarrow{f} \bar{x}$ signifies $x \to \bar{x}$ with $f(x) \to f(\bar{x})$.

Note that in an Euclidean space, the (general) subdifferential also can be constructed via the regular subdifferentials of functions $\widehat{\partial} f$ [1, Definition 8.3]. It is well-known that these two definitions are equivalent (see e.g., [1, page 345] or [57, Theorem 1.89] for details).

Let $\mathcal{U}$ be an open set in $\mathbb{X}$ and $f : \mathcal{U} \subseteq \mathbb{X} \to \mathbb{Z}$ be a locally Lipschitz continuous function on the open set $\mathcal{U}$. It follows from the Rademacher’ s theorem [1, Section 9.J] that f is Fréchet differentiable almost everywhere in $\mathcal{U}$. Let $\mathcal{U}_f$ be the set of points in $\mathcal{U}$ that f is F-differentiable. Define the B-subdifferential of f at $\bar{x}$ as

$$\partial_B f(\bar{x}) := \{V \mid V = \lim_{k\to\infty} f'(x^k), x^k \to \bar{x}, x^k \in \mathcal{U}_f\}.$$

Clarke’s generalized Jacobian [60] of f at $\bar{x}$ is

$$\partial f(\bar{x}) = \text{conv}\{\partial_B f(\bar{x})\},$$

where “conv" is the convex hull of a set. The following important chain rule of the generalized Jacobian for composite functions is firstly given in [11, Lemma 2.1].

Lemma 2.4. *Let $G : \mathbb{X} \to \mathbb{Y}$ be a continuously differentiable function on an open neighborhood $\mathcal{X}$ of $\bar{x}$ and $\theta : \mathcal{U} \subseteq \mathbb{Y} \to \mathbb{Z}$ be a locally Lipschitz continuous function on an open set $\mathcal{U}$ containing $\bar{y} = G(\bar{x})$. Suppose that θ is directionally differentiable at every point in $\mathcal{U}$ and that $G'(\bar{x}) : \mathbb{X} \to \mathbb{Y}$ is onto. Then it holds that*

$$\partial_B \Phi(\bar{x}) = \partial_B \theta(\bar{y}) G'(\bar{x}),$$

where $\Phi : \mathcal{X} \to \mathbb{Z}$ is defined by $\Phi(x) := \theta(G(x))$, $x \in \mathcal{X}$.

In particular, when f is convex and locally Lipschitz, the proximal, (general) subdifferentials and Clarke’s generalized Jacobian coincide with the subdifferential in the sense of convex analysis [59]. When $\mathcal{S}$ is a convex set, it is easy to verify that $\mathcal{N}_{\mathcal{S}}(\cdot) = \partial \delta_{\mathcal{S}}(\cdot)$. Also, we can apply the regular subdifferential mentioned above to define the horizon subdifferential of f at $\bar{x}$, i.e., $\partial^{\infty} f(\bar{x}) := \limsup_{x \xrightarrow{f} \bar{x}, t\downarrow 0} t\, \widehat{\partial} f(x)$.

For a set-valued mapping $\mathcal{F} : \mathbb{X} \rightrightarrows \mathbb{Y}$, denote $\text{gph}\, \mathcal{F} := \{(x, y) \in \mathbb{X} \times \mathbb{Y} \mid y \in \mathcal{F}(x)\}$. The corresponding coderivative of $\mathcal{F}$ can be defined via normal cone as follows.

Definition 2.5. *(see [1, Definition 8.33]) The limiting coderivative of $\mathcal{F}$ at $(\bar{x}, \bar{y}) \in \operatorname{gph}\mathcal{F}$ is*

$$\mathcal{D}^*\mathcal{F}(\bar{x}, \bar{y})(w) := \{z \in \mathbb{X} \mid (z, -w) \in \mathcal{N}_{\operatorname{gph}\mathcal{F}}(\bar{x}, \bar{y})\} \quad \text{for all } w \in \mathbb{Y}.$$

2.1.1 Moreau-Yosida regularization

Assume $f : \mathbb{X} \to (-\infty, \infty]$ is a proper closed convex function. The Moreau-Yosida regularization and proximal mapping of f are defined by

$$ef_\rho(y) := \min_{x\in\mathbb{X}} \left\{ f(x) + \frac{1}{2\rho}\|x - y\|^2 \right\} \quad \text{and} \quad \operatorname{Pr}_{f,\rho}(y) := \arg\min_{x\in\mathbb{X}} \left\{ f(x) + \frac{1}{2\rho}\|x - y\|^2 \right\}.$$

The conjugate of f is $f^*(x^*) = \sup\limits_{x\in\mathbb{X}}\{\langle x, x^*\rangle - f(x)\}$. If $\rho = 1$, we denote $\operatorname{Pr}_{f,1}(y)$ as $\operatorname{Pr}_f(y)$ for simplicity. Specifically, if f is an indicator function of some convex set $\mathcal{S} \subseteq \mathbb{X}$, i.e., $f = \delta_{\mathcal{S}}$, we call the proximal mapping of f at y with $\rho = 1/2$ as the projection of y onto $\mathcal{S}$ and denote it as $\Pi_{\mathcal{S}}(y)$. It is known that $\operatorname{Pr}_{f,\rho}$ is globally Lipschitz continuous with modulus 1 [1, Proposition 12.19]. The directional derivative of the proximal mapping is closely related to the critical cone associated with the generalized equation $\bar{z} \in \partial f(z)$ [8, Section 7.2, 7.3], whose definition is stated below.

Definition 2.6. *For a convex function $f : \mathbb{X} \to (-\infty, \infty]$, its critical cone at z for a subgradient $\bar{z} \in \partial f(z)$ is defined as*

$$\mathcal{C}_f(z, \bar{z}) := \left\{ d \in \mathbb{X} \mid \mathrm{d}f(z)(d) = \langle \bar{z}, d\rangle \right\}. \tag{2-3}$$

where $\mathrm{d}f(z)(d)$ is the subderivative defined in [1, 7(21)]. It is easy to check that if f is Lipschitz continuous around $\bar{x}$, then its semidifferentiability [1, Definition 7.20] at this point is equivalent to its directional differentiability [1, 7(20)] at $\bar{x}$. Thus when f is equal to ϖ_1 (1-4) or θ_1 (1-5), we can replace the subderivative in the above definition by directional derivative, i.e., $\mathcal{C}_f(z, \bar{z}) := \{d \in \mathbb{X} \mid f'(z; d) = \langle \bar{z}, d\rangle\}$. When f is the indicator function of some convex set $\mathcal{S}$, the above definition is reduced to the critical cone of a set defined in [61, Page 98]. The critical cone of a convex set $\mathcal{S}$ at $z \in \mathcal{S}$ for $\bar{z} \in \mathcal{N}_{\mathcal{S}}(z)$ is defined as

$$\mathcal{C}_{\mathcal{S}}(z, \bar{z}) := \mathcal{T}_{\mathcal{S}}(z) \cap \bar{z}^{\perp},$$

where $\mathcal{N}_{\mathcal{S}}(\,\cdot\,)$ is the normal cone defined in Definition 2.1 and $\mathcal{T}_{\mathcal{S}}(\,\cdot\,)$ is the tangent cone defined in Definition 2.2.

From [62, Proposition 3.2], we know that the proximal mapping $\operatorname{Pr}_{\theta_1} : \mathbb{S}^m \to \mathbb{S}^m$ (respectively $\Pi_{\mathcal{K}} : \mathbb{S}^n \to \mathbb{S}^n$) is a spectral operator [63, Definition 2] of $\operatorname{Pr}_{\varpi_1}(\cdot)$ (respectively $\Pi_{\widehat{\mathcal{K}}}(\cdot)$). Following the proof path of [64, Theorem 4.1.1], we can easily know that $\operatorname{Pr}_{\varpi_1}$ and $\Pi_{\widehat{\mathcal{K}}}$ are calmly B-differentiable, i.e., for any given $\bar{x}$, there exists a neighborhood $\mathcal{U}$ of $\bar{x}$ such that for all $\bar{x} + h \in \mathcal{U}$,

$$\operatorname{Pr}_{\varpi_1}(\bar{x} + h) - \operatorname{Pr}_{\varpi_1}(\bar{x}) - \operatorname{Pr}'_{\varpi_1}(\bar{x}; h) = 0 \ \text{ and } \ \Pi_{\widehat{\mathcal{K}}}(\bar{x} + h) - \Pi_{\widehat{\mathcal{K}}}(\bar{x}) - \Pi'_{\widehat{\mathcal{K}}}(\bar{x}; h) = 0,$$

where $\mathrm{Pr}'_{\phi}(\mathrm{x};\mathrm{h})$ and $\Pi'_{\widehat{\mathcal{K}}}(\mathrm{x};\mathrm{h})$ are the directional derivatives defined in [1, Page 257, 7(20)]. Thus by applying [65, Theorem 4.1], we have the following lemma.

Lemma 2.7. *The spectral operator* $\mathrm{Pr}_{\theta_1}(\cdot)$ *(respectively* $\Pi_{\mathcal{K}}(\cdot)$*) is calmly B-differentiable for any given* $\overline{X}_1 \in \mathbb{S}^m$ *(respectively* $\overline{X}_2 \in \mathbb{S}^n$*), i.e., for all* $\mathbb{S}^n \ni H \to 0$ *(respectively* $\mathbb{S}^m \ni H \to 0$*),*

$$\mathrm{Pr}_{\theta_1}(\overline{X}_1 + H) - \mathrm{Pr}_{\theta_1}(\overline{X}_1) - \mathrm{Pr}'_{\theta_1}(\overline{X}_1; H) = O(\|H\|^2).$$

(respectively $\Pi_{\mathcal{K}}(\overline{X}_2 + H) - \Pi_{\mathcal{K}}(\overline{X}_2) - \Pi'_{\mathcal{K}}(\overline{X}_2; H) = O(\|H\|^2)$*.)*

This result is not only of its own interest, but also crucial for the study of the proximal and limiting normal cone of $\partial\theta_1$ and $\mathcal{N}_{\mathcal{K}}$ in Section 3.1. Also, by Moreau decomposition (see e.g., [59, Theorem 31.5]), we have the following lemma.

Lemma 2.8. *Let* $f : \mathbb{S}^n \to \mathbb{R}$ *be a general closed proper convex function. The graph of the set-valued map* ∂f *can be written as*

$$\begin{aligned}\mathrm{gph}\,\partial f = \{(X,Y) \mid Y \in \partial f(X)\} &= \{(X,Y) \mid \mathrm{Pr}_f(X+Y) = X\} \\ &= \{(X,Y) \mid \mathrm{Pr}_{f^*}(X+Y) = Y\}.\end{aligned}$$

2.1.2 Full stability

Next, we summarize some useful results on the Lipschitzian full stability for a general optimization problem obtained recently in [16]. Let $\phi : \mathbb{X} \times \mathbb{P} \to (-\infty, +\infty]$ be a given lower semicontinuous function. Consider the following general optimization problem

$$\min_{x \in \mathbb{X}} \ \phi(x, \bar{p}) \tag{2-4}$$

where $\bar{p} \in \mathbb{P}$ is the normal basic parameter, and its perturbed form

$$\mathcal{P}(p, v) \qquad \min_{x \in \mathbb{X}} \ \phi(x, p) - \langle v, x \rangle,$$

with the basic parameter perturbation $p \in \mathbb{P}$ and the tilt one $v \in \mathbb{X}$. The Lipschitzian full stability of (2-4) is first introduced in [14, 66] and the following definition is taken from [16, Definition 3.2].

Definition 2.9. *Given* $\phi : \mathbb{X} \times \mathbb{P} \to (-\infty, +\infty]$ *and* $(\bar{x}, \bar{p}) \in \mathrm{dom}\,\phi$*, we say that the point* $\bar{x}$ *is a Lipschitzian fully stable local minimizer of* $\mathcal{P}(\bar{p}, \bar{v})$ *corresponding to* $(\bar{p}, \bar{v})$ *with a modulus* $(l, k) > 0$*, if there are a number* $\omega > 0$ *and a neighborhood* $\mathcal{U}_{\bar{p}} \times \mathcal{U}_{\bar{v}}$ *of* $(\bar{p}, \bar{v}) \in \mathbb{P} \times \mathbb{X}$ *such that* $\mathcal{M}_{\omega}(p, v) := \mathrm{argmin}\{\phi(x, p) - \langle x, v \rangle : \|x - \bar{x}\| \leq \omega\}$ *is single-valued on* $\mathcal{U}_{\bar{p}} \times \mathcal{U}_{\bar{v}}$ *with* $\mathcal{M}_{\omega}(\bar{p}, \bar{v}) = \bar{x}$ *satisfying the Lipschitz condition*

$$\|\mathcal{M}_{\omega}(p_1, v_1) - \mathcal{M}_{\omega}(p_2, v_2)\| \leq l\|p_1 - p_2\| + k\|v_1 - v_2\| \quad \forall\, v_1, v_2 \in \mathcal{U}_{\bar{v}} \text{ and } p_1, p_2 \in \mathcal{U}_{\bar{p}}$$

and $m_{\omega}(p, v) := \inf\{\phi(x, p) - \langle x, v \rangle : \|x - \bar{x}\| \leq \omega\}$ *is also Lipschitz continuous around* $(\bar{p}, \bar{v})$*.*

The following definitions on the (continuous) prox-regularity of a lower semi-continuous function and the basic constraint qualification (BCQ) of problem (2-4) are taken from [14].

Definition 2.10. *A lower semi-continuous function $\phi : \mathbb{X} \times \mathbb{P} \to (-\infty, +\infty]$ is prox-regular in x at $\bar{x}$ for $\bar{v}$ with compatible parameterization by p at $\bar{p}$ if $\bar{v} \in \partial_x \phi(\bar{x}, \bar{p})$ and there exist neighborhoods $\mathcal{U}_{\bar{x}}$ of $\bar{x}$, $\mathcal{U}_{\bar{v}}$ of $\bar{v}$ and $\mathcal{U}_{\bar{p}}$ of $\bar{p}$ along with $\varepsilon > 0$ and $t > 0$ such that if $v \in \partial_x \phi(x, p) \cap \mathcal{U}_{\bar{v}}$, $x \in \mathcal{U}_{\bar{x}}$, $p \in \mathcal{U}_{\bar{p}}$ and $\phi(x, p) \le \phi(\bar{x}, \bar{p}) + \varepsilon$, then*

$$\phi(x', p) \ge \phi(x, p) + \langle v, x' - x \rangle - \frac{t}{2}\|x - x'\|^2 \quad \forall\, x' \in \mathbb{X}.$$

It is subdifferentially continuous in x at $\bar{x}$ for $\bar{v}$ with compatible parameterization by p at $\bar{p}$ if function $(x, p, v) \mapsto \phi(x, p)$ is continuous relative to $\operatorname{gph} \partial_x \phi$ at $(\bar{x}, \bar{p}, \bar{v})$, where $\partial_x \phi$ denotes the partial limiting subdifferential of ϕ with respect to x. ϕ is said to be parametrically continuously prox-regular at $(\bar{x}, \bar{p})$ for $\bar{v}$ when ϕ is prox-regular and subdifferentially continuous in x at $\bar{x}$ for $\bar{v}$ with compatible parameterization by p at $\bar{p}$.

Definition 2.11. *We say that the basic constraint qualification (BCQ) of problem* (2-4) *holds at $(\bar{x}, \bar{p}) \in \operatorname{dom} \phi$ if*

$$(0, p) \in \partial^{\infty} \phi(\bar{x}, \bar{p}) \implies p = 0.$$

It is worth to note that for (1-6), BCQ and prox-regularity hold by [14, Proposition 2.2] under Robinson constraint qualification (see (3-61)). The following characterization of Lipschitzian full stability for the general optimization problem (2-4) is established by Mordukhovich et al. [16, Theorem 4.1].

Theorem 2.12. *Suppose that BCQ holds at $(\bar{x}, \bar{p}) \in \operatorname{dom} \phi$ and that ϕ is parametrically continuously prox-regular at $(\bar{x}, \bar{p})$ for $\bar{v} \in \partial_x \phi(\bar{x}, \bar{p})$. Then the following assertions are equivalent:*

(i) The point $\bar{x}$ is a Lipschitzian fully stable local minimizer of problem $\mathcal{P}(\bar{p}, \bar{v})$ with a positive modulus pair $(s, t) > 0$.

(ii) There are neighborhoods $\mathcal{U}_{\bar{x}}$ of $\bar{x}$, $\mathcal{U}_{\bar{p}}$ of $\bar{p}$ and $\mathcal{U}_{\bar{v}}$ of $\bar{v}$ such that the mapping $\mathcal{S}(p, v) := \{x \in \mathbb{X} : v \in \partial_x \phi(x, p)\}$ admits a single valued localization ϑ relative to $\mathcal{U}_{\bar{p}} \times \mathcal{U}_{\bar{v}} \times \mathcal{U}_{\bar{x}}$ such that for any triple $(p, v, u) \in \operatorname{gph} \vartheta = \operatorname{gph} \mathcal{S} \cap (\mathcal{U}_{\bar{p}} \times \mathcal{U}_{\bar{v}} \times \mathcal{U}_{\bar{x}})$ we have the uniform second order growth condition

$$\phi(x, p) \ge \phi(u, p) + \langle v, x - u \rangle + \frac{1}{2t}\|x - u\|^2 \text{ whenever } x \in \mathcal{U}_{\bar{x}}.$$

together with the condition

$$(0, q) \in \mathcal{D}^* \partial_x \phi(\bar{x}, \bar{p}, \bar{v})(0) \Rightarrow q = 0. \tag{2-5}$$

Moreover, [16, Corollary 4.5] gives the following equivalent condition for Lipschitzian full stability.

Theorem 2.13. *In the setting of Theorem 2.12, we have the equivalent statements:*

(i) The point $\bar{x}$ is a Lipschitzian fully stable local minimizer of problem $\mathcal{P}(\bar{p}, \bar{v})$.

(ii) Condition (2-5) *is satisfied together with the inequality*

$$\inf\{\langle z, w\rangle : (z, p) \in \mathcal{D}^*\phi(\bar{x}, \bar{p}, \bar{v})(\mathrm{w})\} > 0 \quad \textit{for all} \quad w \neq 0,$$

where we use the convention that $\inf \varnothing := \infty$.

In the following, we mainly consider the case where $\bar{v} = 0$. It is worth to note that the uniform second order growth condition and (2-5) in Theorem 2.12 are difficult to be verified for the general problems. However, the above two results inspire us to focus on how to obtain the explicit form of the coderivative, which might be of some help in attaining a more explicit characterization of Lipschitzian full stability.

For the CMatOP (1-6), we are able to establish a more explicit characterization of the Lipschitzian full stability at a local minimizer $\bar{x}$ with respect to the multiplier $\overline{Y} \in \partial\theta_1(g_1(\bar{x}, \bar{p}))$ and $\overline{Z} \in \mathcal{N}_{\mathcal{K}}(g_2(\bar{x}, \bar{p}))$ in terms of the coderivatives $\mathcal{D}^*\partial\theta_1(g_1(\bar{x}, \bar{p}), \overline{Y})(\cdot)$ of $\partial\theta_1$ and $\mathcal{D}^*\mathcal{N}_{\mathcal{K}}(g_2(\bar{x}, \bar{p}), \overline{Z})(\cdot)$ of $\mathcal{N}_{\mathcal{K}}$. More details will be discussed in Section 3.2, Theorem 3.13. Since the explicit form of the coderivatives $\mathcal{D}^*\partial\theta_1(g_1(\bar{x}, \bar{p}), \overline{Y})\big((g_1)'_x(\bar{x}, \bar{p})\mathrm{w}\big)$ and $\mathcal{D}^*\mathcal{N}_{\mathcal{K}}(g_2(\bar{x}, \bar{p}), \overline{Z})\big((g_2)'_x(\bar{x}, \bar{p})w\big)$ is such important, we will study the explicit formulas of the limiting normal cones of $\mathrm{gph}\partial\theta_1$ and $\mathrm{gph}\,\mathcal{N}_{\mathcal{K}}$ in Section 3.1 according to the definition of coderivative (Definition 2.5). Moreover, for general constrained optimization problem with a smooth objective function, we can refer to [16, Theorem 5.6] to get a similar result.

2.2 Eigenvalue decomposition properties

Let $X \in \mathbb{S}^n$ be an arbitrary symmetric matrix. Suppose that X has the following eigenvalue decomposition

$$X = U\,\mathrm{Diag}\,(\lambda_1(X), \cdots, \lambda_n(X))\; U^\top, \tag{2-6}$$

where $\lambda_1(X) \geq \cdots \geq \lambda_n(X)$ are the eigenvalues of X arranged in the non-increasing order and U is a matrix of the corresponding orthonormal eigenvectors. We denote the set of all matrices U satisfying (2-6) as $\mathcal{O}^n(X)$ and $\mathcal{O}^n$ as the set of all n-dimensional orthogonal matrix. We also use $\lambda(X)$ to denote the vector whose i-th entry is $\lambda_i(X)$. Denote $\Lambda(X) = \mathrm{Diag}\,(\lambda_1(X), \cdots, \lambda_n(X))$. Let $v_1(X) > v_2(X) > \ldots > v_r(X)$ be the distinct eigenvalues of X arranged in the decreasing order. Define

$$\alpha^l := \big\{i \in \{1, \ldots, n\} \,|\, \lambda_i(X) = v_l(X)\big\}, \quad l = 1, \ldots, r. \tag{2-7}$$

For each $i \in \{1, \ldots, n\}$, we denote $k_i(X)$ as the number of eigenvalues that equal to $\lambda_i(X)$ but are ranked before i (including i) and $o_i(X)$ as the number of eigenvalues that equal to $\lambda_i(X)$ but are ranked after i (excluding i). That is, the scalars $k_i(X)$ and $o_i(X)$ satisfy

$$\begin{aligned}&\lambda_1(X) \geq \ldots \geq \lambda_{i-k_i(X)}(X) > \lambda_{i-k_i(X)+1}(X) = \ldots = \lambda_i(X) = \ldots = \lambda_{i+o_i(X)}(X) \\ &> \lambda_{i+o_i(X)+1}(X) \geq \ldots \geq \lambda_n(X).\end{aligned} \tag{2-8}$$

In the subsequent discussions, when the dependence of k_i and o_i on X can be easily seen from the context, we often drop X for simplicity.

In the following, we summarize some results about the properties of the eigenvalues that are essential in our subsequent discussions. The first result is Ky Fan's inequality [67].

Lemma 2.14. *Let Y and Z be two matrices in $\mathbb{S}^n$. Then*

$$\langle Y, Z\rangle \leq \lambda(Y)^\top \lambda(Z),$$

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

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

The next lemma is about the directional differentiability of the eigenvalue function, which can be found in, for example, [68, Theorem 7] and [69, Proposition 1.4].

Lemma 2.15. *Let $X \in \mathbb{S}^n$ have the eigenvalue decomposition in (2-6). Then for any $\mathbb{S}^n \ni H \to 0$, we have*

$$\lambda_i(X + H) - \lambda_i(X) - \lambda_{k_i}(U_{\alpha^l}^\top H U_{\alpha^l}) = O(\|H\|^2), \quad i \in \alpha^l,\ l = 1, \ldots, r,$$

where for each $i \in \{1, \ldots, n\}$, k_i is defined in (2-8). Hence, for any given direction $H \in \mathbb{S}^n$, the eigenvalue function $\lambda_i(\cdot)$ is directionally differentiable at X with the directional derivative $\lambda_i'(X; H) = \lambda_{k_i}(U_{\alpha^l}^\top H U_{\alpha^l})$ for any $i \in \alpha^l$, $l = 1, \ldots, r$.

Let $l \in \{1, \ldots, r\}$ be fixed. Consider the following eigenvalue decomposition of the symmetric matrix $U_{\alpha^l}^\top H U_{\alpha^l} \in \mathbb{S}^{|\alpha^l|}$:

$$U_{\alpha^l}^\top H U_{\alpha^l} = R\Lambda(U_{\alpha^l}^\top H U_{\alpha^l})R^\top,$$

where $R \in \mathcal{O}^{|\alpha^l|}$. Denote the distinct eigenvalues of $U_{\alpha^l}^\top H U_{\alpha^l}$ by $\tilde{v}_1 > \tilde{v}_2 > \ldots > \tilde{v}_{\tilde{r}}$. Define

$$\tilde{\alpha}^j := \left\{ i \in \{1, \ldots, |\alpha^l|\} \mid \lambda_i(U_{\alpha^l}^\top H U_{\alpha^l}) = \tilde{v}_j \right\}, \quad j = 1, \ldots, \tilde{r}.$$

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

$$\tilde{k}_i := k_{k_i}(U_{\alpha^l}^\top H U_{\alpha^l}) \quad \text{and} \quad \tilde{k}_i \in \tilde{\alpha}^{\tilde{l}},$$

where k_i is defined by (2-8).

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

$$\lim_{t \downarrow 0} \frac{\Phi(z + th + \frac{1}{2}t^2 w) - \Phi(z) - t\,\Phi'(z; h)}{\frac{1}{2}t^2} \quad \text{exists.}$$

The above limit is said to be the (parabolic) second-order directional derivative of Φ at z along the directions h and w, which we denote as $\Phi''(z; h, w)$. The following proposition, which has its source from [69, Proposition 2.2], provides an explicit formula of the (parabolic) second-order directional derivative of the eigenvalue function.

Lemma 2.16. *Let $X \in \mathbb{S}^n$ have the eigenvalue decomposition* (2-42)*. Then for any $H, W \in \mathbb{S}^n$,*

$$\lambda_i''(X; H, W) = \lambda_{\tilde{k}_i}\left(R_{\tilde{\alpha}^{\tilde{l}}}^\top U_{\alpha^l}^\top \left[W - 2H(X - \lambda_i I_n)^\dagger H\right] U_{\alpha^l} R_{\tilde{\alpha}^{\tilde{l}}}\right), \quad i \in \alpha^l,\ l \in \{1, \dots, r\},$$

where $Z^\dagger \in \mathbb{R}^{p \times p}$ is the Moore-Penrose pseudo-inverse of the square matrix $Z \in \mathbb{R}^{p \times p}$.

The following lemma on the subdifferential of spectral functions can be found in [70, 71].

Lemma 2.17. *Let $\phi : \mathbb{R}^n \to (-\infty, +\infty]$ be a proper closed convex and symmetric function. Suppose $X \in \mathbb{S}^n$ has the vector of eigenvalues $\lambda(X)$ in $\operatorname{dom}\phi$. Let $W \in \mathbb{S}^n$. Then $W \in \partial(\phi \circ \lambda)(X)$ if and only if $\lambda(W) \in \partial\phi(\lambda(X))$ and there exists $U \in \mathcal{O}^n(X) \cap \mathcal{O}^n(W)$. In fact, $\partial(\phi \circ \lambda)(X) = \{U \mathrm{Diag}(\mu) U^\top \mid \mu \in \partial\phi(\lambda(X)),\ U \in \mathcal{O}^n(X)\}$.*

2.3 Variational analysis of composite matrix optimization problem

This section lists some preliminaries on the variational properties of composite matrix optimization problem and positive semidefinite cone. Detailed discussions on these subjects can be found in [1, 4, 57, 60].

We denote the set of all $n \times n$ permutation matrices as $\mathcal{P}^n$. Recall that the function ϖ is called symmetric over $\mathbb{R}^n$ if for any $Q \in \mathcal{P}^n$ and any $x \in \mathbb{R}^n$, it holds that $\varpi(x) = \varpi(Qx)$. The following proposition [4, Proposition 1] characterizes the symmetric of piecewise affine function.

Proposition 2.18. *Let $\varpi : \mathbb{R}^n \to (-\infty, \infty]$ be a given proper convex piecewise affine function. Then the function ϖ is symmetric over $\mathbb{R}^n$ if and only if its decomposed* *components* $\varpi_1 : \mathbb{R}^n \to \mathbb{R}$ *and* $\varpi_2 : \mathbb{R}^n \to (-\infty, \infty]$ *in* (1-4) *satisfy the following conditions:*

$$\begin{cases} \varpi_1(x) = \max_{1\le i\le p} \left\{ \max_{Q\in\mathcal{P}^n} \left\{ \langle Qa^i, x\rangle - c_i \right\} \right\}, \\ \operatorname{dom}\varpi = \left\{ x\in\mathbb{R}^n \mid \max_{1\le i\le q} \left\{ \max_{Q\in\mathcal{P}^n} \left\{ \langle Qb^i, x\rangle - d_i \right\} \right\} \le 0 \right\}, \end{cases} \quad \forall\, x\in\mathbb{R}^n. \tag{2-9}$$

Remark 2.1. It is easy to verify from Proposition 2.18 that if a proper convex piecewise affine function ϖ is symmetric, then both of its components ϖ_1 and ϖ_2 in (1-4) are symmetric.

To proceed, we denote

$$\mathcal{D}_i := \left\{ x\in\operatorname{dom}\varpi \mid \langle a^j, x\rangle - c_j \le \langle a^i, x\rangle - c_i,\ \forall\, j = 1,\dots,p \right\}, \quad i = 1,\dots,p.$$

It follows from [72, Proposition 3.2] that $\operatorname{dom}\varpi = \bigcup_{i=1,\dots,p} \mathcal{D}_i$. For any $\bar{x}\in\operatorname{dom}\varpi$, we further denote the following two index sets:

$$\iota_1(\bar{x}) := \{1\le i\le p \mid \bar{x}\in\mathcal{D}_i\} \tag{2-10}$$

and

$$\iota_2(\bar{x}) := \{1\le i\le q \mid \langle b^i, \bar{x}\rangle - d_i = 0\}. \tag{2-11}$$

The notation $\iota_1(\bar{x})$ is adopted from [72, Proposition 3.2], which represents the set of active indices defined in [5, Example 2.68]. It is known that the pointwise-max function ϖ_1 in (1-4) and ψ in (1-3) are directionally differentiable everywhere with the following directional derivatives (see, e.g., [5, Example 2.68])

$$\varpi_1'(\bar{x}; h) = \max_{i\in\iota_1(\bar{x})} \langle a^i, h\rangle \quad \text{and} \quad \psi'(\bar{x}; h) = \max_{i\in\iota_2(\bar{x})} \langle b^i, h\rangle, \quad h\in\mathbb{R}^n. \tag{2-12}$$

Denote ∂f as the subgradient of a convex function f. It also holds that

$$\partial\varpi_1(\bar{x}) = \operatorname{conv}\{a^i,\ i\in\iota_1(\bar{x})\} \quad \text{and} \quad \partial\varpi_2(\bar{x}) = \mathcal{N}_{\operatorname{dom}\varpi}(\bar{x}) = \operatorname{cone}\{b^i,\ i\in\iota_2(\bar{x})\}, \tag{2-13}$$

where "$\operatorname{conv}\mathcal{C}$" and "$\operatorname{cone}\mathcal{C}$" stand for the convex hull and conic hull of a given nonempty closed set $\mathcal{C}$, respectively (if $\iota_2(\bar{x}) = \emptyset$, then $\partial\varpi_2(\bar{x}) = \{0\}$). Therefore, for any given $\bar{y}\in\partial\varpi_1(\bar{x})$ and $\bar{z}\in\partial\varpi_2(\bar{x})$, we are able to define the following two index sets

$$\eta_1(\bar{x}, \bar{y}) := \left\{ i\in\iota_1(\bar{x}) \mid \sum_{i\in\iota_1(\bar{x})} u_i\, a^i = \bar{y},\ \sum_{i\in\iota_1(\bar{x})} u_i = 1,\ 0 < u_i \le 1 \right\}, \tag{2-14}$$

$$\eta_2(\bar{x}, \bar{z}) := \left\{ i\in\iota_2(\bar{x}) \mid \sum_{i\in\iota_2(\bar{x})} u_i b^i = \bar{z},\ u_i > 0 \right\}. \tag{2-15}$$

The following corollary is a direct consequence of Proposition 2.18.

Corollary 2.19. *The following two statements hold.*

(i) For any $i \in \iota_1(\bar{x})$*,* $j \in \iota_2(\bar{x})$ *and* $Q \in \mathcal{P}^n_{\bar{x}}$ *(i.e.,* $Q\bar{x} = \bar{x}$*), there exist* $i' \in \iota_1(\bar{x})$ *and* $j' \in \iota_2(\bar{x})$ *such that* $a^{i'} = Qa^i$ *and* $b^{j'} = Qb^j$*, respectively.*

(ii) For any $i \in \eta_1(\bar{x}, \bar{y})$*,* $j \in \eta_2(\bar{x}, \bar{z})$*,* $Q^1 \in \mathcal{P}^n_{\bar{x}} \cap \mathcal{P}^n_{\bar{y}}$ *and* $Q^2 \in \mathcal{P}^n_{\bar{x}} \cap \mathcal{P}^n_{\bar{z}}$*, there exist* $i' \in \eta_1(\bar{x}, \bar{y})$ *and* $j' \in \eta_2(\bar{x}, \bar{z})$ *such that* $a^{i'} = Q^1 a^i$ *and* $b^{j'} = Q^2 b^j$*, respectively.*

The following proposition, which is adopted from [4, Proposition 2], shows the relationship between the directional derivative of the proximal mappings associated with ϖ_1 and ϖ_2 and the corresponding critical cone. Necessary and sufficient conditions for them to be F(réchet)-differentiable are also provided.

Proposition 2.20. *Let* $\varpi_1 : \mathbb{R}^n \to \mathbb{R}$ *and* $\varpi_2 : \mathbb{R}^n \to (-\infty, \infty]$ *be given by* (1-4)*. Then the following two properties hold for the corresponding proximal mappings* $\mathrm{Pr}_{\varpi_1} : \mathbb{R}^n \to \mathbb{R}^n$ *and* $\mathrm{Pr}_{\varpi_2} : \mathbb{R}^n \to \mathbb{R}^n$*.*

(i) Pr_{ϖ_1} *and* Pr_{ϖ_2} *are directionally differentiable everywhere with the directional derivatives*

$$\begin{cases} \mathrm{Pr}'_{\varpi_1}(x;h) = \underset{d \in \mathbb{R}^n}{\mathrm{argmin}} \left\{ \|d-h\|^2 \mid d \in \mathcal{C}_{\varpi_1}(\mathrm{Pr}_{\varpi_1}(x), x - \mathrm{Pr}_{\varpi_1}(x)) \right\} \\ \mathrm{Pr}'_{\varpi_2}(x;h) = \underset{d \in \mathbb{R}^n}{\mathrm{argmin}} \left\{ \|d-h\|^2 \mid d \in \mathcal{C}_{\varpi_2}(\mathrm{Pr}_{\varpi_2}(x), x - \mathrm{Pr}_{\varpi_2}(x)) \right\} \end{cases} \quad x, h \in \mathbb{R}^n. \tag{2-16}$$

(ii) Let the index sets ι_1*,* ι_2*,* η_1 *and* η_2 *be given by* (2-10)*,* (2-11)*,* (2-14) *and* (2-15)*, respectively. Then* Pr_{ϖ_1} *is F-differentiable at* x *if and only if*

$$\eta_1\left(\mathrm{Pr}_{\varpi_1}(x),\, x - \mathrm{Pr}_{\varpi_1}(x)\right) = \iota_1\left(\mathrm{Pr}_{\varpi_1}(x)\right);$$

similarly, Pr_{ϖ_2} *is F-differentiable at* x *if and only if*

$$\eta_2\left(\mathrm{Pr}_{\varpi_2}(x),\, x - \mathrm{Pr}_{\varpi_2}(x)\right) = \iota_2\left(\mathrm{Pr}_{\varpi_2}(x)\right).$$

Moreover, under the above two conditions, the derivatives $\mathrm{Pr}'_{\varpi_1}(x)$ *and* $\mathrm{Pr}'_{\varpi_2}(y)$ *are given by*

$$\begin{cases} \mathrm{Pr}'_{\varpi_1}(x)\,h = \underset{d \in \mathbb{R}^n}{\mathrm{argmin}} \left\{ \|d-h\|^2 \mid \langle d, a^i - a^j \rangle = 0,\ i,j \in \iota_1\left(\mathrm{Pr}_{\varpi_1}(x)\right) \right\} \\ \mathrm{Pr}'_{\varpi_2}(x)\,h = \underset{d \in \mathbb{R}^n}{\mathrm{argmin}} \left\{ \|d-h\|^2 \mid \langle d, b^i \rangle = 0,\ i \in \iota_2\left(\mathrm{Pr}_{\varpi_2}(x)\right) \right\} \end{cases} \quad x, h \in \mathbb{R}^n. \tag{2-17}$$

Then, we list several important variational properties of the spectral function $\theta \equiv \varpi \circ \lambda$ for a symmetric piecewise affine function ϖ. Most of them are established in [4, Section 4]. Due to the symmetry, properness and convexity of ϖ, its induced spectral function θ is also proper and convex ([73] and [70, page 164]). For more details about the relationship between a symmetric function and its corresponding spectral function, one may refer to [70, 71, 73, 74]. By using the Lipschitz continuity of ψ and ϖ_1, and the well-known Ky-Fan's inequality [67], it is easy to see that θ_1 and ζ are also Lipschitz continuous.

In the following two subsections, the tangent sets, critical cones and the so-called sigma term associated with θ_1 and $\theta_2 := \delta_{\mathcal{K}}$, respectively are characterized.

2.3.1 Variational properties of θ_1

We first introduce the variational properties of the spectral function $\theta_1 = \varpi_1 \circ \lambda$. The results in this part are directly adopted from [4, Section 3.1]. Firstly, we adopt the explicit form of the tangent set and its lineality space. Let $\overline{X} \in \mathbb{S}^n$ be given. One may know from [75, Theorem 3] and the directional differentiability of ϖ_1 that θ_1 is also directionally differentiable. Since θ_1 is Lipschitz continuous on $\mathbb{S}^n$, it follows from [5, Proposition 2.58; page 42, the paragraph under (2.70)] that the tangent cone $\mathcal{T}_{\operatorname{epi}\theta_1}(\overline{X}, \theta_1(\overline{X}))$ of the epigraph $\operatorname{epi}\theta_1$ at $(\overline{X}, \theta_1(\overline{X})) \in \operatorname{epi}\theta_1$ is given by

$$\mathcal{T}_{\operatorname{epi}\theta_1}(\overline{X}, \theta_1(\overline{X})) = \operatorname{epi}\theta_1'(\overline{X};\cdot) = \left\{(H,h) \in \mathbb{S}^n \times \mathbb{R} \mid \theta_1'(\overline{X};H) \le h\right\}.$$

The corresponding lineality space [4, (17)] is defined as

$$\mathcal{T}_{\theta_1}^{\text{lin}}(\overline{X}) := \left\{H \in \mathbb{S}^n \mid \theta_1'(\overline{X};H) = -\theta_1'(\overline{X};-H)\right\}. \tag{2-18}$$

Let $\{\alpha^l\}_{l=1}^r$ be the index sets given by (2-7) with respect to $\overline{X}$. Define the index set

$$\widetilde{\mathcal{E}} := \{l \in \{1,\ldots,r\} \mid \exists\, i,j \in \alpha^l \text{ such that } (a^w)_i \neq (a^w)_j \text{ for some } w \in \iota_1(\lambda(\overline{X}))\},$$

where $\iota_1(\lambda(\overline{X}))$ is the set defined in (2-10) with respect to $\lambda(\overline{X})$. The following proposition characterizes $\mathcal{T}_{\theta_1}^{\text{lin}}(\overline{X})$.

Proposition 2.21. *Let $H \in \mathbb{S}^n$. Then $H \in \mathcal{T}_{\theta_1}^{\text{lin}}(\overline{X})$ implies the existence of scalars $\{\widehat{\rho}_l\}_{l\in\widetilde{\mathcal{E}}}$ such that $\overline{U}_{\alpha^l}^{\top} H \overline{U}_{\alpha^l} = \widehat{\rho}_l\, I_{|\alpha^l|}$ for any $\overline{U} \in \mathcal{O}^n(\overline{X})$. In fact,*

$$H \in \mathcal{T}_{\theta_1}^{\text{lin}}(\overline{X}) \iff \left[\ \langle \lambda'(\overline{X};H), a^i - a^j\rangle = 0, \quad \forall\, i,j \in \iota_1(\lambda(\overline{X}))\ \right].$$

The critical cone of θ_1 is stated as follows. Suppose that $\overline{Y} \in \partial\theta_1(\overline{X})$. Then the critical cone to θ_1 at $\overline{X}$ for $\overline{Y}$ is defined as

$$\mathcal{C}_{\theta_1}(\overline{X}, \overline{Y}) := \left\{H \in \mathbb{S}^n \mid \theta_1'(\overline{X};H) = \langle \overline{Y}, H\rangle\right\}. \tag{2-19}$$

For each $l \in \{1,\ldots,r\}$, we further partition the index set α^l into $\{\beta_k^l\}_{k=1}^{s_l}$ such that each β_k^l contains one distinct eigenvalue of $\overline{Y}$, i.e.,

$$\begin{cases} \lambda_i(\overline{Y}) = \lambda_j(\overline{Y}) & \text{if } i,j \in \beta_k^l, \\ \lambda_i(\overline{Y}) > \lambda_j(\overline{Y}) & \text{if } i \in \beta_k^l,\ j \in \beta_{k'}^l \text{ and } k,k' \in \{1,\ldots,s_l\} \text{ with } k < k'. \end{cases} \tag{2-20}$$

We also denote

$$\mathcal{E}^l := \{k \in \{1,\ldots,s_l\} \mid \exists\, i,j \in \beta^l_k \text{ such that } (a^w)_i \neq (a^w)_j \text{ for some } w \in \eta_1(\lambda(\overline{X}),\lambda(\overline{Y}))\}, \tag{2-21}$$

where $\eta_1(\lambda(\overline{X}),\lambda(\overline{Y}))$ is the index set defined in (2-14) with respect to $\lambda(\overline{X})$ and $\lambda(\overline{Y})$. The characterization of the critical cone $\mathcal{C}_{\theta_1}(\overline{X},\overline{Y})$ is provided in the following proposition.

Proposition 2.22. *Suppose that* $(\overline{X},\overline{Y}) \in \operatorname{gph}\partial\theta_1$ *and* $\overline{U} \in \mathcal{O}^n(\overline{X}) \cap \mathcal{O}^n(\overline{Y})$. *If* $H \in \mathcal{C}_{\theta_1}(\overline{X},\overline{Y})$, *then the following three properties hold:*

(i) for each $l \in \{1,\ldots,r\}$, $\overline{U}_{\alpha^l}^\top H \overline{U}_{\alpha^l}$ *has the following block diagonal structure:*

$$\overline{U}_{\alpha^l}^\top H \overline{U}_{\alpha^l} = \operatorname{Diag}\left((\overline{U}_{\alpha^l}^\top H\overline{U}_{\alpha^l})_{\beta^l_1\beta^l_1},\ldots,(\overline{U}_{\alpha^l}^\top H\overline{U}_{\alpha^l})_{\beta^l_{s_l}\beta^l_{s_l}}\right);$$

(ii) $\langle \lambda'(\overline{X};H), a^i\rangle = \langle \lambda'(\overline{X};H), a^j\rangle = \max\limits_{l\in\iota_1(\lambda(\overline{X}))} \langle \lambda'(\overline{X};H), a^l\rangle, \quad \forall\, i,j \in \eta_1(\lambda(\overline{X}),\lambda(\overline{Y}))$;

(iii) for each $l \in \{1,\ldots,r\}$ *and* $k \in \mathcal{E}^l$, *there exists a scalar* ρ^l_k *such that* $\left(\overline{U}_{\alpha^l}^\top H\overline{U}_{\alpha^l}\right)_{\beta^l_k\beta^l_k} = \rho^l_k I_{|\beta^l_k|}$.

In fact, $H \in \mathcal{C}_{\theta_1}(\overline{X},\overline{Y})$ *if and only if for any* $i,j \in \eta_1(\lambda(\overline{X}),\lambda(\overline{Y}))$,

$$\langle \operatorname{diag}(\overline{U}^\top H\overline{U}),\, a^i\rangle = \langle \operatorname{diag}(\overline{U}^\top H\overline{U}),\, a^j\rangle = \max_{\kappa\in\iota_1(\lambda(\overline{X}))} \langle \lambda'(\overline{X};H), a^\kappa\rangle, \tag{2-22}$$

where the index sets η_1 *and* ι_1 *are defined in* (2-14) *and* (2-10), *respectively.*

Remark 2.2. The convexity of $\mathcal{C}_{\theta_1}(\overline{X},\overline{Y})$ can be directly verified. By[4, page 10], we have that for all $i \in \eta_1(\lambda(\overline{X}),\lambda(\overline{Y}))$, $\langle \operatorname{diag}(\overline{U}^\top H\overline{U}),\, a^i\rangle = \langle \lambda'(\overline{X};H), a^i\rangle$ holds. Pick any $H_1, H_2 \in \mathcal{C}_{\theta_1}(\overline{X},\overline{Y})$ and$c \in [0,1]$. It follows from [4, page 9] that

$$\begin{aligned}
\langle \overline{Y}, cH_1 + (1-c)H_2\rangle &= c\langle \lambda(\overline{Y}), \lambda'(\overline{X},H_1)\rangle + (1-c)\langle \lambda(\overline{Y}), \lambda'(\overline{X},H_2)\rangle \\
&= c\theta_1'(\overline{X},H_1) + (1-c)\theta_1'(\overline{X},H_2) \\
&\le \theta_1'(\overline{X}, cH_1 + (1-c)H_2).
\end{aligned}$$

By the convexity of θ_1, we have for all $t > 0$,

$$c\frac{\theta_1(\overline{X}+tH_1)-\theta_1(\overline{X})}{t} + (1-c)\frac{\theta_1(\overline{X}+tH_2)-\theta_1(\overline{X})}{t} \ge \frac{\theta_1(\overline{X}+t(cH_1+(1-c)H_2))-\theta_1(\overline{X})}{t}.$$

Taking $t \to 0$ on both sides, we have

$$c\theta_1'(\overline{X},H_1) + (1-c)\theta_1'(\overline{X},H_2) \ge \theta_1'(\overline{X}, cH_1 + (1-c)H_2).$$

Thus we have verified $cH_1 + (1-c)H_2 \in \mathcal{C}_{\theta_1}(\overline{X},\overline{Y})$, which implies $\mathcal{C}_{\theta_1}(\overline{X},\overline{Y})$ is convex.

Based on the above proposition, we can directly characterize the affine hull of $\mathcal{C}_{\theta_1}(\overline{X},\overline{Y})$, which we denoted as $\operatorname{aff}(\mathcal{C}_{\theta_1}(\overline{X},\overline{Y}))$.

Proposition 2.23. *Suppose that* $(\overline{X},\overline{Y}) \in \operatorname{gph}\partial\theta_1$ *and* $\overline{U} \in \mathcal{O}^n(\overline{X}) \cap \mathcal{O}^n(\overline{Y})$. *Then* $H \in \operatorname{aff}(\mathcal{C}_{\theta_1}(\overline{X},\overline{Y}))$ *if and only if it satisfies the the properties (i) and (iii) in Proposition 2.22, and*

$$\langle \operatorname{diag}(\overline{U}^\top H\overline{U}), a^i\rangle = \langle \operatorname{diag}(\overline{U}^\top H\overline{U}), a^j\rangle, \quad \forall\, i,j \in \eta_1(\lambda(\overline{X}),\lambda(\overline{Y})). \tag{2-23}$$

As for the sigma term, suppose that $\overline{Y} \in \partial\theta_1(\overline{X})$. Let $H \in \mathcal{C}_{\theta_1}(\overline{X},\overline{Y})$ be arbitrarily given. Since ϖ_1 is Lipschitz continuous, we know from [76, Lemma 3.1] that θ_1 is (parabolic) second-order directionally differentiable with the second-order directional derivative

$$F_{\overline{X},H}(W) := \theta_1''(\overline{X};H,W) = \varpi_1''(\lambda(\overline{X});\lambda'(\overline{X};H),\lambda''(\overline{X};H,W)), \quad W \in \mathbb{S}^n. \tag{2-24}$$

Moreover, it is easy to see that $F_{\overline{X},H} : \mathbb{S}^n \to \mathbb{R}$ is convex. We define the sigma term associated with the spectral function $\theta_1 = \varpi_1 \circ \lambda$ at $\overline{Y} \in \partial\theta_1(\overline{X})$ as the conjugate function (cf. [59] for the definition) of $F_{\overline{X},H}$ at $\overline{Y}$, that is, we consider the function

$$F^*_{\overline{X},H}(\overline{Y}) = \sup_{W\in\mathbb{S}^n}\left\{\langle W,\overline{Y}\rangle - F_{\overline{X},H}(W)\right\}.$$

The proposition below characterizes the property of $F^*_{\overline{X},H}$.

Proposition 2.24. *Suppose that* $(\overline{X},\overline{Y}) \in \operatorname{gph}\partial\theta_1$ *and* $\overline{U} \in \mathcal{O}^n(\overline{X}) \cap \mathcal{O}^n(\overline{Y})$. *Denote* $\bar{v}_1 > \bar{v}_2 > \cdots > \bar{v}_r$ *as the distinct eigenvalues of* $\overline{X}$. *Let* $H \in \mathcal{C}_{\theta_1}(\overline{X},\overline{Y})$ *be given. Then*

$$F^*_{\overline{X},H}(\overline{Y}) = 2\sum_{l=1}^{r}\langle \Lambda(\overline{Y})_{\alpha^l\alpha^l}\,,\ \overline{U}_{\alpha^l}^\top H(\overline{X}-\bar{v}_l I)^\dagger H\overline{U}_{\alpha^l}\rangle, \tag{2-25}$$

where for each $l \in \{1,\ldots,r\}$, $(\overline{X}-\bar{v}_l I)^\dagger$ *is the Moore-Penrose pseudo-inverse of* $\overline{X}-\bar{v}_l I$.

In fact, for any given $\overline{X} \in \mathbb{S}^n$ and any $\overline{Y}, H \in \mathbb{S}^n$ (not necessary in $\mathcal{C}(\overline{X}+\overline{Y},\partial\theta_1(\overline{X}))$), we can define the function $\Upsilon^1_{\overline{X}} : \mathbb{S}^n\times\mathbb{S}^n \to \mathbb{R}$ as the right side of (2-25), i.e.,

$$\Upsilon^1_{\overline{X}}\left(\overline{Y},H\right) := 2\sum_{l=1}^{r}\langle \Lambda(\overline{Y})_{\alpha^l\alpha^l}, \overline{U}_{\alpha^l}^\top H(\overline{X}-\bar{v}_l I)^\dagger H\overline{U}_{\alpha^l}\rangle, \tag{2-26}$$

where $\overline{U} \in \mathcal{O}^n(\overline{X}) \cap \mathcal{O}^n(\overline{Y})$. Notice that if $\overline{Y} \in \partial\theta_1(\overline{X})$, it holds that

$$\Upsilon^1_{\overline{X}}\left(\overline{Y},H\right) = -2\sum_{1\le l<l'\le r}\sum_{i\in\alpha^l}\sum_{j\in\alpha^{l'}}\frac{\lambda_i(\overline{Y})-\lambda_j(\overline{Y})}{\lambda_i(\overline{X})-\lambda_j(\overline{X})}(\overline{U}_{\alpha^l}^\top H\overline{U}_{\alpha^{l'}})^2_{ij}. \tag{2-27}$$

To elaborate the above abstract results more clearly, we provide a simple example for better comprehension.

Example 2.1. Consider the following largest eigenvalue problem in Example 1.1. This corresponds to a special case of problem (1-1) where

$$\varpi(x) = \max_{1\le i\le p} \left\{\langle e^i, x\rangle\right\}, \quad x \in \mathbb{R}^n$$

with e^i being the unit vector whose i-th component is 1 and others are zero. By applying the above results, we can get the explicit forms of tangent cone, critical cone and sigma term correspond to the largest eigenvalue problem (1-7).

The tangent cone and its lineality space. Recall the definitions of $\{\alpha^l\}$ in (2-7) and $\iota_1(\lambda(\overline{X}))$ in (2-10). We have

$$\iota_1(\lambda(\overline{X})) = \left\{1 \le i \le n \mid \lambda_i(\overline{X}) = \bar{v}_1\right\} = \alpha^1.$$

It follows from Proposition 2.21 that

$$H \in \mathcal{T}^{\rm lin}_{\theta_1}(\overline{X}) \iff \left[\lambda_i'(\overline{X};H) = \lambda_j'(\overline{X};H), \quad \forall\, i,j \in \alpha^1\right].$$

Based on Lemma 2.15, the above right-side is further equivalent to the existence of a scalar $\widehat{\rho}$ such that

$$\overline{U}_{\alpha^1}^{\top} H \overline{U}_{\alpha^1} = \widehat{\rho} I_{|\alpha^1|} \ \text{ for some } \overline{U} \in \mathcal{O}^n(\overline{X}).$$

The value of $\widehat{\rho}$ is independent of the selected orthogonal matrix $\overline{U}$ in $\mathcal{O}^n(\overline{X})$ ([75, Proposition 2]).

The critical cone. Given $(\overline{X},\overline{Y}) \in \operatorname{gph}\partial\lambda_1$ and let $\overline{U} \in \mathcal{O}^n(\overline{X}) \cap \mathcal{O}^n(\overline{Y})$. It follows from Lemma 2.17 and [77, Lemma 2.2] (see also [78, Lemma 3.1]) that

$$\begin{cases} 0 \le \lambda_i(\overline{Y}) \le 1, \quad \forall\, i \in \alpha^1 \quad \text{and} \quad \displaystyle\sum_{i\in\alpha^1} \lambda_i(\overline{Y}) = 1, \\ \lambda_i(\overline{Y}) = 0, \quad \forall\, i \in \alpha^l, \quad l = 2,\ldots,r. \end{cases} \tag{2-28}$$

Denote

$$\mu := \{i \in \alpha^1 \mid \lambda_i(\overline{Y}) > 0\} \quad \text{and} \quad \nu := \{i \in \alpha^1 \mid \lambda_i(\overline{Y}) = 0\}. \tag{2-29}$$

For each $l \in \{1,\ldots,r\}$, we further partition the index set α^l by $\{\beta_k^l\}_{k=1}^{s_l}$ as in (2-20) based on $\lambda(\overline{Y})$. We then obtain from (2-28) that

$$\iota_1(\lambda(\overline{X})) = \alpha^1 = \mu \cup \nu, \quad \mu = \bigcup_{k=1}^{s^1-1} \beta_k^1, \quad \nu = \beta_{s^1}^1 \quad \text{and} \quad \alpha^l = \beta_1^l \quad \text{for} \quad l = 2,\ldots,r.$$

Recall the index set $\eta_1(\lambda(\overline{X}),\lambda(\overline{Y})) \subseteq \iota_1(\lambda(\overline{X}))$ defined in (2-14). Obviously $\eta_1(\lambda(\overline{X}),\lambda(\overline{Y})) = \mu$. It follows from Proposition 2.22 that

$$H \in \mathcal{C}_{\lambda_1}(\overline{X},\overline{Y}) \iff \left[\left(\operatorname{diag}(\overline{U}^{\top} H \overline{U})\right)_i = \lambda_1(\overline{U}_{\alpha^1}^{\top} H \overline{U}_{\alpha^1}), \quad \forall\, i \in \mu\right].$$

One can also derive from Proposition 2.23 that

$$H \in \operatorname{aff}(\mathcal{C}_{\lambda_1}(\overline{X},\overline{Y})) \iff \overline{U}_{\alpha^1}^{\top} H \overline{U}_{\alpha^1} = \begin{bmatrix} \rho I_{|\mu|} & 0 \\ 0 & \overline{U}_{\nu}^{\top} H \overline{U}_{\nu} \end{bmatrix} \quad \text{for some } \rho \in \mathbb{R}. \tag{2-30}$$

<u>**The sigma term**</u>. Also given $(\overline{X},\overline{Y}) \in \operatorname{gph}\partial\lambda_1$ and let $\overline{U} \in \mathcal{O}^n(\overline{X}) \cap \mathcal{O}^n(\overline{Y})$. Denote $\omega := \bigcup_{l=2}^{r} \alpha^l$. By noting that $\lambda_i(\overline{Y}) = 0$ for any $i \in \nu \cup \omega$, we derive from (2-27) and (2-28) that

$$\Upsilon^1_{\overline{X}}(\overline{Y},H) = -2\sum_{i\in\mu}\sum_{j\in\omega}\frac{\lambda_i(\overline{Y})}{\lambda_i(\overline{X})-\lambda_j(\overline{X})}(\overline{U}_{\mu}^{\top} H \overline{U}_{\omega})_{ij}^2 \le 0, \quad H \in \mathbb{S}^n. \tag{2-31}$$

2.3.2 Variational properties of θ_2

In this subsection, we present analogue results with respect to the function θ_2. Recall the definition of the convex piecewise affine function ψ in (1-3). For notational simplicity, we denote $\zeta := \psi \circ \lambda$ as the spectral function associated with ψ. Thus, the function θ_2 can be viewed as the the indictor function of the closed convex set $\mathcal{K}$ that is defined in the following way

$$\mathcal{K} := \left\{X \in \mathbb{S}^n \mid \lambda(X) \in \operatorname{dom}\varpi\right\} = \left\{X \in \mathbb{S}^n \mid \zeta(X) \le 0\right\}. \tag{2-32}$$

Let $\overline{X} \in \mathcal{K}$ be given. Denote $\mathcal{N}_{\mathcal{K}}(\overline{X})$ as the normal cone of $\mathcal{K}$ at $\overline{X} \in \mathbb{S}^n$ in the sense of convex analysis [59]. In the rest of the subsection, we assume the following Slater condition for the closed convex set $\mathcal{K}$.

Assumption 1. There exists $\widetilde{X} \in \mathcal{K}$ such that $\zeta(\widetilde{X}) < 0$.

It is worth mentioning that the above assumption automatically holds for many interesting matrix optimization problems, such as the negative semidefinite programming (where $\mathcal{K}$ is the negative semidefinite matrix cone). Recall the index sets $\{\alpha^l\}_{l=1}^r$ given by (2-7) with respect to $\overline{X}$.

The characterization of the tangent cone and its lineality space of $\mathcal{K}$ is stated below. Let $\overline{X} \in \mathcal{K}$ be such that $\zeta(\overline{X}) = 0$. It follows from [75, Theorem 3] and the directional differentiability of ψ that ζ is also directionally differentiable. Since $\zeta : \mathbb{S}^n \to \mathbb{R}$ is a closed convex function, it follows from [5, Proposition 2.61; page 42, the paragraph under (2.70)] that the tangent cone $\mathcal{T}_{\mathcal{K}}(\overline{X})$ of the closed convex set $\mathcal{K}$ is given by

$$\mathcal{T}_{\mathcal{K}}(\overline{X}) := \left\{H \in \mathbb{S}^n \mid \zeta'(\overline{X};H) \le 0\right\}.$$

As shown in [4, (30)], it can be re-written as

$$\mathcal{T}_{\mathcal{K}}(\overline{X}) = \left\{H \in \mathbb{S}^n \mid \langle b^i, \lambda'(\overline{X};H)\rangle \le 0, \ \forall\, i \in \iota_2(\lambda(\overline{X}))\right\}. \tag{2-33}$$

Moreover, the corresponding lineality space $\mathrm{lin}(\mathcal{T}_{\mathcal{K}}(\overline{X}))$ of $\mathcal{T}_{\mathcal{K}}(\overline{X})$ is given by

$$\mathrm{lin}(\mathcal{T}_{\mathcal{K}}(\overline{X})) \quad = \quad \left\{H \in \mathbb{S}^n \mid \zeta'(\overline{X}; H) = -\zeta'(\overline{X}; -H) = 0\right\}. \tag{2-34}$$

Define the index set

$$\widetilde{\mathcal{F}} := \{l \in \{1, \ldots, r\} \mid \exists\, i, j \in \alpha^l \text{ such that } (b^w)_i \neq (b^w)_j \text{ for some } w \in \iota_2(\lambda(\overline{X}))\}.$$

Then, we obtain the following characterization of $\mathrm{lin}(\mathcal{T}_{\mathcal{K}}(\overline{X}))$ based on Corollary 2.19.

Proposition 2.25. *Let $H \in \mathbb{S}^n$. Then $H \in \mathrm{lin}(\mathcal{T}_{\mathcal{K}}(\overline{X}))$ implies the existence of scalars $\{\widetilde{\rho}_l\}_{l \in \widetilde{F}}$ such that for any $\overline{V} \in \mathcal{O}^n(\overline{X})$,*

$$\overline{V}_{\alpha^l}^{\top} H \overline{V}_{\alpha^l} = \widetilde{\rho}_l I_{|\alpha^l|}.$$

In fact,

$$H \in \mathrm{lin}(\mathcal{T}_{\mathcal{K}}(\overline{X})) \iff \left[\langle \lambda'(\overline{X}; H), b^i \rangle = 0, \quad \forall\, i \in \iota_2(\lambda(\overline{X}))\right].$$

As for the critical cone, let $\overline{X} \in \mathcal{K}$ and $\overline{Z} \in \mathcal{N}_{\mathcal{K}}(\overline{X})$. The critical cone of $\mathcal{K}$ at $\overline{X}$ for $\overline{Z}$ is defined by

$$\mathcal{C}_{\mathcal{K}}(\overline{X}, \overline{Z}) := \mathcal{T}_{\mathcal{K}}(\overline{X}) \cap \overline{Z}^{\perp} = \left\{H \in \mathbb{S}^n \mid \zeta'(\overline{X}; H) \le 0,\ \langle \overline{Z}, H \rangle = 0\right\}. \tag{2-35}$$

For each $l \in \{1, \ldots, r\}$, similar to the definition of the index sets β_k^l in (2-20), we use the notation $\{\gamma_k^l\}_{k=1}^{t_l}$ to further partition the set α^l based on the eigenvalue of $\overline{Z}$ as

$$\begin{cases} \lambda_i(\overline{Z}) = \lambda_j(\overline{Z}) & \text{if } i, j \in \gamma_k^l \text{ and } k \in \{1, \ldots, t_l\}, \\ \lambda_i(\overline{Z}) > \lambda_j(\overline{Z}) & \text{if } i \in \gamma_k^l,\ j \in \gamma_{k'}^l \text{ and } k, k' \in \{1, \ldots, t_l\} \text{ with } k < k'. \end{cases} \tag{2-36}$$

Recall the index set $\eta_2(\lambda(\overline{X}), \lambda(\overline{Z}))$ defined in (2-15) with respect to $\lambda(\overline{X})$ and $\lambda(\overline{Z})$. For each $l \in \{1, \ldots, r\}$, define the index set

$$\mathcal{F}_l := \{k \in \{1, \ldots, t_l\} \mid \exists\, i, j \in \gamma_k^l \text{ such that } (b^w)_i \neq (b^w)_j \text{ for some } w \in \eta_2(\lambda(\overline{X}), \lambda(\overline{Z}))\}. \tag{2-37}$$

The following result on the characterization of $\mathcal{C}_{\mathcal{K}}(\overline{X}, \overline{Z})$ can be obtained similarly as Proposition 2.22.

Proposition 2.26. *Suppose that $(\overline{X}, \overline{Z}) \in \mathrm{gph}\, \mathcal{N}_{\mathcal{K}}$ and $\overline{V} \in \mathcal{O}^n(\overline{X}) \cap \mathcal{O}^n(\overline{Z})$. If $H \in \mathcal{C}_{\mathcal{K}}(\overline{X}, \overline{Z})$, then the following three conditions hold:*

(i) for each $l \in \{1, \ldots, r\}$, $\overline{V}_{\alpha^l}^{\top} H \overline{V}_{\alpha^l}$ has the following block diagonal structure, i.e.,

$$\overline{V}_{\alpha^l}^{\top} H \overline{V}_{\alpha^l} = \mathrm{Diag}\left((\overline{V}_{\alpha^l}^{\top} H \overline{V}_{\alpha^l})_{\gamma_1^l \gamma_1^l}, \cdots, (\overline{V}_{\alpha^l}^{\top} H \overline{V}_{\alpha^l})_{\gamma_{t_l}^l \gamma_{t_l}^l}\right);$$

(ii) $\langle \lambda'(\overline{X}; H), b^i \rangle = \max_{j \in \iota_2(\lambda(\overline{X}))} \langle \lambda'(\overline{X}; H), b^j \rangle = 0, \quad \forall\, i \in \eta_2(\lambda(\overline{X}), \lambda(\overline{Y}))$;

(iii) for each $l \in \{1,\ldots,r\}$ and $k \in \mathcal{F}_l$, there exists a scalar $\rho_k^l \in \mathbb{R}$ such that $(\overline{V}_{\alpha^l}^\top H\overline{V}_{\alpha^l})_{\gamma_k^l\gamma_k^l} = \rho_k^l I_{|\gamma_k^l|}$.

In fact, $H \in \mathcal{C}_{\mathcal{K}}(\overline{X},\overline{Z})$ if and only if for any $i \in \eta_2(\lambda(\overline{X}),\lambda(\overline{Z}))$,

$$\langle \operatorname{diag}(\overline{V}^\top H\overline{V}), b^i\rangle = \max_{j\in\iota_2(\lambda(\overline{X}))} \langle \lambda'(\overline{X};H), b^j\rangle = 0,$$

The results below on the characterization of the affine hull of the critical cone $\mathcal{C}_{\mathcal{K}}(\overline{X},\overline{Z})$ follows from Proposition 2.26.

Proposition 2.27. *Suppose that $(\overline{X},\overline{Z}) \in \operatorname{gph}\mathcal{N}_{\mathcal{K}}$. Let $\overline{V} \in \mathcal{O}^n(\overline{X}) \cap \mathcal{O}^n(\overline{Z})$. Then $H \in \operatorname{aff}(\mathcal{C}_{\mathcal{K}}(\overline{X},\overline{Z}))$ if and only if it satisfies the properties (i) and (iii) in Proposition 2.26, and for any $i \in \eta_2(\lambda(\overline{X}),\lambda(\overline{Z}))$,*

$$\langle \operatorname{diag}(\overline{V}^\top H\overline{V}), b^i\rangle = 0.$$

When it comes to the sigma term, suppose that $(\overline{X},\overline{Z}) \in \operatorname{gph}\mathcal{N}_{\mathcal{K}}$. Let $H \in \mathcal{C}_{\mathcal{K}}(\overline{X},\overline{Z})$ be arbitrarily given. As in the conventional conic programming, the sigma term associated with $\mathcal{K}$ is defined as the support function of its second-order tangent set $\mathcal{T}_{\mathcal{K}}^2(\overline{X},H)$ [5, Definition 3.28], whose explicit expression is given in the following proposition.

Proposition 2.28. *Suppose that $(\overline{X},\overline{Z}) \in \operatorname{gph}\mathcal{N}_{\mathcal{K}}$ and $\overline{V} \in \mathcal{O}^n(\overline{X}) \cap \mathcal{O}^n(\overline{Z})$. Let $H \in \mathcal{C}_{\mathcal{K}}(\overline{X},\overline{Z})$ be given. Then the support function of $\mathcal{T}_{\mathcal{K}}^2(\overline{X},H)$ at $\overline{Z}$ takes the following form*

$$\delta^*_{\mathcal{T}_{\mathcal{K}}^2(\overline{X},H)}(\overline{Z}) = 2\sum_{l=1}^{r}\langle \Lambda(\overline{Z})_{\alpha^l\alpha^l}, \overline{V}_{\alpha^l}^\top H(\overline{X}-\bar{v}_l I)^\dagger H\overline{V}_{\alpha^l}\rangle. \tag{2-38}$$

Similarly as that for θ_1, for any given $\overline{X} \in \mathbb{S}^n$, define the function $\Upsilon_{\overline{X}}^2 : \mathcal{N}_{\mathcal{K}}(\overline{X}) \times \mathbb{S}^n \to \mathbb{R}$ as the value of the right side of (2-38), i.e.,

$$\Upsilon_{\overline{X}}^2\left(\overline{Z},H\right) := 2\sum_{l=1}^{r}\langle \Lambda(\overline{Z})_{\alpha^l\alpha^l}, \overline{V}_{\alpha^l}^\top H(\overline{X}-\bar{v}_l I)^\dagger H\overline{V}_{\alpha^l}\rangle, \quad \overline{Z} \in \mathcal{N}_{\mathcal{K}}(\overline{X}) \quad \text{and} \quad H \in \mathbb{S}^n, \tag{2-39}$$

where $\overline{V} \in \mathcal{O}^n(\overline{X})$. If $\overline{Z} \in \mathcal{N}_{\mathcal{K}}(\overline{X})$, then for $\overline{V} \in \mathcal{O}^n(\overline{X}) \cap \mathcal{O}^n(\overline{Z})$,

$$\Upsilon_{\overline{X}}^2\left(\overline{Z},H\right) = -2\sum_{1\le l<l'\le r}\sum_{i\in\alpha^l}\sum_{j\in\alpha^{l'}} \frac{\lambda_i(\overline{Z})-\lambda_j(\overline{Z})}{\lambda_i(\overline{X})-\lambda_j(\overline{X})}(\overline{V}_{\alpha^l}^\top H\overline{V}_{\alpha^{l'}})_{ij}^2. \tag{2-40}$$

Example 2.2. Consider the following problem in Example 1.2. This problem is a special case of problem (1-1) with $\varpi(x) = \delta_{\mathbb{R}^n_+}(x)$ and $\operatorname{dom}\varpi = \left\{x \in \mathbb{R}^n \mid \max_{1\le i\le p}\langle -e^i, x\rangle \le 0\right\}$, where e^i is the unit vector whose i-th component is 1 and others are zero. One may verify that the explicit forms of the tangent cone, the critical cone and the sigma term using the formulas derived in this section are consistent with the results derived in [11, equation (17)-(20) and Lemma 3.1].

2.3.3 A particular case of CMatOP: NLSDP

As a particular case of CMatOP, nonlinear semidefinite programming (NLSDP) is studied more maturely and possesses fruitful results. NLSDP takes the following form

$$\begin{aligned} \min_{x\in\mathbb{X}} \quad & f(x) \\ \text{s.t.} \quad & G(x) \in \mathbb{S}^n_+, \\ & h(x) = 0 \end{aligned} \tag{2-41}$$

Consider the eigenvalue decomposition of $A \in \mathbb{S}^n$, i.e., $A = P\Lambda(A)P^T$, where $P \in \mathcal{O}^n(A)$ is a corresponding orthogonal matrix of the orthonormal eigenvectors. By considering the index sets of positive, zero, and negative eigenvalues of A, we are able to write A in the following form

$$A = \begin{bmatrix} P_\alpha & P_\beta & P_\gamma \end{bmatrix} \begin{bmatrix} \Lambda(A)_{\alpha\alpha} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Lambda(A)_{\gamma\gamma} \end{bmatrix} \begin{bmatrix} P_\alpha^T \\ P_\beta^T \\ P_\gamma^T \end{bmatrix}. \tag{2-42}$$

where $\alpha := \{i : \lambda_i(A) > 0\}$, $\beta := \{i : \lambda_i(A) = 0\}$ and $\gamma := \{i : \lambda_i(A) < 0\}$. It is known from [5, Example 3.140] that the n-dimensional positive semidefinite cone $\mathbb{S}^n_+$ is C^∞-cone reducible. We use $\Pi_{\mathbb{S}^n_+}(A)$ to represent the projection from A to $\mathbb{S}^n_+$ and it follows from [11, p.5] that

$$\Pi_{\mathbb{S}^n_+}(A) = \begin{bmatrix} P_\alpha & P_\beta & P_\gamma \end{bmatrix} \begin{bmatrix} \Lambda(A)_{\alpha\alpha} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} P_\alpha^T \\ P_\beta^T \\ P_\gamma^T \end{bmatrix}.$$

From [79, Theorem 4.7] we know that the metric projection operator $\Pi_{\mathbb{S}^n_+}(\cdot)$ is directionally differentiable at any $A \in \mathbb{S}^n$ and the directional derivative of $\Pi_{\mathbb{S}^n_+}(\cdot)$ at A along direction $H \in \mathbb{S}^n$ is given by

$$\Pi'_{\mathbb{S}^n_+}(A; H) = P \begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & \Sigma_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} \\ \widetilde{H}_{\alpha\beta}^T & \Pi_{\mathbb{S}^{|\beta|}_+}(\widetilde{H}_{\beta\beta}) & 0 \\ \Sigma_{\alpha\gamma}^T \circ \widetilde{H}_{\alpha\gamma}^T & 0 & 0 \end{bmatrix} P^T, \tag{2-43}$$

where $\widetilde{H} := P^T H P$, "$\circ$" is the Hadamard product and

$$\Sigma_{ij} := \frac{\max\{\lambda_i(A), 0\} - \max\{\lambda_j(A), 0\}}{\lambda_i(A) - \lambda_j(A)}, \quad i, j = 1, \ldots, n, \tag{2-44}$$

where $0/0$ is defined to be 1. Moreover, we have

$$\Pi'_{\mathbb{S}^n_-}(A; H) = H - \Pi'_{\mathbb{S}^n_+}(A; H) = P \begin{bmatrix} 0 & 0 & (E - \Sigma)_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} \\ 0 & \Pi_{\mathbb{S}^{|\beta|}_-}(\widetilde{H}_{\beta\beta}) & \widetilde{H}_{\beta\gamma} \\ (E - \Sigma)_{\alpha\gamma}^T \circ \widetilde{H}_{\alpha\gamma}^T & \widetilde{H}_{\beta\gamma}^T & \widetilde{H}_{\gamma\gamma} \end{bmatrix} P^T. \tag{2-45}$$

The tangent cone of SDP cone at $X \in \mathbb{S}^n_+$ is

$$T_{\mathbb{S}^n_+}(X) = \{H \in \mathbb{S}^n \mid H = \Pi'_{\mathbb{S}^n_+}(X; H)\} = \{H \in \mathbb{S}^n \mid P^T_{\alpha^c} H P_{\alpha^c} \in \mathbb{S}^{|\alpha^c|}_+\},$$

where $\alpha^c = \{1, \cdots, n\} \backslash \alpha$. The lineality space of $T_{\mathbb{S}^n_+}(X)$, i.e., the largest linear space in $T_{\mathbb{S}^n_+}(X)$, denoted by $\mathrm{lin}\big(T_{\mathbb{S}^n_+}(X)\big)$, takes the following form:

$$\mathrm{lin}\big(T_{\mathbb{S}^n_+}(X)\big) = \{H \in \mathbb{S}^n \mid P^T_{\alpha^c} H P_{\alpha^c} = 0\}.$$

The critical cone of $\mathbb{S}^n_+$ at $Y \in \mathcal{N}_{\mathbb{S}^n_+}(X)$ with $A = X + Y$ is given by (cf. [12, equations (11) and (12)])

$$\mathcal{C}_{\mathbb{S}^n_+}(X, Y) = \{H \in \mathbb{S}^n \mid P^T_\beta H P_\beta \in \mathbb{S}^{|\beta|}_+, \ P^T_\beta H P_\gamma = 0, \ P^T_\gamma H P_\gamma = 0\} \tag{2-46}$$

and the affine hull of $\mathcal{C}_{\mathbb{S}^n_+}(X, Y)$ is

$$\mathrm{aff}\, \mathcal{C}_{\mathbb{S}^n_+}(X, Y) = \{H \in \mathbb{S}^n \mid P^T_\beta H P_\gamma = 0, \ P^T_\gamma H P_\gamma = 0\}. \tag{2-47}$$

For more details on the properties of SDP cone, we refer the reader to [11, 50, 80].

Without loss of generality, we can omit the equality constraint, then the NLSDP (2-41) can be written in the following form

$$\min \qquad f(x) + \delta_{\mathbb{S}^n_+}(G(x)), \tag{2-48}$$

where $\delta_{\mathbb{S}^n_+} : \mathbb{S}^n \to (-\infty, \infty]$ is the indicator function of positive semidefinite cone $\mathbb{S}^n_+$. Denote $S_{KKT}(u, v)$ the solution set of the KKT optimality condition for problem (2-48), i.e.,

$$S_{KKT}(u, v) = \left\{ (x, Y) \in \mathbb{X} \times \mathbb{S}^n \;\middle|\; \begin{array}{l} L'_x(x, Y) - u = 0, \\ \mathbb{S}^n_+ \ni (G(x) - v) \perp Y \in \mathbb{S}^n_-. \end{array} \right\}. \tag{2-49}$$

For any KKT pair $(\bar{x}, \overline{Y})$ that satisfies the KKT condition with $(u, v) = (0, 0)$, we call $\bar{x}$ a stationary point. Suppose $\bar{x}$ is a stationary point. Define $\mathcal{M}(\bar{x})$ as the set of all multipliers $Y \in \mathbb{S}^n$ satisfying the KKT condition (3-91), i.e.,

$$\mathcal{M}(\bar{x}) = \{Y \in \mathbb{S}^n \mid (\bar{x}, Y) \in S_{KKT}(0, 0)\}. \tag{2-50}$$

The definition of strong SOSC for NLSDP, which demands the supremum of (2-51) over $Y \in \mathcal{M}(\bar{x})$ holds, is originally given by [11] as an analogue for Robinson [13] for the nonlinear programming (NLP). However, the strong SOSC mentioned here is slightly different in only requiring the validity of (2-51) at $\overline{Y}$. It is worth to note when $\mathcal{M}(\bar{x})$ is a singleton, both of them are the same.

Definition 2.29. *[11, Definition 3.2] Let $\bar{x}$ be a stationary point of NLSDP* (2-48) *and $\overline{Y} \in \mathcal{M}(\bar{x})$. We say the strong second order sufficient condition (SOSC) holds at $(\bar{x}, \overline{Y})$ if*

$$\langle L''_{xx}(\bar{x}, \overline{Y})d, d\rangle - \Upsilon_{G(\bar{x})}\big(\overline{Y}, G'(\bar{x})d\big) > 0 \quad \forall\, 0 \neq G'(\bar{x})d \in \operatorname{aff} \mathcal{C}_{\mathbb{S}^n_+}(G(\bar{x}), \overline{Y}). \tag{2-51}$$

where $\Upsilon_{G(\bar{x})}\big(\overline{Y}, G'(\bar{x})d\big) = 2\langle \overline{Y}, (G'(\bar{x})d)G(\bar{x})^{\dagger}(G'(\bar{x})d)\rangle$ is the σ-term and $G(\bar{x})^{\dagger}$ is the generalized inverse matrix of $G(\bar{x})$.

Suppose $\overline{A} = G(\bar{x}) + \overline{Y}$ possesses the decomposition (2-42) with $\overline{P} \in \mathcal{O}^n(\overline{A})$. From [12, page 386], we know the σ-term takes the explicit form of

$$\Upsilon_{G(\bar{x})}\big(\overline{Y}, H\big) = 2 \sum_{i\in\alpha, j\in\gamma} \frac{\lambda_j(\overline{A})}{\lambda_i(\overline{A})}(\widetilde{H}_{ij})^2, \tag{2-52}$$

where $\widetilde{H} = \overline{P}^T H \overline{P}$.

Next, we introduce some essential notation for our main result. The (second order) generalized differentiability of the augmented Lagrangian function $\mathcal{L}_\rho$ defined by (1-12) for NLSDP (2-48) has been explicitly established in [50]. In fact, it is well-known that $\mathcal{L}_\rho$ is continuously differentiable with

$$(\mathcal{L}_\rho)'_x(x, Y) = f'(x) + \rho G'(x)^* \Pi_{\mathbb{S}^n_-}(G(x) + \rho^{-1}Y), \tag{2-53}$$

where $\Pi_{\mathbb{S}^n_-}(\cdot)$ is the metric projection over $\mathbb{S}^n_-$ and $G'(x)^*$ denotes the adjoint of the corresponding linear mapping. Moreover, since $(\mathcal{L}_\rho)'_x$ is Lipschitz continuous, by using Rademacher's theorem, the B(ouligand)-subdifferential of $(\mathcal{L}_\rho)'_x$ at (x, Y) is given by

$$\partial_B\big((\mathcal{L}_\rho)'_x\big)(x, Y) := \{\lim_{k\to\infty}((\mathcal{L}_\rho)'_x)'(x^k, Y^k) \mid (x^k, Y^k) \in \mathcal{U},\ (x^k, Y^k) \to (x, Y)\}, \tag{2-54}$$

where $\mathcal{U}$ is the set of Fréchet-differentiable points of $(\mathcal{L}_\rho)'_x$. It follows from [50, (18)] (using [50, Lemma 2 and 3]) that

$$\begin{aligned}&\pi_x \partial_B\big((\mathcal{L}_\rho)'_x\big)(x, Y)(\Delta x)\\ &= L''_{xx}(x, \rho\Pi_{\mathbb{S}^n_-}(G(x) + \rho^{-1}Y))(\Delta x) + \rho G'(x)^* \partial_B \Pi_{\mathbb{S}^n_-}(G(x) + \rho^{-1}Y)G'(x)(\Delta x),\end{aligned}$$

where $\pi_x \partial_B\big((\mathcal{L}_\rho)'_x\big)(x, Y)$ is the projection of $\partial_B\big((\mathcal{L}_\rho)'_x\big)(x, Y)$ onto the space $\mathbb{X}$ and $\partial_B \Pi_{\mathbb{S}^n_-}(G(x) + \rho^{-1}Y)$ is the B-subdifferential of $\Pi_{\mathbb{S}^n_-}(\cdot)$ at $G(x) + \rho^{-1}Y$. Actually, the set $\pi_x \partial_B\big((\mathcal{L}_\rho)'_x\big)(x, Y)$ is recently defined as the x part of the Hessian bundle [55, (3.1)] for the augmented Lagrangian function $\mathcal{L}_\rho$ at (x, Y) (see [55, (3.6)] for detail).

Let $\bar{x}$ be the stationary point of (2-48). For each $\overline{Y} \in \mathcal{M}(\bar{x})$ and $W \in \partial_B \Pi_{\mathbb{S}^n_-}(G(\bar{x}) + \rho^{-1}\overline{Y})$, define the following mapping

$$\mathcal{A}_\rho(\overline{Y}, W) := L''_{xx}(\bar{x}, \overline{Y}) + \rho G'(\bar{x})^* W G'(\bar{x}). \tag{2-55}$$

Let $\overline{A} = G(\bar{x}) + \overline{Y}$ possesses decomposition (2-42) with $\overline{P} \in \mathcal{O}^n(\overline{A})$. It follows from [50, Lemma 5] that $W \in \partial_B \Pi_{\mathbb{S}^n_-}(\overline{A})$ if and only if there exists $W_0 \in \partial_B \Pi_{\mathbb{S}^{|\beta|}_-}(0)$ such that for all $H \in \mathbb{S}^n$

$$W(H) = \overline{P} \begin{bmatrix} 0 & 0 & \Sigma_{\alpha\gamma} \circ \overline{P}_\alpha^T H \overline{P}_\gamma \\ 0 & W_0(\overline{P}_\beta^T H \overline{P}_\beta) & \overline{P}_\beta^T H \overline{P}_\gamma \\ \Sigma_{\gamma\alpha} \circ \overline{P}_\gamma^T H \overline{P}_\alpha & \overline{P}_\gamma^T H \overline{P}_\beta & \overline{P}_\gamma^T H \overline{P}_\gamma \end{bmatrix} \overline{P}^T,$$

where

$$\begin{cases} \Sigma_{ij} = 1 - \frac{\max\{\lambda_i, 0\} + \max\{\lambda_j, 0\}}{|\lambda_i| + |\lambda_j|}, & (i,j) \notin \beta \times \beta \\ \Sigma_{ij} \in [0,1] & (i,j) \in \beta \times \beta \end{cases}$$

with $\lambda_i := \lambda_i(\overline{A})$ for short. Let $\mathcal{O}^n$ denote the set of n-dimensional orthogonal matrix. Also, $W_0 \in \partial_B \Pi_{\mathbb{S}^{|\beta|}_-}(0)$ if and only if there exist $Q \in \mathcal{O}^{|\beta|}$ and $\Omega \in \mathbb{S}^{|\beta|}$ with entries $\Omega_{ij} \in [0,1]$ such that for all $H \in \mathbb{S}^{|\beta|}$,

$$W_0(H) = Q(\Omega \circ (Q^T H Q))Q^T.$$

By [50, Lemma 9], we have

$$\begin{aligned} \langle d, \mathcal{A}_\rho(\overline{Y}, W)d \rangle =& \langle d, L''_{xx}(\bar{x}, \overline{Y})d \rangle + \rho \sum_{i,j \in \gamma} (\overline{P}^T (G'(\bar{x})d)\overline{P})^2_{ij} + 2\rho \sum_{i \in \beta, j \in \gamma} (\overline{P}^T (G'(\bar{x})d)\overline{P})^2_{ij} \\ &+ 2\rho \sum_{i \in \alpha, j \in \gamma} \frac{-\lambda_j}{\rho\lambda_i - \lambda_j} (\overline{P}^T (G'(\bar{x})d)\overline{P})^2_{ij} + \rho \sum_{i,j \in \beta} (\Omega_\rho)_{ij} (P^T (G'(\bar{x})d)P)^2_{ij}, \end{aligned} \tag{2-56}$$

where $\Omega_\rho \in \mathbb{S}^{|\beta|}$ with entries $(\Omega_\rho)_{ij} \in [0,1]$, $P = [\overline{P}_\alpha \; \overline{P}_\beta Q \; \overline{P}_\gamma]$ with $Q \in \mathcal{O}^{|\beta|}$.

Chapter 3 Perturbation analysis of composite matrix optimization problem

In this chapter, we mainly study several perturbation properties of CMatOP (1-6): Lischitzian full stability, isolated calmness and semi-isolated calmness. All of them are of great importance in sensitivity analysis as they can deeply characterize the difficulty of a problem. On the other hand, in the study of convergence analysis of ALM, a lot of early works usually involve both strong second order sufficient condition and the constraint nondegenerate condition. To understand why these two conditions play such an important role, it is necessary to discuss the relationships of the two conditions with the aforementioned perturbation properties. Although some results may have been established for certain kinds of CMatOP, e.g., NLSDP, several years ago, we believe it is necessary to develop a uniform approach for general CMatOP as it includes so many important practical cases. Moreover, the general CMatOP may lose some very good properties that NLSDP possesses. Thus, how to characterize the perturbation properties of general CMatOP is demanding.

Our main contribution mainly lies in the following two points. Firstly, we characterize the explicit form of the corresponding coderivative for a class of non-polyhedral set and furthermore characterizing Lipschitzian full stability of (1-6) under nondegeneracy condition. To the best of our knowledge, this is the first time explicit expressions for coderivatives of CMatOP are presented. Secondly, we give the equivalence between the strong second order sufficient condition with constraint nondegeneracy and strong regularity under checkable conditions for CMatOP, which extends the result attained in [4] as they only obtain one direction. Similar results for NLSDP can be seen in [11] and [16]. Moreover, ignoring the basic parametric perturbation p provides a complete characterization of tilt stability.

As illustrated in Theorem 2.13, the Lipschitz full stability is equivalent to the positive definiteness of the coderivative. Thus our first step is to provide an explicit characterization of the coderivatives $\mathcal{D}^*\partial\theta_1(g_1(\bar{x},\bar{p}),\overline{Y})(\cdot)$ of $\partial\theta_1$ and $\mathcal{D}^*\mathcal{N}_{\mathcal{K}}(g_2(\bar{x},\bar{p}),\overline{Z})(\cdot)$ of $\mathcal{N}_{\mathcal{K}}$.

3.1 Characterizations of coderivatives

For any given $A \in \mathbb{S}^n$, suppose that A has the following eigenvalue decomposition

$$A = P\,\mathrm{Diag}(\lambda_1(A),\ldots,\lambda_n(A))\,P^T, \tag{3-1}$$

recalling the corresponding notations in the paragraph under (2-6). In later discussions, when the dependence of λ and v_i, $i = 1,\ldots,r$, on A can be seen clearly from the context, we often drop A from these notations. From Definition 2.5 and 2.1, to characterize the coderivatives, we begin with the characterizations of proximal normal cones.

Firstly, for a convex function f, we have the following result on the characterization of the proximal normal cone of $\mathrm{gph}\,\partial f$, which is a generalization of the corresponding result [80, Proposition 3.1] for SDP. The proof is similar to that of [80, Proposition 3.1].

Proposition 3.1. *Given Euclidean space $\mathbb{X}$. Suppose $f:\mathbb{X}\to\mathbb{R}$ is a convex function and its proximal operator $\mathrm{Pr}_f(\cdot)$ is calmly B-differentiable for any given $\overline{X}\in\mathbb{X}$. $(X^*,Y^*)\in\mathcal{N}^{\pi}_{\mathrm{gph}\,\partial f}(\overline{X},\overline{Y})$ if and only if $(X^*,Y^*)\in\mathbb{X}\times\mathbb{X}$ satisfies*

$$\langle X^*,\mathrm{Pr}_f'(\overline{X}+\overline{Y};H)\rangle+\langle Y^*,\mathrm{Pr}_{f^*}'(\overline{X}+\overline{Y};H)\rangle\le 0\quad\forall\, H\in\mathbb{X}.$$

Proof. Since Pr_f is calmly B-differentiable, we know that for any given $A^*\in\mathbb{X}$ and any fixed $\overline{A}\in\mathbb{X}$ there exists $m_1,m_2>0$ (depending on $\overline{A}$ and A^* only) such that for any $A\in\mathbb{X}$ sufficiently close to $\overline{A}$,

$$\langle A^*,\mathrm{Pr}_f(A)-\mathrm{Pr}_f(\overline{A})\rangle\le\langle A^*,\mathrm{Pr}_f'(\overline{A};A-\overline{A})\rangle+m_1\|A-\overline{A}\|^2,\tag{3-2}$$

$$\langle A^*,\mathrm{Pr}_{f^*}(A)-\mathrm{Pr}_{f^*}(\overline{A})\rangle\le\langle A^*,\mathrm{Pr}_{f^*}'(\overline{A};A-\overline{A})\rangle+m_2\|A-\overline{A}\|^2.\tag{3-3}$$

"$\Longleftarrow$": Suppose that $(X^*,Y^*)\in\mathbb{X}\times\mathbb{X}$ is given and satisfies the inequality. By Lemma 2.8, (3-2) and (3-3), we know that there exist a constant $c>0$ and a constant $\widetilde{m}>0$ such that for any $(X,Y)\in\mathrm{gph}\,\partial f$ and $\|(X,Y)-(\overline{X},\overline{Y})\|\le c$,

$$\begin{aligned}&\langle(X^*,Y^*),(X,Y)-(\overline{X},\overline{Y})\rangle\\&=\langle(X^*,Y^*),(\mathrm{Pr}_f(X+Y),\mathrm{Pr}_{f^*}(X+Y))-(\mathrm{Pr}_f(\overline{X}+\overline{Y}),\mathrm{Pr}_{f^*}(\overline{X}+\overline{Y}))\rangle\\&=\langle X^*,\mathrm{Pr}_f(X+Y)-\mathrm{Pr}_f(\overline{X}+\overline{Y})\rangle+\langle Y^*,\mathrm{Pr}_{f^*}(X+Y)-\mathrm{Pr}_{f^*}(\overline{X}+\overline{Y})\rangle\\&\le\widetilde{m}\|(X,Y)-(\overline{X},\overline{Y})\|^2.\end{aligned}$$

By taking $m=\max\{\widetilde{m},\frac{\|(X^*,Y^*)\|}{c}\}$, we know that for any $(X,Y)\in\mathrm{gph}\,\partial f$,

$$\langle(X^*,Y^*),(X,Y)-(\overline{X},\overline{Y})\rangle\le m\|(X,Y)-(\overline{X},\overline{Y})\|^2,\tag{3-4}$$

which implies, by the definition of the proximal normal cone, that $(X^*,Y^*)\in\mathcal{N}^{\pi}_{\mathrm{gph}\,\partial f}(\overline{X},\overline{Y})$.
"$\Longrightarrow$": Let $(X^*,Y^*)\in\mathcal{N}^{\pi}_{\mathrm{gph}\,\partial f}(\overline{X},\overline{Y})$ be given. Then there exists $m>0$ such that for any $(X,Y)\in\mathrm{gph}\,\partial f$,

$$\langle(X^*,Y^*),(X,Y)-(\overline{X},\overline{Y})\rangle\le m\|(X,Y)-(\overline{X},\overline{Y})\|^2.\tag{3-5}$$

Let $H\in\mathbb{X}$ be arbitrary but fixed. For any $t\downarrow 0$, let

$$X_t=\mathrm{Pr}_f(\overline{X}+\overline{Y}+tH)\quad\text{and}\quad Y_t=\mathrm{Pr}_{f^*}(\overline{X}+\overline{Y}+tH).$$

By noting that $(X_t, Y_t) \in \operatorname{gph} \partial f$ and $\operatorname{Pr}_f(\cdot)$, $\operatorname{Pr}_{f^*}(\cdot)$ are globally Lipschitz continuous with modulus 1, we obtain from (3-5) that

$$
\begin{aligned}
&\langle X^*, \operatorname{Pr}_f'(\overline{X}+\overline{Y}; H)\rangle + \langle Y^*, \operatorname{Pr}_{f^*}'(\overline{X}+\overline{Y}; H)\rangle \\
&\le m \lim_{t\downarrow 0} \frac{1}{t}(\|X_t - \overline{X}\|^2 + \|Y_t - \overline{Y}\|^2) \le m \lim_{t\downarrow 0} \frac{1}{t}(2t^2\|H\|^2) = 0.
\end{aligned}
$$

Hence, the proof is completed. □

It is easy to see that this characterization can be applied for θ_1 and $\delta_{\mathcal{K}}$.

3.1.1 Characterizations of proximal normal cones

Recalling the definition of critical cone Definition 2.6 and $\theta_1 = \varpi_1 \circ \lambda$, $\delta_{\mathcal{K}} = \delta_{\widehat{\mathcal{K}}} \circ \lambda$ in (1-6).

Proximal normal cone for $\operatorname{gph} \partial\theta_1$: Given $(\overline{X}_1, \overline{Y}) \in \operatorname{gph} \partial\theta_1$, let $\overline{A} = \overline{X}_1 + \overline{Y}$. Denote $\overline{\lambda} = \lambda(\overline{A})$ and $\overline{v}_i = v_i(\overline{A})$, $i = 1, \ldots, r_1$. Define

$$
\alpha^l := \{1 \le i \le m \mid \overline{\lambda}_i = \overline{v}_l\}, \quad l = 1, \ldots, r_1. \tag{3-6}
$$

Assume $\overline{A}$ has the following eigenvalue decomposition:

$$
\overline{A} = \begin{bmatrix} \overline{P}_{\alpha^1} & \cdots & \overline{P}_{\alpha^{r_1}} \end{bmatrix} \begin{bmatrix} \Lambda(\overline{A})_{\alpha^1\alpha^1} & & 0 \\ & \ddots & \\ 0 & & \Lambda(\overline{A})_{\alpha^{r_1}\alpha^{r_1}} \end{bmatrix} \begin{bmatrix} \overline{P}_{\alpha^1}^T \\ \vdots \\ \overline{P}_{\alpha^{r_1}}^T \end{bmatrix}. \tag{3-7}
$$

Denote $\mathbb{W}_1 = \mathbb{S}^{|\alpha^1|} \times \cdots \times \mathbb{S}^{|\alpha^{r_1}|}$. Define $\mathscr{D} : \mathbb{S}^m \to \mathbb{W}_1$ by $\mathscr{D}(H) = (H_{\alpha^1\alpha^1}, \ldots, H_{\alpha^{r_1}\alpha^{r_1}})$. Let $\kappa(W) := (\lambda(W_1), \ldots, \lambda(W_{r_1}))$ for all $W = (W_1, \ldots, W_{r_1}) \in \mathbb{W}_1$. By employing [63, Theorem 6] and the Moreau decomposition (see e.g., [59, Theorem 31.5]), we have following explicit formulas of $\operatorname{Pr}_{\theta_1}'(\overline{A}; \cdot)$ and $\operatorname{Pr}_{\theta_1^*}'(\overline{A}; \cdot)$ at $\overline{A}$, i.e., for any direction $H \in \mathbb{S}^m$,

$$
\operatorname{Pr}_{\theta_1}'(\overline{A}; H) = \overline{P}\left[\overline{D} \circ \widehat{H} + \operatorname{Diag}\big(\Psi_1(\mathscr{D}(\widehat{H})), \ldots, \Psi_{r_1}(\mathscr{D}(\widehat{H}))\big)\right]\overline{P}^T \tag{3-8}
$$

and

$$
\operatorname{Pr}_{\theta_1^*}'(\overline{A}; H) = H - \operatorname{Pr}_{\theta_1}'(\overline{A}; H),
$$

where $\overline{P} \in \mathcal{O}^m(\overline{A})$, $\widehat{H} = \overline{P}^T H \overline{P}$, $\Psi_l(\mathscr{D}(\widehat{H})) = J_l \operatorname{Diag}\left(\mathrm{z}_{\alpha^l}(\kappa(\mathscr{D}(\widehat{H})))\right) J_l^T$, $\mathrm{z}(\cdot) = \operatorname{Pr}_{\varpi_1}'(\overline{\lambda}; \cdot) \in \mathbb{R}^m$, $J_l \in \mathcal{O}^{|\alpha^l|}(\widehat{H}_{\alpha^l\alpha^l})$ and

$$
\overline{D}_{ij} = D_{ij}(\overline{\lambda}) := \begin{cases} \frac{(\operatorname{Pr}_{\varpi_1}(\overline{\lambda}))_i - \operatorname{Pr}_{\varpi_1}(\overline{\lambda}))_j}{\overline{\lambda}_i - \overline{\lambda}_j} = \frac{\lambda_i(\overline{X}_1) - \lambda_j(\overline{X}_1)}{\overline{\lambda}_i - \overline{\lambda}_j}, & \overline{\lambda}_i \neq \overline{\lambda}_j, \\ 0, & \text{otherwise.} \end{cases} \tag{3-9}
$$

Define matrix $\overline{F}$ in the following form

$$\overline{F}_{ij} = F_{ij}(\overline{\lambda}) := \begin{cases} (E - \overline{D})_{ij}, & i \in \alpha^l, j \in \alpha^{l'}, l \neq l', \\ 0, & \text{otherwise.} \end{cases} \tag{3-10}$$

Moreover, Lemma 2.17 tells us that $Y \in \partial\theta_1(X_1)$ if and only if $\lambda(Y) \in \partial\varpi_1(\lambda(X_1))$ and there exists $P \in \mathcal{O}^m(X_1) \cap \mathcal{O}^m(Y)$.

By substitution, we obtain the following explicit characterization of the proximal normal cone $\mathcal{N}^{\pi}_{\mathrm{gph}\,\partial\theta_1}$ of $\mathrm{gph}\,\partial\theta_1$ from Proposition 3.1, immediately.

Proposition 3.2. *For any* $(\overline{X}_1, \overline{Y}) \in \mathrm{gph}\,\partial\theta_1$, *we have*

$$\mathcal{N}^{\pi}_{\mathrm{gph}\,\partial\theta_1}(\overline{X}_1, \overline{Y}) = \left\{ (X^*, Y^*) \in \mathbb{S}^m \times \mathbb{S}^m \;\middle|\; \begin{array}{l} \overline{D} \circ \widehat{X}^* + \overline{F} \circ \widehat{Y}^* = 0 \\ \kappa(\mathscr{D}(\widehat{X}^*)) \in \mathcal{C}^{\circ}_{\varpi_1}(\lambda(\overline{X}_1), \lambda(\overline{Y})) \\ \kappa(\mathscr{D}(\widehat{Y}^*)) \in \mathcal{C}_{\varpi_1}(\lambda(\overline{X}_1), \lambda(\overline{Y})) \end{array} \right\}, \tag{3-11}$$

where $\widehat{X}^* = \overline{P}^T X^* \overline{P}$, $\widehat{Y}^* = \overline{P}^T Y^* \overline{P}$.

Proof. It follows from Proposition 3.1 that $(X^*, Y^*) \in \mathcal{N}^{\pi}_{\mathrm{gph}\,\partial\theta_1}(\overline{X}_1, \overline{Y})$ if and only if for any $H \in \mathbb{S}^m$,

$$\langle X^*, \mathrm{Pr}'_{\theta_1}(\overline{A}; H)\rangle + \langle Y^*, \mathrm{Pr}'_{\theta_1{}^*}(\overline{A}; H)\rangle \le 0.$$

We know from (3-8) that if $(X^*, Y^*) \in \mathcal{N}^{\pi}_{\mathrm{gph}\,\partial\theta_1}(\overline{X}_1, \overline{Y})$, then for any $H \in \mathbb{S}^m$,

$$\begin{aligned} &\langle \overline{P}^T X^* \overline{P}, \overline{D} \circ \widehat{H} + \mathrm{Diag}\big(\Psi_1(\mathscr{D}(\widehat{H})), \dots, \Psi_{r_1}(\mathscr{D}(\widehat{H}))\big)\rangle \\ &+ \langle \overline{P}^T Y^* \overline{P}, \widehat{H} - \Big[\overline{D} \circ \widehat{H} + \mathrm{Diag}\big(\Psi_1(\mathscr{D}(\widehat{H})), \dots, \Psi_{r_1}(\mathscr{D}(\widehat{H}))\big)\Big]\rangle \le 0, \end{aligned}$$

which implies that

$$\begin{aligned} &\langle \overline{D} \circ \overline{P}^T X^* \overline{P}, \widehat{H}\rangle + \langle \overline{F} \circ \overline{P}^T Y^* \overline{P}, \widehat{H}\rangle + \langle \overline{P}^T X^* \overline{P}, \mathrm{Diag}\big(\Psi_1(\mathscr{D}(\widehat{H})), \dots, \Psi_{r_1}(\mathscr{D}(\widehat{H}))\big)\rangle \\ &+ \langle \overline{P}^T Y^* \overline{P}, \mathrm{Diag}\,(\mathscr{D}(\widehat{H})) - \mathrm{Diag}\big(\Psi_1(\mathscr{D}(\widehat{H})), \dots, \Psi_{r_1}(\mathscr{D}(\widehat{H}))\big)\rangle \le 0. \end{aligned} \tag{3-12}$$

By Ky Fan's inequality [67], we know that for any $H \in \mathbb{S}^m$,

$$\begin{aligned} &\langle \widehat{X}^*, \mathrm{Diag}\big(\Psi_1(\mathscr{D}(\widehat{H})), \dots, \Psi_{r_1}(\mathscr{D}(\widehat{H}))\big)\rangle = \langle \mathscr{D}(\widehat{X}^*), \big(\Psi_1(\mathscr{D}(\widehat{H})), \dots, \Psi_{r_1}(\mathscr{D}(\widehat{H}))\big)\rangle \\ &\le \langle \kappa(\mathscr{D}(\widehat{X}^*)), \kappa\big(\Psi_1(\mathscr{D}(\widehat{H})), \dots, \Psi_{r_1}(\mathscr{D}(\widehat{H}))\big)\rangle = \langle \kappa(\mathscr{D}(\widehat{X}^*)), \mathrm{Pr}'_{\varpi_1}\big(\overline{\lambda}; \kappa(\mathscr{D}(\widehat{H}))\big)\rangle \end{aligned} \tag{3-13}$$

and

$$\begin{aligned} &\langle \widehat{Y}^*, (\mathscr{D}(\widehat{H})) - \mathrm{Diag}\big(\Psi_1(\mathscr{D}(\widehat{H})), \dots, \Psi_{r_1}(\mathscr{D}(\widehat{H}))\big)\rangle \\ &= \langle \mathscr{D}(\widehat{Y}^*), \mathscr{D}(\widehat{H}) - \big(\Psi_1(\mathscr{D}(\widehat{H})), \dots, \Psi_{r_1}(\mathscr{D}(\widehat{H}))\big)\rangle \\ &\le \langle \kappa(\mathscr{D}(\widehat{Y}^*)), \kappa(\mathscr{D}(\widehat{H})) - \kappa\big(\Psi_1(\mathscr{D}(\widehat{H})), \dots, \Psi_{r_1}(\mathscr{D}(\widehat{H}))\big)\rangle \\ &= \langle \kappa(\mathscr{D}(\widehat{Y}^*)), \mathrm{Pr}'_{\varpi_1{}^*}\big(\overline{\lambda}; \kappa(\mathscr{D}(\widehat{H}))\big)\rangle. \end{aligned} \tag{3-14}$$

From Proposition 2.20, we know that $\mathrm{Pr}'_{\varpi_1}(\overline{\lambda}, \cdot)$ is the projection over $\mathcal{C}_{\varpi_1}(\lambda(\overline{X}_1), \lambda(\overline{Y}))$ and $\mathrm{Pr}'_{\varpi_1^*}(\overline{\lambda}, \cdot)$ is the projection over $\mathcal{C}^\circ_{\varpi_1}(\lambda(\overline{X}_1), \lambda(\overline{Y}))$. Since (3-12) holds for all $H \in \mathbb{S}^n$, combining (3-13) and (3-14), we need

$$\overline{D} \circ \widehat{X}^* + \overline{F} \circ \widehat{Y}^* = 0,$$

$$\kappa(\mathscr{D}(\widehat{X}^*)) \in \mathcal{C}^\circ_{\varpi_1}(\lambda(\overline{X}_1), \lambda(\overline{Y})),$$

$$\kappa(\mathscr{D}(\widehat{Y}^*)) \in \mathcal{C}_{\varpi_1}(\lambda(\overline{X}_1), \lambda(\overline{Y})).$$

Therefore, we have (X^*, Y^*) belongs the right-hand side of (3-11).

"$\supseteq$": For all (X^*, Y^*) belongs to the right-hand side of (3-11), we have

$$\langle \overline{D} \circ \overline{P}^T X^* \overline{P}, \widehat{H} \rangle + \langle \overline{F} \circ \overline{P}^T Y^* \overline{P}, \widehat{H} \rangle + \langle \kappa(\mathscr{D}(\widehat{X}^*)), \mathrm{Pr}'_{\varpi_1}(\overline{\lambda}; \kappa(\mathscr{D}(\widehat{H}))) \rangle$$
$$+ \langle \kappa(\mathscr{D}(\widehat{Y}^*)), \mathrm{Pr}'_{\varpi_1^*}(\overline{\lambda}; \kappa(\mathscr{D}(\widehat{H}))) \rangle \le 0.$$

holds for all $H \in \mathbb{S}^n$. By using (3-13) and (3-14), we know that (3-12) holds for all $H \in \mathbb{S}^n$. Thus $(X^*, Y^*) \in \mathcal{N}^\pi_{\mathrm{gph}\, \partial\theta_1}(\overline{X}_1, \overline{Y})$. □

Proximal normal cone for $\mathrm{gph}\, \mathcal{N}_{\mathcal{K}}$: Given $(\overline{X}_2, \overline{Z}) \in \mathrm{gph}\, \mathcal{N}_{\mathcal{K}}$, let $\overline{B} = \overline{X}_2 + \overline{Z}$. Denote $\overline{\lambda} = \lambda(\overline{B})$ and $\overline{v}_i = v_i(\overline{B})$, $i = 1, \ldots, r_2$. Define

$$\varsigma^l := \{1 \le i \le n \mid \overline{\lambda}_i = \overline{v}_l\}, \quad l = 1, \ldots, r_2. \tag{3-15}$$

Assume $\overline{B}$ has the following eigenvalue decomposition:

$$\overline{B} = \begin{bmatrix} \overline{V}_{\varsigma^1} & \cdots & \overline{V}_{\varsigma^{r_2}} \end{bmatrix} \begin{bmatrix} \Lambda(\overline{B})_{\varsigma^1\varsigma^1} & & 0 \\ & \ddots & \\ 0 & & \Lambda(\overline{B})_{\varsigma^{r_2}\varsigma^{r_2}} \end{bmatrix} \begin{bmatrix} \overline{V}^T_{\varsigma^1} \\ \vdots \\ \overline{V}^T_{\varsigma^{r_2}} \end{bmatrix}. \tag{3-16}$$

From Lemma 2.17, we have $Z \in \mathcal{N}_{\mathcal{K}}(X_2)$ if and only if $\lambda(Z) \in \mathcal{N}_{\widehat{\mathcal{K}}}(\lambda(X_2))$ and there exists $V \in \mathcal{O}^n(X_2) \cap \mathcal{O}^n(Z)$.

Denote $\mathbb{W}_2 = \mathbb{S}^{|\varsigma^1|} \times \cdots \times \mathbb{S}^{|\varsigma^{r_2}|}$. Define $\mathscr{D} : \mathbb{S}^n \to \mathbb{W}_2$ by $\mathscr{D}(H) = (H_{\varsigma^1\varsigma^1}, \ldots, H_{\varsigma^{r_2}\varsigma^{r_2}})$. Let $\kappa(W) := (\lambda(W_1), \ldots, \lambda(W_{r_2}))$, $W = (W_1, \ldots, W_{r_2}) \in \mathbb{W}_2$. By employing [63, Theorem 6] and the Moreau decomposition (see e.g., [59, Theorem 31.5]), we have following explicit formulas of $\Pi'_{\mathcal{K}}(\overline{B}; \cdot)$ and $\Pi'_{\mathcal{K}^\circ}(\overline{B}; \cdot)$ at $\overline{B}$, i.e., for any direction $H \in \mathbb{S}^n$,

$$\Pi'_{\mathcal{K}}(\overline{B}; H) = \overline{V} \left[\overline{D} \circ \widehat{H} + \mathrm{Diag}\big(\Psi_1(\mathscr{D}(\widehat{H})), \ldots, \Psi_{r_2}(\mathscr{D}(\widehat{H}))\big) \right] \overline{V}^T \tag{3-17}$$

and

$$\Pi'_{\mathcal{K}^\circ}(\overline{B}; H) = H - \Pi'_{\mathcal{K}}(\overline{B}; H),$$

where $\overline{V} \in \mathcal{O}^n(\overline{B})$, $\widehat{H} = \overline{V}^T H \overline{V}$, $\Psi_l(\mathscr{D}(\widehat{H})) = J_l \mathrm{Diag}\left(\mathrm{z}_{\varsigma^l}(\kappa(\mathscr{D}(\widehat{H})))\right) J_l^T$, $\mathrm{z}(\cdot) = \Pi'_{\widehat{\mathcal{K}}}(\overline{\lambda};\cdot) \in \mathbb{R}^n$, $J_l \in \mathcal{O}^{|\varsigma^l|}(\widehat{H}_{\varsigma^l\varsigma^l})$ and

$$\overline{D}_{ij} = D_{ij}(\overline{\lambda}) := \begin{cases} \frac{(\Pi_{\widehat{\mathcal{K}}}(\overline{\lambda}))_i - (\Pi_{\widehat{\mathcal{K}}}(\overline{\lambda}))_j}{\overline{\lambda}_i - \overline{\lambda}_j} = \frac{\lambda_i(\overline{X}_2) - \lambda_j(\overline{X}_2)}{\overline{\lambda}_i - \overline{\lambda}_j}, & \overline{\lambda}_i \neq \overline{\lambda}_j, \\ 0, & \text{otherwise.} \end{cases} \tag{3-18}$$

Define matrix $\overline{F}$ in the following form

$$\overline{F}_{ij} = F_{ij}(\overline{\lambda}) = \begin{cases} (E - \overline{D})_{ij}, & i \in \varsigma^l, j \in \varsigma^{l'}, l \neq l', \\ 0, & \text{otherwise.} \end{cases} \tag{3-19}$$

Similar to Proposition 3.2, we have the following explicit characterization on the proximal normal cone $\mathcal{N}^{\pi}_{\mathrm{gph}\,\mathcal{N}_{\mathcal{K}}}$.

Proposition 3.3. *For any* $(\overline{X}_2, \overline{Z}) \in \mathrm{gph}\,\mathcal{N}_{\mathcal{K}}$, *we have*

$$\mathcal{N}^{\pi}_{\mathrm{gph}\,\mathcal{N}_{\mathcal{K}}}(\overline{X}_2, \overline{Z}) = \left\{ (X^*, Z^*) \in \mathbb{S}^n \times \mathbb{S}^n \,\middle|\, \begin{array}{l} \overline{D} \circ \widehat{X}^* + \overline{F} \circ \widehat{Z}^* = 0 \\ \kappa(\mathscr{D}(\widehat{X}^*)) \in \mathcal{C}^{\circ}_{\widehat{\mathcal{K}}}(\lambda(\overline{X}_2), \lambda(\overline{Z})) \\ \kappa(\mathscr{D}(\widehat{Z}^*)) \in \mathcal{C}_{\widehat{\mathcal{K}}}(\lambda(\overline{X}_2), \lambda(\overline{Z})) \end{array} \right\}, \tag{3-20}$$

where $\widehat{X}^* = \overline{V}^T X^* \overline{V}$, $\widehat{Z}^* = \overline{V}^T Z^* \overline{V}$.

3.1.2 Characterizations of limiting normal cones

In this subsection, we shall characterize the limiting normal cones of $\mathcal{N}_{\mathrm{gph}\,\partial\theta_1}$ to $\mathrm{gph}\,\partial\theta_1$ and $\mathcal{N}_{\mathrm{gph}\,\mathcal{N}_{\mathcal{K}}}$ to $\mathrm{gph}\,\mathcal{N}_{\mathcal{K}}$, respectively. Recalling the index set (2-10), (2-11), (2-14) and (2-15).

For any given $(\bar{s}, \bar{y}) \in \mathrm{gph}\,\partial\varpi_1$ and $(\bar{t}, \bar{z}) \in \mathrm{gph}\,\mathcal{N}_{\widehat{\mathcal{K}}}$, from [81, Proposition 3.2], we know the critical cones $\mathcal{C}_{\varpi_1}(\bar{s}, \bar{y})$ and $\mathcal{C}_{\widehat{\mathcal{K}}}(\bar{t}, \bar{z})$ have the following forms

$$\mathcal{C}_{\varpi_1}(\bar{s}, \bar{y}) = \left\{ d \in \mathbb{R}^m \,\middle|\, \begin{array}{l} \langle d, a^j - a^i \rangle = 0, \ \forall\, i, j \in \eta_1(\bar{s}, \bar{y}) \\ \langle d, a^j - a^i \rangle \leq 0, \ \forall\, i \in \eta_1(\bar{s}, \bar{y}), j \in \iota_1(\bar{s}) \backslash \eta_1(\bar{s}, \bar{y}) \end{array} \right\} \tag{3-21}$$

and

$$\mathcal{C}_{\widehat{\mathcal{K}}}(\bar{t}, \bar{z}) = \left\{ d \in \mathbb{R}^n \,\middle|\, \begin{array}{l} \langle d, b^i \rangle = 0, \ \forall\, i \in \eta_2(\bar{t}, \bar{z}) \\ \langle d, b^j \rangle \leq 0, \forall\, j \in \iota_2(\bar{t}) \backslash \eta_2(\bar{t}, \bar{z}) \end{array} \right\}. \tag{3-22}$$

It follows from the generalized Farkas lemma (cf. e.g., [5, Proposition 2.201]) the polars of (3-21) and (3-22) are

$$\mathcal{C}^{\circ}_{\varpi_1}(\bar{s}, \bar{y}) = \{ \sum_{i,j \in \eta_1(\bar{s},\bar{y})} u_{ji}(a^j - a^i) + \sum_{i \in \eta_1(\bar{s},\bar{y}), j \in \iota_1(\bar{s}) \backslash \eta_1(\bar{s},\bar{y})} g_{ji}(a^j - a^i) \mid u_{ji} \in \mathbb{R},\ g_{ji} \geq 0 \}$$

and

$$\mathcal{C}^{\circ}_{\widehat{\mathcal{K}}}(\bar{t}, \bar{z}) = \{ \sum_{i \in \iota_2(\bar{t})} u_i b^i + \sum_{j \in \iota_2(\bar{t}) \backslash \eta_2(\bar{t},\bar{z})} g_j b^j \mid u_i \in \mathbb{R},\ g_j \geq 0 \}.$$

3.1.2.1 Characterization of $\mathcal{N}_{\mathrm{gph}\,\partial\theta_1}(\overline{X}_1, \overline{Y})$

Let $(\overline{X}_1, \overline{Y}) \in \mathrm{gph}\,\partial\theta_1$, $\overline{A} = \overline{X}_1 + \overline{Y}$, $\overline{\lambda} = \lambda(\overline{A})$, $\overline{s} = \lambda(\overline{X}_1)$ and $\bar{y} = \lambda(\overline{Y})$. Suppose that $\overline{A}$ has the eigenvalue decomposition (3-7). Let $\mathcal{R}^m_{\gtrsim}(\overline{\lambda})$ be the set of all vectors in $\mathbb{R}^m$ whose components being arranged in block non-increasing order, i.e.,

$$\mathcal{R}^m_{\gtrsim}(\overline{\lambda}) := \{s \in \mathbb{R}^m \mid s_1 \geq \cdots \geq s_{|\alpha^1|}; \ldots ; s_{m-|\alpha^{r_1}|+1} \geq \cdots \geq s_m\},$$

where for each $l \in \{1, \ldots, r_1\}$, α^l denotes the index set defined in (3-6) with respect to $\overline{\lambda}$.

Recalling the critical cone $\mathcal{C}_{\varpi_1}(\bar{s}, \bar{y})$ defined in (3-21). For any $w \in \mathcal{R}^m_{\gtrsim}(\overline{\lambda})$, define $W^1_l(w) \in \mathbb{S}^{|\alpha^l|}$, $l = 1, \ldots, r_1$ by

$$[W^1_l(w)]_{ij} = \begin{cases} \frac{(p_{\alpha^l})_i - (p_{\alpha^l})_j}{(w_{\alpha^l})_i - (w_{\alpha^l})_j} \in [0,1], & \text{if } (w_{\alpha^l})_i \neq (w_{\alpha^l})_j, \\ 0, & \text{otherwise}, \end{cases} \quad i, j \in \{1, \ldots, |\alpha^l|\},$$

where $p := \Pi_{\mathcal{C}_{\varpi_1}(\bar{s},\bar{y})}(w) \in \mathbb{R}^m$. Let $W^1(w) = \mathrm{Diag}\,(W^1_1(w), \ldots, W^1_{r_1}(w)) \in \mathbb{S}^m$ and $\mathcal{U}^1_m$ be the limits set defined by

$$\mathcal{U}^1_m := \left\{ W \in \mathbb{S}^m \mid W = \lim_{k\to\infty} W^1(w^k),\ w^k \to 0, w^k \in \mathcal{R}^m_{\gtrsim}(\overline{\lambda}) \right\}. \tag{3-23}$$

Let $\Xi_1 \in \mathcal{U}^1_m$ be arbitrarily given and $\{w^k\}$ be the corresponding sequence in $\mathcal{R}^m_{\gtrsim}(\overline{\lambda})$ converging to 0 such that $\Xi_1 = \lim_{k\to\infty} W^1(w^k)$. For each k, let $\lambda^k = \overline{\lambda} + w^k$. By taking a subsequence if necessary, we may assume that for sufficiently large k, there exist index sets $\{\rho^l_t\}^{s^l}_{t=1}$ such that for each $l \in \{1, \ldots, r_1\}$,

$$\begin{cases} \lambda^k_i = \lambda^k_j, & \text{if } i, j \in \rho^l_t,\ t = 1, \ldots, s^l, \\ \lambda^k_i \neq \lambda^k_j, & \text{otherwise}, \end{cases} \quad i, j \in \alpha^l. \tag{3-24}$$

It is easy to see for sufficiently large k, we have $\lambda^k_i > \lambda^k_j$ provided $i \in \rho^l_t, j \in \rho^l_{t'}$ and $t, t' \in \{1, \ldots, s^l\}$ with $t < t'$. For any $\Xi_1 \in \mathcal{U}^1_m$, we define $\Xi_2 \in \mathbb{S}^m$ by

$$(\Xi_2)_{ij} = \begin{cases} 1 - (\Xi_1)_{ij}, & i \in \rho^l_t, j \in \rho^l_{t'}, t \neq t', \\ 0, & \text{otherwise}, \end{cases} \quad i, j \in \{1, \ldots, m\}. \tag{3-25}$$

In addiction, denote $s^k := \bar{s} + \Pi_{\mathcal{C}_{\varpi_1}(\bar{s},\bar{y})}(w^k)$ and $y^k := \bar{y} + \Pi_{\mathcal{C}^\circ_{\varpi_1}(\bar{s},\bar{y})}(w^k)$. By taking a subsequence if necessary, we may conclude that the components of s^k and y^k are arranged in a non-increasing order for k sufficiently large. The validity of this claim is stated below. It follows from the B-differentiability of $\mathrm{Pr}_{\varpi_1}(\cdot)$ and Proposition 2.20 that $s^k = \mathrm{Pr}_{\varpi_1}(\overline{\lambda}) + \Pi_{\mathcal{C}_{\varpi_1}(\bar{s},\bar{y})}(w^k) = \mathrm{Pr}_{\varpi_1}(\overline{\lambda} + w^k)$. Therefore, by the symmetry of ϖ_1 and Ky Fan's inequality, we know that the components of s^k is in a non-increasing order for k sufficiently large.

Let $\overline{\eta}_1 := \eta_1(\bar{s}, \bar{y})$ and $\bar{\iota}_1 := \iota_1(\bar{s})$ (recall (2-10) and (2-14)). For each k, we denote $\eta_1^k := \eta_1(s^k, y^k)$ and $\iota_1^k := \iota_1(s^k)$. Denote the partition of $\bar{\iota}_1$ with respect to s^k and y^k as

$$\pi(\bar{\iota}_1, \Xi_1) := \left\{ (\beta_+ \ \beta_0 \ \beta_-) \left| \begin{array}{l} \beta_+ \cup \beta_0 \cup \beta_- = \bar{\iota}_1, \ \beta_+ \cap \beta_0 \cap \beta_- = \varnothing \\ \Xi_1 = \lim_{k\to\infty} W^1(w^k) \text{ such that } s^k \in \mathcal{H}_{\beta_+\cup\beta_0}, \ y^k \in \mathcal{G}_{\beta_+}. \end{array} \right. \right\}, \tag{3-26}$$

where $\mathcal{H}_{\beta_+\cup\beta_0} = \{s \mid \iota_1(s) = \beta_+ \cup \beta_0\}$ and $\mathcal{G}_{\beta_+} = \{s \mid s = \sum_{i\in\beta_+} u_i a^i, \sum_{i\in\beta_+} u_i = 1, u_i > 0\}$. We also give the following notations: $\mathcal{O}_1 := \{Q \in \mathbb{R}^{m\times m} \mid Q = \text{Diag}\,(Q_1, Q_2, \ldots, Q_{r_1}),\ \ Q_l \in \mathcal{O}^{|\alpha^l|}\}$ and for all $H = (H_1 \ldots, H_{r_1}) \in \mathbb{W}_1$,

$$\mathscr{D}^\pi(H) := \left[(H_1)_{\rho_1^1\rho_1^1}, \ldots, (H_1)_{\rho_{s^1}^1\rho_{s^1}^1}, \ldots, (H_{r_1})_{\rho_1^{r_1}\rho_1^{r_1}}, \ldots, (H_{r_1})_{\rho_{s^{r_1}}^{r_1}\rho_{s^{r_1}}^{r_1}} \right].$$

In order to state the characterization of limiting normal cone more explicitly, we introduce the following set

$$\mathcal{N}_1(Q) = \bigcup_{\Xi_1 \in \mathcal{U}_m^1} \left\{ (S^*, T^*) \in \mathbb{W}_1 \times \mathbb{W}_1 \left| \begin{array}{l} \mathscr{D}(\Xi_1) \circ \widetilde{S}^* + \mathscr{D}(\Xi_2) \circ \widetilde{T}^* = 0, \\ \kappa(\mathscr{D}^\pi(\widetilde{S}^*)) \in \{ \sum_{i,j\in\beta_+} u_{ji}(a^j - a^i) + \sum_{i\in\beta_+, j\in\beta_0} g_{ji}(a^j - a^i) : g_{ji} \geq 0 \} \\ \langle \kappa(\mathscr{D}^\pi(\widetilde{T}^*)), a^j - a^i \rangle = 0, i, j \in \beta_+ \\ \langle \kappa(\mathscr{D}^\pi(\widetilde{T}^*)), a^j - a^i \rangle \leq 0, i \in \beta_+, j \in \beta_0 \end{array} \right. \right\},$$

where $\mathbb{W}_1$ is given in the paragraph under (3-7), $S^* = (S_1^*, \ldots, S_{r_1}^*)$, $T^* = (T_1^*, \ldots, T_{r_1}^*)$, $\widetilde{S}^* = \left(Q_1^T S_1^* Q_1, \ldots, Q_{r_1}^T S_{r_1}^* Q_{r_1}\right)$, $\widetilde{T}^* = \left(Q_1^T T_1^* Q_1, \ldots, Q_{r_1}^T T_{r_1}^* Q_{r_1}\right)$, $Q_l \in \mathcal{O}^{|\alpha^l|}$.

Theorem 3.4. *For any $(\overline{X}_1, \overline{Y}) \in \text{gph}\,\partial\theta_1$, $\overline{A} = \overline{X}_1 + \overline{Y}$ have the eigenvalue decomposition (3-7). We have $(X^*, Y^*) \in \mathcal{N}_{\text{gph}\,\partial\theta_1}(\overline{X}_1, \overline{Y})$ if and only if there exists $Q \in \mathcal{O}_1$ such that*

$$\overline{D} \circ Q^T \widehat{X}^* Q + \overline{F} \circ Q^T \widehat{Y}^* Q = 0,$$

$$(\mathscr{D}(\widehat{X}^*), \mathscr{D}(\widehat{Y}^*)) \in \mathcal{N}_1(Q).$$

where $\widehat{X}^* = \overline{P}^T X^* \overline{P}$, $\widehat{Y}^* = \overline{P}^T Y^* \overline{P}$. $\overline{D}$ and $\overline{F}$ are defined in (3-9) *and* (3-10).

Proof. "$\Longrightarrow$": Suppose that $(X^*, Y^*) \in \mathcal{N}_{\text{gph}\,\partial\theta_1}(\overline{X}_1, \overline{Y})$. By the definition of the limiting normal cone, we know that $(X^*, Y^*) = \lim_{k\to\infty} (X^{k^*}, Y^{k^*})$ with

$$(X^{k^*}, Y^{k^*}) \in \mathcal{N}^\pi_{\text{gph}\,\partial\theta_1}(X^k, Y^k) \quad k = 1, 2, \ldots,$$

where $(X^k, Y^k) \to (\overline{X}_1, \overline{Y})$ and $(X^k, Y^k) \in \text{gph}\,\partial\theta_1$. For each k, denote $A^k := X^k + Y^k$ and let $A^k = P^k \Lambda(A^k)(P^k)^T$ be the eigenvalue decomposition of A^k. Denote $\text{s}^k = \lambda(X^k)$, $y^k = \lambda(Y^k)$, $\lambda^k = \lambda(A^k)$ and $w^k = \lambda^k - \overline{\lambda}$. Since $\Lambda(A) = \lim_{k\to\infty} \Lambda(A^k)$, without loss of generality, we assume that the partitions $\bigcup_{t=1}^{s^l} \rho_t^l = \alpha^l$, $l = 1, \ldots, r_1$ are all same for each A^k. Since $\{P^k\}_{k=1}^\infty$ is uniformly bounded, by taking a subsequence if necessary, we may assume that $\{P^k\}_{k=1}^\infty$ converges to an orthogonal matrix $\widetilde{P} \in \mathcal{O}^m$ ([82],[79, Lemma 4.12]). We can write

$$\widetilde{P} = \overline{P}Q = \left[\ \overline{P}_{\alpha^1} Q_1 \quad \overline{P}_{\alpha^2} Q_2 \quad \cdots \quad \overline{P}_{\alpha^{r_1}} Q_{r_1} \ \right],$$

where $Q = \mathrm{Diag}(Q_1, Q_2, \ldots, Q_{r_1})$ and $Q_l \in \mathcal{O}^{|\alpha^l|}, l = 1, \ldots, r_1$. By further taking a subsequence if necessary, we may also assume that there exists a partition $\pi(\bar{\iota}_1, \Xi_1) = (\beta_+ \quad \beta_0 \quad \beta_-)$ of $\bar{\iota}_1$ defined in (3-26) such that for each k,

$$\beta_+ = \eta_1^k, \quad \beta_0 = \iota_1^k \backslash \eta_1^k, \quad \beta_- = \bar{\iota}_1 \backslash \iota_1^k$$

and the three index sets are independent of k.

Then, for each k, since $(X^{k^*}, Y^{k^*}) \in \mathcal{N}^{\pi}_{\mathrm{gph}\,\partial\theta_1}(X^k, Y^k)$, we know from Proposition 3.2 that there exist

$$D_{ij}^k = \begin{cases} \frac{(\mathrm{Pr}_{\varpi_1}(\lambda^k))_i - (\mathrm{Pr}_{\varpi_1}(\lambda^k))_j}{\lambda_i^k - \lambda_j^k}, & \lambda_i^k \neq \lambda_j^k \\ 0, & \text{otherwise} \end{cases}$$

and

$$F_{ij}^k = \begin{cases} (E - D^k)_{ij}, & i \in \rho_t^l, j \in \rho_{t'}^{l'}, t \neq t' \\ 0, & \text{otherwise} \end{cases} \tag{3-27}$$

such that

$$D^k \circ \widehat{X}^{k^*} + F^k \circ \widehat{Y}^{k^*} = 0, \tag{3-28}$$

$$\kappa(\mathscr{D}^{\pi}(\widehat{X}^{k^*})) \in \mathcal{C}^{\circ}_{\varpi_1}(x^k, y^k), \tag{3-29}$$

$$\kappa(\mathscr{D}^{\pi}(\widehat{Y}^{k^*})) \in \mathcal{C}_{\varpi_1}(x^k, y^k), \tag{3-30}$$

where $\widehat{X}^{k^*} = (P^k)^T X^{k^*} P^k, \widehat{Y}^{k^*} = (P^k)^T Y^{k^*} P^k$. By taking limits as $k \to \infty$, we obtain that

$$\widehat{X}^{k^*} \to \widetilde{P}^T X^* \widetilde{P} := \widetilde{X}^* = Q^T \widehat{X}^* Q \quad \text{and} \quad \widehat{Y}^{k^*} \to \widetilde{P}^T Y^* \widetilde{P} := \widetilde{Y}^* = Q^T \widehat{Y}^* Q.$$

Obviously, if $i \in \alpha^l, j \in \alpha^{l'}, l \neq l'$, we have $D_{ij}^k \to \overline{D}_{ij}$ as $k \to \infty$; and for each k, if $i, j \in \rho_t^l$, $D_{ij}^k = 0$. The remaining case is when $i \in \rho_t^l, j \in \rho_{t'}^l, t \neq t'$. Since ϖ_1 is symmetric, we know from [62, Proposition 3.2] that for all permutation matrix $U \in \mathcal{P}^m$,

$$U\mathrm{Pr}_{\varpi_1}(x) = \mathrm{Pr}_{\varpi_1}(Ux), \quad \forall\, x \in \mathbb{R}^m.$$

Combining this with the Lipschitzian continuity of the proximal mapping, we have

$$(D_{ij}^k)^2 = \frac{|(\mathrm{Pr}_{\varpi_1}(\lambda^k))_i - (\mathrm{Pr}_{\varpi_1}(\lambda^k))_j|^2}{|\lambda_i^k - \lambda_j^k|^2} = \frac{\|\mathrm{Pr}_{\varpi_1}(\lambda^k) - \mathrm{Pr}_{\varpi_1}(U\lambda^k)\|^2}{\|\lambda^k - U\lambda^k\|^2} \leq 1,$$

where U is the permutation matrix that only exchange the i-th and j-th component. For all $i \in \rho_t^l, j \in \rho_{t'}^l, t \neq t'$, we have $\overline{\lambda}_i = \overline{\lambda}_j$ and $\bar{s}_i = \bar{s}_j$. From the B-differentiable of ϖ_1 before Lemma 2.7 and [4, Proposition 2], we have

$$D_{ij}^k = \frac{[\Pi_{\mathcal{C}_{\varpi_1}(\bar{s}, \bar{y})}(w^k)]_i - [\Pi_{\mathcal{C}_{\varpi_1}(\bar{s}, \bar{y})}(w^k)]_j}{w_i - w_j}.$$

Taking a subsequence if necessary, by the 1-modulus Lipschitz continuity of projection mapping, it follows that there exist $\Xi_1 \in \mathcal{U}_m^1$ and the corresponding Ξ_2 defined in (3-25) such that

$$\lim_{k\to\infty} D^k = \overline{D} + \Xi_1 \quad \text{and} \quad \lim_{k\to\infty} F^k = \overline{F} + \Xi_2 .$$

Meanwhile, by taking limits in (3-28) as $k \to \infty$, we obtain that

$$\overline{D} \circ \widetilde{X}^* + \overline{F} \circ \widetilde{Y}^* = 0 \quad \text{and} \quad \Xi_1 \circ \widetilde{X}^* + \Xi_2 \circ \widetilde{Y}^* = 0. \tag{3-31}$$

Since $\kappa(\mathscr{D}^\pi(\widehat{X}^{k^*})) \to \kappa(\mathscr{D}^\pi(\widetilde{X}^*))$ and for all k, β_+, β_0 are the same, by taking limits in (3-29) and (3-30) as $k \to \infty$, we have

$$\kappa(\mathscr{D}^\pi(\widetilde{X}^*)) \in \left\{ \sum_{i,j\in\beta_+} u_{ji}(a^j - a^i) + \sum_{i\in\beta_+, j\in\beta_0} g_{ji}(a^j - a^i) \mid g_{ji} \geq 0 \right\}$$

and

$$\langle \kappa(\mathscr{D}^\pi(\widetilde{Y}^*)), a^j - a^i \rangle = 0, i, j \in \beta_+ \quad \text{with} \quad \langle \kappa(\mathscr{D}^\pi(\widetilde{Y}^*)), a^j - a^i \rangle \leq 0, i \in \beta_+, j \in \beta_0.$$

It follows that $(\mathscr{D}(\widehat{X}^*), \mathscr{D}(\widehat{Y}^*)) \in \mathcal{N}_1(Q)$. Thus we have finished the first part of the proof.

"$\Longleftarrow$": From the above discussion, we already obtain "$\subseteq$" in equation

$$\mathcal{N}_{\mathrm{gph}\,\partial\theta}(\overline{X}_1, \overline{Y}) = \bigcup_{\substack{Q \in \mathcal{O}_1 \\ \Xi_1 \in \mathcal{U}_m^1}} \left\{ (X^*, Y^*) \in \mathbb{S}^m \times \mathbb{S}^m \,\middle|\, \begin{array}{l} \overline{D} \circ \widetilde{X}^* + \overline{F} \circ \widetilde{Y}^* = 0, \\ \Xi_1 \circ \widetilde{X}^* + \Xi_2 \circ \widetilde{Y}^* = 0, \\ \kappa(\mathscr{D}^\pi(\mathscr{D}(\widetilde{X}^*))) \in \mathcal{C}_1, \\ \langle \kappa(\mathscr{D}^\pi(\mathscr{D}(\widetilde{Y}^*))), a^i - a^j \rangle = 0, i, j \in \beta_+, \\ \langle \kappa(\mathscr{D}^\pi(\mathscr{D}(\widetilde{Y}^*))), a^i - a^j \rangle \leq 0, i \in \beta_+, j \in \beta_0. \end{array} \right\}, \tag{3-32}$$

where $\mathcal{C}_1 = \{ \sum_{i,j\in\beta_+} u_{ji}(a^i - a^j) + \sum_{i\in\beta_+, j\in\beta_0} g_{ji}(a^i - a^j) \mid g_{ji} \geq 0 \}$. Denote the right hand side of (3-32) as $\mathcal{N}$. Let $(X^*, Y^*) \in \mathcal{N}$. We shall show that there exist two sequences: $\{(X^k, Y^k)\}$ converging to $(\overline{X}_1, \overline{Y})$ and $\{(X^{k^*}, Y^{k^*})\}$ converging to (X^*, Y^*) with $(X^k, Y^k) \in \mathrm{gph}\,\partial\theta_1$ and $(X^{k^*}, Y^{k^*}) \in \mathcal{N}^\pi_{\mathrm{gph}\,\partial\theta_1}(X^k, Y^k)$ for each k.

For each k, let $X^k = \widetilde{P}\mathrm{Diag}(\bar{s} + \Pi_{\mathcal{C}_{\varpi_1}(\bar{s},\bar{y})}(w^k))\widetilde{P}^T$, $Y^k = \widetilde{P}\mathrm{Diag}(\bar{y} + \Pi_{\mathcal{C}^\circ_{\varpi_1}(\bar{s},\bar{y})}(w^k))\widetilde{P}^T$ such that $\Xi_1 = \lim_{k\to\infty} W(w^k)$. By the definition of Ξ_1, we know that $(X^k, Y^k) \to (\overline{X}_1, \overline{Y})$ and $Y^k \in \partial\theta_1(X^k)$. Taking a subsequence such that for all k, we have $\iota_1^k = \beta_+ \cup \beta_0, \eta_1^k = \beta_0$. For each (X^k, Y^k), we get the corresponding D^k, F^k. Denote $c^k = D^k_{ij}$ or $(\Xi_1^k)_{ij}$; $c = D_{ij}$ or $(\Xi_1)_{ij}$, $c \in [0, 1]$.

If $i, j \in \rho_t^l$, we define $(\widetilde{X}^{k^*})_{ij} = (\widetilde{X}^*)_{ij}$, $(\widetilde{Y}^{k^*})_{ij} = (\widetilde{Y}^*)_{ij}$. If i, j are not in the same ρ_t^l and $c \neq 1$, it follows that $c^k \neq 1$ for all k large enough. We define

$$(\widetilde{X}^{k^*})_{ij} = (\widetilde{X}^*)_{ij} \quad \text{and} \quad (\widetilde{Y}^{k^*})_{ij} = \frac{c^k}{c^k - 1}(\widetilde{X}^{k^*})_{ij}.$$

Otherwise, we define

$$(\widetilde{Y}^{k^*})_{ij} = (\widetilde{Y}^*)_{ij} \quad \text{and} \quad (\widetilde{X}^{k^*})_{ij} = \frac{c^k - 1}{c^k}(\widetilde{Y}^{k^*})_{ij}.$$

So $X^{k^*} = \widetilde{P}\widetilde{X}^{k^*}\widetilde{P}^T$, $Y^{k^*} = \widetilde{P}\widetilde{Y}^{k^*}\widetilde{P}^T$.

Then we can easily see that $(X^k, Y^k) \in \operatorname{gph} \partial\theta_1$ and $(X^{k^*}, Y^{k^*}) \in \mathcal{N}^{\pi}_{\operatorname{gph} \partial\theta_1}(X^k, Y^k)$ and $(X^{k^*}, Y^{k^*}) \to (X^*, Y^*)$. We have completed the proof. □

3.1.2.2 Characterization of $\mathcal{N}_{\operatorname{gph} \mathcal{N}_{\mathcal{K}}}(\overline{X}_2, \overline{Z})$

Let $(\overline{X}_2, \overline{Z}) \in \operatorname{gph} \mathcal{N}_{\mathcal{K}}$, $\overline{B} = \overline{X}_2 + \overline{Z}$, $\overline{\lambda} = \lambda(\overline{B})$, $\bar{t} = \lambda(\overline{X}_2)$ and $\bar{z} = \lambda(\overline{Z})$. Suppose $\overline{B}$ has the eigenvalue decomposition (3-16). Let $\mathcal{R}^n_{\gtrsim}(\overline{\lambda})$ be the set of all vectors in $\mathbb{R}^n$ whose components being arranged in block non-increasing order, i.e.,

$$\mathcal{R}^n_{\gtrsim}(\overline{\lambda}) := \{t \in \mathbb{R}^n \mid t_1 \geq \cdots \geq t_{|\varsigma^1|}; \ldots ; t_{n-|\varsigma^{r_2}|+1} \geq \cdots \geq t_n\},$$

where for each $l \in \{1, \ldots, r_2\}$, ς^l denotes the index set defined in (3-15) with respect to $\overline{\lambda}$.

Recalling the critical cone $\mathcal{C}_{\widehat{\mathcal{K}}}(\bar{t}, \bar{z})$ defined in (3-22). For any $w \in \mathcal{R}^n_{\gtrsim}(\overline{\lambda})$, define $W_l^2(w) \in \mathbb{S}^{|\varsigma^l|}$, $l \in \{1, \ldots, r_2\}$ by

$$[W_l^2(w)]_{ij} = \begin{cases} \frac{(p_{\varsigma^l})_i - (p_{\varsigma^l})_j}{(w_{\varsigma^l})_i - (w_{\varsigma^l})_j}, & (w_{\varsigma^l})_i \neq (w_{\varsigma^l})_j, \\ 0, & \text{otherwise}, \end{cases} \quad i, j \in \{1, \ldots, |\varsigma^l|\},$$

where $p := \Pi_{\mathcal{C}_{\widehat{\mathcal{K}}}(\bar{t},\bar{z})}(w) \in \mathbb{R}^n$. Let $W^2(w) = \operatorname{Diag}(W_1^2(w), \ldots, W_{r_2}^2(w))$ and $\mathcal{U}_n^2$ be the limit set defined by

$$\mathcal{U}_n^2 = \left\{ W \in \mathbb{S}^n \mid W = \lim_{k\to\infty} W^2(w^k),\ w^k \to 0, w^k \in \mathcal{R}^n_{\gtrsim}(\overline{\lambda}) \right\}. \tag{3-33}$$

Let $\Xi_1 \in \mathcal{U}_n^2$ be arbitrarily given and $\{w^k\}$ be the corresponding sequence in $\mathcal{R}^n_{\gtrsim}(\overline{\lambda})$ converging to 0 such that $\Xi_1 = \lim_{k\to\infty} W^1(w^k)$. For each k, let $\lambda^k = \overline{\lambda} + w^k$. By taking a subsequence if necessary, we may assume that for sufficiently large k, there exists index set $\{\varrho_t^l\}_{t=1}^{e^l}$ such that for each $l \in \{1, \ldots, r_2\}$,

$$\begin{cases} \lambda_i^k = \lambda_j^k, & i, j \in \varrho_t^l, t \in \{1, \ldots, e^l\}, \\ \lambda_i^k > \lambda_j^k, & \text{otherwise}, \end{cases} \quad i, j \in \varsigma^l. \tag{3-34}$$

It is easy to see for sufficiently large k, we have $\lambda_i^k > \lambda_j^k$ provided $i \in \varrho_t^l$, $j \in \varrho_{t'}^l$ and $t, t' \in \{1, \ldots, e^l\}$ with $t < t'$. For any $\Xi_1 \in \mathcal{U}_n^2$, we define $\Xi_2 \in \mathbb{S}^n$ by

$$(\Xi_2)_{ij} = \begin{cases} 1 - (\Xi_1)_{ij}, & i \in \varrho_t^l, j \in \varrho_{t'}^l, t \neq t', \\ 0, & \text{otherwise} \end{cases} \quad i, j \in \{1, \ldots, n\} \tag{3-35}$$

In addition, denote $t^k := \bar{t} + \Pi_{\triangle_2}(w^k)$ and $z^k := \bar{z} + \Pi_{\triangle_2^\circ}(w^k)$. When k is large enough, by taking a subsequence if necessary, we know that the components of t^k and z^k are arranged in nonincreasing order by the discussion under (3-25).

Let $\overline{\eta}_2 := \eta_2(\bar{t}, \bar{z})$ and $\bar{\iota}_2 := \iota_2(\bar{t})$ (recall (2-11) and (2-15)). For each k, we denote $\eta_2^k := \eta_2(t^k, z^k)$ and $\iota_2^k := \iota_2(t^k)$. Denote the partition of $\bar{\iota}_2$ as

$$\pi(\bar{\iota}_2, \Xi_1) := \left\{ (\beta_+ \ \beta_0 \ \beta_-) \ \middle| \ \begin{array}{l} \beta_+ \cup \beta_0 \cup \beta_- = \bar{\iota}_2, \ \beta_+ \cap \beta_0 \cap \beta_- = \emptyset, \\ \Xi_1 = \lim\limits_{k\to\infty} W^2(w^k) \text{ such that } t^k \in \mathcal{H}_{\beta_+\cup\beta_0}, \ z^k \in \mathcal{G}_{\beta_+}. \end{array} \right\},$$

where $\mathcal{H}_{\beta_+\cup\beta_0} = \{t \mid \iota_2(t) = \beta_+ \cup \beta_0\}$ and $\mathcal{G}_{\beta_+} = \{t \mid t = \sum\limits_{i\in\beta_+} u_i b^i, u_i > 0\}$. We also give the following notations: $\mathcal{O}_2 = \{Q \in \mathbb{R}^{n\times n} \mid Q = \mathrm{Diag}\,(Q_1, Q_2, \ldots, Q_{r_2}), \ \ Q_l \in \mathcal{O}^{|\varsigma^l|}\}$ and

$$\mathscr{D}^\pi(H) = \left[(H_1)_{\varrho_1^1\varrho_1^1}, \ldots, (H_1)_{\varrho_{e^1}^1\varrho_{e^1}^1}, \ldots, (H_{r_2})_{\varrho_1^{r_2}\varrho_1^{r_2}}, \ldots, (H_{r_2})_{\varrho_{e^{r_2}}^{r_2}\varrho_{e^{r_2}}^{r_2}} \right] \ \forall H \in \mathbb{W}_2.$$

To state the characterization of limiting normal cone more explicitly, we introduce the following set

$$\mathcal{N}_2(Q) = \bigcup_{\Xi_1 \in \mathcal{U}_n^2} \left\{ (S^*, T^*) \in \mathbb{W}_2 \times \mathbb{W}_2 \ \middle| \ \begin{array}{l} \mathscr{D}(\Xi_1) \circ \widetilde{S}^* + \mathscr{D}(\Xi_2) \circ \widetilde{T}^* = 0, \\ \kappa(\mathscr{D}^\pi(\widetilde{S}^*)) \in \{ \sum\limits_{i\in\beta_+} u_i b^i + \sum\limits_{i\in\beta_0} g_i b^i : g_i \geq 0\} \\ \langle \kappa(\mathscr{D}^\pi(\widetilde{T}^*)), b^i \rangle = 0, i \in \beta_+ \\ \langle \kappa(\mathscr{D}^\pi(\widetilde{T}^*)), b^j \rangle \leq 0, j \in \beta_0 \end{array} \right\},$$

where $\mathbb{W}_2$ is given in the paragraph under (3-16), $S^* = (S_1^*, \ldots, S_{r_2}^*)$, $T^* = (T_1^*, \ldots, T_{r_2}^*)$, $\widetilde{S}^* = \left(Q_1^T S_1^* Q_1, \ldots, Q_{r_2}^T S_{r_2}^* Q_{r_2}\right)$, $\widetilde{T}^* = \left(Q_1^T T_1^* Q_1, \ldots, Q_{r_2}^T T_{r_2}^* Q_{r_2}\right)$, $Q_l \in \mathcal{O}^{|\varsigma^l|}$.

Theorem 3.5. *For any $(\overline{X}_2, \overline{Z}) \in \mathrm{gph}\, \mathcal{N}_{\mathcal{K}}$, $\overline{B} = \overline{X}_2 + \overline{Z}$ have the eigenvalue decomposition (3-16). We have $(X^*, Z^*) \in \mathcal{N}_{\mathrm{gph}\,\mathcal{N}_{\mathcal{K}}}(\overline{X}_2, \overline{Z})$ if and only if there exists $Q \in \mathcal{O}_2$ such that*

$$\overline{D} \circ Q^T \widehat{X}^* Q + \overline{F} \circ Q^T \widehat{Z}^* Q = 0,$$

$$(\mathscr{D}(\widehat{X}^*), \mathscr{D}(\widehat{Z}^*)) \in \mathcal{N}_2(Q),$$

where $\widehat{X}^* = \overline{V}^T X^* \overline{V}$, $\widehat{Z}^* = \overline{V}^T Z^* \overline{V}$. $\overline{D}$ and $\overline{F}$ are defined in (3-18) *and* (3-19).

Proof. The proof is similar to that of Theorem 3.4, we omit it here for simplicity. □

3.2 Lipschitzian full stability for nonsmooth composite matrix optimization problem

In this section, we shall study the characterizations of Lipschitzian full stability for the local minimizers of CMatOP (1-6). Moreover, as a by-product, we will show the equivalence between the strong regularity and Lipschitzian full stability of the local minimizer of CMatOP (1-6) under the constraint nondegeneracy condition (see Definition 3.7).

Recall the CMatOP (1-6):

$$\begin{array}{ll}\min\limits_{x\in\mathbb{X}} & f(x,\bar{p})+\theta_1(g_1(x,\bar{p}))\\ \text{s.t.} & h(x,\bar{p})=0,\\ & g_2(x,\bar{p})\in\mathcal{K}.\end{array} \tag{3-36}$$

The Lagrangian function of the above problem can be written as $L(x,\bar{p};Y,y,Z) = f(x,\bar{p})+\langle Y,g_1(x,\bar{p})\rangle+\langle y,h(x,\bar{p})\rangle+\langle Z,g_2(x,\bar{p})\rangle$, yielding the KKT optimality condition

$$\begin{cases} L'_x(x,\bar{p};Y,y,Z)=0,\\ Y\in\partial\theta_1(g_1(x,\bar{p})),\\ h(x,\bar{p})=0,\\ Z\in\mathcal{N}_{\mathcal{K}}(g_2(x,\bar{p})).\end{cases} \tag{3-37}$$

Moreover, we can write the KKT system into the following generalized equation:

$$0\in\begin{bmatrix} L'_x(x,\bar{p};Y,y,Z)\\ -g_1(x,\bar{p})\\ h(x,\bar{p})\\ -g_2(x,\bar{p})\end{bmatrix}+\begin{bmatrix}\{0\}\\ \partial{\theta_1}^*(Y)\\ \{0\}\\ \partial\delta^*_{\mathcal{K}}(Z)\end{bmatrix}. \tag{3-38}$$

The parametric perturbation of (3-36) is constructed as

$$\begin{array}{ll}\min & f(x,p)+\theta_1(g_1(x,p))-\langle x,v\rangle\\ \text{s.t.} & h(x,p)=0,\\ & g_2(x,p)\in\mathcal{K}.\end{array} \tag{3-39}$$

If $(p,v)=(\bar{p},0)$, assume $(\bar{x},\overline{Y},\bar{y},\overline{Z})\in\mathbb{X}\times\mathbb{S}^m\times\mathbb{Y}\times\mathbb{S}^n$ satisfying (3-37). We call $\bar{x}$ a stationary point, $(\overline{Y},\bar{y},\overline{Z})$ the corresponding multiplier and $(\bar{x},\overline{Y},\bar{y},\overline{Z})$ a KKT point of (3-36). We also use $\mathcal{M}(\bar{x})$ to denote the set of multipliers $(\overline{Y},\bar{y},\overline{Z})$ for any stationary point $\bar{x}$ such that $(\bar{x},\overline{Y},\bar{y},\overline{Z})$ is a KKT point. For this kind of CMatOP, the definition of strongly regular is adopted from [13].

Definition 3.6. *We say $(\bar{x},\overline{Y},\bar{y},\overline{Z})\in\mathbb{X}\times\mathbb{S}^m\times\mathbb{Y}\times\mathbb{S}^n$ is a strongly regular solution of the generalized equation (3-38), if there exist neighborhoods $\mathcal{B}$ of the origin 0 and $\mathcal{V}$ of $(\bar{x},\overline{Y},\bar{y},\overline{Z})$ such that for every $\xi\in\mathcal{B}$, the following generalized equation*

$$\xi\in\begin{bmatrix} L'_x(x,\bar{p};Y,y,Z)\\ -g_1(x,\bar{p})\\ h(x,\bar{p})\\ -g_2(x,\bar{p})\end{bmatrix}+\begin{bmatrix}\{0\}\\ \partial{\theta_1}^*(Y)\\ \{0\}\\ \partial\delta^*_{\mathcal{K}}(Z)\end{bmatrix}$$

has a unique solution in $\mathcal{V}$, denote by $\mathcal{S}_{\mathcal{V}}(\xi)$, and the mapping $\mathcal{S}_{\mathcal{V}}:\mathcal{B}\to\mathcal{V}$ is Lipschitz continuous.

Suppose $(\bar{x},\overline{Y},\bar{y},\overline{Z})$ to be the KKT point of problem (3-36). Denote $\overline{X}_1=g_1(\bar{x},\bar{p})$ and assume it has d_1 different eigenvalues $v_1(\overline{X}_1)>\cdots>v_{d_1}(\overline{X}_1)$, define

$$\mu^s:=\{1\le i\le m\mid \lambda_i(\overline{X}_1)=v_s(\overline{X}_1)\},\ s=1,\ldots,d_1. \tag{3-40}$$

For each $s\in\{1,\ldots,d_1\}$, we use the notation $\{\chi_t^s\}_{t=1}^{u^s}$ to further partition the set μ^s based on the eigenvalue of $\overline{Y}$ as

$$\begin{cases}\lambda_i(\overline{Y})=\lambda_j(\overline{Y}), & \text{if } i,j\in\chi_t^s \text{ and } t\in\{1,\ldots,u^s\},\\ \lambda_i(\overline{Y})>\lambda_j(\overline{Y}), & \text{if } i\in\chi_t^s, j\in\chi_{t'}^s \text{ and } t,t'\in\{1,\ldots,u^s\} \text{ with } t<t'.\end{cases}$$

Recalling the eigenvalue decomposition for $\overline{A}=\overline{X}_1+\overline{Y}$, see (3-7). We have $\sum\limits_{s=1}^{d_1}u^s=r_1$ and every χ_t^s corresponds to some α^l (3-6). Also, let $\bar{\iota}_1=\iota_1(\lambda(\overline{X}_1))$ and $\overline{\eta}_1=\eta_1(\lambda(\overline{X}_1),\lambda(\overline{Y}))$ be the index sets defined in (2-10) and (2-14) with respect to $\lambda(\overline{X}_1)$ and $\lambda(\overline{Y})$, correspondingly. For each $s\in\{1,\ldots,d_1\}$, define the index set

$$\mathcal{E}^s:=\left\{1\le t\le u^s\mid \exists\, i,j\in\chi_t^s \text{ such that } (a^w)_i\ne(a^w)_j \text{ for some } w\in\overline{\eta}_1\right\}. \tag{3-41}$$

Similarly, denote $\overline{X}_2=g_2(\bar{x},\bar{p})$ and assume it has d_2 different eigenvalues $v_1(\overline{X}_2)>\cdots>v_{d_2}(\overline{X}_2)$, define

$$\nu^s:=\{1\le i\le n\mid \lambda_i(\overline{X}_2)=v_s(\overline{X}_2)\},\ s=1,\ldots,d_2.$$

For each $s\in\{1,\ldots,d_2\}$, we use the notation $\{\zeta_t^s\}_{t=1}^{q^s}$ to further partition the set ν^s based on the eigenvalue of $\overline{Z}$ as

$$\begin{cases}\lambda_i(\overline{Z})=\lambda_j(\overline{Z}), & \text{if } i,j\in\zeta_t^s \text{ and } t\in\{1,\ldots,q^s\},\\ \lambda_i(\overline{Z})>\lambda_j(\overline{Z}), & \text{if } i\in\zeta_t^s, j\in\zeta_{t'}^s \text{ and } t,t'\in\{1,\ldots,q^s\} \text{ with } t<t'.\end{cases} \tag{3-42}$$

Recalling the eigenvalue decomposition for $\overline{B}=\overline{X}_2+\overline{Z}$, see (3-16). We have that $\sum\limits_{s=1}^{d_2}q^s=r_2$ and every ζ_t^s corresponds to some ς^l (3-15). Also, let $\bar{\iota}_2=\iota_2(\lambda(\overline{X}_2))$ and $\overline{\eta}_2=\eta_2(\lambda(\overline{X}_2),\lambda(\overline{Z}))$ be the index sets defined in (2-11) and (2-15) with respect to $\lambda(\overline{X}_2)$ and $\lambda(\overline{Z})$, correspondingly. For each $s\in\{1,\ldots,d_2\}$, define the index set

$$\mathcal{F}^s:=\left\{1\le t\le q^s\mid \exists\, i,j\in\zeta_t^s \text{ such that } (b^w)_i\ne(b^w)_j \text{ for some } w\in\overline{\eta}_2\right\}.$$

To explore the relationship between strong regularity and Lipschitzian full stability, we always need the following definition of constraint nondegeneracy condition and C^2-cone reducible for (3-36) (cf. e.g. [83] and [5, Definition 3.135]). It is well known that C^2-cone reducibility holds for most problems we care about, e.g., SDP [5, Example 3.140].

Definition 3.7. *The constraint nondegeneracy condition of problem (3-36) at x is defined as*

$$\begin{bmatrix} (g_1)'_x(x,\bar{p}) \\ h'_x(x,\bar{p}) \\ (g_2)'_x(x,\bar{p}) \end{bmatrix} \mathbb{X} + \begin{bmatrix} \mathcal{T}^{\text{lin}}_{\theta_1}(g_1(x,\bar{p})) \\ \{0\} \\ \text{lin}\big(\mathcal{T}_{\mathcal{K}}(g_2(x,\bar{p}))\big) \end{bmatrix} = \begin{bmatrix} \mathbb{S}^m \\ \mathbb{Y} \\ \mathbb{S}^n \end{bmatrix} \tag{3-43}$$

where $\mathcal{T}^{\text{lin}}_{\theta_1}(X) := \{H \in \mathbb{S}^n \mid {\theta_1}'(X;H) = -{\theta_1}'(X;-H)\}$ *and* $\text{lin}\big(\mathcal{T}_{\mathcal{K}}(g_2(x,\bar{p}))\big)$ *is the largest linear subspace of* $\mathbb{S}^n$ *that is contained in* $\mathcal{T}_{\mathcal{K}}(g_2(\bar{x},\bar{p}))$ *of convex analysis.*

Lemma 3.8. *Let* $f : \mathbb{X} \to \mathbb{R}$ *be a proper convex function. For any* $X \in \text{ri}\,(\text{dom}\, f)$ *and* $Y \in \partial f(X)$*, we have* $\mathcal{T}^{\text{lin}}_f(X) \subseteq \mathcal{C}_f(X,Y)$*.*

Proof. For all $H \in \mathcal{T}^{\text{lin}}_f(X)$, we have $f'(X;H) = -f'(X;-H)$. It follows from [59, Theorem 23.4] that $f'(X;H) = \sup\{\langle H,Y\rangle : Y \in \partial f(X)\}$. Then we have

$$\begin{aligned} f'(X;H) = \sup\{\langle H,Y\rangle : Y \in \partial f(X)\} &= -\sup\{\langle -H,Y\rangle : Y \in \partial f(X)\} \\ &= \inf\{\langle H,Y\rangle : Y \in \partial f(X)\}, \end{aligned}$$

which implies $f'(X;H) = \langle H,Y\rangle$. Thus we have $H \in \mathcal{C}_f(X,Y)$. □

Remark 3.1. The result that is similar to Lemma 3.8 is well known for convex set $\mathcal{K}$, i.e., for any $X \in \text{ri}\,\mathcal{K}$ and $Y \in \mathcal{N}_{\mathcal{K}}(X)$, we have $\text{lin}(\mathcal{T}_{\mathcal{K}}(X)) \subseteq \mathcal{C}_{\mathcal{K}}(X,Y)$.

Definition 3.9. *Let* $\mathcal{Q} \subseteq \mathbb{X}$ *be a pointed convex closed cone (a cone is said to be pointed if* $z \in \mathcal{Q}$ *and* $-z \in \mathcal{Q}$ *implies that* $z = 0$*). The closed convex set* $\mathcal{K} \subseteq \mathbb{Y}$ *is said to be* C^2*-cone reducible at* $\overline{X} \in \mathcal{K}$ *to the cone* $\mathcal{Q}$*, if there exist an open neighborhood* $\mathcal{W} \subseteq \mathbb{Y}$ *of* $\overline{X}$ *and a twice continuously differentiable mapping* $\Xi : \mathcal{W} \to \mathbb{X}$ *such that: (i)* $\Xi(\overline{X}) = 0 \in \mathbb{X}$*; (ii) the derivative mapping* $\Xi'(\overline{X}) : \mathbb{Y} \to \mathbb{X}$ *is onto; (iii)* $\mathcal{K} \cap \mathcal{W} = \{X \in \mathcal{W} \mid \Xi(X) \in \mathcal{Q}\}$*. We say that* $\mathcal{K}$ *is* C^2*-cone reducible if* $\mathcal{K}$ *is* C^2*-cone reducible at every* $\overline{X} \in \mathcal{K}$*.*

Notice that for the CMatOP (3-36), it follows from [84, Proposition 3.3] that $\mathcal{K}$ satisfies C^2-cone reducible since the polyhedron $\widehat{\mathcal{K}}$ is C^2-cone reducible. We can also write the constraint of (3-36) into the Cartesian product form $(h(x,\bar{p}), g_2(x,\bar{p})) \in \{0\} \times \mathcal{K}$. From [1, Proposition 6.41], we have that

$$\mathcal{N}_{\{0\}\times\mathcal{K}}(h(x,\bar{p}), g_2(x,\bar{p})) = \mathcal{N}_{\{0\}}(h(x,\bar{p})) \times \mathcal{N}_{\mathcal{K}}(g_2(x,\bar{p})) = \mathbb{Y} \times \mathcal{N}_{\mathcal{K}}(g_2(x,\bar{p})).$$

If $(\bar{x}, \overline{Y}, \bar{y}, \overline{Z})$ is a KKT point of problem (3-36), it follows from the definition of coderivative (Definition 2.5) that $x^* \in \mathcal{D}^*\partial\delta_{\{0\}\times\mathcal{K}}(h(\bar{x},\bar{p}), g_2(\bar{x},\bar{p}), \bar{y}, \overline{Z})(y^*)$ if and only if

$$\begin{aligned} (x^*, -y^*) &\in \mathcal{N}_{\text{gph}\,\mathcal{N}_{\{0\}\times\mathcal{K}}}(h(\bar{x},\bar{p}), g_2(\bar{x},\bar{p}), \bar{y}, \overline{Z}) \\ &= \mathcal{N}_{\text{gph}\,\mathcal{N}_{\{0\}}}(h(\bar{x},\bar{p}), \bar{y}) \times \mathcal{N}_{\text{gph}\,\mathcal{N}_{\mathcal{K}}}(g_2(\bar{x},\bar{p}), \overline{Z}). \end{aligned}$$

Moreover, we have $\mathcal{N}_{\mathrm{gph}\,\mathcal{N}_{\{0\}}}(h(\bar{x},\bar{p}),\bar{y}) = \mathbb{Y}\times\{0\}$. So,

$$\begin{aligned}&\mathcal{D}^*\partial\delta_{\{0\}\times\mathcal{K}}(h(\bar{x},\bar{p}),g_2(\bar{x},\bar{p}),\bar{y},\overline{Z})(y_1^*,y_2^*)\\ &=\mathcal{D}^*\mathcal{N}_{\{0\}}(h(\bar{x},\bar{p}),\bar{y})(y_1^*)\times\mathcal{D}^*\mathcal{N}_{\mathcal{K}}(g_2(\bar{x},\bar{p}),\overline{Z})(y_2^*).\end{aligned} \tag{3-44}$$

In this subsection, $\phi(x,\bar{p})$ in (2-4) equals $f(x,\bar{p})+\theta_1(g_1(x,\bar{p}))+\delta_{\{0\}\times\mathcal{K}}(h(x,\bar{p}),g_2(x,\bar{p}))$. Suppose that $\bar{x}\in\mathbb{X}$ is a feasible point of problem (3-36) with $\bar{p}$. Then, it follows from [14, Proposition 2.2] that the lower semi-continuous function $\phi(x,p)$ is parametrically continuously prox-regular at $(\bar{x},\bar{p})$ for $\bar{v}\equiv 0$ and the basic constraint qualification (BCQ) of problem (3-36) holds at $(\bar{x},\bar{p})$ (recall Definition 2.10 and 2.11) under Robinson constraint qualification (see (3-61)). The following lemma which is an analogy to [16, Lemma 5.5] is critical in establishing the main result of this section. We write down the proof here for convenient reference.

Lemma 3.10. *Let $\bar{x}$ be a stationary point of (3-36). Suppose the nondegeneracy condition is satisfied at $\bar{x}$. Recall $\overline{X}_1=g_1(\bar{x},\bar{p})$ and $\overline{X}_2=g_2(\bar{x},\bar{p})$. Then for all $w\in\mathbb{X}$, we have*

$$\begin{aligned}\mathcal{D}^*\partial_x\phi(\bar{x},\bar{p},0)(w)=&\big(L''_{xx}(\bar{x},\bar{p},\overline{Y},\bar{y},\overline{Z})w,L''_{xp}(\bar{x},\bar{p},\overline{Y},\bar{y},\overline{Z})w\big)\\ &+g_1'(\bar{x},\bar{p})^*\mathcal{D}^*\partial\theta_1(\overline{X}_1,\overline{Y})\big((g_1)'_x(\bar{x},\bar{p})w\big)\\ &+G_2'(\bar{x},\bar{p})^*\mathcal{D}^*\partial\delta_{\{0\}\times\mathcal{K}}(G_2(\bar{x},\bar{p}),\bar{y},\overline{Z})\big((G_2)'_x(\bar{x},\bar{p})w\big),\end{aligned} \tag{3-45}$$

where $(\overline{Y},\bar{y},\overline{Z})$ is the multiplier with respect to $\bar{x}$, $G_2(\bar{x},\bar{p})=(h(\bar{x},\bar{p}),\overline{X}_2)$ and $g_1'(\bar{x},\bar{p})^*(\cdot)$ denotes the adjoint operator of $g_1'(\bar{x},\bar{p})(\cdot)$. Moreover, the condition (2-5) in Theorem 2.12 holds for ϕ with $\bar{v}=0$.

Proof. The details of the proof of (3-45) is exactly the same as in [16, Lemma 5.5]. The subdifferential sum rule to ϕ together with the coderivative sum rule [57, Theorem 1.62] implies

$$\begin{aligned}\mathcal{D}^*\partial_x\phi(\bar{x},\bar{p},0)(w)=&\big(f''_{xx}(\bar{x},\bar{p})w,f''_{xp}(\bar{x},\bar{p})w\big)\\ &+\mathcal{D}^*\partial_x(\theta_1\circ g_1+\delta_{\{0\}\times\mathcal{K}}\circ G_2)\big(\bar{x},\bar{p},-f'_x(\bar{x},\bar{p})\big)(w).\end{aligned} \tag{3-46}$$

Let $\Gamma(X_1,X_2)=\theta_1(X_1)+\delta_{\{0\}\times\mathcal{K}}(X_2)$ and $G(\bar{x},\bar{p})=(g_1(\bar{x},\bar{p}),G_2(\bar{x},\bar{p}))$. It follows that

$$\mathcal{D}^*\partial_x(\theta_1\circ g_1+\delta_{\{0\}\times\mathcal{K}}\circ G_2)\big(\bar{x},\bar{p},-f'_x(\bar{x},\bar{p})\big)(w)=\mathcal{D}^*\partial_x(\Gamma\circ G)\big(\bar{x},\bar{p},-f'_x(\bar{x},\bar{p})\big)(w).$$

Since $\mathcal{K}$ is C^2-cone reducible and the nondegeneracy conditions of problem (3-36) hold at $\bar{x}$, we can apply [85, Theorem 3.6] to get

$$\begin{aligned}\mathcal{D}^*\partial_x(\Gamma\circ G)(\bar{x},\bar{p},-f'_x(\bar{x},\bar{p}))(w)=&(\nabla^2_{xx}\langle\overline{\Sigma},G\rangle(\bar{x},\bar{p})w,\nabla^2_{xp}\langle\overline{\Sigma},G\rangle(\bar{x},\bar{p})w)\\ &+G'(\bar{x},\bar{p})^*\mathcal{D}^*\partial\Gamma(G(\bar{x},\bar{p}),\overline{\Sigma})(G'_x(\bar{x},\bar{p})w),\end{aligned} \tag{3-47}$$

where $\overline{\Sigma} = (\overline{Y}, \bar{y}, \overline{Z})$. It is easy to see that

$$\begin{aligned}&\mathcal{D}^*\partial\Gamma(G(\bar{x},\bar{p}),\overline{\Sigma})(G'_x(\bar{x},\bar{p})w)\\&\quad = \mathcal{D}^*\partial\Gamma\big((g_1(\bar{x},\bar{p}),G_2(\bar{x},\bar{p})),(\overline{Y},\bar{y},\overline{Z})\big)\big((g_1)'_x(\bar{x},\bar{p})w,(G_2)'_x(\bar{x},\bar{p})w\big).\end{aligned}$$

By [1, Proposition 10.5], we have $\partial\Gamma(X_1,X_2) = \partial\theta_1(X_1)\times\partial\delta_{\{0\}\times\mathcal{K}}(X_2)$. So for all $S \in \mathcal{D}^*\partial\Gamma\big((g_1(\bar{x},\bar{p}),G_2(\bar{x},\bar{p})),(\overline{Y},\bar{y},\overline{Z})\big)\big((g_1)'_x(\bar{x},\bar{p})w,(G_2)'_x(\bar{x},\bar{p})w\big)$, we have

$$\begin{aligned}(S,-(g_1)'_x(\bar{x},\bar{p})w,-(G_2)'_x(\bar{x},\bar{p})w) &\in \mathcal{N}_{\mathrm{gph}\,\partial\Gamma}\big((g_1(\bar{x},\bar{p}),G_2(\bar{x},\bar{p})),(\overline{Y},\bar{y},\overline{Z})\big)\\&= \mathcal{N}_{\mathrm{gph}\,\partial\theta_1}(g_1(\bar{x},\bar{p}),\overline{Y})\times\mathcal{N}_{\mathrm{gph}\,\partial\delta_{\{0\}\times\mathcal{K}}}(G_2(\bar{x},\bar{p}),\bar{y},\overline{Z}).\end{aligned}$$

It follows that

$$\begin{aligned}&\mathcal{D}^*\partial\Gamma\big((g_1(\bar{x},\bar{p}),G_2(\bar{x},\bar{p})),(\overline{Y},\bar{y},\overline{Z})\big)\big((g_1)'_x(\bar{x},\bar{p})w,(G_2)'_x(\bar{x},\bar{p})w\big)\\&\quad = \mathcal{D}^*\partial\theta_1(g_1(\bar{x},\bar{p}),\overline{Y})\big((g_1)'_x(\bar{x},\bar{p})w\big)\times\mathcal{D}^*\partial\delta_{\{0\}\times\mathcal{K}}(G_2(\bar{x},\bar{p}),\bar{y},\overline{Z})\big((G_2)'_x(\bar{x},\bar{p})w\big).\end{aligned}$$

The above equation together with (3-47) implies that

$$\begin{aligned}\mathcal{D}^*\partial_x(\Gamma\circ G)(\bar{x},\bar{p},-f'_x(\bar{x},\bar{p}))(w) =&\big(\nabla^2_{xx}\langle\overline{\Sigma},G\rangle(\bar{x},\bar{p})w,\nabla^2_{xp}\langle\overline{\Sigma},G\rangle(\bar{x},\bar{p})w\big)\\&+g'_1x(\bar{x},\bar{p})^*\mathcal{D}^*\partial\theta_1(g_1(\bar{x},\bar{p}),\overline{Y})\big((g_1)'_x(\bar{x},\bar{p})w\big)\\&+G'_2x(\bar{x},\bar{p})^*\mathcal{D}^*\partial\delta_{\{0\}\times\mathcal{K}}(G_2(\bar{x},\bar{p}),\bar{y},\overline{Z})\big((G_2)'_x(\bar{x},\bar{p})w\big).\end{aligned} \tag{3-48}$$

Combining (3-46) and (3-48), we have (3-45).

Next we only need to verify the condition (2-5), i.e., if $(0,q)\in\mathcal{D}^*\partial\phi_x(\bar{x},\bar{p},0)(0)$, then $q=0$. It follows from (3-45) that we shall show $q=0$ if

$$(0,q)\in g'_1(\bar{x},\bar{p})^*\mathcal{D}^*\partial\theta_1(\overline{X}_1,\overline{Y})(0)+G'_2(\bar{x},\bar{p})^*\mathcal{D}^*\partial\delta_{\{0\}\times\mathcal{K}}(G_2(\bar{x},\bar{p}),\bar{y},\overline{Z})(0). \tag{3-49}$$

From (3-49), we know $(0,q)=(w,q_1)+(-w,q_2)$, where $(w,q_1)\in g'_1(\bar{x},\bar{p})^*\mathcal{D}^*\partial\theta_1(\overline{X}_1,\overline{Y})(0)$ and $(-w,q_2)\in G'_2(\bar{x},\bar{p})^*\mathcal{D}^*\partial\delta_{\{0\}\times\mathcal{K}}(G_2(\bar{x},\bar{p}),\bar{y},\overline{Z})(0)$. By (3-49), we know that there exists $S_1\in\mathcal{D}^*\partial\theta_1(\overline{X}_1,\overline{Y})(0)$ and $S_2\in\mathcal{D}^*\partial\delta_{\{0\}\times\mathcal{K}}(G_2(\bar{x},\bar{p}),\bar{y},\overline{Z})(0)$ such that

$$\begin{cases}0=(g_1)'_x(\bar{x},\bar{p})^*S_1+(G_2)'_x(\bar{x},\bar{p})^*S_2,\\ q=(g_1)'_p(\bar{x},\bar{p})^*S_1+(G_2)'_p(\bar{x},\bar{p})^*S_2.\end{cases} \tag{3-50}$$

Since $S_1\in\mathcal{D}^*\partial\theta_1(\overline{X}_1,\overline{Y})(0)$ and $S_2\in\mathcal{D}^*\partial\delta_{\{0\}\times\mathcal{K}}(G_2(\bar{x},\bar{p}),\bar{y},\overline{Z})(0)$, we have $(S_2,0)\in\mathcal{N}_{\mathrm{gph}\,\partial\theta_1}(\overline{X}_1,\overline{Y})$ and $(S_2,0)\in\mathcal{N}_{\mathrm{gph}\,\partial\delta_{\{0\}\times\mathcal{K}}}(G_2(\bar{x},\bar{p}),\bar{y},\overline{Z})$. It then follows from the nondegeneracy condition (Definition 3.7) that there exist $z\in\mathbb{X}$, $\mathrm{H}:=(H_1,H_2)\in\mathcal{T}^{\mathrm{lin}}_{\theta_1}(\overline{X}_1)\times\big(\{0\}\times\mathrm{lin}(\mathcal{T}_{\mathcal{K}}(X_2))\big)$ such that

$$\mathrm{S}:=\begin{pmatrix}S_1\\S_2\end{pmatrix}=\begin{pmatrix}(g_1)'_x(\bar{x},\bar{p})\\(G_2)'_x(\bar{x},\bar{p})\end{pmatrix}z+\mathrm{H},$$

which together with the first equality in (3-50), yields

$$\begin{aligned}\|\mathrm{S}\|^2 &= \langle \mathrm{S}, \begin{pmatrix} (g_1)'_x(\bar{x},\bar{p}) \\ (G_2)'_x(\bar{x},\bar{p}) \end{pmatrix} z + \mathrm{H}\rangle \\ &= \langle (g_1)'_x(\bar{x},\bar{p})^* S_1 + (G_2)'_x(\bar{x},\bar{p})^* S_2, z\rangle + \langle \mathrm{S},\mathrm{H}\rangle = \langle \mathrm{S},\mathrm{H}\rangle. \\ &= \langle S_1, H_1\rangle + \langle S_2, H_2\rangle \end{aligned} \tag{3-51}$$

Therefore, the desired result then follows if we show $\langle \mathrm{S},\mathrm{H}\rangle \leq 0$. We only show $\langle S_1, H_1\rangle \leq 0$ here since the proof of $\langle S_2, H_2\rangle \leq 0$ is established in a similar way. For notation simplicity, we drop the lower index 1 of S_1 and H_1 during the following proof.

By noting $(S,0) \in \mathcal{N}_{\mathrm{gph}\,\partial\theta_1}(\overline{X}_1,\overline{Y})$, it follows from Theorem 3.4 that there exist $Q \in \{Q \in \mathbb{R}^{m\times m} \mid Q = \mathrm{Diag}\,(Q_1, Q_2, \ldots, Q_{r_1}),\ Q_l \in \mathcal{O}^{|\alpha^l|}\}$ and $\Xi_1 = \lim\limits_{k\to\infty} W(w^k)$ such that $X^k \to \overline{X}_1$ and $Y^k \to \overline{Y}$ with $X^k = Q^T\overline{P}^T \mathrm{Diag}(\lambda(\overline{X}_1) + \Pi_{\mathcal{C}_{\varpi_1}(\bar{s},\bar{y})}(w^k))\overline{P}Q$, $Y^k = Q^T\overline{P}^T \mathrm{Diag}(\lambda(\overline{Y}) + \Pi_{\mathcal{C}^\circ_{\varpi_1}(\bar{s},\bar{y})}(w^k))\overline{P}Q$. It is clear that $\langle S, H\rangle = \langle \widetilde{S}, \widetilde{H}\rangle$ with $\widetilde{S} = Q^T\overline{P}^T S\overline{P}Q, \widetilde{H} = Q^T\overline{P}^T H\overline{P}Q$. From Theorem 3.4, we know $\overline{D}\circ\widetilde{S} + \overline{F}\circ 0 = 0$, which implies $\widetilde{S} = \mathrm{Diag}\,(\widetilde{S}_{\mu^1\mu^1}, \ldots, \widetilde{S}_{\mu^{d_1}\mu^{d_1}})$, where μ^s is given in (3-40). By using Proposition 2.21 and $H \in \mathcal{T}^{\mathrm{lin}}_{\theta_1}(\overline{X}_1)$, we have

$$\begin{aligned}\langle S, H\rangle = \langle \widetilde{S}, \widetilde{H}\rangle &= \sum_{s=1}^{d_1}\langle \widetilde{S}_{\mu^s\mu^s}, \widetilde{H}_{\mu^s\mu^s}\rangle = \sum_{s=1}^{d_1}\langle \widetilde{S}_{\mu^s\mu^s}, \widehat{\rho}_s I_{|\mu^s|}\rangle \\ &= \sum_{l=1}^{r_1}\sum_{t=1}^{s^l}\langle \widetilde{S}_{\rho^l_t\rho^l_t}, \widetilde{H}_{\rho^l_t\rho^l_t}\rangle \\ &\leq \sum_{l=1}^{r_1}\sum_{t=1}^{s^l}\langle \lambda(\widetilde{S}_{\rho^l_t\rho^l_t}), \lambda(\widetilde{H}_{\rho^l_t\rho^l_t})\rangle = \langle \kappa(\mathscr{D}^\pi(\widetilde{S})), \kappa(\mathscr{D}^\pi(\widetilde{H}))\rangle = 0,\end{aligned}$$

where the definition of ρ^l_t is given in (3-24). The last equality holds since $\kappa(\mathscr{D}^\pi(\widetilde{S})) = \sum\limits_{i,j\in\eta^k_1} u^{ji}(a^j - a^i) + \sum\limits_{i\in\eta^k_1, j\in\iota^k_1\setminus\eta^k_1} g^{ji}(a^j - a^i)$, where $g^{ji} \geq 0$, $\eta^k_1 = \eta_1(\lambda(X^k), \lambda(Y^k))$, $\iota^k_1 = \iota_1(\lambda(X^k))$. By using Proposition 2.21 and $H \in \mathcal{T}^{\mathrm{lin}}_{\theta_1}(\overline{X}_1)$ again, we have $\langle \kappa(\mathscr{D}^\pi(\widetilde{H})), a^j - a^i\rangle = \langle \lambda'(\overline{X}_1; H), a^j - a^i\rangle = 0$ for any $i, j \in \bar{\iota}_1$. Then we obtain the last equality. Thus, we know from (3-51) that $\mathrm{S} = 0$, which implies $q = 0$.

□

Let $\bar{x} \in \mathbb{X}$ be a stationary point of problem (3-36) and $(\overline{Y}, \bar{y}, \overline{Z}) \in \mathcal{M}(\bar{x})$. Since $\mathcal{M}(\bar{x})$ is nonempty and single valued under the condition of nondegeneracy, the critical cone of (3-36) can be defined as

$$\mathcal{C}(\bar{x}) = \{d \in \mathbb{X} \mid h'_x(\bar{x},\bar{p})d = 0,\ (g_1)'_x(\bar{x},\bar{p})d \in \mathcal{C}_{\theta_1}(\overline{X}_1,\overline{Y}),\ (g_2)'_x(\bar{x},\bar{p})d \in \mathcal{C}_{\mathcal{K}}(\overline{X}_2,\overline{Z})\},$$

where $\mathcal{C}_{\theta_1}(\overline{X}_1,\overline{Y})$ and $\mathcal{C}_{\mathcal{K}}(\overline{X}_2,\overline{Z})$ are the critical cone defined in (2-3) for functions θ_1 and $\delta_{\mathcal{K}}$. Here we assume it to be non-empty. We defined the outer approximation set to

$\mathcal{C}(\bar{x})$ with respect to $(\overline{Y}, \bar{y}, \overline{Z}) \in \mathcal{M}(\bar{x})$ as

$$\operatorname{app}(\overline{Y}, \bar{y}, \overline{Z}) = \left\{ d \in \mathbb{X} \;\middle|\; \begin{array}{l} h'_x(\bar{x}, \bar{p})d = 0, \quad (g_1)'_x(\bar{x}, \bar{p})d \in \operatorname{aff} \mathcal{C}_{\theta_1}(\overline{X}_1, \overline{Y}), \\ (g_2)'_x(\bar{x}, \bar{p})d \in \operatorname{aff} \mathcal{C}_{\mathcal{K}}(\overline{X}_2, \overline{Z}) \end{array} \right\}. \tag{3-52}$$

The definition of strong second order sufficient condition (SOSC) for problem (3-36) is stated in the following sense.

Definition 3.11. *Let $\bar{x} \in \mathbb{X}$ be a stationary point of problem (3-36). We say the strong second order sufficient condition (SOSC) holds at $\bar{x} \in \mathbb{X}$ if for all $0 \neq d \in \bigcap_{(\overline{Y},\bar{y},\overline{Z}) \in \mathcal{M}(\bar{x})} \operatorname{app}(\overline{Y}, \bar{y}, \overline{Z})$*

$$\sup_{(\overline{Y},\bar{y},\overline{Z}) \in \mathcal{M}(\bar{x})} \left\{ \langle d, L''_{xx}(\bar{x}, \bar{p}, \overline{Y}, \bar{y}, \overline{Z})d \rangle - \Upsilon^1_{\overline{X}_1}\big(\overline{Y}, (g_1)'_x(\bar{x}, \bar{p})d\big) - \Upsilon^2_{\overline{X}_2}\big(\overline{Z}, (g_2)'_x(\bar{x}, \bar{p})d\big) \right\} > 0,$$

where the function $\Upsilon^1_{\overline{X}_1}(\overline{Y}, H) : \partial\theta_1(\overline{X}_1) \times \mathbb{S}^m \to \mathbb{R}$ and $\Upsilon^2_{\overline{X}_2}(\overline{Z}, H) : \mathcal{N}_{\mathcal{K}}(\overline{X}_2) \times \mathbb{S}^n \to \mathbb{R}$ are the σ-term for problem (3-36). From (2-27) *and* (2-40)*, we know that*

$$\Upsilon^1_{\overline{X}_1}(\overline{Y}, H) = -2 \sum_{1 \le p \le p' \le d_1} \sum_{i \in \mu^p} \sum_{j \in \mu^{p'}} \frac{\lambda_i(\overline{Y}) - \lambda_j(\overline{Y})}{\lambda_i(\overline{X}_1) - \lambda_j(\overline{X}_1)} (\overline{P}^T_{\mu^p} H \overline{P}_{\mu^{p'}})^2_{ij} \tag{3-53}$$

and

$$\Upsilon^2_{\overline{X}_2}(\overline{Z}, H) = -2 \sum_{1 \le p \le p' \le d_2} \sum_{i \in \nu^p} \sum_{j \in \nu^{p'}} \frac{\lambda_i(\overline{Z}) - \lambda_j(\overline{Z})}{\lambda_i(\overline{X}_2) - \lambda_j(\overline{X}_2)} (\overline{V}^T_{\nu^p} H \overline{V}_{\nu^p})^2_{ij}, \tag{3-54}$$

where $\overline{P} \in \mathcal{O}^m(\overline{X}_1) \cap \mathcal{O}^m(\overline{Y})$ and $\overline{V} \in \mathcal{O}^n(\overline{X}_2) \cap \mathcal{O}^n(\overline{Z})$.

The next result tells us that the "inf" term in second order subdifferential condition (see Theorem 3.13) is exactly the σ-term in strong SOSC. Denote

$$\mathcal{C}_1(\overline{X}_1, \overline{Y}) = \left\{ d \in \mathbb{R}^m \;\middle|\; \begin{array}{l} \langle d, a^j - a^i \rangle = 0, i, j \in \bar{\eta}_1, \\ \langle d, a^j - a^i \rangle < 0, i \in \bar{\eta}_1, j \in \bar{\iota}_1 \backslash \bar{\eta}_1 \end{array} \right\}$$

and

$$\mathcal{C}_2(\overline{X}_2, \overline{Z}) = \left\{ d \in \mathbb{R}^m \;\middle|\; \begin{array}{l} \langle d, b^i \rangle = 0, i \in \bar{\eta}_2, \\ \langle d, b^j \rangle < 0, j \in \bar{\iota}_2 \backslash \bar{\eta}_2 \end{array} \right\}.$$

For any matrix $H \in \mathbb{S}^m$, we apply the following notations, $\widehat{H} = \overline{P}^T H \overline{P}$, $\widetilde{H} = Q^T \widehat{H} Q$, where $\overline{P} \in \mathcal{O}^m(\overline{X}_1) \cap \mathcal{O}^m(\overline{Y})$, $Q \in \mathcal{O}_1 = \{Q \in \mathbb{R}^{m \times m} \mid Q = \operatorname{Diag}(Q_1, Q_2, \ldots, Q_{r_1}),\; Q_l \in \mathcal{O}^{|\alpha^l|}\}$. Also let $\widetilde{P} = \overline{P}Q$. Similarly, for any matrix $S \in \mathbb{S}^n$, let $\widehat{S} = \overline{V}^T S \overline{V}$, $\widetilde{S} = Q^T \widehat{S} Q$, where $\overline{V} \in \mathcal{O}^n(\overline{X}_2) \cap \mathcal{O}^n(\overline{Z})$, $Q \in \mathcal{O}_2 = \{Q \in \mathbb{R}^{n \times n} \mid Q = \operatorname{Diag}(Q_1, Q_2, \ldots, Q_{r_2}),\; Q_l \in \mathcal{O}^{|\varsigma^l|}\}$. Also let $\widetilde{V} = \overline{V}Q$.

Proposition 3.12. *Let $\bar{x}$ be a stationary point of problem (3-36) and $(\overline{Y}, \bar{y}, \overline{Z})$ be a unique Lagrange multiplier of the corresponding KKT system under the validity of the constraint nondegeneracy condition* (5-12) *at $\bar{x}$. Recall $\overline{X}_1 = g_1(\bar{x}, \bar{p})$ and $\overline{X}_2 = g_2(\bar{x}, \bar{p})$. If $\mathcal{C}_1(\overline{X}_1, \overline{Y}) \neq \emptyset$ and $\mathcal{C}_2(\overline{X}_2, \overline{Z}) \neq \emptyset$, then*

(i) $\mathrm{dom}\,\mathcal{D}^*\partial\theta_1(\overline{X}_1, \overline{Y})\big((g_1)'_x(\bar{x}, \bar{p})(\cdot)\big) \cap \mathrm{dom}\,\mathcal{D}^*\mathcal{N}_{\{0\}\times\mathcal{K}}\big(G_2(\bar{x}, \bar{p}), \bar{y}, \overline{Z}\big)\big(d(\cdot)\big) \supseteq \mathrm{app}\,(\overline{Y}, \bar{y}, \overline{Z})$*, where* $G_2(\bar{x}, \bar{p}) = \big(h(\bar{x}, \bar{p}), \overline{X}_2\big)$ *and* $d(\cdot) = \big(h'_x(\bar{x}, \bar{p})(\cdot), (g_2)'_x(\bar{x}, \bar{p})(\cdot)\big)$;

(ii) for any $w \in \mathrm{app}(\overline{Y}, \bar{y}, \overline{Z})$, we have

$$-\Upsilon^1_{\overline{X}_1}\big(\overline{Y}, (g_1)'_x(\bar{x}, \bar{p})w\big) = \inf\Big\{\langle T, (g_1)'_x(\bar{x}, \bar{p})w\rangle : T \in \mathcal{D}^*\partial\theta_1(\overline{X}_1, \overline{Y})\big((g_1)'_x(\bar{x}, \bar{p})w\big)\Big\} \tag{3-55}$$

and

$$\begin{aligned} -\Upsilon^2_{\overline{X}_2}\big(\overline{Z}, (g_2)'_x(\bar{x}, \bar{p})w\big) &= \inf\big\{\langle T, (G_2)'_x(\bar{x}, \bar{p})w\rangle : T \in \mathcal{D}^*\mathcal{N}_{\{0\}\times\mathcal{K}}\big(G_2(\bar{x}, \bar{p}), \bar{y}, \overline{Z}\big)\big(d(w)\big)\big\} \\ &= \inf\big\{\langle T, (g_2)'_x(\bar{x}, \bar{p})w\rangle : T \in \mathcal{D}^*\mathcal{N}_{\mathcal{K}}\big(\overline{X}_2, \overline{Z}\big)\big((g_2)'_x(\bar{x}, \bar{p})w\big)\big\}. \end{aligned} \tag{3-56}$$

Proof. To justify (i), pick $w \in \mathrm{app}\,(\overline{Y}, \bar{y}, \overline{Z})$ and define $W = (g_1)'_x(\bar{x}, \bar{p})w$. For any $w \in \overline{\eta}_1$, we first show that

$$\langle \kappa(\mathscr{D}^{\pi}(-\widetilde{W})), a^w\rangle = \langle \mathrm{diag}\,(-\widehat{W}), a^w\rangle. \tag{3-57}$$

In fact, for each $p \in \{1, \ldots, d_1\}$, if $q \in \mathcal{E}^p$ (recall (3-41)), we derive from Proposition 2.23 that there exists a scalar $\widetilde{v}^p_q$ such that $(\overline{P}^T_{\mu^p} W \overline{P}_{\mu^p})_{\chi^p_q\chi^p_q} = \widetilde{v}^p_q I_{|\chi^p_q|}$. If $q \notin \mathcal{E}^p$, then for any $w \in \overline{\eta}_1$, there exists a scalar $\widetilde{u}^q_p$ such that

$$(a^w)_i = (a^w)_j = \widetilde{u}^q_p \quad \forall\, i, j \in \chi^p_q,$$

which yields that for any $w \in \overline{\eta}_1$,

$$\begin{aligned} \langle \mathrm{diag}\,(-\widehat{W}), a^w\rangle &= \sum_{p=1}^{d_1}\Big(\sum_{q\in\mathcal{E}^p}\langle \mathrm{diag}\,(-\widehat{W})_{\chi^p_q}, (a^w)_{\chi^p_q}\rangle + \sum_{q\notin\mathcal{E}^p}\langle \mathrm{diag}\,(-\widehat{W})_{\chi^p_q}, (a^w)_{\chi^p_q}\rangle\Big) \\ &= \sum_{p=1}^{d_1}\Big(\sum_{q\in\mathcal{E}^p}\langle \widetilde{v}^p_q 1_{|\chi^p_q|}, (a^w)_{\chi^p_q}\rangle + \sum_{q\notin\mathcal{E}^p}\langle \mathrm{diag}\,(-\widehat{W})_{\chi^p_q}, \widetilde{u}^q_p 1_{|\chi^p_q|}\rangle\Big) \\ &= \langle \kappa(\mathscr{D}^{\pi}(-\widetilde{W})), a^w\rangle, \end{aligned}$$

where $1_{|\chi^p_q|} = (1, \ldots, 1)^T \in \mathbb{R}^{|\chi^p_q|}$. Thus, we obtain (3-57).

Since $\mathcal{C}_1(\overline{X}_1, \overline{Y}) \neq \emptyset$, we know that there exists a sequence $\{w^k\}$ converges to 0 such that for each k, $w^k \in \mathrm{Ker}\,\{a^j - a^i\}_{i,j\in\overline{\eta}_1}$ and $\langle w^k, a^j - a^i\rangle < 0$, $i \in \overline{\eta}_1$, $j \in \bar{\iota}_1\backslash\overline{\eta}_1$. From Corollary 2.19, we have for all permutation matrix $U \in \mathcal{P}^m_{\lambda(\overline{X}_1)} \cap \mathcal{P}^m_{\lambda(\overline{Y})}$ ($\lambda(\overline{X}_1) = U\lambda(\overline{X}_1)$ and $\lambda(\overline{Y}) = U\lambda(\overline{Y})$) and $i \in \overline{\eta}_1$, there exists $i' \in \overline{\eta}_1$ and $U \in \mathcal{P}^m_{\lambda(\overline{X}_1)}$ such that $a^{i'} = Ua^i$. Moreover, for all $j \in \bar{\iota}_1\backslash\overline{\eta}_1$, we know there exists $j' \in \bar{\iota}_1$ such that $a^{j'} = Ua^j$ and

$j' \in \bar{\iota}_1 \backslash \bar{\eta}_1$. Thus if $w^k \in \mathcal{C}_1(\overline{X}_1, \overline{Y})$, it follows that for all $U \in \mathcal{P}^m_{\lambda(\overline{X}_1)} \cap \mathcal{P}^m_{\lambda(\overline{Y})}$, $Uw^k \in \mathcal{C}_1(\overline{X}_1, \overline{Y})$. Therefore, we may assume $w^k \in \mathcal{R}^m_{\gtrsim}(\overline{A})$, where $\overline{A} = \overline{X}_1 + \overline{Y}$. Moreover, we have $\Pi_{\mathcal{C}_{\varpi_1}(\lambda(\overline{X}_1), \lambda(\overline{Y}))}(w^k) = w^k$ is satisfied.

For each k, let $x^k := \lambda(\overline{X}_1) + w^k$. It follows from the B-differentiability of $\mathrm{Pr}_{\varpi_1}(\cdot)$ that $x^k = \mathrm{Pr}_{\varpi_1}(\lambda(\overline{A})) + \Pi_{\mathcal{C}_{\varpi_1}(\lambda(\overline{X}_1), \lambda(\overline{Y}))}(w^k) = \mathrm{Pr}_{\varpi_1}(\lambda(\overline{A}) + w^k)$. Therefore, by taking a subsequence if necessary, it then follows from the symmetry of ϖ_1 and Ky Fan's inequality that the components of x^k are actually in a non-increasing order for k sufficiently large. For each k, define $X^k = \overline{P}\mathrm{Diag}(x^k)\overline{P}^T$. Thus, we have $\iota_1(\lambda(X^k)) = \bar{\eta}_1$. Assuming $\lambda(\overline{Y}) = \sum_{i \in \bar{\eta}_1} g_i a^i$, and then we let $\lambda(Y^k) = \lambda(\overline{Y})$. When k is sufficiently large, we know $\bar{\eta}_1 = \eta_1(\lambda(X^k), \lambda(Y^k))$. By choosing Ξ_1 such that $\Xi_1 = \lim_{k \to \infty} W(w^k)$. It follows from (3-57) and Proposition 2.23 that $\langle \kappa(\mathscr{D}^\pi(-\widetilde{W})), a^i - a^j \rangle = 0, \forall (i,j) \in \beta_+ = \bar{\eta}_1$. Find a $\widetilde{T}$ satisfying the first three relationships in (3-32), specifically let $\mathscr{D}^\pi(\widetilde{T}) = 0$. It follows from Theorem 3.4 that $T \in \mathcal{D}^*\partial\theta_1(\overline{X}_1, \overline{Y})(W)$. Thus we have $w \in \mathrm{dom}\, \mathcal{D}^*\partial\theta_1(\overline{X}_1, \overline{Y})\big((g_1)'_x(\bar{x}, \bar{p})(\cdot)\big)$.

Then we prove $w \in \mathrm{dom}\, \mathcal{D}^*\mathcal{N}_{\{0\}\times\mathcal{K}}(G_2(\bar{x}, \bar{p}), \bar{y}, \overline{Z})\big(d(\cdot)\big)$. From (3-44), we have

$$
\begin{aligned}
&\mathcal{D}^*\partial\delta_{\{0\}\times\mathcal{K}}(h(\bar{x}, \bar{p}), \overline{X}_2, \bar{y}, \overline{Z})\big(h'_x(\bar{x}, \bar{p})w, (g_2)'_x(\bar{x}, \bar{p})w\big) \\
&= \mathcal{D}^*\mathcal{N}_{\{0\}}(h(x, \bar{p}), \bar{y})(h'_x(\bar{x}, \bar{p})w) \times \mathcal{D}^*\mathcal{N}_{\mathcal{K}}(\overline{X}_2, \overline{Z})\big((g_2)'_x(\bar{x}, \bar{p})w\big).
\end{aligned} \tag{3-58}
$$

From (3-52), we know that $h'_x(\bar{x}, \bar{p})w = 0$. It follows that $T \in \mathcal{D}^*\mathcal{N}_{\{0\}}(h(x, \bar{p}), \bar{y})(h'_x(\bar{x}, \bar{p})w)$ for all $T \in \mathbb{Y}$. The verification of $\mathcal{D}^*\mathcal{N}_{\mathcal{K}}(\overline{X}_2, \overline{Z})\big((g_2)'_x(\bar{x}, \bar{p})w\big) \neq \emptyset$ is similar to the discussion above. The remaining proof is also in a similar approach to that of case θ_1, we omit the details of the proof for simplicity. Thus we have completed the proof of (i).

Next, we shall show (ii) holds. We first prove that the inequality "$\leq$" of (3-55) holds. For all $w \in \mathrm{app}(\overline{Y}, \bar{y}, \overline{Z})$, we observe that for $T \in \mathcal{D}^*\partial\theta_1(\overline{X}_1, \overline{Y})(W)$,

$$
\begin{aligned}
\langle T, W \rangle &= \mathrm{Tr}\,(\widetilde{P}\widetilde{T}^T\widetilde{P}^T\widetilde{P}\widetilde{W}\widetilde{P}^T) = \mathrm{Tr}\,(\widetilde{T}^T\widetilde{W}) \\
&= \sum_{p=1}^{d_1}\sum_{p'=1}^{d_1} \mathrm{Tr}\,(\widetilde{T}_{\mu^p\mu^{p'}}\widetilde{W}_{\mu^{p'}\mu^p}) \\
&= \sum_{p=1}^{d_1} \mathrm{Tr}\,(\widetilde{T}_{\mu^p\mu^p}\widetilde{W}_{\mu^p\mu^p}) + 2\sum_{1 \leq p < p' \leq d_1} \mathrm{Tr}\,(\widetilde{T}_{\mu^p\mu^{p'}}\widetilde{W}_{\mu^{p'}\mu^p}).
\end{aligned} \tag{3-59}
$$

From Theorem 3.4, we have for any $i \in \mu^p$ and $j \in \mu^{p'}$,

$$
\frac{\lambda_i(\overline{X}_1) - \lambda_j(\overline{X}_1)}{\lambda_i(\overline{A}) - \lambda_j(\overline{A})}\widetilde{T}_{ij} - \frac{\lambda_i(\overline{Y}) - \lambda_j(\overline{Y})}{\lambda_i(\overline{A}) - \lambda_j(\overline{A})}\widetilde{W}_{ij} = 0,
$$

which ensures therefore the equalities

$$
\begin{aligned}
2\sum_{1\le p<p'\le d_1}\mathrm{Tr}\,(\widetilde{T}_{\mu^p\mu^{p'}}\widetilde{W}_{\mu^{p'}\mu^p}) &= 2\sum_{1\le p<p'\le d_1}\mathrm{Tr}\,(\widehat{T}_{\mu^p\mu^{p'}}\widehat{W}_{\mu^{p'}\mu^p}) \\
&= 2\sum_{1\le p<p'\le d_1}\sum_{i\in\mu^p,\, j\in\mu^{p'}}\frac{\lambda_i(\overline{Y})-\lambda_j(\overline{Y})}{\lambda_i(\overline{X}_1)-\lambda_j(\overline{X}_1)}(\widehat{W}_{ij})^2 \\
&= -\Upsilon^1_{\overline{X}_1}\big(\overline{Y},(g_1)'_x(\bar{x},\bar{p})w\big). \qquad (3\text{-}60)
\end{aligned}
$$

As for the other part, from Proposition 2.23, we know $(\widehat{W})_{\mu^p\mu^p}$ is block diagonal, which implies $(\widetilde{W})_{\mu^p\mu^p}$ is also block diagonal. Thus

$$
\begin{aligned}
\sum_{p=1}^{d_1}\mathrm{Tr}\,(\widetilde{T}_{\mu^p\mu^p}(-\widetilde{W})_{\mu^p\mu^p}) &= \sum_{p=1}^{d_1}\sum_{q=1}^{u_q^p}\mathrm{Tr}\,(\widetilde{T}_{\chi_q^p\chi_q^p}(-\widetilde{W})_{\chi_q^p\chi_q^p}) = \sum_{l=1}^{r_1}\mathrm{Tr}\,(\widetilde{T}_{\alpha^l\alpha^l}(-\widetilde{W})_{\alpha^l\alpha^l}) \\
&= \sum_{l=1}^{r_1}\Big(\sum_{t=1}^{s^l}\mathrm{Tr}\,(\widetilde{T}_{\rho_t^l\rho_t^l}(-\widetilde{W})_{\rho_t^l\rho_t^l}) + 2\sum_{1\le t<t'\le s^l}\mathrm{Tr}\,(\widetilde{T}_{\rho_t^l\rho_{t'}^l}(-\widetilde{W})_{\rho_{t'}^l\rho_t^l})\Big) \\
&\le \langle\kappa(\mathscr{D}^\pi(\widetilde{T})),\kappa(\mathscr{D}^\pi(-\widetilde{W}))\rangle + 2\sum_{l=1}^{r_1}\sum_{1\le t<t'\le s^l}\mathrm{Tr}\,(\widetilde{T}_{\rho_t^l\rho_{t'}^l}(-\widetilde{W})_{\rho_{t'}^l\rho_t^l}) \\
&= \langle\kappa(\mathscr{D}^\pi(\widetilde{T})),\kappa(\mathscr{D}^\pi(-\widetilde{W}))\rangle - 2\sum_{l=1}^{r_1}\sum_{1\le t<t'\le s^l}\sum_{i\in\rho_t^l,\, j\in\rho_{t'}^l} f_{ij}.
\end{aligned}
$$

where

$$
f_{ij} = \begin{cases} \frac{(\Xi_1)_{ij}}{(\Xi_2)_{ij}}(\widetilde{T}_{ij})^2, & \text{if } (\Xi_2)_{ij}\neq 0 \\ \frac{(\Xi_2)_{ij}}{(\Xi_1)_{ij}}(\widetilde{W}_{ij})^2, & \text{otherwise.}\end{cases}
$$

Obviously, the second part in the last equation ≤ 0. The first part in the last equation ≤ 0 because $\kappa(\mathscr{D}^\pi(\widetilde{T}))\in\mathcal{C}^\circ_{\varpi_1}\big(\lambda(\overline{X}_1),\lambda(\overline{Y})\big)$ and $\kappa(\mathscr{D}^\pi(-\widetilde{W}))\in\mathcal{C}_{\varpi_1}\big(\lambda(\overline{X}_1),\lambda(\overline{Y})\big)$. Therefore, we know that

$$
\sum_{l=1}^{d_1}\mathrm{Tr}\,(\widetilde{T}_{\mu^l\mu^l}\widetilde{W}_{\mu^l\mu^l})\ge 0,
$$

which together with (3-59) and (3-60) implies that for any $w\in\mathrm{app}(\overline{Y},\bar{y},\overline{Z})$,

$$
-\Upsilon^1_{\overline{X}_1}\big(\overline{Y},(g_1)'_x(\bar{x},\bar{p})w\big)\le\inf\Big\{\langle T,(g_1)'_x(\bar{x},\bar{p})w\rangle : T\in\mathcal{D}^*\partial\theta_1(\overline{X}_1,\overline{Y})\big((g_1)'_x(\bar{x},\bar{p})w\big)\Big\}.
$$

The inequality "$\ge$" follows from the above discussion when we pick $(\Xi_1)_{ij}=1,\forall i\in\rho_t^l, j\in\rho_{t'}^l, t\neq t', 1\le l\le r_1$; otherwise $(\Xi_1)_{ij}=0$. Let $\mathscr{D}^\pi(\widetilde{T})=0$. It follows that $\widetilde{T}_{\alpha^l\alpha^l}=0=\widetilde{T}_{\chi_q^p\chi_q^p}$ for some $1\le p\le d_1, 1\le q\le u^p$. Since for all $i\in\chi_q^p, j\in\chi_{q'}^p, q\neq q'$, we have $(\overline{D})_{ij}=0$, which implies the corresponding $\widetilde{T}_{ij}$ can be free. Thus we can assume $\widetilde{T}_{\mu^p\mu^p}=0$, which implies $\sum_{p=1}^{d_1}\mathrm{Tr}\,(\widetilde{T}_{\mu^p\mu^p}(-\widetilde{W})_{\mu^p\mu^p})=0$. Thus, we have find a T such that $\langle T,W\rangle=-\Upsilon^1_{\overline{X}_1}\big(\overline{Y},(g_1)'_x(\bar{x},\bar{p})w\big)$. From (3-59), the inequality "$\ge$" holds. Thus we have proved (3-55).

The proof of first equation in (3-56) is similar to that of (3-55). Its second equation comes from (4-50) and condition $h'_x(\bar{x},\bar{p})w = 0$. □

Remark 3.2. We claim that $\mathcal{C}_1(\overline{X}_1,\overline{Y}) \neq \emptyset$ (respectively $\mathcal{C}_2(\overline{X}_2,\overline{Z}) \neq \emptyset$) holds at least in the following two circumstance.

(1) When $\overline{\eta}_1 = \overline{\iota}_1$ (respectively $\overline{\eta}_2 = \overline{\iota}_2$), this condition trivially holds. ($\overline{\eta}_1 = \overline{\iota}_1$ implies the validity of the strict complementarity condition, e.g., in SDP, this means $\beta = \emptyset$, i.e, the eigenvalue function at this point is differentiable).

(2) This condition holds for largest eigenvalue problem, Ky Fan k-norm (respectively SDP case) and so on. Taking $\mathcal{C}_1(\overline{X}_1,\overline{Y})$ as an example, we can show the nonempty condition holds for CMatOP that satisfy the following two conditions: (1) $\{a^i\}_{i=1}^p$ can be generated by only one a^{i_1}; (2) there exists $d \in \mathcal{C}_{\varpi_1}(\lambda(\overline{X}_1),\lambda(\overline{Y}))$ that satisfies $d_{\alpha^l} \neq d_{\alpha^{l'}}$ if $l \neq l'$; and for all $1 \leq l \leq d_1$, there exists some constant s_l such that $d_{\alpha^l} = s_l 1_{|\alpha^l|}$.

To verify this, we can assume $\overline{\iota}_1 = \{i_1,\dots,i_{|\overline{\iota}_1|}\}$ and the components of a^{i_1} are arranged in some order. Let $\mathcal{P}^m_{\lambda(\overline{X}_1)}$ be the set of permutation matrix such that the order of $\lambda(\overline{X}_1)$ remains under this permutation, i.e., for all $Q \in \mathcal{P}^m_{\lambda(\overline{X}_1)}$, we have $\lambda(\overline{X}_1) = Q\lambda(\overline{X}_1)$. Combining this with Proposition 2.22 and Corollary 2.19, we know that for all $d \in \mathcal{C}_{\varpi_1}(\lambda(\overline{X}_1),\lambda(\overline{Y}))$, $\max\limits_{i\in\overline{\iota}_1}\langle a^i, d\rangle = \max\limits_{Q\in\mathcal{P}^m_{\lambda(\overline{X}_1)}} \langle a^{i_1}, Qd\rangle$. On the other hand, if d satisfies condition (2), it is easy to see that for all $Q \in \mathcal{P}^m_{\lambda(\overline{Y})} \cap \mathcal{P}^m_{\lambda(\overline{X}_1)}$, the order of Qd remains and it is the same as that of a^{i_1}; for all $Q \notin \mathcal{P}^n_{\lambda(\overline{Y})}$ and $Q \in \mathcal{P}^n_{\lambda(\overline{X}_1)}$, the order of Qd changes strictly. So condition $\mathcal{C}_1(\overline{X}_1,\overline{Y}) \neq \emptyset$ holds.

Remark 3.3. The " ⊇ " in Proposition 3.12 (i) reduces to "=" when we have $\overline{\eta}_1 \subseteq \eta_1^k$ and $\overline{\eta}_2 \subseteq \eta_2^k$ for each k. Appendix 2 gives a sufficient condition of when $\overline{\eta}_1 \subseteq \eta_1^k$ holds. Also, the proof of "=" under condition $\overline{\eta}_1 \subseteq \eta_1^k$ and $\overline{\eta}_2 \subseteq \eta_2^k$ is also shown in Appendix 2.

The main result of this section is presented below. This theorem reveals the relationship of several critical variational properties of the solution of (3-36).

Theorem 3.13. *Let $\bar{x}$ be a stationary point of problem (3-36) and RCQ holds, i.e.,*

$$0 \in \operatorname{int}\{g(\bar{x},\bar{p}) + g'_x(\bar{x},\bar{p})\mathbb{X} - \mathcal{K}\}. \tag{3-61}$$

Let $(\overline{Y},\bar{y},\overline{Z})$ be the corresponding Lagrangian multiplier and denote $\overline{X}_1 = g_1(\bar{x},\bar{p})$, $\overline{X}_2 = g_2(\bar{x},\bar{p})$. Assume $\mathcal{C}_1(\overline{X}_1,\overline{Y}) \neq \emptyset$ and $\mathcal{C}_2(\overline{X}_2,\overline{Z}) \neq \emptyset$. Then the following assertions are equivalent:

(i) The pair $(\bar{x},\overline{Y},\bar{y},\overline{Z})$ is a strongly regular solution and $\bar{x}$ is a local minimizer of problem (3-36).

(ii) The constraint nondegeneracy condition holds at $\bar{x}$ and the point $\bar{x}$ is a Lipschitzian fully stable local minimizer of problem (3-36).

(iii) The constraint nondegeneracy condition holds at $\bar{x}$ and the second order subdifferential condition holds at $(\bar{x}, \overline{Y}, \bar{y}, \overline{Z})$, i.e., for all $w \neq 0$,

$$\langle L''_{xx}(\bar{x}, \bar{p}, \overline{Y}, \bar{y}, \overline{Z})w, w\rangle + \inf\left\{\langle T, (g_1)'_x(\bar{x}, \bar{p})w\rangle : T \in \mathcal{D}^*\partial\theta_1(\overline{X}_1, \overline{Y})\big((g_1)'_x(\bar{x}, \bar{p})w\big)\right\}$$
$$+ \inf\left\{\langle T, (g_2)'_x(\bar{x}, \bar{p})w\rangle : T \in \mathcal{D}^*\mathcal{N}_{\mathcal{K}}\big(\overline{X}_2, \overline{Z}\big)\big((g_2)'_x(\bar{x}, \bar{p})w\big)\right\} > 0.$$

(iv) The constraint nondegeneracy condition holds at $\bar{x}$ and the strong SOSC holds at $(\bar{x}, \overline{Y}, \bar{y}, \overline{Z})$, i.e., for all $0 \neq w \in \operatorname{app}(\overline{Y}, \bar{y}, \overline{Z})$,

$$\langle L''_{xx}(\bar{x}, \bar{p}, \overline{Y}, \bar{y}, \overline{Z})w, w\rangle - \Upsilon^1_{\overline{X}_1}\big(\overline{Y}, (g_1)'_x(\bar{x}, \bar{p})w\big) - \Upsilon^2_{\overline{X}_2}\big(\overline{Z}, (g_2)'_x(\bar{x}, \bar{p})w\big) \geq 0.$$

(v) Every element in $\partial_c F(\bar{x}, \overline{Y}, \bar{y}, \overline{Z})$ is nonsingular, where $\partial_c F(\bar{x}, \overline{Y}, \bar{y}, \overline{Z})$ is the Clarke subdifferential of F at $(\bar{x}, \overline{Y}, \bar{y}, \overline{Z})$(see [60]) and

$$F(x; Y, y, Z) := \begin{bmatrix} L'_x(x, \bar{p}; Y, y, Z) \\ g_1(x, \bar{p}) - \operatorname{Pr}_{\theta_1}(g_1(x, \bar{p}) + Y) \\ h(x, \bar{p}) \\ g_2(x, \bar{p}) - \Pi_{\mathcal{K}}(g_2(x, \bar{p}) + Z) \end{bmatrix}.$$

Proof. From [4, Theorem 1], we have $(iv) \Rightarrow (v) \Rightarrow (i)$. Then we prove the remaining of the theorem by $(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (iv)$.
$(i) \Longrightarrow (ii)$: See Appendix 1.
$(ii) \Longrightarrow (iii)$: The framework of the proof is directly taken from [16, Theorem 5.6]. Assuming (ii), from [16, Corollary 4.5] we have

$$\inf\{\langle G, w\rangle : (G, q) \in \mathcal{D}^*\partial_x\phi(\bar{x}, \bar{p}, 0)(w)\} > 0 \quad \text{for all } w \neq 0.$$

This together with the second-order representation from Lemma 3.10 and (3-44) tell us that

$$\begin{aligned}
0 < &\inf\Big\{\langle G, w\rangle : G \in L''_{xx}(\bar{x}, \bar{p}, \overline{Y}, \bar{y}, \overline{Z})w + (g_1)'_x(\bar{x}, \bar{p})^*\mathcal{D}^*\partial\theta_1(\overline{X}_1, \overline{Y})\big((g_1)'_x(\bar{x}, \bar{p})w\big) \\
&\quad + (G_2)'_x(\bar{x}, \bar{p})^*\mathcal{D}^*\partial\delta_{\{0\}\times\mathcal{K}}(G_2(\bar{x}, \bar{p}), \bar{y}, \overline{Z})\big((G_2)'_x(\bar{x}, \bar{p})w\big)\Big\} \\
= &\langle L''_{xx}(\bar{x}, \bar{p}, \overline{Y}, \bar{y}, \overline{Z})w, w\rangle + \inf\left\{\langle (g_1)'_x(\bar{x}, \bar{p})^*T, w\rangle : T \in \mathcal{D}^*\partial\theta_1(\overline{X}_1, \overline{Y})\big((g_1)'_x(\bar{x}, \bar{p})w\big)\right\} \\
&\quad + \inf\left\{\langle (g_2)'_x(\bar{x}, \bar{p})^*T, w\rangle : T \in \mathcal{D}^*\mathcal{N}_{\mathcal{K}}\big(\overline{X}_2, \overline{Z}\big)\big((g_2)'_x(\bar{x}, \bar{p})w\big)\right\} \\
= &\langle L''_{xx}(\bar{x}, \bar{p}, \overline{Y}, \bar{y}, \overline{Z})w, w\rangle + \inf\left\{\langle T, (g_1)'_x(\bar{x}, \bar{p})w\rangle : T \in \mathcal{D}^*\partial\theta_1(\overline{X}_1, \overline{Y})\big((g_1)'_x(\bar{x}, \bar{p})w\big)\right\} \\
&\quad + \inf\left\{\langle T, (g_2)'_x(\bar{x}, \bar{p})w\rangle : T \in \mathcal{D}^*\mathcal{N}_{\mathcal{K}}\big(\overline{X}_2, \overline{Z}\big)\big((g_2)'_x(\bar{x}, \bar{p})w\big)\right\}
\end{aligned}$$

for any $w \neq 0$, which shows that the inequality in (iii) holds.
$(iii) \Longrightarrow (iv)$: (iii) implies that second order subdifferential condition holds for all $0 \neq w \in \operatorname{app}(\overline{Y}, \bar{y}, \overline{Z})$. Applying Proposition 3.12, we get (iv). Thus we have completed the whole proof. □

Example 3.1. Largest eigenvalue minimization: Recalling Example 1.1, we have $\theta_1 = \varpi_1 \circ \lambda$. Let $\overline{X} = g(\bar{x}, \bar{p})$ for a given $\bar{x}$. Assume $\overline{Y} \in \partial\theta_1(\overline{X})$. Recall the definition of $\{\mu^p\}$ in (3-40). Then we have the $\iota_1(\lambda(\overline{X})) = \{1 \le i \le m \mid \lambda_i(\overline{X}) = v_1(\overline{X})\} = \mu^1$. From [4, Section 5], we know that given $(\overline{X}, \overline{Y}) \in \operatorname{gph} \partial\theta_1$, and let $\overline{P} \in \mathcal{O}^m(\overline{X}) \cap \mathcal{O}^m(\overline{Y})$,

$$\begin{cases} 0 \le \lambda_i(\overline{Y}) \le 1, \quad \forall i \in \mu^1 \text{ and } \sum\limits_{i \in \mu^1} \lambda_i(\overline{Y}) = 1, \\ \lambda_i(\overline{Y}) = 0, \quad \forall i \in \mu^p, \quad p = 2, \dots, d_1. \end{cases} \tag{3-62}$$

Denote $\gamma := \{i \in \mu^1 \mid \lambda_i(\overline{Y}) > 0\}$ and $\beta := \{i \in \nu^1 \mid \lambda_i(\overline{Y}) = 0\}$. From (3-62), we obtain

$$\iota_1(\lambda(\overline{X})) = \mu^1 = \gamma \cup \beta, \quad \gamma = \bigcup_{q=1}^{s^1-1} \zeta_q^1, \quad \beta = \zeta_{s^1}^1 \text{ and } \mu^p = \zeta_1^p \text{ for } p = 2, \dots, d_1.$$

Obviously $\eta_1(\lambda(\overline{X}), \lambda(\overline{Y})) = \gamma$. From Theorem 3.4, we can easily write down its limiting normal cone:

$$\mathcal{N}_{\operatorname{gph} \partial\theta_1}(\overline{X}, \overline{Y}) = \bigcup_{\substack{Q \in \mathcal{O}^n \\ \Xi_1 \in \mathcal{U}_m^1}} \left\{ (X^*, Y^*) \in \mathbb{S}^m \times \mathbb{S}^m \;\middle|\; \begin{array}{l} \overline{D} \circ \widetilde{X}^* + \overline{F} \circ \widetilde{Y}^* = 0, \\ \Xi_1 \circ \widetilde{X}^* + \Xi_2 \circ \widetilde{Y}^* = 0, \\ \kappa(\mathscr{D}^\pi(\widetilde{X}^*)) \in \mathcal{C}_1 \\ \langle \kappa(\mathscr{D}^\pi(\widetilde{Y}^*)), e^i - e^j \rangle = 0, i, j \in \beta_+ \\ \langle \kappa(\mathscr{D}^\pi(\widetilde{Y}^*)), e^i - e^j \rangle \le 0, i \in \beta_+, j \in \beta_0 \end{array} \right\}.$$

where $\mathcal{C}_1 = \{ \sum\limits_{i,j \in \beta_+} u_{ji}(e^j - e^i) + \sum\limits_{i \in \beta_+, j \in \beta_0} g_{ji}(e^i - e^j) \mid g_{ji} \ge 0\}$. Moreover, $\mathcal{C}_1(\lambda(\overline{X}), \lambda(\overline{Y})) \neq \emptyset$ holds from Remark 3.2. The equivalence among strong regularity, Lipschitz full stability and strong SOSC can be obtained from Theorem 3.13 and this result can be seen in [4, Theorem 2].

Example 3.2. SDP: Recalling Example 1.2. Assume $(\bar{x}, \overline{Z})$ such that $\overline{Z} \in \mathcal{N}_{\mathbb{S}^n_+}(g(\bar{x}, \bar{p}))$. Clearly we have

$$\eta_2(\lambda(g(\bar{x}, \bar{p})), \lambda(\overline{Z})) = \gamma, \;\; \iota_2(\lambda(g(\bar{x}, \bar{p}))) = \beta \cup \gamma.$$

where $\alpha = \{i \mid \lambda_i(g(\bar{x}, \bar{p}) + \overline{Z}) > 0\}, \beta = \{i \mid \lambda_i(g(\bar{x}, \bar{p}) + \overline{Z}) = 0\}, \gamma = \{i \mid \lambda_i(g(\bar{x}, \bar{p}) + \overline{Z}) < 0\}$. Assume x^k such that $X^k := g(x^k, \bar{p}) \to g(\bar{x}, \bar{p})$ and $Z^k \to \overline{Z}$ with $Z^k \in \mathcal{N}_{\mathbb{S}^n_+}(g(x^k, \bar{p}))$ as $k \to \infty$. By taking a subsequence if necessary, we assume β can be divided into three parts such that $\beta = \beta_+ \cup \beta_0 \cup \beta_-$, where $\beta_+ = \{i \in \beta \mid \lambda_i(X^k + Z^k) > 0\}$, $\beta_0 = \{i \in \beta \mid \lambda_i(X^k + Z^k) = 0\}$, $\beta_- = \{i \in \beta \mid \lambda_i(X^k + Z^k) < 0\}$. So

$$\eta_2(\lambda(X^k), \lambda(Z^k)) = \beta_- \cup \gamma, \quad \iota_2(\lambda(X^k)) = \beta_0 \cup \beta_- \cup \gamma.$$

Let $\overline{B} = g(\bar{x}, \bar{p}) + \overline{Z}$,

$$\overline{D}_{ij} = \Sigma_{ij} := \frac{\max\{\lambda_i(\overline{B}), 0\} - \max\{\lambda_j(\overline{B}), 0\}}{\lambda_i(\overline{B}) - \lambda_j(\overline{B})}, \quad i, j = 1, \dots, n, \tag{3-63}$$

Clearly, we have $\mathcal{C}_2(\lambda(g(\bar{x}, \bar{p}), \lambda(\overline{Z})) \neq \emptyset$. After complicated calculations, we derive [80, Theorem 3.1] from Theorem 3.4. Similarly, strong regularity, Lipschitz full stability with nondegeneracy and nondegeneracy with SSOSC are equivalent from Theorem 3.13. For more details, please see [11] and [16].

Remark 3.4. Similarly as in [16], note that ignoring the basic parametric perturbation p provides a complete characterization of tilt stability for CMatOP.

3.3 Isolated calmness for CMatOP

In this section, we will focus on another perturbation property named isolated calmness for CMatOP. Before we step any further, it is necessary to list some very important stability properties.

Definition 3.14. *[86] A set-valued map $S : \mathbb{P} \rightrightarrows \mathbb{Z}$ is said to be upper-Lipschitz at $\bar{p} \in \mathbb{P}$ if there exist a neighborhood $\mathcal{U}$ of $\bar{p}$ and $\kappa > 0$ such that*

$$S(p) \subseteq S(\bar{p}) + \kappa \|p - \bar{p}\| \mathcal{B} \quad \forall\, p \in \mathcal{U}. \tag{3-64}$$

Definition 3.15. *[87] A set-valued map $S : \mathbb{P} \rightrightarrows \mathbb{Z}$ is said to be pseudo-Lipschitz (or locally Lipschitz like or have the Aubin property) at $(\bar{p}, \bar{z}) \in \operatorname{gph} S$ if there exist a neighborhood $\mathcal{U}$ of $\bar{p}$, a neighborhood $\mathcal{V}$ of $\bar{z}$ and $\kappa > 0$ such that*

$$S(p) \cap \mathcal{V} \subseteq S(p') + \kappa \|p - p'\| \mathcal{B} \quad \forall\, p, p' \in \mathcal{U}. \tag{3-65}$$

Equivalently, S is pseudo-Lipschitz at $(\bar{p}, \bar{z}) \in \operatorname{gph} S$ if there exist a neighborhood $\mathcal{U}$ of $\bar{p}$, a neighborhood $\mathcal{V}$ of $\bar{z}$ and $\kappa > 0$ such that

$$\operatorname{dist}(z, S(p')) \le \kappa \operatorname{dist}(p', S^{-1}(z)) \quad \forall\, p' \in \mathcal{U}, z \in \mathcal{V},$$

i.e., the inverse map S^{-1} is metrically regular around $(\bar{z}, \bar{p})$.

Definition 3.16. *[1, 88] A set-valued map $S : \mathbb{P} \rightrightarrows \mathbb{Z}$ is said to be calm (or pseudo upper-Lipschitz continuous) at $(\bar{p}, \bar{z}) \in \operatorname{gph} S$ if there exist a neighborhood $\mathcal{U}$ of $\bar{p}$, a neighborhood $\mathcal{V}$ of $\bar{z}$ and $\kappa > 0$ such that*

$$S(p) \cap \mathcal{V} \subseteq S(\bar{p}) + \kappa \|p - \bar{p}\| \mathcal{B} \quad \forall\, p \in \mathcal{U}. \tag{3-66}$$

Equivalently, S is calm at $(\bar{p}, \bar{z}) \in \operatorname{gph} S$ if there exist a neighborhood $\mathcal{U}$ of $\bar{z}$ and $\kappa > 0$ such that

$$\operatorname{dist}(z, S(\bar{p})) \le \kappa \operatorname{dist}(\bar{p}, S^{-1}(z)) \quad \forall\, z \in \mathcal{V}, \tag{3-67}$$

i.e., the inverse map S^{-1} is metrically subregular at $(\bar{z}, \bar{p})$.

Definition 3.17. *[61] A set-valued map $S : \mathbb{P} \rightrightarrows \mathbb{Z}$ is said to be isolated calm at $(\bar{p}, \bar{z}) \in \operatorname{gph} S$ if there exist a neighborhood $\mathcal{U}$ of $\bar{p}$, a neighborhood $\mathcal{V}$ of $\bar{z}$ and $\kappa > 0$ such that*

$$S(p) \cap \mathcal{V} \subseteq \{\bar{z}\} + \kappa \|p - \bar{p}\| \mathcal{B} \quad \forall\, p \in \mathcal{U}. \tag{3-68}$$

Equivalently, S is isolated calm at $(\bar{p}, \bar{z}) \in \operatorname{gph} S$ if there exist a neighborhood $\mathcal{V}$ of $\bar{z}$ and $\kappa > 0$ such that

$$\|z - \bar{z}\| \leq \kappa \operatorname{dist}(\bar{p}, \mathcal{S}^{-1}(z)) \quad \forall\, z \in \mathcal{V},$$

i.e., the inverse map S^{-1} is strongly metrical subregular around $(\bar{z}, \bar{p})$. Moreover, S is said to be robustly isolated calm at $\bar{p}$ for $\bar{z}$ if (3-68) holds and for each $p \in \mathcal{U}$, $S(p) \cap \mathcal{V} \neq \emptyset$.

Let $\mathbb{F}$ be an Euclidean space, $P(z) : \mathbb{Z} \to \mathbb{F}$ be a locally Lipschitz function and $\mathcal{D} \subset \mathbb{F}$ be a closed subset. For convenience we summarize some verifiable sufficient conditions for the *calmness* of the set-valued map $S(p) := \{z \in \mathbb{Z} \mid p \in -P(z) + \mathcal{D}\}$.

Proposition 3.18. *Let $\Omega := \{z \in \mathbb{Z} \mid P(z) \in \mathcal{D}\}$ and $\bar{z} \in \Omega$. Suppose that $P(z)$ is continuously differentiable at $\bar{z}$. Let $T_{\Omega}^{lin}(\bar{z}) := \{w \in \mathbb{Z} | P'(\bar{z})w \in T_{\mathcal{D}}(P(\bar{z}))\}$ be the linearized cone of Ω at $\bar{z}$. Then the set-valued map $S(p) = \{z \in \mathbb{Z} \mid p \in -P(z) + \mathcal{D}\}$ is calm at $(0, \bar{z})$ if one of the following conditions holds.*

1. Linear CQ holds (see, e.g., [89, Theorem 4.3]): $P(z)$ is piecewise affine and D is the union of finitely many convex polyhedra sets.

2. No nonzero abnormal multiplier constraint qualification (NNAMCQ) holds at $\bar{z}$ (see, e.g., [89, Theorem 4.4])

$$P'(\bar{z})^*\lambda = 0,\ \lambda \in \mathcal{N}_{\mathcal{D}}(P(\bar{z})) \quad \Longrightarrow \quad \lambda = 0.$$

3. First-order sufficient condition for metric subregularity (FOSCMS) at $\bar{z}$ for the system $P(z) \in \mathcal{D}$ at $\bar{z}$ [90, Corollary 1]: for every $0 \neq w \in \mathcal{T}_{\Omega}^{lin}(\bar{z})$, one has

$$P'(\bar{z})^*\lambda = 0,\ \lambda \in \mathcal{N}_{\mathcal{D}}(P(\bar{z}); P'(\bar{z})w) \quad \Longrightarrow \quad \lambda = 0.$$

Suppose a general optimization problem

$$\begin{array}{ll} \min & f(x) \\ s.t. & G(x) \in \mathcal{C}, \end{array} \tag{3-69}$$

where $f : \mathbb{X} \to \mathbb{R}$, $G : \mathbb{X} \to \mathbb{Y}$ are twice continuously differentiable, $\mathcal{C}$ is a closed convex set.

Definition 3.19. *[5, Definition 2.86] Let $\bar{x}$ be a feasible solution to optimization problem (3-69). Robinson's constraint qualification (RCQ) is said to hold at $\bar{x}$ if*

$$0 \in \operatorname{int} \{G(\bar{x}) + G'(\bar{x})\mathbb{X} - \mathcal{C}\},$$

where "int" denotes the topological interior part of a given set.

It is well known from [5, Proposition 2.97, Corollary 2.98] that the following conditions are equivalent to each other and to RCQ

- $G'(\bar{x})\mathbb{X} + \mathcal{T}_{\mathcal{C}}(G(\bar{x})) = \mathbb{Y}$
- $\operatorname{Ker} G'(\bar{x})^* \cap \mathcal{N}_{\mathcal{C}}(G(\bar{x})) = \{0\}$

The constraint nondegenerate condition is defined as

$$G'(\bar{x})\mathbb{X} + \operatorname{lin}(\mathcal{T}_{\mathcal{C}}(G(\bar{x}))) = \mathbb{Y}.$$

If $\bar{x}$ is a locally optimal solution and $\mathcal{M}(\bar{x}) \neq \emptyset$ with $\overline{Y} \in \mathcal{M}(\bar{x})$, the strict Robinson's constraint qualification (SRCQ), i.e.,

$$G'(\bar{x})\mathbb{X} + \mathcal{T}_{\mathcal{C}}(G(\bar{x})) \cap [\overline{Y}]^{\perp} = \mathbb{Y},$$

is a sufficient condition for the uniqueness of multiplier.

In [91], the authors establish that under RCQ, the KKT solution mapping is robustly isolated calm if and only if both SRCQ and the second order sufficient condition (SOSC) hold. In the following part, we will prove a similar result for CMatOP (1-1). Moreover, we will prove that for certain kind of CMatOP, the strong SOSC of primal problem is equivalent to the SRCQ of its dual problem.

3.3.1 The equivalence between SOSC and SRCQ for quadratic CMatOP

Recalling that $\theta_1 = \varpi_1 \circ \lambda$ and $\mathcal{K} = \{X \in \mathbb{S}^n : \lambda(X) \in \widehat{\mathcal{K}}\}$ in general CMatOP. We consider the following quadratic CMatOP.

$$\text{(QCMatOP)} \quad \begin{array}{ll} \min\limits_{X \in \mathbb{S}^n} & \frac{1}{2}\langle X, \mathcal{Q}X\rangle + \langle C, X\rangle + \theta_1(X_1) + \delta_{\mathcal{K}}(X_2) \\ \text{s.t.} & \mathcal{A}X = w, \end{array} \tag{3-70}$$

where $X = (X_1, X_2) \in \mathbb{X} := \mathbb{S}^m \times \mathbb{S}^n$, $\delta_{\mathcal{K}} : \mathbb{S}^n \to (-\infty, \infty]$ is the indicator function of closed convex cone $\mathcal{K}$, $\mathcal{Q} : \mathbb{X} \to \mathbb{X}$ is a positively semidefinite self-adjoint linear operator, $\mathcal{A} : \mathbb{X} \to \mathbb{R}^s$ is a linear operator, $C \in \mathbb{X}$ and $w \in \mathbb{R}^s$ are given data. The restricted Wolfe dual of (3-70) with respect to linear subspace $\mathbb{T}$ is given by

$$\text{(Dual)} \quad \begin{array}{ll} \max & w^T y - \frac{1}{2}\langle T, \mathcal{Q}T\rangle - {\theta_1}^*(Y) + \delta_{\mathcal{K}^\circ}(Z) \\ \text{s.t.} & \mathcal{A}^* y - \mathcal{Q}T - S = C \\ & T \in \mathbb{T} \supseteq \operatorname{Range}(\mathcal{Q}), \end{array} \tag{3-71}$$

where $S = (Y, Z)$.

Definition 3.20. *Let $\overline{X} \in \mathbb{X}$ be a stationary point of problem (3-70). Denote its corresponding multiplier as $\mathcal{M}_P(\overline{X})$ with $(\overline{Y}, \bar{y}, \overline{Z}) \in \mathcal{M}_P(\overline{X})$. We say the second order* *sufficient condition (SOSC) holds at* $\overline{X} \in \mathbb{X}$ *if for all* $0 \neq H = (H_1, H_2) \in \mathcal{C}(\overline{X}) := \{H \in \mathbb{X} \mid \mathcal{A}H = 0, H_1 \in \mathcal{C}_{\theta_1}(\overline{X}_1, \overline{Y}), H_2 \in \mathcal{C}_{\mathcal{K}}(\overline{X}_2, \overline{Z})\}$*, we have*

$$\sup_{(Y,y,Z)\in\mathcal{M}_P(\overline{X})} \left\{\langle H, \mathcal{Q}H\rangle - \Upsilon^1_{\overline{X}_1}\big(Y, H_1\big) - \Upsilon^2_{\overline{X}_2}\big(Z, H_2\big)\right\} > 0,$$

where the function $\Upsilon^1_{\overline{X}_1}(Y, H_1) : \partial\theta_1(\overline{X}_1)\times\mathbb{S}^m \to \mathbb{R}$ *and* $\Upsilon^2_{\overline{X}_2}(Z, H_2) : \mathcal{N}_{\mathcal{K}}(\overline{X}_2)\times\mathbb{S}^n \to \mathbb{R}$ *are the* σ*-term of SOSC for problem (3-70). The explicit forms of* σ*-terms are given in* (3-53) *and* (3-54).

The SRCQ of problem (3-71) is defined in a similar manner to [92, Definition 5.1].

Definition 3.21. *Suppose that* $\mathcal{M}_P(\overline{X}) \neq \emptyset$*. We say that the extended SRCQ for the dual problem* (3-71) *holds at* $\mathcal{M}_P(\overline{X})$ *with respect to* $\overline{X}$ *if*

$$conv\left\{\cup_{(y,S)\in\mathcal{M}_P(\overline{X})}\begin{bmatrix}\mathcal{C}_{\theta_1^*}(Y,\overline{X}_1)\\ \mathcal{T}_{\mathcal{K}^\circ}(Z)\cap(\overline{X}_2)^\perp\end{bmatrix}\right\} + \mathcal{A}^*\mathbb{R}^s - \mathcal{Q}\mathbb{T} = \mathbb{X} \tag{3-72}$$

where "conv" denotes the convex hull of a set.

Similar to [91, Lemma 10], we have the following lemma for the proximal mapping $\mathrm{Pr}_{\theta_1}(\cdot)$

Lemma 3.22. *Let* $C \in \mathbb{Y}$*,* $A = \mathrm{Pr}_{\theta_1}(C)$ *and* $B = C - A$*.*

(i) Let $G, H \in \mathbb{Y}$*.* $G - \mathrm{Pr}'_{\theta_1}(C; G + H) = 0$ *if and only if*

$$\begin{cases} G \in \mathcal{C}_{\theta_1}(A, B),\\ H - \frac{1}{2}\nabla_G\Upsilon^1_A(B, G) \in [\mathcal{C}_{\theta_1}(A, B)]^\circ,\\ \langle G, H\rangle = -\Upsilon^1_A(B, G).\end{cases} \tag{3-73}$$

(ii) Let $\mathcal{A} : \mathbb{X} \to \mathbb{Y}$ *be a linear operator. Then the following two statements are equivalent:*

(a) $H \in \mathbb{Y}$ *is a solution to the following system of equations*

$$\begin{cases} \mathcal{A}^*H = 0\\ \mathrm{Pr}'_{\theta_1}(C; H) = 0\end{cases} \tag{3-74}$$

(b) $H \in [\mathcal{A}\mathbb{X} + \mathcal{C}_{\theta_1}(A, B)]^\circ$*.*

Proof. We define $h : \mathbb{Y} \to (-\infty, \infty]$ by

$$h(D) := -\Upsilon^1_A(B, D) + \delta_{\mathcal{C}_{\theta_1}(A,B)}(D), \quad D \in \mathbb{Y}.$$

It follows from the explicit form of $\Upsilon^1_A(B, D)$ (3-53) and Remark 2.2 that $h(\cdot)$ is a proper closed convex function. Then we have $\overline{D} \in \mathcal{C}_{\theta_1}(A, B)$ is the unique optimal solution to problem

$$\min\left\{\|D - (H + G)\|^2 - \Upsilon^1_A(B, D) \mid D \in \mathcal{C}_{\theta_1}(A, B)\right\},$$

if and only if $\overline{D}$ is the unique optimal solution to the following strongly convex optimization problem

$$\min\{\|D-(H+G)\|^2+h(D)\},$$

or equivalently,

$$0\in 2(\overline{D}-(H+G))+\nabla_G\Upsilon_A^1(B,\overline{D})+\mathcal{N}_{\mathcal{C}_{\theta_1}(A,B)}(\overline{D}).$$

Combining this with [8, Proposition 7.1], we have $G=\mathrm{Pr}'_{\theta_1}(C;G+H)$ if and only if

$$\begin{cases} G\in\mathcal{C}_{\theta_1}(A,B), \\ H-\frac{1}{2}\nabla_G\Upsilon_A^1(B,G)\in[\mathcal{C}_{\theta_1}(A,B)]^\circ, \\ \langle G,H-\frac{1}{2}\nabla_G\Upsilon_A^1(B,G)\rangle=0. \end{cases} \tag{3-75}$$

It can be checked directly that for each $G\in\mathcal{C}_{\theta_1}(A,B)$,

$$\langle G,\nabla_G\Upsilon_A^1(B,G\rangle=2\Upsilon_A^1(B,G)$$

is satisfied. Thus we have (3-73) holds.

By noting that $0\in\mathcal{C}_{\theta_1}(A,B)$ and taking $G=0$ in (i), we obtain (ii) immediately. □

Remark 3.5. Following the same proof sketch, we can obtain a similar result for $\mathrm{Pr}_{\theta_1+\delta_{\mathcal{K}}}$ with $C:=(C_1,C_2)\in\mathbb{S}^n\times\mathbb{S}^m$. We omit it here for simplicity.

Proposition 3.23. *Let* $\overline{X}\in\mathbb{X}$ *be a locally optimal solution to problem* (3-70) *with* $\mathcal{M}_P(\overline{X})\neq\emptyset$. *Let* $\mathbb{T}\subseteq\mathbb{X}$ *be any linear subspace that is contained in Range*$\mathcal{Q}$. *Then the following two conditions are equivalent:*

(i) The SOSC for the primal problem (3-70) *holds at* $\overline{X}$*:*

$$\sup_{(y,Y,Z)\in\mathcal{M}_P(\overline{X})}\left\{\langle H,\mathcal{Q}H\rangle-\Upsilon_{\overline{X}_1}^1(Y,H_1)-\Upsilon_{\overline{X}_2}^2(Z,H_2)\right\}>0,\quad\forall\,0\neq H\in\mathcal{C}(\overline{X}).$$

(ii) The extended SRCQ for the dual problem (3-71) *holds at* $\mathcal{M}_P(\overline{X})$ *with respect to* $\overline{X}$.

Proof. For notational convenience, denote

$$\Gamma:=\mathrm{conv}\left\{\cup_{(y,Y,Z)\in\mathcal{M}_P(\overline{X})}\begin{bmatrix}\mathcal{C}_{\theta_1^*}(X,\overline{X}_1)\\ \mathcal{T}_{\mathcal{K}^\circ}(Y)\cap(\overline{X}_2)^\perp\end{bmatrix}\right\}\quad\text{and}\quad\mathcal{D}:=\Gamma+\mathcal{A}^*\mathbb{R}^s-\mathcal{Q}\mathbb{T}.$$

"$(i)\Longrightarrow(ii)$" We prove this part by contradiction. Suppose that the extended SRCQ for the dual problem (3-71) does not hold at $\mathcal{M}_P(\overline{X})$ with respect to $\overline{X}$. Then $\mathcal{D}\neq\mathbb{X}$. Let $\mathrm{cl}(\mathcal{D})$ denote the closure of $\mathcal{D}$. Since $\mathrm{cl}(\mathcal{D})\neq\mathbb{X}$ (cf. [59, Theorem 6.3]), there exists a point $A=(A_1,A_2)\in\mathbb{X}$ but $A\notin\mathrm{cl}(\mathcal{D})$.

Let $G = A - \Pi_{\mathrm{cl}(\mathcal{D})}(A)$. By using the fact that $\mathrm{cl}(\mathcal{D})$ is a closed convex cone in $\mathbb{X}$, we have

$$\langle G, W\rangle \le 0, \quad \forall\, W \in \mathrm{cl}(\mathcal{D}),$$

which, together with the assumption $\mathrm{Range}\mathcal{Q} \subseteq \mathbb{T}$, implies that $\mathcal{A}G = 0$, $\mathcal{Q}G = 0$ and

$$\langle G, W\rangle \le 0, \quad \forall\, W \in \Gamma. \tag{3-76}$$

Let (y, Y, Z) be an arbitrary point in $\mathcal{M}_P(\overline{X})$. Then $Z \in \mathcal{N}_{\mathcal{K}}(\overline{X}_2)$. Then, by using (3-76) and [91, Lemma 10 (ii)], we obtain

$$\Pi'_{\mathcal{K}^\circ}(Z + \overline{X}_2, G_2) = G_2 - \Pi'_{\mathcal{K}}(Z + \overline{X}_2, G_2) = 0.$$

By using (3-17), we know that $(\overline{V}^T G_2 \overline{V})_{ij} = 0$, $\forall\, i \in \nu^l,\ j \in \nu^{l'}$. It follows from the explicit form of the σ-term in Definition 3.20 that

$$\Upsilon^2_{\overline{X}_2}(Z, G_2) = 0.$$

Then we only need to prove $\Upsilon^1_{\overline{X}_1}(Y, G_1) = 0$. By using (3-76) and Lemma 3.22, we obtain

$$\mathrm{Pr}'_{\theta_1^*}(Y + \overline{X}_1, G_1) = G_1 - \mathrm{Pr}'_{\theta_1}(Y + \overline{X}_1, G_1) = 0.$$

By using (3-8), we know that $(\overline{P}^T G_1 \overline{P})_{ij} = 0$, $\forall\, i \in \mu^l,\ j \in \mu^{l'}$. It follows from the explicit form of the σ-term in Definition 3.20 that

$$\Upsilon^1_{\overline{X}_1}(Y, G_1) = 0.$$

It follows that for all $(y, Y, Z) \in \mathcal{M}_P(\overline{X})$, we have found an $0 \ne G \in \mathcal{C}(\overline{X})$ such that

$$\langle G, \mathcal{Q}G\rangle - \Upsilon^1_{\overline{X}_1}(Y, G_1) - \Upsilon^2_{\overline{X}_2}(Z, G_2) = 0,$$

which is a contradiction to SOSC. Then we have proved this part.

"$(ii) \Longrightarrow (i)$" For the sake of contradiction, we suppose that SOSC for the primal problem (3-70) does not hold at $\overline{X}$. It follows that there exists $0 \ne G \in \mathcal{C}(\overline{X})$ such that

$$\sup_{(y,Y,Z)\in\mathcal{M}_P(\overline{X})} \left\{\langle G, \mathcal{Q}G\rangle - \Upsilon^1_{\overline{X}_1}(Y, G_1) - \Upsilon^2_{\overline{X}_2}(Z, G_2)\right\} = 0, \quad \forall\, 0 \ne G \in \mathcal{C}(\overline{X}),$$

which implies for all $(y, Y, Z) \in \mathcal{M}_P(\overline{X})$,

$$\langle G, \mathcal{Q}G\rangle = 0,\ \Upsilon^1_{\overline{X}_1}(Y, G_1) = 0,\ \text{and } \Upsilon^2_{\overline{X}_2}(Z, G_2) = 0.$$

Let (y, Y, Z) be an arbitrary point in $\mathcal{M}_P(\overline{X})$. Then, from $\Upsilon^2_{\overline{X}_2}(Z, G_2) = 0$, we know that $(\overline{V}^T G_2 \overline{V})_{ij} = 0$, $\forall\, i \in \nu^l,\ j \in \nu^{l'},\ l < l'$. It follows from [4, Proposition 8(i)] that

$$(\widehat{G}_2)_{\nu^l\nu^l} = \mathrm{Diag}((\widehat{G}_{\nu^l\nu^l})_{\chi_1^l\chi_1^l}, \ldots, (\widehat{G}_{\nu^l\nu^l})_{\chi_{s^l}^l\chi_{s^l}^l}) := \mathscr{D}(\widehat{G}_2).$$

Also, we have

$$\kappa(\mathscr{D}(\widehat{G}_2)) \in \mathcal{C}_{\widehat{\mathcal{K}}}(\lambda(\overline{X}_2), \lambda(Z)),$$

which implies $\Pi'_{\widehat{\mathcal{K}}}(\lambda(X_2+Z); \kappa(\mathscr{D}(\widehat{G}_2))) = \Pi_{\mathcal{C}_{\widehat{\mathcal{K}}}(\lambda(\overline{X}_2),\lambda(Z))}(\kappa(\mathscr{D}(\widehat{G}_2))) = \kappa(\mathscr{D}(\widehat{G}_2))$. Combining this with (3-17), we have

$$\Pi'_{\mathcal{K}^\circ}(Z+\overline{X}_2, G_2) = G_2 - \Pi'_{\mathcal{K}}(Z+\overline{X}_2, G_2) = 0.$$

By using [91, Lemma 10 (ii)] again, we know that $G_2 \in [\mathcal{C}_{\mathcal{K}^\circ}(\overline{X}_2, Z)]^\circ$.

Similarly, from $\Upsilon^1_{\overline{X}_1}(Y, G_1) = 0$, we know that $(\overline{P}^T G_1 \overline{P})_{ij} = 0,\ \forall\, i \in \mu^l,\ j \in \mu^{l'},\ l < l'$. By using [4, Proposition 4(i)], we can similarly prove that

$$\mathrm{Pr}'_{\varpi_1}(\lambda(X_1+Y); \kappa(\mathscr{D}(\widehat{G}_1))) = \Pi_{\mathcal{C}_{\varpi_1}(\lambda(\overline{X}_1),\lambda(Y))}(\kappa(\mathscr{D}(\widehat{G}_1))) = \kappa(\mathscr{D}(\widehat{G}_1)).$$

Combining this with (3-8), we have

$$\mathrm{Pr}'_{\theta_1^*}(Y+\overline{X}_1; G_1) = G_1 - \mathrm{Pr}'_{\theta_1}(Y+\overline{X}_1; G_1) = 0.$$

Using Lemma 3.22 again, we obtain $G_1 \in [\mathcal{C}_{\theta_1^*}(\overline{X}_1, Y)]^\circ$.

Since the extended SRCQ (3-72) holds, there exists $\hat{y} \in \mathbb{R}^s$, $\hat{w} \in \mathbb{T}$ and $\hat{d} \in \Gamma$ such that $G = \hat{d} + \mathcal{A}^*\hat{y} - \mathcal{Q}\hat{w}$. By Carathéodory's theorem, there exist a positive integer k, scalars $a_i \geq 0$, $i = 1, \ldots, k$, with $\sum_{i=1}^k a_i = 1$, and points

$$\hat{d}^i \in \cup_{(y,S)\in\mathcal{M}_P(\overline{X})} \begin{bmatrix} \mathcal{C}_{\theta_1^*}(Y, \overline{X}_1) \\ \mathcal{T}_{\mathcal{K}^\circ}(Z) \cap (\overline{X}_2)^\perp \end{bmatrix}, \quad i = 1, \cdots, k$$

such that $\hat{d} = \sum_{i=1}^k a_i \hat{d}^i$. Thus we have

$$\langle G, G\rangle = \langle \hat{d} + \mathcal{A}^*\hat{y} - \mathcal{Q}\hat{w}, G\rangle = \langle \hat{d}, G\rangle \leq 0.$$

This contradiction shows the validity of this part. □

3.3.2 Characterization of isolated calmness for CMatOP

Consider the following canonically perturbed CMatOP:

$$\begin{array}{ll} \min\limits_{x\in\mathbb{X}} & f(x) - \langle v, x\rangle + \theta_1(g_1(x)+u_1) \\ \text{s.t.} & h(x) = u_2\,, \\ & g_2(x) + u_3 \in \mathcal{K}\,. \end{array} \tag{3-77}$$

Let $\bar{x} \in \mathbb{X}$ be a stationary point of problem (3-77). Denote $\mathcal{M}(\bar{x})$ as the set of Lagrangian multipliers at $\bar{x}$ with respect to $(v, u_1, u_2, u_3) = (0,0,0,0)$. Let $(\overline{Y}, \bar{y}, \overline{Z}) \in \mathcal{M}(\bar{x})$. Since $\mathcal{M}(\bar{x})$ is nonempty, the critical cone of (3-77) with $(v, u_1, u_2, u_3) = (0,0,0,0)$ can be defined as

$$\mathcal{C}(\bar{x}) := \left\{ d \in \mathbb{X} \;\middle|\; \begin{array}{l} h'(\bar{x})d = 0, \quad g_1'(\bar{x})d \in \mathcal{C}_{\theta_1}(g_1(\bar{x}), \overline{Y}), \\ g_2'(\bar{x})d \in \mathcal{C}_{\mathcal{K}}(g_2(\bar{x}), \overline{Z}) \end{array} \right\}. \tag{3-78}$$

where $\mathcal{C}_{\theta_1}(g_1(\overline{x}), \overline{Y})$ and $\mathcal{C}_{\mathcal{K}}(g_2(\overline{x}), \overline{Z})$ are the critical cones defined in (2-19) and (2-35), respectively.

Definition 3.24. *Let $\bar{x} \in \mathbb{X}$ be a stationary point of problem (3-77) with $(v, u_1, u_2, u_3) = (0, 0, 0, 0)$. We say the second order sufficient condition (SOSC) holds at $\bar{x} \in \mathbb{X}$ if for all $0 \neq d \in \mathcal{C}(\bar{x})$,*

$$\sup_{(\overline{Y},\bar{y},\overline{Z})\in\mathcal{M}(\bar{x})} \left\{\langle d, L''_{xx}(\bar{x}, \bar{p}, \overline{Y}, \bar{y}, \overline{Z})d\rangle - \Upsilon^1_{\overline{X}_1}\big(\overline{Y}, (g_1)'_x(\bar{x}, \bar{p})d\big) - \Upsilon^2_{\overline{X}_2}\big(\overline{Z}, (g_2)'_x(\bar{x}, \bar{p})d\big)\right\} > 0,$$

where the function $\Upsilon^1_{\overline{X}_1}(\overline{Y}, H) : \partial\theta_1(\overline{X}_1) \times \mathbb{S}^m \to \mathbb{R}$ and $\Upsilon^2_{\overline{X}_2}(\overline{Z}, H) : \mathcal{N}_{\mathcal{K}}(\overline{X}_2) \times \mathbb{S}^n \to \mathbb{R}$ are given in (3-53) *and* (3-54).

Definition 3.25. *The Robinson's constraint qualification (RCQ) of problem (3-77) with $(v, u_1, u_2, u_3) = (0, 0, 0, 0)$ at $\bar{x}$ is defined as*

$$\begin{bmatrix} (g_1)'_x(x) \\ h'_x(x) \\ (g_2)'_x(x) \end{bmatrix} \mathbb{X} + \begin{bmatrix} \mathcal{T}_{\theta_1}(g_1(x)) \\ \{0\} \\ \mathcal{T}_{\mathcal{K}}(g_2(x)) \end{bmatrix} = \begin{bmatrix} \mathbb{S}^m \\ \mathbb{Y} \\ \mathbb{S}^n \end{bmatrix}. \tag{3-79}$$

Definition 3.26. *The strict Robinson's constraint qualification (SRCQ) of problem (3-77) with $(v, u_1, u_2, u_3) = (0, 0, 0, 0)$ at $\bar{x}$ with respect to $(\overline{Y}, \bar{y}, \overline{Z}) \in \mathcal{M}(\bar{x}) \neq \emptyset$ is defined as*

$$\begin{bmatrix} (g_1)'_x(x) \\ h'_x(x) \\ (g_2)'_x(x) \end{bmatrix} \mathbb{X} + \begin{bmatrix} \mathcal{T}_{\theta_1}(g_1(x)) \\ \{0\} \\ \mathcal{T}_{\mathcal{K}}(g_2(x)) \end{bmatrix} \cap \begin{bmatrix} \overline{Y} \\ \bar{y} \\ \overline{Z} \end{bmatrix}^{\perp} = \begin{bmatrix} \mathbb{S}^m \\ \mathbb{Y} \\ \mathbb{S}^n \end{bmatrix}. \tag{3-80}$$

Proposition 3.27. *Let $\bar{x}$ be a stationary point for problem* (3-77) *with $(v, u_1, u_2, u_3) = (0, 0, 0, 0)$. If the SOSC (Definition 3.24) for problem* (3-77) *holds at $\bar{x}$ and the SRCQ (Definition 3.26) holds at $\bar{x}$ with respect to $(\overline{Y}, \bar{y}, \overline{Z}) \in \mathcal{M}(\bar{x})$, then F^{-1} is isolated calm at the origin for $(\bar{x}, \overline{Y}, \bar{y}, \overline{Z})$, where F is the natural mapping defined by*

$$F(x; Y, y, Z) := \begin{bmatrix} L'_x(x, Y, y, Z) \\ g_1(x) - \mathrm{Pr}_{\theta_1}(g_1(x) + Y) \\ h(x) \\ g_2(x) - \Pi_{\mathcal{K}}(g_2(x) + Z) \end{bmatrix}.$$

Proof. Let $(\Delta x, \Delta Y, \Delta y, \Delta Z) \in \mathbb{X} \times \mathbb{S}^m \times \mathbb{Y} \times \mathbb{S}^n$ be arbitrarily chosen such that

$$\begin{cases} L''_{xx}(\bar{x}, \overline{Y}, \bar{y}, \overline{Z})\Delta x + h'(\bar{x})^*\Delta y + g_1'(\bar{x})^*\Delta Y + g_2'(\bar{x})^*\Delta Z = 0 \\ g_1'(\bar{x})\Delta x - \mathrm{Pr}'_{\theta_1}(g_1(\bar{x}) + \overline{Y}; g_1'(\bar{x})\Delta x + \Delta Y) = 0 \\ g_2'(\bar{x})\Delta x - \Pi'_{\mathcal{K}}(g_2(\bar{x}) + \overline{Z}; g_2'(\bar{x})\Delta x + \Delta Z) = 0 \\ h'(\bar{x})\Delta x - \Pi'_{\{0\}}(h(\bar{x}) + \bar{y}; h'(\bar{x})\Delta x + \Delta y) = 0 \end{cases}. \tag{3-81}$$

By part (i) of Lemma 3.22 and [91, Lemma 10], we know from the last three relationships in (3-81) that

$$g_1'(\bar{x})\Delta x \in \mathcal{C}_{\theta_1}(g_1(\bar{x}), \overline{Y}) \quad \text{and} \quad \langle g_1'(\bar{x})\Delta x, \Delta Y\rangle = -\Upsilon^1_{g_1(\bar{x})}(\overline{Y}, g_1'(\bar{x})\Delta x)$$

$$g_2'(\bar{x})\Delta x \in \mathcal{C}_{\mathcal{K}}(g_2(\bar{x}), \overline{Z}) \quad \text{and} \quad \langle g_2'(\bar{x})\Delta x, \Delta Z\rangle = -\Upsilon^2_{g_2(\bar{x})}(\overline{Z}, g_2'(\bar{x})\Delta x)$$
$$h'(\bar{x})\Delta x \in \mathcal{C}_{\{0\}}(h(\bar{x}), \bar{y}) \quad \text{and} \quad \langle h'(\bar{x})\Delta x, \Delta y\rangle = 0$$

Thus we have $\Delta x \in \mathcal{C}(\bar{x})$. By taking the inner product between Δx and both sides of the first equation of (3-81), respectively, we obtain that

$$\begin{aligned} 0 &= \langle \Delta x, L''_{xx}(\bar{x}, \overline{Y}, \bar{y}, \overline{Z})\Delta x\rangle + \langle \Delta x, h'(\bar{x})^*\Delta y\rangle + \langle \Delta x, g_1'(\bar{x})^*\Delta Y\rangle + \langle \Delta x, g_2'(\bar{x})^*\Delta Z\rangle \\ &= \langle \Delta x, L''_{xx}(\bar{x}, \overline{Y}, \bar{y}, \overline{Z})\Delta x\rangle - \Upsilon^1_{g_1(\bar{x})}(\overline{Y}, g_1'(\bar{x})\Delta x) - \Upsilon^2_{g_2(\bar{x})}(\overline{Z}, g_2'(\bar{x})\Delta x). \end{aligned}$$

It follows from the SOSC (Definition 3.24) for problem (3-77) with $(v, u_1, u_2, u_3) = (0, 0, 0, 0)$ that $\Delta x = 0$. Therefore, (3-81) is reduced to

$$\begin{cases} h'(\bar{x})^*\Delta y + g_1'(\bar{x})^*\Delta Y + g_2'(\bar{x})^*\Delta Z = 0 \\ \mathrm{Pr}'_{\theta_1}(g_1(\bar{x}) + \overline{Y}; \Delta Y) = 0 \\ \Pi'_{\mathcal{K}}(g_2(\bar{x}) + \overline{Z}; \Delta Z) = 0 \\ \Pi'_{\{0\}}(h(\bar{x}) + \bar{y}; \Delta y) = 0 \end{cases} \tag{3-82}$$

By part (ii) of Lemma 3.22 and Remark 3.5, we have

$$\begin{bmatrix} \Delta Y \\ \Delta y \\ \Delta Z \end{bmatrix} \in \left(\begin{bmatrix} (g_1)'_x(x) \\ h'_x(x) \\ (g_2)'_x(x) \end{bmatrix} \mathbb{X} + \begin{bmatrix} \mathcal{T}_{\theta_1}(g_1(x)) \\ \{0\} \\ \mathcal{T}_{\mathcal{K}}(g_2(x)) \end{bmatrix} \cap \begin{bmatrix} \overline{Y} \\ \bar{y} \\ \overline{Z} \end{bmatrix}^{\perp} \right)^{\circ}.$$

Combining this with SRCQ, we know that $(\Delta Y, \Delta y, \Delta Z) = 0$. Thus, the only $(\Delta x, \Delta Y, \Delta y, \Delta Z) \in \mathbb{X} \times \mathbb{S}^m \times \mathbb{Y} \times \mathbb{S}^n$ satisfying (3-81) is $(\Delta x, \Delta Y, \Delta y, \Delta Z) = (0, 0, 0, 0)$. It follows from the C^2-cone reducibility of θ_1 and $\mathcal{K}$ that F is locally Lipschitz continuous around $(\bar{x}, \overline{Y}, \bar{y}, \overline{Z})$ and is directionally differentiable at $(\bar{x}, \overline{Y}, \bar{y}, \overline{Z})$. Then we know from [91, Lemma 4] and [1, 8(19)] that F^{-1} is isolated calm at the origin for $(\bar{x}, \overline{Y}, \bar{y}, \overline{Z})$ if and only if $F'(\bar{x}, \overline{Y}, \bar{y}, \overline{Z}; (\Delta x, \Delta Y, \Delta y, \Delta Z)) = 0$ implies $(\Delta x, \Delta Y, \Delta y, \Delta Z) = 0$, i.e.,

$$\text{(3-81)} \quad \text{implies} \quad (\Delta x, \Delta Y, \Delta y, \Delta Z) = 0. \tag{3-83}$$

Thus F^{-1} is isolated calm at the origin for $(\bar{x}, \overline{Y}, \bar{y}, \overline{Z})$. □

To explore the converse implication of the above proposition, we introduce the following lemma first.

Lemma 3.28. *Let $\bar{x}$ be a stationary point for problem* (3-77) *with* $(v, u_1, u_2, u_3) = (0, 0, 0, 0)$. *SRCQ (Definition 3.26) holds at $\bar{x}$ with respect to $(\overline{Y}, \bar{y}, \overline{Z}) \in \mathcal{M}(\bar{x})$. Suppose F^{-1} is isolated calm at the origin for $(\bar{x}, \overline{Y}, \bar{y}, \overline{Z})$ and there exists $\Delta x \in \mathcal{C}(\bar{x})\backslash\{0\}$ such that*

$$\langle \Delta x, L''_{xx}(\bar{x}, \overline{Y}, \bar{y}, \overline{Z})\Delta x\rangle - \Upsilon^1_{g_1(\bar{x})}(\overline{Y}, g_1'(\bar{x})\Delta x) - \Upsilon^2_{g_2(\bar{x})}(\overline{Z}, g_2'(\bar{x})\Delta x) = 0. \tag{3-84}$$

Then there exists $d \in \mathcal{C}(\bar{x})\backslash\{0\}$ *such that*

$$\langle d, L''_{xx}(\bar{x},\overline{Y},\bar{y},\overline{Z})\Delta x\rangle - \frac{1}{2}\langle d, g_1'(\bar{x})^*(\Upsilon^1_{g_1(\bar{x})})'(\overline{Y}, g_1'(\bar{x})\Delta x)\rangle - \frac{1}{2}\langle d, g_2'(\bar{x})^*(\Upsilon^2_{g_2(\bar{x})})'(\overline{Z}, g_2'(\bar{x})\Delta x)\rangle < 0, \tag{3-85}$$

where $(\Upsilon^1_{g_1(\bar{x})})'(\overline{Y}, H)$ *denotes the derivative of* $\Upsilon^1_{g_1(\bar{x})}(\overline{Y},\cdot)$ *on* H.

Proof. Suppose for every $d \in \mathcal{C}(\bar{x})$, the inequality (3-85) fails. By (3-84) and noting that

$$\langle(\Upsilon^1_{g_1(\bar{x})})'(\overline{Y}, g_1'(\bar{x})\Delta x), g_1'(\bar{x})\Delta x\rangle = 2\Upsilon^1_{g_1(\bar{x})}(\overline{Y}, g_1'(\bar{x})\Delta x),$$
$$\langle(\Upsilon^1_{g_2(\bar{x})})'(\overline{Z}, g_2'(\bar{x})\Delta x), g_2'(\bar{x})\Delta x\rangle = 2\Upsilon^2_{g_2(\bar{x})}(\overline{Z}, g_2'(\bar{x})\Delta x),$$

we know that Δx is an optimal solution to the following linear conic programming problem

$$\begin{array}{ll} \min & \langle d, L''_{xx}(\bar{x},\overline{Y},\bar{y},\overline{Z})\Delta x\rangle - \frac{1}{2}\langle d, g_1'(\bar{x})^*(\Upsilon^1_{g_1(\bar{x})})'(\overline{Y}, g_1'(\bar{x})\Delta x)\rangle \\ & -\frac{1}{2}\langle d, g_2'(\bar{x})^*(\Upsilon^2_{g_2(\bar{x})})'(\overline{Z}, g_2'(\bar{x})\Delta x)\rangle \\ \text{s.t.} & g_1'(\bar{x})d \in \mathcal{C}_{\theta_1}(g_1(\bar{x}),\overline{Y}) \\ & g_2'(\bar{x})d \in \mathcal{C}_{\theta_1}(g_2(\bar{x}),\overline{Z}) \\ & h'(\bar{x})d = 0. \end{array} \tag{3-86}$$

It is clear that the RCQ of (3-86) holds at Δx, since SRCQ is satisfied at $\bar{x}$ with respect to $(\overline{Y},\bar{y},\overline{Z})$. Let $G(x) = (h(x), g_1(x), g_2(x))$. Therefore, we know that there exists $\Delta\mu = (\Delta z, \Delta W_1, \Delta W_2) \in \mathbb{S}^m \times \mathbb{Y} \times \mathbb{S}^n$ such that

$$\begin{cases} L''_{xx}(\bar{x},\overline{Y},\bar{y},\overline{Z})\Delta x - \frac{1}{2}g_1'(\bar{x})^*(\Upsilon^1_{g_1(\bar{x})})'(\overline{Y}, g_1'(\bar{x})\Delta x) - \frac{1}{2}g_2'(\bar{x})^*(\Upsilon^2_{g_2(\bar{x})})'(\overline{Z}, g_2'(\bar{x})\Delta x) + G'(\bar{x})^*\Delta\mu = 0 \\ \Delta W_1 \in \mathcal{N}_{\mathcal{C}_{\theta_1}(g_1(\bar{x}),\overline{Y})}(g_1'(\bar{x})\Delta x) \\ \Delta W_2 \in \mathcal{N}_{\mathcal{C}_{\mathcal{K}}(g_2(\bar{x}),\overline{Z})}(g_2'(\bar{x})\Delta x) \\ h'(\bar{x})\Delta x = 0. \end{cases} \tag{3-87}$$

Denote $\Delta A_1 := \Delta W_1 + \frac{1}{2}(\Upsilon^1_{g_1(\bar{x})})'(\overline{Y}, g_1'(\bar{x})\Delta x)$ and $\Delta A_2 := \Delta W_2 + \frac{1}{2}(\Upsilon^2_{g_2(\bar{x})})'(\overline{Z}, g_2'(\bar{x})\Delta x)$. By noting that $\Delta x \in \mathcal{C}(\bar{x})$ and $\langle(\Upsilon^1_{g_1(\bar{x})})'(\overline{Y}, g_1'(\bar{x})\Delta x), g_1'(\bar{x})\Delta x\rangle = 2\Upsilon^1_{g_1(\bar{x})}(\overline{Y}, g_1'(\bar{x})\Delta x)$, $\langle(\Upsilon^1_{g_2(\bar{x})})'(\overline{Z}, g_2'(\bar{x})\Delta x), g_2'(\bar{x})\Delta x\rangle = 2\Upsilon^2_{g_2(\bar{x})}(\overline{Z}, g_2'(\bar{x})\Delta x)$, we obtain that

$$\langle g_1'(\bar{x})\Delta x, \Delta A_1\rangle = \Upsilon^1_{g_1(\bar{x})}(\overline{Y}, g_1'(\bar{x})\Delta x), \quad \langle g_2'(\bar{x})\Delta x, \Delta A_2\rangle = \Upsilon^2_{g_2(\bar{x})}(\overline{Y}, g_2'(\bar{x})\Delta x).$$

Therefore, it follows from part (i) of Lemma 3.22 and [91, Lemma 10] that

$$g_1'(\bar{x})\Delta x - \mathrm{Pr}'_{\theta_1}(g_1(\bar{x}) + \overline{Y}; g_1'(\bar{x})\Delta x + \Delta A_1) = 0,$$

$$g_2'(\bar{x})\Delta x - \Pi'_{\mathcal{K}}(g_2(\bar{x}) + \overline{Z}; g_2'(\bar{x})\Delta x + \Delta A_2) = 0,$$

which, together with the first equations of (3-87), implies that $0 \neq (\Delta x, \Delta A_1, \Delta A_2)$ satisfies (3-83). This contradicts the isolated calmness of F^{-1} at origin for $(\bar{x},\overline{Y},\bar{y},\overline{Z})$. This contradiction completes the proof. □

Then, we can obtain the converse implication of Proposition 3.27.

Proposition 3.29. *Let $\bar{x}$ be a stationary point for problem* (3-77) *with* $(v, u_1, u_2, u_3) = (0, 0, 0, 0)$. *RCQ (Definition 3.25) holds at $\bar{x}$. If F^{-1} is isolated calm at the origin for $(\bar{x}, \overline{Y}, \bar{y}, \overline{Z})$, then*

(i) the SRCQ (Definition 3.26) holds at $\bar{x}$ with respect to $(\overline{Y}, \bar{y}, \overline{Z}) \in \mathcal{M}(\bar{x})$.

(ii) the SOSC (Definition 3.24) for problem (3-77) *holds at $\bar{x}$ with respect to* $(v, u_1, u_2, u_3) = (0, 0, 0, 0)$.

Proof. We first prove part (i). Assume, on the contrary, that the SRCQ does not hold at $\bar{x}$ with respect to $(\overline{Y}, \bar{y}, \overline{Z})$. Then, there exists $0 \neq (\Delta z, \Delta W_1, \Delta W_2) \in \mathbb{S}^m \times \mathbb{Y} \times \mathbb{S}^n$ such that

$$\begin{bmatrix} \Delta W_1 \\ \Delta z \\ \Delta W_2 \end{bmatrix} \in \left(\begin{bmatrix} (g_1)'_x(x) \\ h'_x(x) \\ (g_2)'_x(x) \end{bmatrix} \mathbb{X} + \begin{bmatrix} \mathcal{T}_{\theta_1}(g_1(x)) \\ \{0\} \\ \mathcal{T}_{\mathcal{K}}(g_2(x)) \end{bmatrix} \cap \begin{bmatrix} \overline{Y} \\ \bar{y} \\ \overline{Z} \end{bmatrix}^{\perp} \right)^{\circ}.$$

By part (ii) of Lemma 3.22 and Remark 3.5,

$$\begin{cases} h'(\bar{x})^* \Delta z + g_1'(\bar{x})^* \Delta W_1 + g_2'(\bar{x})^* \Delta W_2 = 0 \\ \mathrm{Pr}'_{\theta_1}(g_1(\bar{x}) + \overline{Y}; \Delta W_1) = 0 \\ \Pi'_{\mathcal{K}}(g_2(\bar{x}) + \overline{Z}; \Delta W_2) = 0 \\ \Pi'_{\{0\}}(h(\bar{x}) + \bar{y}; \Delta z) = 0, \end{cases} \tag{3-88}$$

which implies that $F'((\bar{x}, \overline{Y}, \bar{y}, \overline{Z}); (0, \Delta z, \Delta W_1, \Delta W_2)) = 0$. Since F^{-1} is isolated calm at the origin for $(\bar{x}, \overline{Y}, \bar{y}, \overline{Z})$, we know from (3-83) that $(\Delta z, \Delta W_1, \Delta W_2) = 0$. This contradiction shows that the SRCQ holds at $\bar{x}$ with respect to $(\overline{Y}, \bar{y}, \overline{Z})$.

Next, we shall prove part (ii). Since $\bar{x}$ is a locally optimal solution to problem (3-77) with $(v, u_1, u_2, u_3) = (0, 0, 0, 0)$ and the SRCQ holds at $\bar{x}$, we know from [5, Theorem 3.108] that for all $d \in \mathcal{C}(\bar{x})$,

$$\langle d, L''_{xx}(\bar{x}, \overline{Y}, \bar{y}, \overline{Z}) d \rangle - \Upsilon^1_{g_1(\bar{x})}(\overline{Y}, g_1'(\bar{x})\Delta x) - \Upsilon^2_{g_2(\bar{x})}(\overline{Z}, g_2'(\bar{x})\Delta x) \geq 0. \tag{3-89}$$

Therefore, if we assume that the SOSC does not hold at $\bar{x}$, then there exists $\Delta x \in \mathcal{C}(\bar{x}) \backslash \{0\}$ such that (3-84) holds. Thus, we know from Lemma 3.28 that there exists $d \in \mathcal{C}(\bar{x})$ such that (3-85) holds. Hence, for any $\tau > 0$, we have

$$\begin{aligned} &\langle \Delta x + \tau d, L''_{xx}(\bar{x}, \overline{Y}, \bar{y}, \overline{Z})(\Delta x + \tau d) \rangle - \Upsilon^1_{g_1(\bar{x})}(\overline{Y}, g_1'(\bar{x})(\Delta x + \tau d)) - \Upsilon^2_{g_2(\bar{x})}(\overline{Z}, g_2'(\bar{x})(\Delta x + \tau d)) \\ =& \langle \Delta x, L''_{xx}(\bar{x}, \overline{Y}, \bar{y}, \overline{Z}) \Delta x \rangle - \Upsilon^1_{g_1(\bar{x})}(\overline{Y}, g_1'(\bar{x})\Delta x) - \Upsilon^2_{g_2(\bar{x})}(\overline{Z}, g_2'(\bar{x})\Delta x) \\ &+ \tau^2 \big(\langle d, L''_{xx}(\bar{x}, \overline{Y}, \bar{y}, \overline{Z}) d \rangle - \Upsilon^1_{g_1(\bar{x})}(\overline{Y}, g_1'(\bar{x})\tau d) - \Upsilon^2_{g_2(\bar{x})}(\overline{Z}, g_2'(\bar{x})\tau d) \big) \\ &+ 2\tau \big(\langle d, L''_{xx}(\bar{x}, \overline{Y}, \bar{y}, \overline{Z}) \Delta x \rangle \\ &- \frac{1}{2} \langle d, g_1'(\bar{x})^* (\Upsilon^1_{g_1(\bar{x})})'(\overline{Y}, g_1'(\bar{x})\Delta x) \rangle - \frac{1}{2} \langle d, g_2'(\bar{x})^* (\Upsilon^2_{g_2(\bar{x})})'(\overline{Z}, g_2'(\bar{x})\Delta x) \rangle \big). \end{aligned}$$

Since $\mathcal{C}(\bar{x})$ is a convex cone, it follows from (3-85) that for $\tau > 0$ sufficiently small, $\Delta x + \tau d \in \mathcal{C}(\bar{x})$ and

$$\langle \Delta x+\tau d, L''_{xx}(\bar{x},\overline{Y},\bar{y},\overline{Z})(\Delta x+\tau d)\rangle - \Upsilon^1_{g_1(\bar{x})}(\overline{Y}, g_1'(\bar{x})(\Delta x+\tau d)) - \Upsilon^2_{g_2(\bar{x})}(\overline{Z}, g_2'(\bar{x})(\Delta x+\tau d)) < 0,$$

which is in contradiction with the second-order necessary condition (3-89). This completes the proof. □

Lemma 3.30. *Let* $(0,0,0,0,\bar{x},\overline{Y},\bar{y},\overline{Z}) \in \operatorname{gph} \mathcal{S}_{\rm KKT}$, *where*

$$\begin{aligned}\mathcal{S}_{KKT}(v,u_1,u_2,u_3) = & \{(x, W_1 - \mathrm{Pr}_{\theta_1}(W_1), y, W_2 - \Pi_{\mathcal{K}}(W_2)) \in \mathbb{X}\times\mathbb{S}^m\times\mathbb{Y}\times\mathbb{S}^n \mid \\ & \Psi(x,W_1,y,W_2) = (v,u_1,u_2,u_3)\},\end{aligned}$$

and Ψ *is the Robinson's normal mapping defined by*

$$\Psi(x,W_1,y,W_2) \begin{bmatrix} f'(x) + h'(x)^* y + g_1'(x)^*(W_1 - \mathrm{Pr}_{\theta_1}(W_1)) + g_2'(x)^*(W_2 - \Pi_{\mathcal{K}}(W_2)) \\ h(x) \\ g_1(x) - \mathrm{Pr}_{\theta_1}(W_1) \\ g_2(x) - \Pi_{\mathcal{K}}(W_2) \end{bmatrix}.$$

The set-valued mapping $\mathcal{S}_{KKT}$ *is isolated calm at the origin for* $(\bar{x},\overline{Y},\bar{y},\overline{Z})$ *if and only if the set-valued mapping* F^{-1} *is isolated calm at the origin for* $(\bar{x},\overline{Y},\bar{y},\overline{Z})$.

Proof. The proof is in a similar manner to that of [91, Lemma 18]. We omit it here for simplicity. □

Combining the results in this subsection together, we obtain the main result in this subsection

Theorem 3.31. *Let* $\bar{x}$ *be a stationary point for problem* (3-77) *with* $(v,u_1,u_2,u_3) = (0,0,0,0)$. *RCQ (Definition 3.25) holds at* $\bar{x}$. *Let* $(\overline{Y},\bar{y},\overline{Z}) \in \mathcal{M}(\bar{x}) \neq \emptyset$. *Then the following statements are equivalent:*

(i) the SRCQ holds at $\bar{x}$ *with respect to* $(\overline{Y},\bar{y},\overline{Z})$ *and the SOSC holds at* $\bar{x}$ *for problem* (3-77) *with* $(v,u_1,u_2,u_3) = (0,0,0,0)$*;*

(ii) $\bar{x}$ *is a locally optimal solution to problem* (3-77) *with* $(v,u_1,u_2,u_3) = (0,0,0,0)$ *and* $\mathcal{S}_{KKT}$ *is robustly isolated calm at the origin for* $(\bar{x},\overline{Y},\bar{y},\overline{Z})$*;*

(iii) $\bar{x}$ *is a locally optimal solution to problem* (3-77) *with* $(v,u_1,u_2,u_3) = (0,0,0,0)$ *and* $\mathcal{S}_{KKT}$ *is isolated calm at the origin for* $(\bar{x},\overline{Y},\bar{y},\overline{Z})$*;*

(iv) $\bar{x}$ *is a locally optimal solution to problem* (3-77) *with* $(v,u_1,u_2,u_3) = (0,0,0,0)$ *and* F^{-1} *is isolated calm at the origin for* $(\bar{x},\overline{Y},\bar{y},\overline{Z})$*;*

Proof. We only need to consider the equivalence between (ii) and (iii). We can obtain the similar result of [91, Theorem 17] for (3-77) by considering θ_1 as a constraint, i.e., $(g_1(x), t) \in \mathrm{gph}\theta_1$. Then it follows from the similar results of [91, Remark 1, Theorem 17] that the equivalence holds. □

3.4 A sufficient condition for the semi-isolated calmness of S_{KKT} (3-91)

In this section, we give a sufficient condition for the semi-isolated calmness of KKT pair. Firstly, we introduce the definition of semi-isolated calmness. For given perturbation parameters $(v, u_1, , u_2) \in \mathbb{X}\times\mathbb{Y}\times\mathbb{S}^n$, the corresponding canonical perturbation counterpart of (2-48) is given by

$$\begin{aligned} \min_{x\in\mathbb{X}} \quad & f(x) - \langle v, x\rangle \\ \text{s.t.} \quad & h(x) - u_1 = 0, \\ & G(x) - u_2 \in \mathbb{S}^n_+. \end{aligned} \tag{3-90}$$

Denote $S_{KKT}(v, u_1, , u_2)$ the solution set of the KKT optimality condition for problem (3-90), i.e.,

$$S_{KKT}(v, u_1, , u_2) = \left\{ (x, y, \Gamma) \in \mathbb{X}\times\mathbb{Y}\times\mathbb{S}^n : \begin{array}{l} L'_x(x, y, \Gamma) - v = 0, \\ h(x) - u_1 = 0, \\ \mathbb{S}^n_+ \ni (G(x) - u_2) \perp \Gamma \in \mathbb{S}^n_-. \end{array} \right\}. \tag{3-91}$$

Define the set of all multipliers as

$$\mathcal{M}(\bar{x}) = \{(y, \Gamma) \in \mathbb{Y}\times\mathbb{S}^n \mid (\bar{x}, y, \Gamma) \in S_{KKT}(0, 0, 0)\}. \tag{3-92}$$

It is easy to see $\mathcal{M}(\bar{x})$ is convex.

The definition of semi-isolated calmness of a set-valued mapping is first officially presented by [47, Theorem 4.1], which is an extension of [93, Proposition 1.43], to establish the characterization of noncritical multipliers for the polyhedral problem (to be more specifically, composite piecewise linear quadratic problem). It is worth noting that for the polyhedral case, semi-isolated calmness is equivalent to noncritical [47, Theorem 4.1], though this does not hold for the non-polyhedral case as stated in [94, Theorem 5.6].

Definition 3.32. *The semi-isolated clamness for the mapping S_{KKT} at $((0, 0, 0), (\bar{x}, \bar{y}, \overline{\Gamma}))$ holds if there exists $\kappa > 0$ and open neighborhoods $\mathcal{U}$ of $(0, 0, 0)$ and $\mathcal{V}$ of $(\bar{x}, \bar{y}, \overline{\Gamma})$ such that for all $(v, u_1, , u_2) \in \mathcal{U}$,*

$$\|x - \bar{x}\| + \mathrm{dist}((y, \Gamma), \mathcal{M}(\bar{x})) \le \kappa\|(v, u_1, , u_2)\| \quad \forall\, (x, y, \Gamma) \in S_{KKT}(v, u_1, , u_2) \cap \mathcal{V}.$$

In order to provide a sufficient condition for the semi-isolated calmness of S_{KKT}, we need the definition of bounded linear regularity of a collection of closed convex sets, which can be found in, e.g., [95, Definition 5.6].

Definition 3.33. *Let $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_m \subseteq \mathbb{X}$ be closed convex sets for some positive integer m. Suppose that $\mathcal{D} := \mathcal{D}_1 \cap \mathcal{D}_2 \cap \ldots \cap \mathcal{D}_m$ is non-empty. The collection $\{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_m\}$*

is said to be boundedly linearly regular if for every bounded set $\mathcal{B} \subseteq \mathbb{X}$*, there exists a constant* $\kappa > 0$ *such that*

$$\operatorname{dist}(x, \mathcal{D}) \leqslant \kappa \max\{\operatorname{dist}(x, \mathcal{D}_1), \ldots, \operatorname{dist}(x, \mathcal{D}_m)\}, \forall x \in \mathcal{B}.$$

A sufficient condition to guarantee the property of bounded linear regularity is established in [96, Corollary 3]. Denote $\mathcal{G}_1(\bar{x}) = \{(y, \Gamma) \in \mathbb{Y} \times \mathbb{S}^n \mid f'(\bar{x}) + h'(\bar{x})^* y + G'(\bar{x})^* \Gamma = 0\}$ and $\mathcal{G}_2(\bar{x}) = \{(y, \Gamma) \in \mathbb{Y} \times \mathbb{S}^n \mid \Gamma \in \mathcal{N}_{\mathbb{S}^n_+}(G(\bar{x}))\}$. It is easy to see that $\mathcal{G}_1(\bar{x})$ is a polyhedron and $\mathcal{G}_2(\bar{x})$ is convex. Along with [97, Theorem 3.1], the following result gives a sufficient condition for semi-isolated calmness. Its proof is inspired from [94, Theorem 5.9] by using a similar reduction method and we omit it here for brevity.

Theorem 3.34. *Let* $\bar{x} \in \mathbb{X}$ *be a stationary point to the NLSDP* (3-90) *with* $(v, u_1, , u_2) = (0, 0, 0)$ *and* $(\bar{y}, \overline{\Gamma}) \in \mathcal{M}(\bar{x})$*. Suppose SOSC* (4-25) *holds at* $(\bar{x}, \bar{y}, \overline{\Gamma})$ *and*

$$\mathcal{G}_1(\bar{x}) \cap \operatorname{ri} \mathcal{G}_2(\bar{x}) \neq \emptyset. \tag{3-93}$$

Then there exist a constant $\kappa_1 > 0$*, neighborhoods* $\mathcal{V} := \mathcal{B}_{r_1}(\bar{x}, \bar{y}, \overline{\Gamma})$ *of* $(\bar{x}, \bar{y}, \overline{\Gamma})$ *and* $\mathcal{U}$ *of* $(0, 0, 0)$ *such that for any* $(v, u_1, , u_2) \in \mathcal{U}$*,*

$$\|x - \bar{x}\| + \operatorname{dist}((y, \Gamma), \mathcal{M}(\bar{x})) \leq \kappa_1 \|(v, u_1, , u_2)\| \quad \forall\, (x, y, \Gamma) \in S_{KKT}(v, u_1, , u_2) \cap \mathcal{V}.$$

Regarding to condition (3-93), it suffice to admit a point $(y', \Gamma') \in \mathcal{M}(\bar{x})$ possessing the strict complementarity condition. We have to emphasis that this (y', Γ') can be different from both the reference point $(\bar{y}, \overline{\Gamma})$ and the limit point of the sequence generated by ALM, which is mentioned in Theorem 4.14. In addition, it is worth to note that when $\mathcal{M}(\bar{x})$ is a singleton, the semi-isolated calmness of S_{KKT} is reduced to its isolated calmness.

Remark 3.6. It follows from [98, Proposition 3.2] (see also [84, Proposition 17]) that the bounded linear regularity of the collection $\{\mathcal{G}_1(\bar{x}), \mathcal{G}_2(\bar{x})\}$ holds under one of the following two conditions:

(i) $\mathcal{G}_2(\bar{x})$ is a polyhedron, i.e., $|\gamma| \geq n - 1$, where γ is the negative eigenvalue index set of matrix $G(\bar{x}) + \overline{\Gamma}$;

(ii) there exists a strict complementarity KKT pair $(\bar{x}, \widetilde{y}, \widetilde{\Gamma})$ ($(\widetilde{y}, \widetilde{\Gamma})$ does not have to be $(\bar{y}, \overline{\Gamma})$), i.e., $\operatorname{rank}(G(\bar{x})) + \operatorname{rank}(\widetilde{\Gamma}) = n$.

We can also study the validity of semi-isolated calmness directly from [94, Theorem 5.9]. Suppose $(\bar{x}, \bar{y}, \overline{\Gamma})$ is a KKT pair. It follows that the key of the validity of semi-isolated calmness mainly lies in when S is calm at $(\bar{y}, \overline{\Gamma})$ for $(v, u) = (0, 0)$, where $S(v, u) = \{(y, \Gamma) \in \mathbb{Y} \times \mathbb{S}^n_+ \mid L'_x(\bar{x}, y, \Gamma) = v,\ \Gamma \in \mathcal{N}_{\mathbb{S}^n_+}(G(\bar{x}) + u)\}$. It is easy to see that the strong regularity/isolated calmness/Aubin of S implies the calmness of S.

Chapter 4 Convergence analysis of ALM for NLSDP under semi-isolated calmness

As shown in the above chapter, the constraint nondegenerate condition together with strong SOSC is equivalent to strong regularity for CMatOP, which is a very good stability property that brings a lot of convenience in the theoretical analysis of augmented Lagrangian method (ALM). It can be seen from [51] that the above condition can be relaxed to SOSC with SRCQ, which is equivalent to robust isolated calmness [91]. However, existing results usually require the uniqueness of multipliers for nonpolyhedral nonconvex problems. This condition is of less practical use in the real world. Without loss of generality, we choose NLSDP as an example to illustrate the ALM behavior for nonpolyhedral nonconvex problems in the remaining of this thesis. In this chapter, without requiring the uniqueness of multipliers, the local (asymptotic Q-superlinear) Q-linear convergence rate of the primal-dual sequences generated by ALM for the nonlinear semidefinite programming (NLSDP) is established by assuming the second-order sufficient condition (SOSC) and the semi-isolated calmness of the Karush–Kuhn–Tucker (KKT) solution under some mild conditions.

The following definition is an extension of the second order expansion of functions defined in [99, Definition 1.1].

Definition 4.1. *Consider a function $f : \mathbb{X} \to \mathbb{R}$ and a point $\bar{x}$ where f is differentiable. We say f has a uniform second order expansion at $\bar{x}$ with certain conditions if f satisfies the following two conditions.*

(i) f has a second order expansion at $\bar{x}$, i.e., there exists a finite, continuous and positively homogeneous of degree 2 function g such that

$$f(\bar{x}+cv) = f(\bar{x}) + c\langle f'(\bar{x}), v\rangle + \frac{c^2}{2}g(v) + o(c^2\|v\|^2), \quad c \in \mathbb{R},\ v \in \mathbb{X}.$$

(ii) There exists a constant $r > 0$ such that $o(c^2\|v\|^2)$ is uniform for all $x \in \mathcal{B}_r(\bar{x})$ with certain conditions, i.e., for all $\varepsilon > 0$, there exist positive constants ω, r such that for all $x \in \mathcal{B}_r(\bar{x})$ with certain conditions and all $\|cv\| \le \omega$, we have

$$\frac{f(x+cv) - f(x) - c\langle \nabla f(x), v\rangle - \frac{c^2}{2}g(v)}{\|c^2v^2\|} \le \varepsilon,$$

where ω is uniform for all $x \in \mathcal{B}_r(\bar{x})$ with certain conditions.

4.1 The uniform second order expansion of $\frac{1}{2}\mathrm{dist}^2(\cdot, \mathbb{S}^n_+)$

In this section, we shall establish the uniform expansion of $\frac{1}{2}\mathrm{dist}^2(\cdot, \mathbb{S}^n_+)$, which is crucial for the subsequent analysis of deriving the uniform quadratic growth condition of augmented Lagrangian function. Before we step further, we need to give the following notation. Given two Euclidean spaces $\mathbb{X}$, $\mathbb{Z}$, a positive constant r and $\overline{A} \in \mathbb{X}$. Considering a mapping $\Delta : \mathbb{X} \times \mathbb{X} \to \mathbb{Z}$. We say $\Delta(A, H) = O(\|H\|)$ with $O(\|H\|)$ uniform for all $A \in \mathcal{B}_r(\overline{A})$ with certain conditions if there exist positive constants q, ω, r such that for all $A \in \mathcal{B}_r(\overline{A})$ with certain conditions and all $\|H\| \le \omega$, we have

$$\frac{\|\Delta(A, H)\|}{\|H\|} \le q,$$

where ω and q are uniform for all $A \in \mathcal{B}_r(\overline{A})$ with certain conditions. We call ω the uniform radius and q the uniform constant of $O(\|H\|)$. To obtain the main result of this section, firstly we need the following lemma, which illustrates the uniform expansion for eigenvalue vector matrix. Its non-uniform form was stated in [82] and essentially proved in the derivation of [79, Lemma 4.12].

Lemma 4.2. *Given a fixed $\overline{A} \in \mathbb{S}^n$. Let $0 < r < \min_{i<j}\{v_i(\overline{A}) - v_j(\overline{A})\}/3$. For any $H \in \mathbb{S}^n$ and $A \in \mathcal{B}_r(\overline{A})$, let U be an orthogonal matrix such that*

$$U^T\big(\Lambda(A) + H\big)U = \Lambda\big(\Lambda(A) + H\big). \tag{4-1}$$

Then, for any $H \to 0$, we have

$$\begin{cases} U_{\bar{\alpha}^s\bar{\alpha}^t} = O(\|H\|), & s, t = 1, \cdots, \bar{d},\ s \neq t \\ U_{\bar{\alpha}^s\bar{\alpha}^s}U^T_{\bar{\alpha}^s\bar{\alpha}^s} = I_{|\bar{\alpha}^s|} + O(\|H\|^2), & s = 1, \cdots, \bar{d}, \end{cases} \tag{4-2}$$

where $\bar{\alpha}^s := \alpha^s(\overline{A}) = \{i \mid \lambda_i(\overline{A}) = v_s(\overline{A})\}$, $s = 1, \ldots, \bar{d}$. Furthermore, for each $s \in \{1, \ldots, \bar{d}\}$, there exists $Q_s \in \mathcal{O}^{|\bar{\alpha}^s|}$ such that

$$U_{\bar{\alpha}^s\bar{\alpha}^s} = Q_k + O(\|H\|^2) \tag{4-3}$$

and

$$Q_s^T H_{\bar{\alpha}^s\bar{\alpha}^s} Q_s = \Lambda_{\bar{\alpha}^s\bar{\alpha}^s}(\Lambda(X) + H) - Q_s^T \Lambda(A)_{\bar{\alpha}^s\bar{\alpha}^s} Q_s + O(\|H\|^2). \tag{4-4}$$

It is worth to note that the $O(\|H\|)$ and $O(\|H\|^2)$ above are uniform for all $A \in \mathcal{B}_r(\overline{A})$.

Proof. For each $s \in \{1, \ldots, \bar{d}\}$, let $\Lambda_{\bar{\alpha}^s\bar{\alpha}^s} = \mathrm{Diag}(\lambda_{\bar{\alpha}^s}(A))$ and $\Xi_{\bar{\alpha}^s\bar{\alpha}^s} = \mathrm{Diag}(\lambda_{\bar{\alpha}^s}(A + H))$. We first show that (4-2) holds. If $\bar{d} = 1$, i.e., $\lambda_1(\overline{A}) = \cdots = \lambda_n(\overline{A})$, the first equation in (4-2) trivially holds. Next we assume that $\bar{d} \ge 2$. From (4-1), we have for any $\mathcal{S}^n \ni H \to 0$,

$$\begin{bmatrix} \Lambda_{\bar{\alpha}^1\bar{\alpha}^1} & 0 & \ldots & 0 \\ 0 & \Lambda_{\bar{\alpha}^2\bar{\alpha}^2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \Lambda_{\bar{\alpha}^{\bar{d}}\bar{\alpha}^{\bar{d}}} \end{bmatrix} U + HU = U \begin{bmatrix} \Xi_{\bar{\alpha}^1\bar{\alpha}^1} & 0 & \ldots & 0 \\ 0 & \Xi_{\bar{\alpha}^2\bar{\alpha}^2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \Xi_{\bar{\alpha}^{\bar{d}}\bar{\alpha}^{\bar{d}}} \end{bmatrix}.$$

It follows from $A \in \mathcal{B}_r(\overline{A})$ that $\lambda_i(A) \neq \lambda_j(A)$ whenever $i \in \bar{\alpha}^t$, $j \in \bar{\alpha}^s$ with $s \neq t$. It is easy to see that for all $i \in \bar{\alpha}^s$, $j \in \bar{\alpha}^t$ with $s \neq t$, $U_{ij} = \frac{(HU)_{ij}}{\Lambda_{ii} - \Xi_{jj}}$. Then we have for all $\|H\| \leq \omega := r/6$,

$$\frac{\|U_{\bar{\alpha}^s\bar{\alpha}^t}\|}{\|H\|} \leq \sum_{i\in\bar{\alpha}^s, j\in\bar{\alpha}^t} \frac{1}{(\Lambda_{ii} - \Xi_{jj})^2} \leq \sum_{i\in\bar{\alpha}^k, j\in\bar{\alpha}^l} \frac{1}{(|v_i(\overline{A}) - v_j(\overline{A})| - 2r - 2\omega)^2} := q,$$

where ω and q are independent of A. Hence we obtain that

$$U_{\bar{\alpha}^s\bar{\alpha}^t} = O(\|H\|) \quad \forall\, 1 \leq s \neq t \leq \bar{d},$$

where $O(\|H\|)$ is uniform for all $A \in \mathcal{B}_r(\overline{A})$. By using the fact that U is orthogonal, we obtain directly that the second equation in (4-2) holds. In order to prove (4-3), we consider the SVD of each $U_{\bar{\alpha}^s\bar{\alpha}^s}$, $s = 1, \ldots, \bar{d}$. Fix $s \in \{1, \ldots, \bar{d}\}$. Let W and V be in $\mathcal{O}^{|\bar{\alpha}^s|}$ such that $U_{\bar{\alpha}^s\bar{\alpha}^s} = W\Sigma V^T$, where Σ is a nonnegative diagonal matrix. From (4-2), we obtain that for all $A \in \mathcal{B}_r(\overline{A})$,

$$W\Sigma^2 W^T = I_{|\bar{\alpha}^s|} + O(\|H\|^2),$$

which is equivalent to

$$\Sigma^2 = W^T W + O(\|H\|^2) = I_{|\bar{\alpha}^s|} + O(\|H\|^2).$$

Since Σ is a nonnegative diagonal matrix, we may conclude that

$$\Sigma = \mathrm{Diag}(1 + O(\|H\|^2), \ldots, 1 + O(\|H\|^2)).$$

Therefore, from $U_{\bar{\alpha}^s\bar{\alpha}^s} = W\Sigma V^T$, we have $U_{\bar{\alpha}^s\bar{\alpha}^s} = WV^T + O(\|H\|^2)$. Since $WV^T \in \mathcal{O}^{|\bar{\alpha}^s|}$, we know that for all $A \in \mathcal{B}_r(\overline{A})$, (4-3) holds. Next, we shall show (4-4) holds. For each $s \in \{1, \ldots, \bar{d}\}$ by comparing the s-th diagonal block of both sides of (4-1), we obtain that

$$U_{\bar{\alpha}^s}^T(\Lambda(A) + H)U_{\bar{\alpha}^s} = \Xi_{\bar{\alpha}^s\bar{\alpha}^s}. \tag{4-5}$$

Fix $s \in \{1, \ldots, \bar{d}\}$. From (4-2) and (4-5), we know that

$$\begin{aligned} U_{\bar{\alpha}^s}^T\Lambda(A)U_{\bar{\alpha}^s} &= \begin{bmatrix} O(\|H\|) & U_{\bar{\alpha}^s\bar{\alpha}^s} & O(\|H\|) \end{bmatrix} \begin{bmatrix} \Lambda_1(A) & 0 & 0 \\ 0 & \Lambda(A)_{\bar{\alpha}^s\bar{\alpha}^s} & 0 \\ 0 & 0 & \Lambda_2(A) \end{bmatrix} \begin{bmatrix} O(\|H\|) \\ U_{\bar{\alpha}^s\bar{\alpha}^s} \\ O(\|H\|) \end{bmatrix} \\ &= O(\|H\|^2)\Lambda_1(A) + U_{\bar{\alpha}^s\bar{\alpha}^s}^T\Lambda(A)_{\bar{\alpha}^k\bar{\alpha}^k}U_{\bar{\alpha}^k\bar{\alpha}^k} + O(\|H\|^2)\Lambda_2(A). \end{aligned}$$

It follows that

$$\Xi_{\bar{\alpha}^k\bar{\alpha}^k} - \big(O(\|H\|^2)\Lambda_1(A) + U_{\bar{\alpha}^k\bar{\alpha}^s}^T\Lambda(A)_{\bar{\alpha}^s\bar{\alpha}^s}U_{\bar{\alpha}^s\bar{\alpha}^s} + O(\|H\|^2)\Lambda_2(A)\big) = U_{\bar{\alpha}^s\bar{\alpha}^s}^T H_{\bar{\alpha}^s\bar{\alpha}^s}U_{\bar{\alpha}^s\bar{\alpha}^s} + O(\|H\|^2).$$

Since $U_{\bar{\alpha}^s\bar{\alpha}^s} = Q_s + O(\|H\|^2)$ and $\|\Lambda(A)\| \leq \|\Lambda(\overline{A})\| + r$, we obtain that

$$Q_s^T H_{\bar{\alpha}^s\bar{\alpha}^s} Q_k = \Xi_{\bar{\alpha}^s\bar{\alpha}^s} - Q_s^T\Lambda(A)_{\bar{\alpha}^s\bar{\alpha}^s}Q_s + O(\|H\|^2).$$

Hence (4-4) holds with the uniform $O(\|H\|^2)$ for all $A \in \mathcal{B}_r(\overline{A})$. The proof is completed. □

By applying Lemma 4.2, we can obtain the following result.

Lemma 4.3. *Given* $\overline{A} \in \mathbb{S}^n$ *and let* $0 < r < \min_{i<j}\{v_i(\overline{A}) - v_j(\overline{A})\}/3$*. For any* $H \in \mathbb{S}^n$ *and* $A \in \mathcal{B}_r(\overline{A})$*, let* U *be an orthogonal matrix such that*

$$U^T\big(A+H\big)U = \Lambda(A+H). \tag{4-6}$$

For all $t \in \{1,\dots,\bar{d}\}$*, there exist* $Q_t \in \mathcal{O}^{|\bar{\iota}_t|}$ *(depends on* H*) such that for all* $H \to 0$*,*

$$(P^TU)_{\bar{\alpha}^s\bar{\alpha}^t} = \Theta^{st} \circ (\widetilde{H}_{\bar{\alpha}^s\bar{\alpha}^t}Q_t) + O(\|H\|^2), \quad s \neq t,$$

where $O(\|H\|^2)$ *is uniform for all* $A \in \mathcal{B}_r(\overline{A})$*,* $(\Theta^{st})_{ij} = 1/\big((\Lambda(A)_{\bar{\alpha}^t\bar{\alpha}^t})_{ii} - (\Lambda(A)_{\bar{\alpha}^s\bar{\alpha}^s})_{jj}\big)$ *and* $\widetilde{H} = P^THP$*,* $P \in \mathcal{O}^n(A)$*.*

Proof. We first consider the case where A is diagonal. For notational simplicity, let $\Lambda = \Lambda(A)$ and $\Xi = \Lambda(A+H)$. From (4-6), we have $AU + HU = U\Xi$, which implies

$$\Lambda_{\bar{\alpha}^s\bar{\alpha}^s}U_{\bar{\alpha}^s\bar{\alpha}^t} + (HU)_{\bar{\alpha}^s\bar{\alpha}^t} = U_{\bar{\alpha}^s\bar{\alpha}^t}\Xi_{\bar{\alpha}^t\bar{\alpha}^t}.$$

It follows that

$$\Lambda_{\bar{\alpha}^s\bar{\alpha}^s}U_{\bar{\alpha}^s\bar{\alpha}^t} + \sum_{j=1}^{\bar{d}} H_{\bar{\alpha}^s\bar{\alpha}^j}U_{\bar{\alpha}^j\bar{\alpha}^t} = U_{\bar{\alpha}^s\bar{\alpha}^t}\Xi_{\bar{\alpha}^t\bar{\alpha}^t}.$$

This, together with Lemma 4.2 shows that

$$U_{\bar{\alpha}^s\bar{\alpha}^t} = \Sigma^{st} \circ \sum_{j=1}^{\bar{d}} H_{\bar{\alpha}^s\bar{\alpha}^j}U_{\bar{\alpha}^j\bar{\alpha}^t} = \Sigma^{st} \circ H_{\bar{\alpha}^s\bar{\alpha}^t}Q_t + O(\|H\|^2) \tag{4-7}$$

where $(\Sigma^{st})_{ij} = 1/((\Xi_{\bar{\alpha}^t\bar{\alpha}^t})_i - (\Lambda_{\bar{\alpha}^s\bar{\alpha}^s})_j)$. It is easy to see that $1/\big((\Xi_{\bar{\alpha}^t\bar{\alpha}^t})_i - (\Lambda_{\bar{\alpha}^s\bar{\alpha}^s})_j\big) = 1/\big((\Lambda_{\bar{\alpha}^t\bar{\alpha}^t})_i - (\Lambda_{\bar{\alpha}^s\bar{\alpha}^s})_j\big) + O(\|H\|)$. Combining this with (4-7), we have

$$U_{\bar{\alpha}^s\bar{\alpha}^t} = \Theta^{st} \circ H_{\bar{\alpha}^s\bar{\alpha}^t}Q_t + O(\|H\|^2), \text{ with } O(\|H\|^2) \text{ uniform for all } A \in \mathcal{B}_r(\overline{A}).$$

Next we consider $A = P^T\Lambda(A)P^T$. Re-write (4-6) as

$$\Lambda(A) + P^THP = P^TU\Lambda(A+H)U^TP.$$

Let $\widetilde{H} := P^THP$. Since P is an orthogonal matrix, the following proof is the same as the diagonal case. Thus we have completed the proof. □

For two matrices $A, \overline{A} \in \mathbb{S}^n$, notation $\pi(A) = \pi(\overline{A})$ means these two matrices possess the same index sets of different eigenvalues, i.e, A and $\overline{A}$ both have $\bar{d}$ different eigenvalues with $\alpha^l(A) = \alpha^l(\overline{A}) := \{i \mid \lambda_i(\overline{A}) = v_s(\overline{A})\}$ for all $l = 1,\dots,\bar{d}$ and the numbers of positive, zero, negative eigenvalues equal. Applying Lemma 4.2, we can get the uniform 1-order B-differentiability of projection function for SDP case, which is an enhancement of [80, Proposition 2.6].

Proposition 4.4. *Given a fixed $\overline{A} \in \mathbb{S}^n$ and let $0 < r < \min_{i<j}\{v_i(\overline{A}) - v_j(\overline{A})\}/3$. The metric projection operator is uniformly 1-order B-differentiable for any $A \in \mathcal{B}_r(\overline{A})$ with $\pi(\overline{A}) = \pi(A)$, i.e., for $\mathbb{S}^n \ni H \to 0$,*

$$\Pi_{\mathbb{S}^n_+}(A+H) - \Pi_{\mathbb{S}^n_+}(A) - \Pi'_{\mathbb{S}^n_+}(A;H) = O(\|H\|^2) \tag{4-8}$$

and $O(\|H\|^2)$ is uniform for all $A \in \mathcal{B}_r(\overline{A})$ with $\pi(\overline{A}) = \pi(A)$.

Proof. Firstly, we show (4-8) holds for the case that $A = \Lambda(A)$. For any $H \in \mathcal{S}^n$, denote $Z = A + H$. Let $U \in \mathcal{O}^n$ (depending on H) be such that

$$\Lambda(A) + H = U\Lambda(Z)U^T. \tag{4-9}$$

Let $\omega > 0$ be any fixed number such that $0 \le \omega \le \frac{\lambda_{|\alpha|}(\overline{A})-r}{2}$ if $\alpha \neq \emptyset$ and be any fixed positive number otherwise. Then, define the following continuous scalar function

$$f(t) := \begin{cases} t & \text{if} \quad t > \omega \\ 2t - \omega & \text{if} \quad \frac{\omega}{2} < t < \delta \\ 0 & \text{if} \quad t < \frac{\omega}{2}. \end{cases}$$

Therefore, we have

$$\{\lambda_1(A), \ldots, \lambda_{|\alpha|}(A)\} \in (\omega, +\infty) \quad \text{and} \quad \{\lambda_{|\alpha|+1}(A), \ldots, \lambda_n(A)\} \in (-\infty, \frac{\omega}{2}).$$

For the scalar function f, let $F : \mathcal{S}^n \to \mathcal{S}^n$ be the corresponding Löwner’s operator, i.e., for any $W \in \mathcal{S}^n$,

$$F(W) := \sum_{i=1}^n f(\lambda_i(W))P_iP_i^T,$$

where $P \in \mathcal{O}^n(W)$. Since f is real analytic on the open set $(-\infty, \frac{\omega}{2}) \cup (\omega, +\infty)$, It is well-known that for H sufficiently close to zero,

$$F(A+H) - F(A) - F'(A)H = O(\|H\|^2) \tag{4-10}$$

and

$$F'(A)H = \begin{bmatrix} H_{\alpha\alpha} & H_{\alpha\beta} & \Sigma_{\alpha\gamma} \circ H_{\alpha\gamma} \\ H_{\alpha\beta}^T & 0 & 0 \\ \Sigma_{\alpha\gamma}^T \circ H_{\alpha\gamma}^T & 0 & 0 \end{bmatrix},$$

where $O(\|H\|^2)$ is independent of A for any $A \in \mathcal{B}_r(\overline{A})$ and $\Sigma \in \mathcal{S}^n$ is given by

$$\Sigma_{ij} = \frac{\max\{\lambda_i(A), 0\} - \max\{\lambda_j(A), 0\}}{\lambda_i(A) - \lambda_j(A)}, \quad i, j = 1, \ldots, n.$$

Let $R(\cdot) := \Pi_{\mathcal{S}^n_+}(\cdot) - F(\cdot)$. By the definition of f, we know that $F(A) = \Pi_{\mathcal{S}^n_+}(A)$, which implies that $R(A) = 0$. Meanwhile, it is clear that the matrix valued function

R is directionally differentiable at A, and the directional derivative of R for any given direction $H \in \mathcal{S}^n$ is given by

$$R'(A;H) = \Pi'_{\mathcal{S}^n_+}(A;H) - F'(A)H = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \Pi_{\mathbb{S}^{|\beta|}_+}(H_{\beta\beta}) & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

By the Lipschitz continuity of $\lambda(\cdot)$, we know that for H sufficiently close to zero,

$$\left\{\lambda_1(Z), \ldots, \lambda_{|\alpha|}(Z)\right\} \in (\omega, +\infty), \quad \left\{\lambda_{|\alpha|+1}(Z), \ldots, \lambda_{|\beta|}(Z)\right\} \in (-\infty, \frac{\omega}{2})$$

and

$$\{\lambda_{|\beta|+1}(Z), \ldots, \lambda_n(Z)\} \in (-\infty, 0).$$

Therefore, by the definition of F , we know that for H sufficiently close to zero,

$$R(A+H) = \Pi_{\mathbb{S}^n_+}(A+H) - F(A+H) = U \begin{bmatrix} 0 & 0 & 0 \\ 0 & \Pi_{\mathbb{S}^{|\beta|}_+}(\Lambda(Z)_{\beta\beta}) & 0 \\ 0 & 0 & 0 \end{bmatrix} U^T.$$

Since $U \in \mathcal{O}^n(Z)$, we know from Lemma 4.2 that for any $\mathcal{S}^n \ni H \to 0$, there exists an orthogonal matrix $Q \in \mathcal{O}^{|\beta|}$ such that

$$U_\beta = \begin{bmatrix} O(\|H\|) \\ U_{\beta\beta} \\ O(\|H\|) \end{bmatrix} \quad \text{and} \quad U_{\beta\beta} = Q + O(\|H\|^2), \tag{4-11}$$

Therefore, by noting that $\Pi_{\mathbb{S}^{|\beta|}_+}(\Lambda(Z)_{\beta\beta}) = O(\|H\|)$ and $O(\|H\|)$ is uniform for $A \in \mathcal{B}_r(\overline{A})$ with $\pi(\overline{A}) = \pi(A)$, we obtain from the above discussion that

$$R(A+H) - R(A) - R'(A;H) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & Q\Pi_{\mathbb{S}^{|\beta|}_+}(\Lambda(Z)_{\beta\beta})Q^T - \Pi_{\mathbb{S}^{|\beta|}_+}(H_{\beta\beta}) & 0 \\ 0 & 0 & 0 \end{bmatrix} + O(\|H\|^2)$$

By (4-9) and (4-11), we know that

$$\Lambda(Z)_{\beta\beta} = U_\beta^T \Lambda(A) U_\beta + U_\beta^T H U_\beta = U_{\beta\beta}^T H_{\beta\beta} U_{\beta\beta} + O(\|H\|^2) = Q^T H_{\beta\beta} Q + O(\|H\|^2)\,.$$

Since $Q \in \mathcal{O}^{|\beta|}$, we have

$$H_{\beta\beta} = Q\Lambda(Z)_{\beta\beta}Q^T + O(\|H\|^2)\,,$$

where $O(\|H\|)$ is uniform for $A \in \mathcal{B}_r(\overline{A})$ with $\pi(\overline{A}) = \pi(A)$. Combining this with the globally Lipschitz continuity of $\Pi_{\mathbb{S}^{|\beta|}_+}(\cdot)$ and $\Pi_{\mathbb{S}^{|\beta|}_+}(Q\Lambda(Z)_{\beta\beta}Q^T) = Q\Pi_{\mathbb{S}^{|\beta|}_+}(\Lambda(Z)_{\beta\beta})Q^T$, we obtain that

$$Q\Pi_{\mathbb{S}^{|\beta|}_+}(\Lambda(Z)_{\beta\beta})Q^T - \Pi_{\mathbb{S}^{|\beta|}_+}(H_{\beta\beta}) = O(\|H\|^2)\,.$$

Therefore,

$$R(A+H)-R(A)-R'(A;H)=O(\|H\|^2)\,. \tag{4-12}$$

By combining (4-10) and (4-12), we know that for any $\mathbb{S}^n \ni H \to 0$,

$$\Pi_{\mathbb{S}^n_+}(\Lambda(A)+H)-\Pi_{\mathbb{S}^n_+}(\Lambda(A))-\Pi'_{\mathbb{S}^n_+}(\Lambda(A);H)=O(\|H\|^2) \tag{4-13}$$

and $O(\|H\|^2)$ is uniform for $A \in \mathcal{B}_r(\overline{A})$ with $\pi(\overline{A})=\pi(A)$. Next, we consider the case that $A=P^T\Lambda(A)P$. Re-write (4-9) as

$$\Lambda(A)+P^THP=P^TU\Lambda(Z)U^TP\,.$$

Let $\widetilde{H}:=P^THP$. Then, we have $\Pi_{\mathbb{S}^n_+}(A+H)=P\,\Pi_{\mathbb{S}^n_+}(\Lambda(A)+\widetilde{H})P^T$. Therefore, since $P\in\mathcal{O}^n$, we know from (4-13) and (2-43) that for any $\mathbb{S}^n \ni H\to 0$, (4-8) holds. □

With the help of the properties of the second subderivative $\mathrm{d}^2\delta_{\mathcal{K}}$, whose definition is given in [1, Definition 13.3], we can calculate its corresponding Moreau envelop in the following explicit form. This will be of great use in deriving the main result of this section.

Lemma 4.5. *Suppose $\mathcal{K}=\{0\}\times\mathbb{S}^n_+$. Given $(\Phi(x),\zeta)\in\operatorname{gph}N_{\mathcal{K}}$. Denote $A=G(x)+\Gamma$ and A possesses the eigenvalue decomposition in* (2-42)*. We have that for all $H\in\mathbb{S}^n$, the Moreau envelop of $\mathrm{d}^2\delta_{\mathcal{K}}(\Phi(x),\zeta)(\cdot)$, which is defined as*

$$e_\rho\big(\mathrm{d}^2\delta_{\mathcal{K}}(\Phi(x),\zeta)\big)(b,B):=\inf_{(w,W)\in\mathbb{Y}\times\mathbb{S}^n}\left\{\mathrm{d}^2\delta_{\mathcal{K}}(\Phi(x),\zeta)(w,W)+\frac{1}{2\rho}\|(w,W)-(b,B)\|^2\right\},$$

can be calculated in the following form, i.e., for all $(b,B)\in\mathbb{Y}\times\mathbb{S}^n$,

$$\begin{aligned}&e_{1/(2\rho)}\big(\mathrm{d}^2\delta_{\mathcal{K}}(\Phi(x),\zeta)\big)(b,B)=e_{1/(2\rho)}\big(\mathrm{d}^2\delta_{\mathbb{S}^n_+}(G(x),\Gamma)\big)(B)+\rho\|b\|^2\\ =&2\rho\sum_{i\in\beta,\,j\in\gamma}\widetilde{B}_{ij}^2+\rho\sum_{i,\,j\in\gamma}\widetilde{B}_{ij}^2+\rho\,\mathrm{dist}^2(\widetilde{B}_{\beta\beta},\mathcal{S}^{|\beta|}_+)\\ &+2\rho\sum_{i\in\alpha,\,j\in\gamma}\Big(\frac{\lambda_j(A)/\lambda_i(A)}{-\lambda_j(A)/\lambda_i(A)+\rho}\Big)^2\widetilde{B}_{ij}^2-2\sum_{i\in\alpha,\,j\in\gamma}\frac{\lambda_j(A)}{\lambda_i(A)}\Big(\frac{\rho\widetilde{B}_{ij}}{-\lambda_j(A)/\lambda_i(A)+\rho}\Big)^2+\rho\|b\|^2,\end{aligned}$$

where $\widetilde{B}=P^TBP$ with $P\in\mathcal{O}^n(A)$.

Proof. From [100, Theorem 6.2] and [11, Lemma 3.1], we know that

$$\mathrm{d}^2\delta_{\mathcal{K}}(\Phi(x),\zeta)(z)=-\Upsilon_{G(x)}\big(\Gamma,Z\big)+\delta_{\mathcal{C}_{\mathcal{K}}}(\Phi(x),\zeta)(z),$$

where $z=(w,Z)\in\mathbb{Y}\times\mathbb{S}^n$ and $\Upsilon_{G(x)}\big(\Gamma,Z\big)=2\langle\Gamma,ZG(x)^\dagger Z\rangle$ is the σ-term. Combining this with the definition of Moreau envelop, we have

$$\begin{aligned}e_{1/(2\rho)}\big(\mathrm{d}^2\delta_{\mathcal{K}}(\Phi(x),\zeta)\big)(b,B)&=\inf_z\{\rho\|z-(b,B)\|^2+\mathrm{d}^2\delta_{\mathcal{K}}(\Phi(x),\zeta)(z)\}\\ &=\inf_{z\in\mathcal{C}_{\mathcal{K}}(\Phi(x),\zeta)}\{-\Upsilon_{G(x)}\big(\Gamma,Z\big)+\rho\|z-(b,B)\|^2\}\\ &=\inf_{Z\in\mathcal{C}_{\mathbb{S}^n_+}(G(x),\Gamma)}\{-\Upsilon_{G(x)}\big(\Gamma,Z\big)+\rho\|Z-B\|^2\}+\rho\|b\|^2.\end{aligned} \tag{4-14}$$

From [11, (28)], we have

$$-\Upsilon_{G(x)}(\Gamma, Z) = -2\sum_{i\in\alpha,\, j\in\gamma}\frac{\lambda_j(A)}{\lambda_i(A)}(P^T ZP)_{ij}^2.$$

From [11, (19)], we also have

$$\mathcal{C}_{\mathbb{S}^n_+}(G(x),\Gamma) = \left\{Z \in \mathbb{S}^n : P_\beta^T ZP_\beta \in \mathcal{S}_+^{|\beta|}, P_\beta^T ZP_\gamma = 0, P_\gamma^T ZP_\gamma = 0\right\}.$$

Let $\widetilde{Z} = P^T ZP$, $\widetilde{B} = P^T BP$. Thus

$$\inf_{Z\in\mathcal{C}_{\mathbb{S}^n_+}(G(x),\Gamma)}\{-\Upsilon_{G(x)}(\Gamma, Z) + \rho\|Z - B\|^2\} = \inf_{Z\in\mathcal{C}_{\mathbb{S}^n_+}(G(x),\Gamma)}\{-2\sum_{i\in\alpha,\, j\in\gamma}\frac{\lambda_j(A)}{\lambda_i(A)}(P^T ZP)_{ij}^2 + \rho\|Z - B\|^2\}$$

$$= \inf_{Z\in\mathcal{C}_{\mathbb{S}^n_+}(G(x),\Gamma)}\{-2\sum_{i\in\alpha,\, j\in\gamma}\frac{\lambda_j(A)}{\lambda_i(A)}\widetilde{Z}_{ij}^2 + \rho\sum_{i,j}|\widetilde{Z}_{ij} - \widetilde{B}_{ij}|^2\}. \tag{4-15}$$

For all $i \in \alpha$, $j \in \alpha\cup\beta$ and $i \in \beta$, $j \in \alpha$, to obtain the minmum of (4-15), let $\widetilde{Z}_{ij} = \widetilde{B}_{ij}$. For all $i \in \alpha$, $j \in \gamma$ and $i \in \gamma$, $j \in \alpha$, we get the optimal solution $\widetilde{Z}_{ij} = \frac{\rho\widetilde{B}_{ij}}{-\lambda_j(A)/\lambda_i(A)+\rho}$. For all $i, j \in \beta$, $\widetilde{Z}_{\beta\beta} = \Pi_{\mathbb{S}_+^{|\beta|}}(\widetilde{B}_{\beta\beta})$. Otherwise, $\widetilde{Z}_{ij} = 0$. It follows from (4-15) that

$$\begin{aligned}&\inf_{Z\in\mathcal{C}_{\mathbb{S}^n_+}(G(x),\Gamma)}\{-\Upsilon_{G(x)}(\Gamma, Z) + \rho\|Z - B\|^2\}\\ =&2\rho\sum_{i\in\beta,\, j\in\gamma}\widetilde{B}_{ij}^2 + \rho\sum_{i,\, j\in\gamma}\widetilde{B}_{ij}^2 + \rho\,\mathrm{dist}^2(\widetilde{B}_{\beta\beta}, \mathbb{S}_+^{|\beta|})\\ &+ 2\rho\sum_{i\in\alpha,\, j\in\gamma}\left(\frac{\lambda_j(A)/\lambda_i(A)}{-\lambda_j(A)/\lambda_i(A)+\rho}\right)^2\widetilde{B}_{ij}^2 - 2\sum_{i\in\alpha,\, j\in\gamma}\frac{\lambda_j(A)}{\lambda_i(A)}\left(\frac{\rho\widetilde{B}_{ij}}{-\lambda_j(A)/\lambda_i(A)+\rho}\right)^2\end{aligned} \tag{4-16}$$

Combining (4-14) and (4-16) we have completed the proof. □

The following result illustrates the uniform second-order expansion for $\frac{1}{2}\mathrm{dist}^2(\cdot, \mathbb{S}^n_+)$, which will be used in the derivation of Proposition 4.9. It is worth to note that the second-order expansion for it is firstly studied in [99, Theorem 3.5]. By taking advantage of Lemmas 4.2-4.5, we can provide a direct proof here.

Proposition 4.6. *Given* $(G(\bar{x}), \overline{\Gamma}) \in \mathrm{gph}\, N_{\mathbb{S}^n_+}$. *Denote* $\overline{A} = G(\bar{x}) + \overline{\Gamma}$ *and* $\overline{A}$ *possesses the eigenvalue decomposition in* (2-42) *with* $\overline{P} \in \mathcal{O}^n(\overline{A})$. *Let* $0 < r < \min_{i<j}\{v_i(\overline{A}) - v_j(\overline{A})\}/3$. *For any* $A := G(\bar{x}) + \Gamma \in \mathcal{B}_r(\overline{A})$ *with* $\Gamma \in N_{\mathbb{S}^n_+}(G(\bar{x}))$ *and* $\pi(A) = \pi(\overline{A})$, *we have for all* $H \to 0$,

$$\frac{1}{2}\mathrm{dist}^2(A+H, \mathbb{S}^n_+) - \frac{1}{2}\mathrm{dist}^2(A, \mathbb{S}^n_+) = \langle \Pi_{\mathbb{S}^n_-}(A), H\rangle + \frac{1}{2}e_{1/2}\big(\mathrm{d}^2\delta_{\mathbb{S}^n_+}(G(\bar{x}),\Gamma)\big)(H) + O(\|H\|^3),$$

where $O(\|H\|^3)$ *is uniform for all* $A \in \mathcal{B}_r(\overline{A})$ *with* $\Gamma \in N_{\mathbb{S}^n_+}(G(\bar{x}))$ *and* $\pi(A) = \pi(\overline{A})$, $\mathrm{d}^2\delta_{\mathbb{S}^n_+}(G(\bar{x}),\Gamma)$ *is defined in [1, Definition 13.3].*

Proof. It is well-known that $\frac{1}{2}\text{dist}^2(A,\mathbb{S}^n_+) = e_1\delta_{\mathbb{S}^n_+}(A)$. From [1, Theorem 2.26], we have $\nabla e_1\delta_{\mathbb{S}^n_+}(A) = \Pi_{\mathbb{S}^n_-}(A)$. Denote $Q(A) = \Pi_{\mathbb{S}^n_-}(A)$. It follows that

$$\begin{aligned}
&\frac{1}{2}\text{dist}^2(A+H,\mathbb{S}^n_+) - \frac{1}{2}\text{dist}^2(A,\mathbb{S}^n_+) = \frac{1}{2}\langle Q(A+H)-Q(A), Q(A+H)+Q(A)\rangle\\
=&\frac{1}{2}\langle H, Q(A)+Q(A+H)-Q(A)\rangle + \frac{1}{2}\langle H, Q(A)\rangle\\
&-\frac{1}{2}\langle \Pi_{\mathbb{S}^n_+}(A+H)-\Pi_{\mathbb{S}^n_+}(A), Q(A+H)\rangle - \frac{1}{2}\langle \Pi_{\mathbb{S}^n_+}(A+H)-\Pi_{\mathbb{S}^n_+}(A), Q(A)\rangle\\
=&\langle H, Q(A)\rangle + \frac{1}{2}\langle H, Q'(A,H)+O(\|H\|^2)\rangle\\
&-\frac{1}{2}\langle \Pi_{\mathbb{S}^n_+}(A+H)-\Pi_{\mathbb{S}^n_+}(A), Q(A+H)\rangle - \frac{1}{2}\langle \Pi_{\mathbb{S}^n_+}(A+H)-\Pi_{\mathbb{S}^n_+}(A), Q(A)\rangle
\end{aligned} \tag{4-17}$$

$$\begin{aligned}
=&\langle H, Q(A)\rangle + \frac{1}{2}\langle H, Q'(A,H)\rangle + O(\|H\|^3)\\
&-\frac{1}{2}\langle \Pi_{\mathbb{S}^n_+}(A+H)-\Pi_{\mathbb{S}^n_+}(A), Q(A+H)\rangle - \frac{1}{2}\langle \Pi_{\mathbb{S}^n_+}(A+H)-\Pi_{\mathbb{S}^n_+}(A), Q(A)\rangle
\end{aligned} \tag{4-18}$$

where (4-17) comes from Proposition 4.4. We denote $\lambda_i := \lambda_i(A)$, $\Lambda := \Lambda(A)$ and $\Xi := \Lambda(A+H)$ for short. Let $P \in \mathcal{O}^n(A)$, $\widetilde{H} = P^T HP$ and $E \in \mathcal{S}^n$ be the matrix whose components are all 1. It follows from $H = \Pi'_{\mathbb{S}^n_+}(A,H) + \Pi'_{\mathbb{S}^n_-}(A,H)$, (2-43) and (2-45) that

$$\begin{aligned}
&\langle H, Q'(A,H)\rangle = \langle \Pi'_{\mathbb{S}^n_+}(A,H), \Pi'_{\mathbb{S}^n_-}(A,H)\rangle + \|\Pi'_{\mathbb{S}^n_-}(A,H)\|^2\\
&= \langle \Sigma_{\alpha\gamma}\circ \widetilde{H}_{\alpha\gamma}, (E-\Sigma)^T_{\alpha\gamma}\rangle + \langle \Pi_{\mathbb{S}^{|\beta|}_+}(\widetilde{H}_{\beta\beta}), \Pi_{\mathbb{S}^{|\beta|}_-}(\widetilde{H}_{\beta\beta})\rangle + \langle \Sigma^T_{\alpha\gamma}\circ \widetilde{H}^T_{\alpha\gamma}, (E-\Sigma)_{\alpha\gamma}\circ \widetilde{H}_{\alpha\gamma}\rangle + \|\Pi'_{\mathbb{S}^n_-}(A,H)\|^2\\
&= 2\sum_{i\in\alpha, j\in\gamma} \frac{-\lambda_i\lambda_j}{(\lambda_i-\lambda_j)^2}\widetilde{H}^2_{ij} + \|\Pi'_{\mathbb{S}^n_-}(A,H)\|^2\\
&= 2\sum_{i\in\alpha, j\in\gamma} \frac{\lambda_j^2-\lambda_i\lambda_j}{(\lambda_i-\lambda_j)^2}\widetilde{H}^2_{ij} + \sum_{i,j\in\gamma}\widetilde{H}^2_{ij} + 2\sum_{i\in\beta, j\in\gamma}\widetilde{H}^2_{ij} + \|\Pi_{\mathbb{S}^{|\beta|}_-}(\widetilde{H}_{\beta\beta})\|^2\\
&= 2\sum_{i\in\beta, j\in\gamma}\widetilde{H}^2_{ij} + \sum_{i,j\in\gamma}\widetilde{H}^2_{ij} + \text{dist}^2(\widetilde{H}_{\beta\beta}, \mathbb{S}^{|\beta|}_+) + 2\sum_{i\in\alpha, j\in\gamma} \frac{\lambda_j^2-\lambda_i\lambda_j}{(\lambda_i-\lambda_j)^2}\widetilde{H}^2_{ij}\\
&= \inf_z\{\|z-H\|^2 + \text{d}^2\delta_{\mathcal{K}}(G(\bar{x}),\Gamma)(z)\},
\end{aligned} \tag{4-19}$$

where the last equality comes from Lemma 4.5. It is easy to see from the projection onto SDP (2-42) that

$$\begin{aligned}
&-\langle \Pi_{\mathbb{S}^n_+}(A+H)-\Pi_{\mathbb{S}^n_+}(A), Q(A+H)\rangle = \langle \Pi_{\mathbb{S}^n_+}(A), Q(A+H)\rangle\\
&= \langle P\begin{bmatrix} \Lambda(A)_{\alpha\alpha} & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}P^T, U\Xi_- U^T\rangle\\
&= \text{tr}(\Lambda_{\alpha\alpha}(P^TU)_{\alpha\gamma}\Xi_{\gamma\gamma}(P^TU)^T_{\alpha\gamma}) + \text{tr}(\Lambda_{\alpha\alpha}(P^TU)_{\alpha\beta}\Xi_{\beta\beta}(P^TU)^T_{\alpha\beta}),
\end{aligned} \tag{4-20}$$

where $U \in \mathcal{O}^n(A+H)$ and $(\Xi_-)_{ij} = \min\{0, \Xi_{ij}\}$. It follows from Lemma 4.3 and the Lipschitz continuity of $\lambda(\cdot)$ that

$$\begin{aligned}(P^TU)_{\alpha\gamma}\Xi_{\gamma\gamma}(P^TU)_{\alpha\gamma}^T &= \left[\Theta_{\alpha\gamma} \circ (\widetilde{H}_{\alpha\gamma}Q_\gamma)\right]\Xi_{\gamma\gamma}\left[\Theta_{\alpha\gamma}^T \circ (Q_\gamma^T\widetilde{H}_{\alpha\gamma}^T)\right] + O(\|H\|^3)\\ &= \left[\Theta_{\alpha\gamma} \circ (\widetilde{H}_{\alpha\gamma}Q_\gamma)\right]\Lambda_{\gamma\gamma}\left[\Theta_{\alpha\gamma}^T \circ (Q_\gamma^T\widetilde{H}_{\alpha\gamma}^T)\right] + O(\|H\|^3).\end{aligned}$$

Thus we have

$$\operatorname{tr}(\Lambda_{\alpha\alpha}(P^TU)_{\alpha\gamma}\Xi_{\gamma\gamma}(P^TU)_{\alpha\gamma}^T) = \sum_{i\in\alpha, j\in\gamma} \frac{\lambda_i\lambda_j}{(\lambda_i - \lambda_j)^2}\|(\widetilde{H}_{\alpha\gamma}Q_\gamma)_{ij}\|^2 + O(\|H\|^3),$$

where $O(\|H\|^3)$ is uniform for all $A \in \mathcal{B}_r(\overline{A})$ with $\Gamma \in N_{\mathbb{S}_+^n}(G(\bar{x}))$ and $\pi(A) = \pi(\overline{A})$. Also, it is easy to see that $\operatorname{tr}(\Lambda_{\alpha\alpha}(P^TU)_{\alpha\beta}\Xi_{\beta\beta}(P^TU)_{\alpha\beta}^T) = O(\|H\|^3)$. Taking this into (4-20), we have

$$-\langle \Pi_{\mathbb{S}_+^n}(A+H) - \Pi_{\mathbb{S}_+^n}(A), Q(A+H)\rangle = \sum_{i\in\alpha, j\in\gamma} \frac{\lambda_i\lambda_j}{(\lambda_i - \lambda_j)^2}\|(\widetilde{H}_{\alpha\gamma}Q_\gamma)_{ij}\|^2 + O(\|H\|^3).$$

Similarly, we can compute $\langle \Pi_{\mathbb{S}_+^n}(A+H) - \Pi_{\mathbb{S}_+^n}(A), Q(A)\rangle$ in the exactly same way, i.e.,

$$\begin{aligned}\langle \Pi_{\mathbb{S}_+^n}(A+H) - \Pi_{\mathbb{S}_+^n}(A), Q(A)\rangle &= \operatorname{tr}(\Lambda_{\gamma\gamma}(P^TU)_{\gamma\alpha}\Xi_{\alpha\alpha}(P^TU)_{\gamma\alpha}^T) + \operatorname{tr}(\Lambda_{\gamma\gamma}(P^TU)_{\gamma\beta}\Xi_{\beta\beta}(P^TU)_{\gamma\beta}^T)\\ &= \sum_{i\in\alpha, j\in\gamma} \frac{\lambda_i\lambda_j}{(\lambda_i - \lambda_j)^2}\|(\widetilde{H}_{\gamma\alpha}Q_\alpha)_{ij}\|^2 + O(\|H\|^3).\end{aligned} \tag{4-21}$$

Combining (4-18), (4-19), (4-20) and (4-21) together, we have obtained the result. □

4.2 ALM convergence for NLSDP

In this section, we shall study the convergence of the inexact augmented Lagrangian method (ALM) for NLSDP (2-48) without requiring the uniqueness of Lagrangian multipliers. Recalling the augmented Lagrangian function of NLSDP defined in (1-12). We define the residual function by

$$R(x,\zeta) = \|\nabla_x L(x,\zeta)\| + \|\Phi(x) - \Pi_{\mathcal{K}}(\Phi(x)+\zeta)\|. \tag{4-22}$$

The augmented Lagrangian method is stated below.

Let $\bar{x}$ be a stationary point of NLSDP (2-48). Assume $\mathcal{M}(\bar{x})$ is nonempty with $(\bar{y}, \overline{\Gamma}) \in \mathcal{M}(\bar{x})$, the critical cone of problem (2-48) is adopted from [11, (37)]

$$\mathcal{C}(\bar{x}) = \left\{d \in \mathcal{X} : \nabla h(\bar{x})d = 0, G'(\bar{x})d \in \mathcal{C}_{\mathbb{S}_+^n}(G(\bar{x}), \overline{\Gamma})\right\},$$

where $\mathcal{C}_{\mathbb{S}_+^n}(G(\bar{x}), \overline{\Gamma})$ is the critical cone defined in (2-46). We also need the definition of second order sufficient condition for (2-48), which can be found from, e.g., [43, equation (2.11)].

Algorithm 1 (Augmented Lagrangian method)

Require: Let $(x^0, \zeta^0) \in \mathbb{X} \times \mathbb{Y} \times \mathbb{S}^n$, $\rho^0 > 0$, $\varsigma > 1$, $\{\epsilon_k\}_{k\geq 0}$ with $\epsilon_k > 0$ for all k and $\epsilon_k \to 0$ and set $k := 0$.

1: If (x^k, ζ^k) satisfies a suitable termination criterion: STOP.

2: Compute x^{k+1} such that

$$\|(\mathcal{L}_{\rho^k})'_x(\cdot, \zeta^k)\| \leq \epsilon_k. \tag{4-23}$$

3: Update the vector of multipliers to

$$\zeta^{k+1} := \rho^k \big[\Phi(x^{k+1}) + \frac{\zeta^k}{\rho^k} - \Pi_{\mathcal{K}}(\Phi(x^{k+1}) + \frac{\zeta^k}{\rho^k})\big]. \tag{4-24}$$

4: Update ρ^{k+1} by $\rho^{k+1} = \rho^k$ or $\rho^{k+1} = \varsigma\rho^k$ according to certain rules.

5: Set $k \leftarrow k+1$ and go to 1.

Definition 4.7. *Let $\bar{x}$ be a stationary point of NLSDP* (2-48). *Given $(\bar{y}, \overline{\Gamma}) \in \mathcal{M}(\bar{x})$. We say the second order sufficient condition (SOSC) holds at $(\bar{x}, \bar{y}, \overline{\Gamma})$ if*

$$\langle L''_{xx}(\bar{x}, \bar{y}, \overline{\Gamma})d, d\rangle - \Upsilon_{G(\bar{x})}\big(\overline{\Gamma}, G'(\bar{x})d\big) > 0, \quad \forall\, 0 \neq d \in \mathcal{C}(\bar{x}). \tag{4-25}$$

where $\Upsilon_{G(\bar{x})}\big(\overline{\Gamma}, G'(\bar{x})d\big) = 2\langle\overline{\Gamma}, (G'(\bar{x})d)G(\bar{x})^{\dagger}(G'(\bar{x})d)\rangle$ is the σ-term (cf. [11, Lemma 3.1]) and $G(\bar{x})^{\dagger}$ is the generalized inverse matrix of $G(\bar{x})$.

4.2.1 Properties of augmented Lagrangian function and the solution of subproblem (4-23)

To establish the ALM convergence of NLSDP, we firstly need the quadratic growth condition for augmented Lagrangian function as presented in [50] and [43]. The (uniform) positive definite condition is critical in establishing the (uniform) quadratic growth condition. In [43, Lemma 4.2], they studied the uniform version for linearly quadratic composite optimization problems under only SOSC. We try to extend their result from the polyhedral problem to NLSDP. It is worth noting that similar result for NLSDP was established in [50, Proposition 4], although they supposed nondegeneracy and strongly SOSC. It follows from [100, Theorem 8.3] that for any given $\zeta \in \mathcal{M}(\bar{x})$, function $x \mapsto \mathcal{L}_\rho(x, \zeta)$ is twice semidifferentiable with respect to x at $\bar{x}$. Then we obtain the following lemma.

Lemma 4.8. *Let $\bar{x} \in \mathbb{X}$ be a stationary point to the NLSDP* (2-48) *and $\bar{\zeta} \in \mathcal{M}(\bar{x})$* (3-92). *Then the following conditions are equivalent:*

(i) the SOSC holds at $(\bar{x}, \bar{\zeta})$ (see Definition 4.7);

(ii) there are positive constants ρ_3, ε', l' such that for all $\zeta \in \mathcal{M}(\bar{x}) \cap \mathcal{B}_{\varepsilon'}(\bar{\zeta})$ and all $\rho \geq \rho_3$,

$$\mathrm{d}^2_x\mathcal{L}_\rho(\bar{x}, \zeta)(w) \geq l'\|w\|^2 \quad \forall\, w \in \mathbb{R}^n\backslash\{0\}, \tag{4-26}$$

where $\mathrm{d}_x^2\mathcal{L}_\rho$ *is the second semiderivative on* x *defined in [1, Definition 13.3].*

Proof. "(b)⇒(a)": It follows from [100, Theorem 8.4], directly.

"(a)⇒(b)": First, it follows from [1, Proposition 13.5] that the second semiderivative is lower semicontinuous and positive homogenous of degree 2. Thus, by [100, Theorem 8.4 (ii)], we know that the SOSC is equivalent to the existence of $l' > 0$ such that

$$\mathrm{d}_x^2\mathcal{L}_{\rho_3}\big(\bar{x},\bar{\zeta}\big)(w) \geq 2l' \quad \forall\, w \in \mathcal{B}, \tag{4-27}$$

where $\mathcal{B}$ is the unite sphere of $\mathbb{X}$.

Next, we shall show that there exists an $\varepsilon' > 0$ such that for any $\zeta \in \mathcal{M}(\bar{x}) \cap \mathcal{B}_{\varepsilon'}(\bar{\lambda})$,

$$\mathrm{d}_x^2\mathcal{L}_{\rho_3}\big(\bar{x},\zeta\big)(w) \geq l' \quad \forall\, w \in \mathcal{B}.$$

From [100, Theorem 8.3 (i)], we know that for any $\zeta \in \mathcal{M}(\bar{x})$,

$$\begin{aligned}\mathrm{d}_x^2\mathcal{L}_{\rho_3}(\bar{x},\zeta)(w) =& \langle L''_{xx}(\bar{x},\zeta)w, w\rangle \\ &+ \inf_z\{\rho_3\|z - \Phi'(\bar{x})w\|^2 + \mathrm{d}^2\delta_{\mathcal{K}}(\Phi(\bar{x}),\zeta)(z)\},\end{aligned} \tag{4-28}$$

where $\mathcal{K} = \{0\} \times \mathbb{S}^n_+$. Choose $\varepsilon'' \in (0, l'/(2\|\Phi''(\bar{x})\|))$ if $\|\Phi''(\bar{x})\| \neq 0$ and $\varepsilon'' > 0$ otherwise. For each $w \in \mathcal{B}$, it is clear that for any $\zeta \in \mathcal{B}_{\varepsilon''}(\bar{\zeta})$,

$$\begin{aligned}\langle L''_{xx}(\bar{x},\zeta)w, w\rangle &= \langle L''_{xx}(\bar{x},\bar{\zeta})w, w\rangle + \langle \zeta - \bar{\zeta}, \Phi''(\bar{x})(w,w)\rangle \\ &\geq \langle L''_{xx}(\bar{x},\bar{\zeta})w, w\rangle - \|\Phi''(\bar{x})\| \cdot \|\zeta - \bar{\zeta}\| \\ &\geq \langle L''_{xx}(\bar{x},\bar{\zeta})w, w\rangle - \frac{l'}{2}.\end{aligned} \tag{4-29}$$

Let $A = \Gamma + G(\bar{x})$ and $\overline{A} = \overline{\Gamma} + G(\bar{x})$. Suppose $\overline{P} \in \mathcal{O}^n(\overline{A})$ and $P \in \mathcal{O}^n(A)$. Suppose $\overline{A}$ possesses the eigenvalue decomposition (2-42). Let $0 < r < \min_{i<j}\{v_i(\overline{A}) - v_j(\overline{A})\}/3$. For all $\Gamma \in \mathcal{B}_r(\overline{\Gamma})$, denote $\alpha \cup \beta_+ := \{i \mid \lambda(A) > 0\}$, $\beta_0 := \{i \mid \lambda(A) = 0\}$ and $\gamma \cup \beta_- := \{i \mid \lambda(A) < 0\}$. Since $\Pi_{\mathbb{S}^n_+}(A) = G(\bar{x})$, we have $\beta_+ = \emptyset$ and $\beta = \beta_0 \cup \beta_-$. Let $\widetilde{Z} = P^T Z P$, $\widetilde{B} = P^T(G'(\bar{x})w)P$. It follows from Lemma 4.5 that

$$\begin{aligned}&\inf_z\{\rho_3\|z - \Phi'(\bar{x})w\|^2 + \mathrm{d}^2\delta_{\mathcal{K}}(\Phi(\bar{x}),\zeta)(z)\} \\ =& \inf_{Z \in \mathcal{C}_{\mathbb{S}^n_+}(G(\bar{x}),\Gamma)}\{-\Upsilon_{G(\bar{x})}\big(\Gamma, Z\big) + \rho_3\|Z - G'(\bar{x})w\|^2\} + \rho_3\|h'(\bar{x})w\|^2 \\ =& 2\rho_3 \sum_{i\in\beta_0,\, j\in\gamma\cup\beta_-} \widetilde{B}_{ij}^2 + \rho_3 \sum_{i,\, j\in\gamma\cup\beta_-} \widetilde{B}_{ij}^2 + \rho_3 \mathrm{dist}^2(\widetilde{B}_{\beta_0\beta_0}, \mathbb{S}^{|\beta_0|}_+) \\ &+ 2\rho_3 \sum_{i\in\alpha,\, j\in\gamma\cup\beta_-} \Big(\frac{\lambda_j(A)/\lambda_i(A)}{-\lambda_j(A)/\lambda_i(A) + \rho_3}\Big)^2 \widetilde{B}_{ij}^2 \\ &- 2 \sum_{i\in\alpha,\, j\in\gamma\cup\beta_-} \frac{\lambda_j(A)}{\lambda_i(A)}\Big(\frac{\rho_3 \widetilde{B}_{ij}}{-\lambda_j(A)/\lambda_i(A) + \rho_3}\Big)^2 + \rho_3\|h'(\bar{x})w\|^2.\end{aligned} \tag{4-30}$$

From [62, Proposition 2.6], we know that for Γ sufficiently close to $\overline{\Gamma}$, we have $\mathrm{dist}(P, \mathcal{O}^n(\overline{A})) = O(\|\Gamma - \overline{\Gamma}\|)$, which implies for every A, there exists $Q = \mathrm{Diag}(Q_1, \ldots, Q_{\bar{d}})$, such that

$$P = \overline{P}Q + O(\|\Gamma - \overline{\Gamma}\|), \tag{4-31}$$

where $\bar{d}$ is the number of the different eigenvalues of $\overline{A}$ and $Q_l \in \mathcal{O}^{|\bar{\iota}_l|}$. We can take Γ such that $\|\Gamma - \overline{\Gamma}\| \leq r$. Let $\widehat{B} = \overline{P}^T(G'(\bar{x})w)\overline{P}$ and $\mathcal{C}_{\mathbb{S}_+^{|\beta|}} := \{W \in \mathbb{S}_+^{|\beta|} \mid \overline{P}_{\beta_0}^T W \overline{P}_{\beta_0} \in \mathbb{S}_+^{|\beta|},\ \overline{P}_{\beta_0}^T W \overline{P}_{\beta_-} = 0,\ \overline{P}_{\beta_-}^T W \overline{P}_{\beta_-} = 0\}$. Then we have

$$\begin{aligned}
&2\rho_3 \sum_{i\in\beta_0,\, j\in\gamma\cup\beta_-} \widetilde{B}_{ij}^2 + \rho_3 \sum_{i,\, j\in\gamma\cup\beta_-} \widetilde{B}_{ij}^2 + \rho_3 \mathrm{dist}^2(\widetilde{B}_{\beta_0\beta_0}, \mathbb{S}_+^{|\beta_0|}) \\
=&2\rho_3 \sum_{i\in\beta,\, j\in\gamma} \widetilde{B}_{ij}^2 + \rho_3 \sum_{i,\, j\in\gamma} \widetilde{B}_{ij}^2 + \rho_3 \mathrm{dist}^2(\widetilde{B}_{\beta_0\beta_0}, \mathbb{S}_+^{|\beta_0|}) + 2\rho_3 \sum_{i\in\beta_0,\, j\in\beta_-} \widetilde{B}_{ij}^2 + \rho_3 \sum_{i,\, j\in\beta_-} \widetilde{B}_{ij}^2 \\
\overset{(2\text{-}46)}{=}&2\rho_3 \sum_{i\in\beta,\, j\in\gamma} \widetilde{B}_{ij}^2 + \rho_3 \sum_{i,\, j\in\gamma} \widetilde{B}_{ij}^2 + \rho_3 \mathrm{dist}^2\big(\widetilde{B}_{\beta\beta},, \mathcal{C}_{\mathbb{S}_+^{|\beta|}}\big) = 2\rho_3 \|\widetilde{B}_{\beta\gamma}\|^2 + \rho_3 \|\widetilde{B}_{\gamma\gamma}\|^2 + \rho_3 \mathrm{dist}^2\big(\widetilde{B}_{\beta\beta}, \mathcal{C}_{\mathbb{S}_+^{|\beta|}}\big) \\
=&2\rho_3 \|\widehat{B}_{\beta\gamma}\|^2 + \rho_3 \|\widehat{B}_{\gamma\gamma}\|^2 + \rho_3 \mathrm{dist}^2\big(\widetilde{B}_{\beta\beta}, \mathcal{C}_{\mathbb{S}_+^{|\beta|}}\big) + O(\|\Gamma - \overline{\Gamma}\|),
\end{aligned}$$

where the last equality comes from the block diagonal structure of Q and (4-31). It follows that

$$\begin{aligned}
&2\rho_3 \sum_{i\in\beta_0,\, j\in\gamma\cup\beta_-} \widetilde{B}_{ij}^2 + \rho_3 \sum_{i,\, j\in\gamma\cup\beta_-} \widetilde{B}_{ij}^2 + \rho_3 \mathrm{dist}^2(\widetilde{B}_{\beta_0\beta_0}, \mathbb{S}_+^{|\beta_0|}) \\
=&2\rho_3 \sum_{i\in\beta,\, j\in\gamma} \widehat{B}_{ij}^2 + \rho_3 \sum_{i,\, j\in\gamma} \widehat{B}_{ij}^2 + \rho_3 \mathrm{dist}^2(\widetilde{B}_{\beta\beta}, \mathcal{C}_{\mathbb{S}_+^{|\beta|}}) + O(\|\Gamma - \overline{\Gamma}\|) \\
\geq&\rho_3 \mathrm{dist}^2\big(\widetilde{B}_{\beta\beta}, \mathbb{S}_+^{|\beta|}\big) + 2\rho_3 \sum_{i\in\beta,\, j\in\gamma} \widehat{B}_{ij}^2 + \rho_3 \sum_{i\in\gamma,\, j\in\gamma} \widehat{B}_{ij}^2 + O(\|\Gamma - \overline{\Gamma}\|) \\
=&\rho_3 \mathrm{dist}^2\big(\widehat{B}_{\beta\beta}, \mathbb{S}_+^{|\beta|}\big) + 2\rho_3 \sum_{i\in\beta,\, j\in\gamma} \widehat{B}_{ij}^2 + \rho_3 \sum_{i\in\gamma,\, j\in\gamma} \widehat{B}_{ij}^2 + O(\|\Gamma - \overline{\Gamma}\|),
\end{aligned} \tag{4-32}$$

where the inequality comes from $\mathbb{S}_+^{|\beta|} \supseteq \mathcal{C}_{\mathbb{S}_+^{|\beta|}}$. and the last equality comes from $\mathrm{dist}\big(\widetilde{B}_{\beta\beta}, \mathbb{S}_+^{|\beta|}\big) = \|\widetilde{B}_{\beta\beta} - \Pi_{\mathbb{S}_+^{|\beta|}}(\widetilde{B}_{\beta\beta})\| = \|\widehat{B}_{\beta\beta} - \Pi_{\mathbb{S}_+^{|\beta|}}(\widehat{B}_{\beta\beta})\| + O(\|\Gamma - \overline{\Gamma}\|) = \mathrm{dist}\big(\widehat{B}_{\beta\beta}, \mathbb{S}_+^{|\beta|}\big) + O(\|\Gamma - \overline{\Gamma}\|)$.

We also have

$$
\begin{aligned}
&2\rho_3 \sum_{i\in\alpha,\, j\in\gamma\cup\beta_-} \Big(\frac{\lambda_j(A)/\lambda_i(A)}{-\lambda_j(A)/\lambda_i(A)+\rho_3}\Big)^2 \widetilde{B}_{ij}^2 - 2\sum_{i\in\alpha,\, j\in\gamma\cup\beta_-} \frac{\lambda_j(A)}{\lambda_i(A)}\Big(\frac{\rho_3\widetilde{B}_{ij}}{-\lambda_j(A)/\lambda_i(A)+\rho_3}\Big)^2\\
\geq& 2\rho_3 \sum_{i\in\alpha,\, j\in\gamma} \Big(\frac{\lambda_j(\Gamma)}{-\lambda_j(\Gamma)+\rho_3\lambda_i(G(\bar{x}))}\Big)^2 \widetilde{B}_{ij}^2 - 2\sum_{i\in\alpha,\, j\in\gamma} \frac{\lambda_j(\Gamma)}{\lambda_i(G(\bar{x}))}\Big(\frac{\rho_3\widetilde{B}_{ij}\lambda_i(G(\bar{x}))}{-\lambda_j(\Gamma)+\rho_3\lambda_i(G(\bar{x}))}\Big)^2\\
=& 2\rho_3 \sum_{i\in\alpha,\, j\in\gamma} [(\frac{\lambda_j(\overline{\Gamma})}{-\lambda_j(\overline{\Gamma})+\rho_3\lambda_i(G(\bar{x}))})^2 + O(\|\Gamma-\overline{\Gamma}\|)](\widetilde{B}_{ij})^2\\
&-2\sum_{i\in\alpha,\, j\in\gamma} [\frac{\rho_3^2\lambda_i(G(\bar{x}))\lambda_j(\overline{\Gamma})}{(-\lambda_j(\overline{\Gamma})+\rho_3\lambda_i(G(\bar{x})))^2} + O(\|\Gamma-\overline{\Gamma}\|)](\widetilde{B}_{ij})^2\\
=& 2\rho_3 \sum_{i\in\alpha,\, j\in\gamma} [\frac{\lambda_j(\overline{\Gamma})^2-\rho_3\lambda_j(\overline{\Gamma})\lambda_i(G(\bar{x}))}{(-\lambda_j(\overline{\Gamma})+\rho_3\lambda_i(G(\bar{x})))^2} + O(\|\Gamma-\overline{\Gamma}\|)](\widetilde{B}_{ij})^2
\end{aligned}
$$

where the first equality follows from the Lipschitz continuity of $\lambda(\cdot)$ and the locally Lipschitz continuity of $g_1(z) = (\frac{z}{-z+\rho_3\lambda_i(G(\bar{x}))})^2$, $g_2(z) = \frac{\rho_3^2\lambda_i(G(\bar{x}))z}{(-z+\rho_3\lambda_i(G(\bar{x})))^2}$ over $z < 0$. Suppose $\overline{A}$ has $\bar{d}$ different eigenvalues with $v_1(\overline{A}) > \cdots > v_{d_0}(\overline{A}) > 0 = v_{d_0+1}(\overline{A}) > v_{d_0+2}(\overline{A}) > v_{\bar{d}}(\overline{A})$ and denote $\iota_s = \{i \mid \lambda_i(\overline{A}) = v_s(\overline{A})\}$. Then we have

$$
\begin{aligned}
&2\rho_3 \sum_{i\in\alpha,\, j\in\gamma} [\frac{\lambda_j(\overline{\Gamma})^2-\rho_3\lambda_j(\overline{\Gamma})\lambda_i(G(\bar{x}))}{(-\lambda_j(\overline{\Gamma})+\rho_3\lambda_i(G(\bar{x})))^2} + O(\|\Gamma-\overline{\Gamma}\|)](\widetilde{B}_{ij})^2\\
=& 2\rho_3 \sum_{s=1}^{d_0}\sum_{t=d_0+2}^{\bar{d}} [\frac{v_t(\overline{\Gamma})^2-\rho_3 v_t(\overline{\Gamma})v_s(G(\bar{x}))}{(-v_t(\overline{\Gamma})+\rho_3 v_s(G(\bar{x})))^2} + O(\|\Gamma-\overline{\Gamma}\|)] \sum_{i\in\iota_s, j\in\iota_t}(\widetilde{B}_{ij})^2\\
=& 2\rho_3 \sum_{s=1}^{d_0}\sum_{t=d_0+2}^{\bar{d}} [\frac{v_t(\overline{\Gamma})^2-\rho_3 v_t(\overline{\Gamma})v_s(G(\bar{x}))}{(-v_t(\overline{\Gamma})+\rho_3 v_s(G(\bar{x})))^2}] \sum_{i\in\iota_s, j\in\iota_t}(\widehat{B}_{ij})^2 + O(\|\Gamma-\overline{\Gamma}\|)\\
=& 2\rho_3 \sum_{i\in\alpha,\, j\in\gamma} [\frac{\lambda_j(\overline{\Gamma})^2-\rho_3\lambda_j(\overline{\Gamma})\lambda_i(G(\bar{x}))}{(-\lambda_j(\overline{\Gamma})+\rho_3\lambda_i(G(\bar{x})))^2}](\widehat{B}_{ij})^2 + O(\|\Gamma-\overline{\Gamma}\|),
\end{aligned}
\tag{4-33}
$$

where the last second inequality follows from the block diagonal structure of Q as $\sum\limits_{i\in\iota_s, j\in\iota_t}(\widetilde{B}_{ij})^2 = \|Q_s\widehat{B}_{\iota_s\iota_t}Q_t^T\|^2 = \|\widehat{B}_{\iota_s\iota_t}\|^2 = \sum\limits_{i\in\iota_s, j\in\iota_t}(\widehat{B}_{ij})^2$.

It follows from Lemma 4.5 again that if we combine (4-32) and (4-33) together, the right hand side of the above equation is exactly

$$
\inf_{Z\in\mathcal{C}_{\mathbb{S}^n_+}(G(\bar{x}),\overline{\Gamma})} \{-\Upsilon_{G(\bar{x})}\big(\overline{\Gamma}, Z\big) + \rho_3\|Z - G'(\bar{x})w\|^2\} + \Delta(\Gamma),
$$

where $\Delta(\Gamma) = O(\|\Gamma-\overline{\Gamma}\|)$ is a number, which satisfies that there exist $q > 0$ and $\omega > 0$ such that for all $\Gamma \in \mathcal{B}_\omega(\overline{\Gamma})$, $|\Delta(\Gamma)| \leq q\|\Gamma-\overline{\Gamma}\|$ as it originates from (4-31). Combining this fact with (4-30), we have for all Γ such that $\|\Gamma - \overline{\Gamma}\| \leq \min\{r, l'/q, \omega\}$,

$$
\begin{aligned}
&\inf_z\{\rho_3\|z - \Phi'(\bar{x})w\|^2 + \mathrm{d}^2\delta_{\mathcal{K}}(\Phi(\bar{x}), \zeta)(z)\} \\
&= \inf_{Z\in\mathcal{C}_{\mathbb{S}^n_+}(G(\bar{x}),\Gamma)}\{-\Upsilon_{G(\bar{x})}\big(\Gamma, Z\big) + \rho_3\|Z - \nabla G(\bar{x})w\|^2\} + \rho_3\|h'(\bar{x})w\|^2 \\
&\geq \inf_{Z\in\mathcal{C}_{\mathbb{S}^n_+}(G(\bar{x}),\overline{\Gamma})}\{-\Upsilon_{G(\bar{x})}\big(\overline{\Gamma}, Z\big) + \rho_3\|Z - G'(\bar{x})w\|^2\} + \rho_3\|h'(\bar{x})w\|^2 + O(\|\Gamma - \overline{\Gamma}\|) \\
&\geq \inf_z\{\rho_3\|z - \Phi'(\bar{x})w\|^2 + \mathrm{d}^2\delta_{\mathcal{K}}(\Phi(\bar{x}), \bar{\zeta})(z)\} - l'.
\end{aligned} \tag{4-34}
$$

Let $\varepsilon' = \min\{\varepsilon'', l'/q, r, \omega\}$. By (4-27), (4-28), (4-29) and (4-34), we have verified that for any $\zeta \in \mathcal{M}(\bar{x}) \cap \mathcal{B}_{\varepsilon'}(\bar{\zeta})$,

$$
\mathrm{d}_x^2\mathcal{L}_{\rho_3}\big(\bar{x}, \zeta\big)(w) \geq \mathrm{d}_x^2\mathcal{L}_{\rho_3}\big(\bar{x}, \bar{\zeta}\big)(w) - l' \geq l' \quad \forall\, w \in \mathcal{B}.
$$

Using the positive homogeneity of the second semiderivative yields (4-26) for $\rho = \rho_3$ and for all $\zeta \in \mathcal{M}(\bar{x}) \cap \mathcal{B}_{\varepsilon'}(\bar{\zeta})$. Recall that the function

$$
\rho \mapsto e_{1/2\rho}\big(\mathrm{d}^2\delta_{\mathcal{K}}(\Phi(\bar{x}), \zeta)\big)(\Phi'(\bar{x})w)
$$

is nondecreasing. Therefor the function $\rho \mapsto \mathrm{d}_x^2\mathcal{L}_\rho\big((\bar{x}, \bar{\zeta}), 0\big)(w)$ is also nondecreasing. This yields (4-26) for all $\zeta \in \mathcal{M}(\bar{x}) \cap \mathcal{B}_{\varepsilon'}(\bar{\zeta})$ and all $\rho \geq \rho_3$, and hence complete the proof. □

We also need the following result on the uniform expansion of augmented Lagrangian function, which can be obtained from Proposition 4.6.

Proposition 4.9. *Let $\bar{x} \in \mathcal{X}$ be a stationary point to the NLSDP* (2-48) *and $\bar{\zeta} \in \mathcal{M}(\bar{x})$. Let $\overline{A} = G(\bar{x}) + \overline{\Gamma}$ and $0 < r < \min_{i<j}\{v_i(\overline{A}) - v_j(\overline{A})\}/3$. For all $\zeta \in \mathcal{M}(\bar{x}) \cap \mathcal{B}_r(\bar{\lambda})$ with $\pi(\Gamma) = \pi(\overline{\Gamma})$ and any $\rho > 0$, we have*

$$
\frac{f(\bar{x}) - \mathcal{L}_\rho(x, \zeta)}{\|x - \bar{x}\|^2} = -\frac{1}{2}\mathrm{d}_x^2\mathcal{L}_\rho(\bar{x}, \zeta)(\frac{x - \bar{x}}{\|x - \bar{x}\|}) + O(\|x - \bar{x}\|), \tag{4-35}
$$

where $O(\|x - \bar{x}\|)$ is uniform for all $\zeta \in \mathcal{M}(\bar{x}) \cap \mathcal{B}_r(\bar{\lambda})$ with $\pi(\Gamma) = \pi(\overline{\Gamma})$.

Proof. From [43, Proposition 3.2], we know that for all $\zeta \in \mathcal{M}(\bar{x})$ and any $\rho > 0$, $f(\bar{x}) = \mathcal{L}_\rho(\bar{x}, \zeta)$. It follows that

$$
\begin{aligned}
\mathcal{L}_\rho(x, \zeta) - f(\bar{x}) =& f(x) - f(\bar{x}) + \langle y, h(x)\rangle + \frac{\rho}{2}\|h(x)\|^2 - \big(\langle y, h(\bar{x})\rangle + \frac{\rho}{2}\|h(\bar{x})\|^2\big) \\
&+ \rho[\frac{1}{2}\mathrm{dist}^2(G(x) + \rho^{-1}\Gamma, \mathbb{S}^n_+) - \frac{1}{2}\mathrm{dist}^2(G(\bar{x}) + \rho^{-1}\Gamma, \mathbb{S}^n_+)].
\end{aligned} \tag{4-36}
$$

It can be checked directly from Proposition 4.6 that

$$\begin{aligned}&\frac{1}{2}\mathrm{dist}^2(G(x)+\rho^{-1}\Gamma,\mathbb{S}^n_+)-\frac{1}{2}\mathrm{dist}^2(G(\bar{x})+\rho^{-1}\Gamma,\mathbb{S}^n_+)\\=&\langle\Pi_{\mathbb{S}^n_-}(G(\bar{x})+\rho^{-1}\Gamma),\frac{1}{2}G''(\bar{x})(x-\bar{x},x-\bar{x})\rangle+O(\|G(x)-G(\bar{x})\|^3)\\&+O(\|x-\bar{x}\|^3)+\frac{1}{2}e_{1/2}\big(\mathrm{d}^2\delta_{\mathbb{S}^n_+}(G(\bar{x}),\Gamma)\big)\big(G'(\bar{x})(x-\bar{x})+O(\|x-\bar{x}\|^2)\big).\end{aligned}\tag{4-37}$$

From the explicit form of $e_{1/2}\big(\mathrm{d}^2\delta_{\mathbb{S}^n_+}(G(\bar{x}),\Gamma)\big)(\cdot)$ in Lemma 4.5, we know that

$$\begin{aligned}&e_{1/2}\big(\mathrm{d}^2\delta_{\mathbb{S}^n_+}(G(\bar{x}),\Gamma)\big)\big(G'(\bar{x})(x-\bar{x})+O(\|x-\bar{x}\|^2)\big)\\&=e_{1/2}\big(\mathrm{d}^2\delta_{\mathbb{S}^n_+}(G(\bar{x}),\Gamma)\big)\big(G'(\bar{x})(x-\bar{x})\big)+O(\|x-\bar{x}\|^3),\end{aligned}\tag{4-38}$$

where $O(\|x-\bar{x}\|^2)$ and $O(\|x-\bar{x}\|^3)$ in (4-37) and (4-38) are uniform for all $\zeta\in\mathcal{M}(\bar{x})\cap\mathcal{B}_r(\bar{\lambda})$ with $\pi(\Gamma)=\pi(\overline{\Gamma})$. Combining the continuity of $f(x)-f(\bar{x})+\langle y,h(x)\rangle+\frac{\rho}{2}\|h(x)\|^2-\big(\langle y,h(\bar{x})\rangle+\frac{\rho}{2}\|h(\bar{x})\|^2\big)$ on x with (4-36), (4-37), (4-38) and [100, Theorem 8.3] with $\bar{\zeta}\in\mathcal{M}(\bar{x})$, we have attained (4-35). □

Combining Lemma 4.8 and Proposition 4.9 together, we are ready to state the uniform quadratic growth condition for augmented Lagrangian function under SOSC. The non-uniform form for NLSDP is firstly studied in [50, Proposition 1] and extended to general C^2-reducible constrained optimization by Mohammadi et al. [100, Theorem 8.4] under weaker condition.

Theorem 4.10. *Let $\bar{x}\in\mathcal{X}$ be a stationary point to the NLSDP* (2-48) *and $\bar{\zeta}\in\mathcal{M}(\bar{x})$* (3-92)*. Then we have the following two results:*

(i) If $\bar{\zeta}\in\mathrm{ri}\,\mathcal{M}(\bar{x})$ (the relative interior of $\mathcal{M}(\bar{x})$), the SOSC (4-25) *holds at $(\bar{x},\bar{\zeta})$ if and only if there are positive constants ρ_3, θ, ε, l such that for all $\rho\geq\rho_3$ and all $\zeta\in\mathcal{M}(\bar{x})\cap\mathcal{B}_\varepsilon(\bar{\zeta})$ the uniform quadratic growth condition*

$$\mathcal{L}_\rho(x,\zeta)\geq f(\bar{x})+l\|x-\bar{x}\|^2\quad\text{for all }x\in\mathcal{B}_\theta(\bar{x})\tag{4-39}$$

is satisfied.

(ii) If $\bar{\zeta}\in\mathrm{rbd}\,\mathcal{M}(\bar{x})$ (the relative boundary of $\mathcal{M}(\bar{x})$), the SOSC holds at $(\bar{x},\bar{\zeta})$ if and only if there are positive constants ρ_3, θ, ε, l such that (4-39) *holds uniformly for all $\rho\geq\rho_3$ and all $\zeta\in\mathcal{M}(\bar{x})\cap\mathcal{B}_\varepsilon(\bar{\zeta})$ with $\pi(\Gamma)=\pi(\overline{\Gamma})$.*

Proof. "$\Leftarrow$" can be obtained from [100, Theorem 8.4]. Then we are going to verify the opposite direction. It follows from Lemma 4.8 that there exist the positive constants l', ε' and ρ_3 for which (4-26) is satisfied for all $\zeta\in\mathcal{M}(\bar{x})\cap\mathcal{B}_{\varepsilon'}(\bar{\zeta})$ and all $\rho\geq\rho_3$. Using this and $f(\bar{x})=\mathcal{L}_{\rho_3}(\bar{x},\zeta)$ for any $\zeta\in\mathcal{M}(\bar{x})$, which can be obtained by the same proof of [43, Proposition 3.2 (a)], we deduce from (4-26) that for any given $\zeta\in\mathcal{M}(\bar{x})\cap\mathcal{B}_{\varepsilon'}(\bar{\zeta})$ there exists $\theta_\zeta>0$ for which we have

$$\mathcal{L}_{\rho_3}(x,\zeta)\geq f(\bar{x})+\frac{l'}{2}\|x-\bar{x}\|^2\quad\text{for all }x\in\mathcal{B}_{\theta_\zeta}(\bar{x}),\tag{4-40}$$

where the constant l' can be chosen the same for all the multipliers $\zeta \in \mathcal{M}(\bar{x}) \cap \mathcal{B}_{\varepsilon'}(\bar{\zeta})$. It can be obtained directly from the definition of the second subderivative. The radii of the balls centered at $\bar{x}$ in (4-40), however, depend on ζ. If $\bar{\zeta} \in \mathrm{ri}\, \mathcal{M}(\bar{x})$, we argue that a common radius can be chosen for all the multipliers $\zeta \in \mathcal{M}(\bar{x})$ that are sufficiently close to $\bar{\zeta}$. Its proof is exactly the same as the proof of [52, Theorem 4.5]. We omit it here for simplicity.

If $\bar{\zeta} \in \mathrm{rbd}\, \mathcal{M}(\bar{x})$, we prove that a common radius can be chosen for all $\zeta \in \mathcal{M}(\bar{x}) \cap \mathcal{B}_{\varepsilon}(\bar{\zeta})$ with $\pi(\Gamma) = \pi(\overline{\Gamma})$. Following the proof of [43, Proposition 3.2], we know that $f(\bar{x}) = \mathcal{L}_{\rho_3}(\bar{x}, \zeta)$. It is easy to see that for all $x \in \mathcal{B}_{\theta_{\bar{\zeta}}}(\bar{x})$, we have

$$\frac{f(\bar{x}) - \mathcal{L}_{\rho_3}(x, \bar{\zeta})}{\|x - \bar{x}\|^2} \leq -\frac{l'}{2}.$$

From Proposition 4.9 with $\zeta \in \mathcal{M}(\bar{x})$ and the positive homogenous of degree 2 of second semiderivative, we have

$$\frac{f(\bar{x}) - \mathcal{L}_{\rho_3}(x, \zeta)}{\|x - \bar{x}\|^2} = -\frac{1}{2}\mathrm{d}_x^2\mathcal{L}_{\rho_3}(\bar{x}, \zeta)(\frac{x - \bar{x}}{\|x - \bar{x}\|}) + O(\|x - \bar{x}\|),$$

where $O(\|x - \bar{x}\|)$ is uniform for all $\zeta \in \mathcal{M}(\bar{x}) \cap \mathcal{B}_r(\bar{\zeta})$ with $\pi(\Gamma) = \pi(\overline{\Gamma})$, with its uniform radius c and uniform constant q_1.

From the proof of the Lemma 4.8, we know for all $\zeta \in \mathcal{M}(\bar{x}) \cap \mathcal{B}_{\varepsilon'}(\bar{\zeta})$,

$$\mathrm{d}_x^2\mathcal{L}_{\rho_3}(\bar{x}, \zeta)(\frac{x - \bar{x}}{\|x - \bar{x}\|}) \geq \mathrm{d}_x^2\mathcal{L}_{\rho_3}(\bar{x}, \bar{\zeta})(\frac{x - \bar{x}}{\|x - \bar{x}\|}) + O(\|\Gamma - \overline{\Gamma}\|).$$

Suppose the uniform constant for $O(\|\Gamma - \overline{\Gamma}\|)$ is q_2. Let $\|x - \bar{x}\| \leq \min\{\theta_{\bar{\lambda}}, c, l'/(16q_1)\} := \theta$ and $\|\zeta - \bar{\zeta}\| \leq \min\{r, \varepsilon', l'/(16q_2)\} := \varepsilon$ with $\zeta \in \mathcal{M}(\bar{x}), \pi(\Gamma) = \pi(\overline{\Gamma})$. It is easy to see that $\|\Gamma - \overline{\Gamma}\| \leq \|\zeta - \bar{\zeta}\| \leq \varepsilon$. Thus we have

$$\begin{aligned}
\frac{f(\bar{x}) - \mathcal{L}_{\rho_3}(x, \zeta)}{\|x - \bar{x}\|^2} &= -\frac{1}{2}\mathrm{d}_x^2\mathcal{L}_{\rho_3}(\bar{x}, \zeta)(\frac{x - \bar{x}}{\|x - \bar{x}\|}) + O(\|x - \bar{x}\|) \\
&\leq -\frac{1}{2}\mathrm{d}_x^2\mathcal{L}_{\rho_3}(\bar{x}, \bar{\zeta})(\frac{x - \bar{x}}{\|x - \bar{x}\|}) + O(\|\Gamma - \overline{\Gamma}\|) + \frac{l'}{16} \\
&\leq -\frac{1}{2}\mathrm{d}_x^2\mathcal{L}_{\rho_3}(\bar{x}, \bar{\zeta})(\frac{x - \bar{x}}{\|x - \bar{x}\|}) + \frac{l'}{8} \\
&= \frac{f(\bar{x}) - \mathcal{L}_{\rho_3}(x, \bar{\zeta})}{\|x - \bar{x}\|^2} + O(\|x - \bar{x}\|) + \frac{l'}{8} \leq \frac{f(\bar{x}) - \mathcal{L}_{\rho_3}(x, \bar{\zeta})}{\|x - \bar{x}\|^2} + \frac{3l'}{16}.
\end{aligned}$$

Taking the supremum of $x \in \mathcal{B}_{\theta}(\bar{x})$ on both sides and we have for all $\zeta \in \mathcal{B}_{\varepsilon}(\bar{\zeta}) \cap \mathcal{M}(\bar{x})$ with $\pi(\Gamma) = \pi(\overline{\Gamma})$,

$$\sup_{x \in \mathcal{B}_{\theta}(\bar{x})} \frac{f(\bar{x}) - \mathcal{L}_{\rho_3}(x, \zeta)}{\|x - \bar{x}\|^2} \leq \sup_{x \in \mathcal{B}_{\theta}(\bar{x})} \frac{f(\bar{x}) - \mathcal{L}_{\rho_3}(x, \bar{\zeta})}{\|x - \bar{x}\|^2} + \frac{3l'}{16} \leq -\frac{5l'}{16}.$$

It follows that for all $x \in \mathcal{B}_{\theta}(\bar{x})$ and $\zeta \in \mathcal{B}_{\varepsilon}(\bar{\zeta}) \cap \mathcal{M}(\bar{x})$ with $\pi(\Gamma) = \pi(\overline{\Gamma})$,

$$\mathcal{L}_{\rho_3}(x, \zeta) \geq f(\bar{x}) + \frac{5l'}{16}\|x - \bar{x}\|^2.$$

From [1, Exercise 11.56], we know that for all $\rho \geq \rho_3$, $\mathcal{L}_\rho(x,\zeta) \geq \mathcal{L}_{\rho_3}(x,\zeta)$. Setting $l = \frac{5l'}{16}$ and we have proved (4-39). □

The locally Lipschitz continuous property of local minimizer of augmented Lagrangian function for NLSDP is established in [50, Theorem 1] under nondegeneracy and strong SOSC. It is worth noting that [43, Proposition 5.2] have verified the uniformly isolated calmness of the local minimizers of the augmented Lagrangian function for composite linear quadratic problems by only requiring SOSC. Here, we aim at extending a similar result to NLSDP, which may relax the conditions in [50, Theorem 1].

Proposition 4.11. *Let $\bar{x} \in \mathbb{X}$ be a stationary point to the NLSDP* (2-48) *and $\bar{\zeta} \in \mathcal{M}(\bar{x})$* (3-92). *Suppose $(\bar{x},\bar{\zeta})$ satisfies SOSC* (4-25). *Then there are positive constants $\tau, \rho_3, \widehat{r} > 0$ such that for every $\rho \geq \rho_3$ and $\zeta \in \mathcal{B}_{\widehat{r}/2\tau}(\bar{\zeta})$, the set of the local minimizers of function $\mathcal{L}_\rho(x,\zeta)$ over $x \in \mathcal{B}_{\widehat{r}}(\bar{x})$, defined by $\mathcal{S}_\rho(\zeta)$, satisfies the uniform isolated calmness property, i.e.,*

$$\mathcal{S}_\rho(\zeta) \subseteq \{\bar{x}\} + \tau\|\zeta - \bar{\zeta}\|\mathcal{B}$$

and satisfies $\emptyset \neq \mathcal{S}_\rho(\zeta) \subseteq \operatorname{int}\mathcal{B}_{\widehat{r}}(\bar{x})$, where $\mathcal{B}$ is the unite ball in $\mathbb{X}$.

Proof. It follows from the continuity of $\mathcal{L}_\rho$ and the compactness of $\mathcal{B}_{\widehat{r}}(\bar{x})$ that $\mathcal{S}_\rho(\zeta) \neq \emptyset$ (cf. [101, Theorem 4.16]). From [1], we know that the Lagrangian function $\mathcal{L}_\rho(x,\zeta)$ is concave on ζ. Combined with [100, Theorem 8.4], we have there are $\rho_3 > 0, \widehat{r} > 0$ and $l > 0$ such that for all $\rho \geq \rho_3$ and $x \in \mathcal{B}_{\widehat{r}}(\bar{x})$

$$\begin{aligned}\mathcal{L}_\rho(x,\zeta) &\geq \mathcal{L}_\rho(x,\bar{\zeta}) - \langle(\mathcal{L}_\rho)'_\zeta(x,\zeta), \bar{\zeta} - \zeta\rangle \\ &= \mathcal{L}_\rho(x,\bar{\zeta}) - \langle\Phi(x) - \Pi_{\mathcal{K}}(\Phi(x) + \rho^{-1}\zeta), \bar{\zeta} - \zeta\rangle \\ &\geq f(\bar{x}) + l\|x - \bar{x}\|^2 - \langle\Phi(x) - \Pi_{\mathcal{K}}(\Phi(x) + \rho^{-1}\zeta), \bar{\zeta} - \zeta\rangle,\end{aligned}$$

where the l, ρ_3 and $\widehat{r}$ is the same as in [100, Theorem 8.4], $\mathcal{K} = \{0\} \times \mathbb{S}^n_+$. Let $u \in \mathcal{S}_\rho(\zeta) := \arg\min\{\mathcal{L}_\rho(x,\zeta) \mid x \in \mathcal{B}_{\widehat{r}}(\bar{x})\}$ and we have

$$\mathcal{L}_\rho(u,\zeta) \leq \mathcal{L}_\rho(\bar{x},\zeta) = f(\bar{x}) + \frac{\rho}{2}\mathrm{dist}(\Phi(\bar{x}) + \rho^{-1}\zeta, \mathcal{K})^2 - \frac{1}{2}\rho^{-1}\|\zeta\|^2 \leq f(\bar{x}).$$

Define $\tau := \frac{\kappa_\Phi}{l} + \sqrt{\frac{\kappa_\Phi^2}{l^2} + \frac{1}{l\rho_3}}$ and fix $\zeta \in \mathcal{B}_{\widehat{r}/2\tau}(\bar{\zeta})$ and $\rho \geq \rho_3$, where κ_Φ is the Lipschitzian constant of Φ. Combining the above two inequalities together, we have

$$\|u - \bar{x}\|^2 \leq \frac{1}{l}\langle\Phi(u) - \Pi_{\mathcal{K}}(\Phi(u) + \rho^{-1}\zeta), \bar{\zeta} - \zeta\rangle. \tag{4-41}$$

Since $\bar{\zeta} \in N_{\mathcal{K}}(\Phi(\bar{x}))$, we have $\Phi(\bar{x}) = \Pi_{\mathcal{K}}(\Phi(\bar{x}) + \rho^{-1}\bar{\zeta})$. It follows that

$$\begin{aligned}&\|\Phi(u) - \Pi_{\mathcal{K}}(\Phi(u) + \rho^{-1}\zeta)\| \\ &= \|\Phi(u) - \Phi(\bar{x}) + \Pi_{\mathcal{K}}(\Phi(\bar{x}) + \rho^{-1}\bar{\zeta}) - \Pi_{\mathcal{K}}(\Phi(u) + \rho^{-1}\zeta)\| \\ &\leq 2\|\Phi(u) - \Phi(\bar{x})\| + \rho^{-1}\|\bar{\zeta} - \zeta\| \leq 2\kappa_\Phi\|u - \bar{x}\| + \rho^{-1}\|\bar{\zeta} - \zeta\|.\end{aligned}$$

Combining the above two inequalities together leads us to

$$\|u-\bar{x}\|^2 \le \frac{1}{l}\big(2\kappa_\Phi\|u-\bar{x}\| + \rho^{-1}\|\bar{\zeta}-\zeta\|\big)\|\zeta-\bar{\zeta}\|.$$

The latter inequality can be written as the following form

$$l\|u-\bar{x}\|^2 - 2\kappa_\Phi\|\zeta-\bar{\zeta}\|\cdot\|u-\bar{x}\| - \rho^{-1}\|\zeta-\bar{\zeta}\|^2 \le 0.$$

It follows that

$$\|u-\bar{x}\| \le \Big(\frac{\kappa_\Phi}{l} + \sqrt{\frac{\kappa_\Phi^2}{l^2}+\frac{1}{l\rho}}\Big)\|\zeta-\bar{\zeta}\| \le \tau\|\zeta-\bar{\zeta}\| \le \tau\widehat{r}/2\tau < \widehat{r}.$$

Then we have completed the proof. □

Remark 4.1. Recalling the definition of residual function $R(x,\zeta)$ (4-22). It is easy to know that for KKT point $(\bar{x},\bar{\zeta})$, there exist r_2 and κ_2 such that for all $(x,\zeta)\in\mathcal{B}_{r_2}(\bar{x},\bar{\zeta})$,

$$R(x,\zeta) \le \kappa_2\big(\|x-\bar{x}\| + \operatorname{dist}(\zeta,\mathcal{M}(\bar{x}))\big). \tag{4-42}$$

Its proof is in the same way as in [43, Proposition 5.4]. Moreover, by Theorem 4.10 and the proof of (4-41), we can prove in the same approach that if $\bar{\zeta}\in\operatorname{ri}\mathcal{M}(\bar{x})$, for all $\mu\in\mathcal{M}(\bar{x})\cap\mathcal{B}_\varepsilon(\bar{\zeta})$, there exist $\rho_3>0$, $\theta>0$ and $l>0$ such that for all $\rho\ge\rho_3$, $x\in\mathcal{B}_\theta(\bar{x})\cap\mathcal{S}_\rho(\zeta)$, $\zeta\in\mathbb{Y}\times\mathbb{S}^n$,

$$\|x-\bar{x}\|^2 \le \frac{1}{l}\langle\Phi(x)-\Pi_{\mathcal{K}}(\Phi(x)+\rho^{-1}\zeta),\mu-\zeta\rangle. \tag{4-43}$$

Similarly, if $\bar{\zeta}\in\operatorname{rbd}\mathcal{M}(\bar{x})$, we have that there also exists $\varepsilon>0$ such that for all $\mu=(y,\Gamma)\in\mathcal{M}(\bar{x})\cap\mathcal{B}_\varepsilon(\bar{\zeta})$ with $\pi(\Gamma)=\pi(\overline{\Gamma})$, (4-43) holds for all $\rho\ge\rho_3$, $x\in\mathcal{B}_\theta(\bar{x})\cap\mathcal{S}_\rho(\zeta)$, $\zeta\in\mathbb{Y}\times\mathbb{S}^n$.

Remark 4.2. It is worth noting that [55, Theorems 1 and 2] obtained augmented tilt stability under the strongly variational sufficient condition. The relationship between augmented tilt stability and uniform isolated calmness of $\mathcal{S}_\rho(\zeta)$ remains unknown to us though it seems the former is stronger. However, as mentioned in [20], the strongly variational sufficient condition used in [55] may fail when SOSC holds. This implies that the approach taken here is different from that of [55].

4.2.2 Local convergence analysis

Now we are going to establish the linear convergence of ALM for NLSDP. The following error bound estimate is an analogy to [43, Theorem 5.5], which mainly focuses on the polyhedron case. We illustrate it in two different cases.

Proposition 4.12. *Let $\bar{x} \in \mathbb{X}$ be a stationary point to the NLSDP* (2-48) *and $\bar{\zeta} \in \mathcal{M}(\bar{x})$* (3-92). *Suppose $(\bar{x}, \bar{\zeta})$ satisfies SOSC* (4-25) *and the semi-isolated calmness (see Definition 3.32) holds for S_{KKT} at $\big((0,0),(\bar{x},\bar{\zeta})\big)$. If $\bar{\zeta} \in \operatorname{ri}\mathcal{M}(\bar{x})$, then there exists positive constants r_3, κ_3 and ρ_3 such that for all $\rho \geq \rho_3$, $(x,\zeta) \in \mathcal{B}_{r_3}(\bar{x},\bar{\zeta})$ with $R(x,\zeta) > 0$, and all the optimal solutions u to problem*

$$\min \mathcal{L}_\rho(w,\zeta) \quad \text{subject to} \quad w \in \mathcal{B}_{\widehat{r}}(\bar{x}) \tag{4-44}$$

with $\widehat{r}$ obtained in Proposition 4.11, the error bound estimate

$$\|u - x\| + \|\rho\Pi_{-\mathcal{K}}(\Phi(u) + \rho^{-1}\zeta) - \zeta\| \leq \kappa_3 R(x,\zeta) \tag{4-45}$$

holds, where $\mathcal{K} = \{0\} \times \mathbb{S}^n_+$. If $\bar{\zeta} \in \operatorname{rbd}\mathcal{M}(\bar{x})$, (4-45) *also holds for all $\rho \geq \rho_3$, $(x,\zeta) \in \mathcal{B}_{r_3}(\bar{x},\bar{\zeta})$ with $\pi(\Gamma_\pi) = \pi(\overline{\Gamma})$ and $R(x,\zeta) > 0$, where $\Pi_{\mathcal{M}(\bar{x})}(\zeta) = (y_\pi, \Gamma_\pi)$.*

Proof. By Proposition 4.11, we know that for every $\zeta \in \mathcal{B}_{\widehat{r}/2\tau}(\bar{\zeta})$ and every $\rho \geq \rho_3$ any optimal solution u to (4-44) satisfies the first-order optimality condition

$$\nabla_x \mathcal{L}_\rho(u,\zeta) = 0. \tag{4-46}$$

If $\bar{\zeta} \in \operatorname{ri}\mathcal{M}(\bar{x})$, assume by contradiction that the error bound estimate (4-45) fails, which implies that there exists a sequence $\{(x^k,\zeta^k,\rho^k)\}_{k=1}^\infty \subset \mathbb{X}\times\mathbb{Y}\times\mathbb{S}^n\times[\rho_3,\infty)$ with $(x^k,\zeta^k) \to (\bar{x},\bar{\zeta})$ as $k \to \infty$ such that for each k,

$$\|u^k - x^k\| + \|d^k - \zeta^k\| > kR_k, \tag{4-47}$$

where $d^k := \rho^k\Pi_{-\mathcal{K}}(\Phi(u^k) + \frac{\zeta^k}{\rho^k})$ and u^k is an optimal solution to (4-44) for $(\zeta,\rho) = (\zeta^k,\rho^k)$ and

$$R_k := R(x^k,\zeta^k) = \|L'_x(x^k,\zeta^k)\| + \|\Phi(x^k) - \Pi_{\mathcal{K}}(\Phi(x^k) + \zeta^k)\|. \tag{4-48}$$

If $\bar{\zeta} \in \operatorname{rbd}\mathcal{M}(\bar{x})$, we also assume by contradiction and the only difference lies in the supposed sequence ζ^k satisfies $\pi(\Gamma^k_\pi) = \pi(\overline{\Gamma})$ in addition. For each k, denote $c_k := \|u^k - x^k\| + \|d^k - \zeta^k\|$. Thus, we know from (4-47) that $R_k = o(c_k)$. It then follows from (4-48) that for $k \to \infty$

$$L'_x(x^k,\zeta^k) + o(c_k) = 0 \quad \text{and} \quad \Phi(x^k) + o(c_k) = \Pi_{\mathcal{K}}(\Phi(x^k) + \zeta^k). \tag{4-49}$$

By passing to a subsequence if necessary, we are able to find $0 \neq (\xi,\eta)$ with $\xi \in \mathbb{X}$ and $\eta := (\eta_0,\eta_1) \in \mathbb{Y}\times\mathbb{S}^n$ such that

$$\frac{u^k - x^k}{c_k} \to \xi \quad \text{and} \quad \frac{d^k - \zeta^k}{c_k} \to \eta \quad \text{as} \quad k \to \infty. \tag{4-50}$$

Since $(x^k, \zeta^k) \to (\bar{x}, \bar{\zeta})$, we obtain from (4-42) that $R_k \to 0$ as $k \to \infty$. Moreover, it follows from the definition of semi-isolated calmness and [97, Theorem 3.1] that there exists $\kappa > 0$ such that for (x, ζ) sufficiently close to $(\bar{x}, \bar{\zeta})$,

$$\|x - \bar{x}\| + \operatorname{dist}(\zeta, \mathcal{M}(\bar{x})) \leq \kappa R(x, \zeta).$$

For each k, let $\mu^k = (\mu_1^k, \mu_2^k) \in \mathbb{Y} \times \mathbb{S}^n$ be the metric projection of ζ^k over the nonempty closed convex set $\mathcal{M}(\bar{x})$, i.e., $\mu^k := \Pi_{\mathcal{M}(\bar{x})}(\zeta^k)$. Thus, without loss of generality, we may assume that $x^k - \bar{x} = O(R_k)$ and $\zeta^k - \mu^k = O(R_k)$ for $k \to \infty$, which in turn results in

$$x^k - \bar{x} = o(c_k) \quad \text{and} \quad \zeta^k - \mu^k = o(c_k) \quad \text{as} \quad k \to \infty. \tag{4-51}$$

The latter along with $\zeta^k \to \bar{\zeta}$ tells us that $\mu^k \to \bar{\zeta}$ as $k \to \infty$, since $\mathcal{M}(\bar{x})$ is closed and convex. Therefore, we get $\mu^k \in \mathcal{M}(\bar{x}) \cap \mathcal{B}_\varepsilon(\bar{\zeta})$ for all k sufficiently large.

If $\bar{\zeta} \in \operatorname{ri} \mathcal{M}(\bar{x})$, then we know from (4-43) in Remark 4.1 that there exists $l > 0$, for all k sufficiently large,

$$\|u^k - \bar{x}\|^2 \leq \frac{1}{\rho^k l} \langle d^k - \zeta^k, \mu^k - \zeta^k \rangle \leq \frac{1}{\rho^k l} \|d^k - \zeta^k\| \cdot \|\mu^k - \zeta^k\|. \tag{4-52}$$

If $\bar{\zeta} \in \operatorname{rbd} \mathcal{M}(\bar{x})$, again from Remark 4.1, we know that (4-52) also holds for all k sufficiently large when $\pi(\mu_2^k) = \pi(\overline{\Gamma})$. For simplicity, we only show the $\bar{\zeta} \in \operatorname{ri} \mathcal{M}(\bar{x})$ case here while the other case can be obtained similarly. By using Proposition 4.11, we obtain that

$$\|u^k - \bar{x}\| \leq \tau \|\zeta^k - \bar{\zeta}\| \to 0 \quad \text{as} \quad k \to \infty. \tag{4-53}$$

Multipling (4-52) by $1/c_k^2$, we know from (4-50) and (4-51) that

$$\frac{\|u^k - \bar{x}\|^2}{c_k^2} \leq \frac{1}{\rho^k l} \frac{\|d^k - \zeta^k\|}{c_k} \cdot \frac{\|\mu^k - \zeta^k\|}{c_k} \to 0 \quad \text{as} \quad k \to \infty, \tag{4-54}$$

which implies for $k \to \infty$, $u^k - \bar{x} = o(c_k)$. This, together with (4-51), yields that

$$\xi = \lim_{k \to \infty} \frac{u^k - x^k}{c_k} = \lim_{k \to \infty} \frac{u^k - \bar{x}}{c_k} - \lim_{k \to \infty} \frac{x^k - \bar{x}}{c_k} = 0 - 0 = 0. \tag{4-55}$$

We also have

$$d^k = \nabla e_{1/\rho_k} \delta_{\mathcal{K}}(\Phi(u^k) + \frac{\zeta^k}{\rho^k}) = \rho^k \big(\Phi(u^k) + \frac{\zeta^k}{\rho^k} - \Pi_{\mathcal{K}}(\Phi(u^k) + \frac{\zeta^k}{\rho^k})\big). \tag{4-56}$$

The desired result then follows by contradiction, if we show that $\eta = 0$. To this end, let us consider the following two cases.

Case 1: either the sequence $\{\rho^k\}_{k=1}^\infty$ or $\{\rho^k / c_k\}_{k=1}^\infty$ is bounded, it follows from (4-54) or (4-53) that $\frac{\rho^k}{c_k}\|u^k - \bar{x}\| \to 0$ as $k \to \infty$. It follows from (4-56) and the twice continuous differentiability of Φ that there exists $\kappa_\Phi > 0$ such that

$$
\begin{aligned}
\frac{\|d^k - \zeta^k\|}{c_k} &= \frac{\rho^k}{c_k}\|\Phi(u^k) - \Pi_{\mathcal{K}}(\Phi(u^k) + (\rho^k)^{-1}\zeta^k)\| \\
&= \frac{\rho^k}{c_k}\|\Phi(u^k) - \Phi(\bar{x}) + \Pi_{\mathcal{K}}(\Phi(\bar{x}) + (\rho^k)^{-1}\mu^k) - \Pi_{\mathcal{K}}(\Phi(u^k) + (\rho^k)^{-1}\zeta^k)\| \\
&\leq \frac{\rho^k}{c_k}(2\|\Phi(u^k) - \Phi(\bar{x})\| + (\rho^k)^{-1}\|\mu^k - \zeta^k\|) \\
&\leq 2\kappa_\Phi \frac{\rho^k}{c_k}\|u^k - \bar{x}\| + \frac{\|\mu^k - \zeta^k\|}{c_k} \to 0 \quad \text{as} \quad k \to \infty.
\end{aligned}
$$

Therefore, by (4-51) we obtain that $\eta = 0$.

Case 2: both sequences $\{\rho^k\}_{k=1}^\infty$ and $\{\rho^k/c_k\}_{k=1}^\infty$ are unbounded. By passing to a subsequence if necessary, we may assume that

$$
\rho^k \to \infty \quad \text{and} \quad \frac{\rho^k}{c_k} \to \infty \quad \text{as} \quad k \to \infty. \tag{4-57}
$$

Since u^k is an optimal solution to (4-44) associated with (ζ^k, ρ^k), we deduce from (4-46) that for each k,

$$
(\mathcal{L}_{\rho^k})'_x(u^k, \zeta^k) = L'_x(u^k, d^k) = 0. \tag{4-58}
$$

From (4-54), we have

$$
\frac{\rho^k}{c_k}\|u^k - \bar{x}\|^2 \leq \frac{1}{l}\frac{\|d^k - \zeta^k\|}{c_k} \cdot \|\mu^k - \zeta^k\| \to 0 \quad \text{as} \quad k \to \infty. \tag{4-59}
$$

For each k, define $z^k := \Phi(u^k) + (\rho^k)^{-1}(\zeta^k - d^k)$. Thus, it follows from (4-56) and [1, Theorem 2.26] that $d^k \in \partial\delta_{\mathcal{K}}(z^k)$. Using (4-50), (4-53) and (4-57) allows us to arrive at

$$
z^k = \Phi(u^k) - \frac{d^k - \zeta^k}{c_k} \cdot \frac{c_k}{\rho^k} \to \Phi(\bar{x}) =: \bar{z} \quad \text{as} \quad k \to \infty.
$$

For each k, by noting that $\mu^k \in \mathcal{M}(\bar{x})$, we have $\mu^k \in N_{\mathcal{K}}(\bar{z})$. By using the firmly non-expansiveness of projection mapping (cf. [102, (1.6)]) and [59, Theorem 31.5], we have for each k,

$$
\|z^k - \bar{z}\|^2 \leq \langle z^k - \bar{z}, z^k + d^k - \bar{z} - \mu^k\rangle = \|z^k - \bar{z}\|^2 + \langle z^k - \bar{z}, d^k - \mu^k\rangle,
$$

which yields $\langle z^k - \bar{z}, d^k - \mu^k\rangle \geq 0$. Therefore, we have for each k,

$$
\frac{\rho^k}{c_k^2}\langle z^k - \bar{z}, \mu^k - d^k\rangle = \langle\frac{\rho^k}{c_k}(\Phi(u^k) - \Phi(\bar{x})) - \frac{d^k - \zeta^k}{c_k}, \frac{\mu^k - d^k}{c_k}\rangle \leq 0,
$$

which implies

$$
\langle -\frac{d^k - \zeta^k}{c_k}, \frac{\mu^k - d^k}{c_k}\rangle \leq -\langle\frac{\rho^k}{c_k}(\Phi(u^k) - \Phi(\bar{x})), \frac{\mu^k - d^k}{c_k}\rangle. \tag{4-60}
$$

Meanwhile, since Φ is twice continuously differentiable and (4-53) holds, we have $\Phi(u^k) - \Phi(\bar{x}) = \Phi'(\bar{x})(u^k - \bar{x}) + O(\|u^k - \bar{x}\|^2)$ and $\Phi(u^k) - \Phi(\bar{x}) = \Phi'(u^k)(u^k - \bar{x}) + O(\|u^k - \bar{x}\|^2)$ as $k \to \infty$. It is worth to note the second $O(\|u^k - \bar{x}\|^2)$ is uniform for k sufficiently large. Thus, since f is twice continuous differentiable, we know that there exists $\kappa_f > 0$ such that for k sufficiently large,

$$
\begin{aligned}
&-\langle \frac{\rho^k}{c_k}(\Phi(u^k) - \Phi(\bar{x})), \frac{\mu^k - d^k}{c_k}\rangle = -\frac{\rho^k}{c_k^2}\big(\langle \Phi(u^k) - \Phi(\bar{x}), \mu^k\rangle - \langle \Phi(u^k) - \Phi(\bar{x}), d^k\rangle\big)\\
&= -\frac{\rho^k}{c_k^2}\big(\langle \Phi'(\bar{x})(u^k - \bar{x}) + O(\|u^k - \bar{x}\|^2), \mu^k\rangle - \langle \Phi'(u^k)(u^k - \bar{x}) + O(\|u^k - \bar{x}\|^2), d^k\rangle\big)\\
&= -\frac{\rho^k}{c_k^2}\big(\langle (u^k - \bar{x}), \Phi'(\bar{x})^*\mu^k\rangle - \langle (u^k - \bar{x}), \Phi'(u^k)^* d^k\rangle + \langle O(\|u^k - \bar{x}\|^2), \mu^k\rangle + \langle O(\|u^k - \bar{x}\|^2), d^k\rangle\big)\\
&= -\frac{\rho^k}{c_k^2}\big(\langle (u^k - \bar{x}), \Phi'(\bar{x})^*\mu^k - \Phi'(u^k)^* d^k\rangle + \langle O(\|u^k - \bar{x}\|^2), \mu^k\rangle + \langle O(\|u^k - \bar{x}\|^2), d^k\rangle\big)\\
&= -\frac{\rho^k}{c_k^2}\big(\langle u^k - \bar{x}, f'(u^k) - f'(\bar{x})\rangle + \langle O(\|u^k - \bar{x}\|^2), \mu^k\rangle + \langle O(\|u^k - \bar{x}\|^2), d^k\rangle\big)\\
&\le \frac{\rho^k}{c_k^2}\kappa_f\|u^k - \bar{x}\|^2 + \langle O(\frac{\rho^k}{c_k^2}\|u^k - \bar{x}\|^2), -\mu^k\rangle + \langle O(\frac{\rho^k}{c_k}\|u^k - \bar{x}\|^2), -\frac{d^k}{c_k}\rangle\\
&= \frac{\rho^k}{c_k^2}\kappa_f\|u^k - \bar{x}\|^2 + \langle O(\frac{\rho^k}{c_k^2}\|u^k - \bar{x}\|^2), -\mu^k\rangle + \langle O(\frac{\rho^k}{c_k}\|u^k - \bar{x}\|^2), \frac{-d^k + \zeta^k}{c_k}\rangle + \langle O(\frac{\rho^k}{c_k^2}\|u^k - \bar{x}\|^2), -\zeta^k\rangle,
\end{aligned}
\tag{4-61}
$$

where the forth equation follows from $L'_x(u^k, d^k) = 0$ (4-58) and $L'_x(\bar{x}, \mu^k) = 0$ for all k. It then follows from (4-54) and (4-59) that

$$
\frac{\rho^k}{c_k^2}\|u^k - \bar{x}\|^2 \to 0 \quad \text{and} \quad \frac{\rho^k}{c_k}\|u^k - \bar{x}\|^2 \to 0 \quad \text{as} \quad k \to \infty.
$$

Combining this with $\zeta^k \to \bar{\zeta}$, $\mu^k \to \bar{\zeta}$, (4-50) and

$$
\lim_{k\to\infty}\frac{\mu^k - d^k}{c_k} = \lim_{k\to\infty}\frac{\mu^k - \zeta^k + \zeta^k - d^k}{c_k} = \lim_{k\to\infty}\frac{\mu^k - \zeta^k}{c_k} + \lim_{k\to\infty}\frac{\zeta^k - d^k}{c_k} = -\eta,
$$

we obtain from (4-60) and (4-61) that $\|\eta\|^2 \le 0$ by taking $k \to \infty$. Thus, we know that $\eta = 0$, which completes the proof. □

Before we put forward the main result of this chapter, we need to propose the following assumption.

Assumption 1. Given $\bar{\zeta} \in \operatorname{rbd}\mathcal{M}(\bar{x})$. There exits $r_5 > 0$, $\chi > 0$ such that for all $(x, \zeta) \in \mathcal{B}_{r_5}(\bar{x}, \bar{\zeta})$ with $\zeta = (y, \Gamma) \notin \mathcal{M}(\bar{x})$, there exist $\widehat{\zeta} \in \mathcal{M}(\bar{x})$ with $\pi(\widehat{\Gamma}) = \pi(\overline{\Gamma})$ such that

$$
\|\Pi_{\mathcal{M}(\bar{x})}(\zeta) - \widehat{\zeta}\| \le \chi R(x, \zeta).
$$

It is worth to note that Assumption 1 is not like any constraint qualification that we are familiar with. However, it is easy to see that it holds at least in the following circumstances.

(a) If for all $\Gamma_1, \Gamma_2 \in \mathcal{M}(\bar{x})$, $\pi(\Gamma_1) = \pi(\Gamma_2)$ holds, then Assumption 1 trivially holds under semi-isolated calmness for S_{KKT} at $\big((0,0),(\bar{x},\bar{\zeta})\big)$ as we only need to pick $\widehat{\zeta} = \Pi_{\mathcal{M}(\bar{x})}(\zeta)$. This circumstance also includes the case where $\mathcal{M}(\bar{x})$ is a singleton. It is worth to note that for non-polyheral cases, the strict Ronbinson condition mentioned in [51, Definition 2.3] is not equivalent to the uniqueness of $\mathcal{M}(\bar{x})$ (cf. [91, Example 3]).

(b) Suppose $A = G(\bar{x}) + \overline{\Gamma}$ possesses the eigenvalue decomposition (2-42). If there is no multiple root in $\beta \cup \gamma$ part eigenvalues of $\overline{\Gamma}$, then for all λ sufficiently close to $\bar{\lambda}$, every Γ also do not possess multiple root in $\beta \cup \gamma$ part for eigenvalues. We can also pick $\widehat{\zeta} = \Pi_{\mathcal{M}(\bar{x})}(\zeta)$ under the semi-isolated calmness for S_{KKT} at $\big((0,0),(\bar{x},\bar{\zeta})\big)$ to show the validity of Assumption 1. A sufficient condition for semi-isolated calmness will be given in Section 3.4. It is worth to note that in this case we only require the existence of such $\overline{\Gamma}$ and an oracle to find a starting point lies in the neighborhood mentioned in Assumption 1 when (4-25) holds for all $\zeta \in \mathcal{M}(\bar{x})$), which is called Robinson's SOSC [103] (see also [91, (25)]). Moreover, in this case, neither the set $\mathcal{M}(\bar{x})$ has to be a polyhedron nor the strict complementarity condition, i.e., $\bar{\zeta} \in \operatorname{ri} \mathcal{M}(\bar{x})$ is satisfied.

Moreover, at the end of this chapter, we give 2 examples (Example 4.1, 4.2) to show the validity of this assumption. By adding Assumption 1 to the conditions in Proposition 4.12, we can get a further result of Proposition 4.12 for the case when $\bar{\zeta} \in \operatorname{rbd} \mathcal{M}(\bar{x})$.

Corollary 4.13. *Besides the conditions in Proposition 4.12, if the Assumption 1 also holds, we have when* $\bar{\zeta} \in \operatorname{rbd} \mathcal{M}(\bar{x})$, (4-45) *also holds for all* $\rho \geq \rho_3$ *and* $(x,\zeta) \in \mathcal{B}_{r_3}(\bar{x},\bar{\zeta})$ *with* $R(x,\zeta) > 0$ *and* $R(\bar{x},\zeta) > 0$.

Proof. The proof is exactly the same as the proof of Proposition 4.12. Suppose the contradiction sequence (x^k, ζ^k) also satisfies $\zeta^k \notin \mathcal{M}(\bar{x})$. The only difference lies in (4-51) and (4-52), as we can find $\widehat{\mu}^k \in \mathcal{M}(\bar{x})$ with $\pi(\widehat{\Gamma}^k) = \pi(\overline{\Gamma})$ satisfies $\|\widehat{\mu}^k - \mu^k\| = O(R_k)$. Then μ^k in (4-51) and (4-52) can be alternated by $\widehat{\mu}^k$. Then we have completed the proof. □

Remark 4.3. Given $(x^k, \zeta^k) \in \mathcal{B}_{r_3}(\bar{x},\bar{\zeta})$ and $\rho^k \geq \rho_3$ with r_3 and ρ_3 taken from Proposition 4.12. Suppose x^{k+1} is the optimal solution to (4-44). Increasing κ_3 if necessary. Similarly as in [43, Remark 5.6], we know that for $\widetilde{x}^{k+1}$ sufficiently close to x^{k+1}, we also have

$$\|(\mathcal{L}_{\rho^k})'_x(\widetilde{x}^{k+1}, \zeta^k)\| \leq \epsilon_k$$

and

$$\|\widetilde{x}^{k+1} - x^k\| + \|\nabla e_{1/\rho}\delta_{\mathcal{K}}(\Phi(\widetilde{x}^{k+1}) + (\rho^k)^{-1}\zeta^k) - \zeta^k\| \leq \kappa_3 R(x^k, \zeta^k).$$

Next we are going to propose the main result of this chapter, which is inspired by [43, Theorem 5.6]. It illustrates the local linear convergence of ALM for NLSDP without requiring the uniqueness of multipliers by applying Proposition 4.12 and (4-43).

Theorem 4.14. *Let $\bar{x} \in \mathbb{X}$ be a stationary point to the NLSDP* (2-48) *and $\bar{\zeta} \in \mathcal{M}(\bar{x})$* (3-92). *Suppose $(\bar{x}, \bar{\zeta})$ satisfies SOSC* (4-25) *and the semi-isolated calmness (see Definition 3.32) holds for S_{KKT} at $\big((0,0), (\bar{x}, \bar{\zeta})\big)$.*

(i) If $\bar{\zeta} \in \mathrm{ri}\,\mathcal{M}(\bar{x})$, then there exist positive constants $\overline{r}$, $\overline{\sigma}$, $\overline{\varrho}$ such that for any starting point $(x^0, \zeta^0) \in \mathcal{B}_{\overline{r}}(\bar{x}, \bar{\zeta})$ the primal-dual sequence $\{(x^k, \zeta^k)\}_{k\geq 0}$ generated by Algorithm 2 with $\rho^k \geq \overline{\varrho}$ and $\epsilon_k = o(R(x^k, \zeta^k))$ for all k satisfies the estimate

$$\|x^{k+1} - x^k\| + \|\zeta^{k+1} - \zeta^k\| \leq \overline{\sigma} R(x^k, \zeta^k). \tag{4-62}$$

(ii) If $\bar{\zeta} \in \mathrm{rbd}\,\mathcal{M}(\bar{x})$ and Assumption 1 holds, then there exist positive constants $\overline{r}$, $\overline{\sigma}$, $\overline{\varrho}$ such that for any starting point $(x^0, \zeta^0) \in \mathcal{B}_{\overline{r}}(\bar{x}, \bar{\zeta})$ the primal-dual sequence $\{(x^k, \zeta^k)\}_{k\geq 0}$ generated by Algorithm 2 with $\rho^k \geq \overline{\varrho}$ and $\epsilon_k = o(R(x^k, \zeta^k))$ and $\zeta^k \notin \mathcal{M}(\bar{x})$ for all k satisfies the estimate (4-62).

Moreover, for each case, the sequence is convergent to $(\bar{x}, \widetilde{\zeta})$ for some $\widetilde{\zeta} \in \mathcal{M}(\bar{x})$ and its rate of convergence is linear, i.e., for k sufficiently large,

$$\|(x^{k+1}, \zeta^{k+1}) - (\bar{x}, \widetilde{\zeta})\| \leq \tau^k \|(x^k, \zeta^k) - (\bar{x}, \widetilde{\zeta})\|, \tag{4-63}$$

where $\tau^k = 2\sqrt{2}\overline{\sigma}\kappa_1\kappa_2^2(R_k^{-1}\epsilon_k + (\rho^k)^{-1}\overline{\sigma})$, $R_k := R(x^k, \zeta^k)$.

Proof. Consider $R_k := R(x^k, \zeta^k)$. If $R_k = 0$ for some k, then the pair (x^k, ζ^k) satisfies the KKT system and the algorithm should stop. Thus we assume $R_k > 0$ for all $k \in \mathbb{N}$. Pick κ_1 and r_1 from Definition 3.32 with $\mathcal{V} = \mathcal{B}_{r_1}(\bar{x}, \bar{\zeta})$ and $\kappa = \kappa_1$, κ_2 and r_2 from (4-42), ρ_3, κ_3 and r_3 from Proposition 4.12 or Corollary 4.13, τ and $\widehat{r}$ from Proposition 4.11. By the definition of ϵ_k, we can find $r_4 > 0$ such that

$$\epsilon(x, \zeta) \leq \frac{1}{4\kappa_1\kappa_2} R(x, \zeta) \quad \text{whenever} \quad (x, \zeta) \in \mathcal{B}_{r_4}(\bar{x}, \bar{\zeta}). \tag{4-64}$$

Define $\overline{\sigma} = \kappa_3$ and

$$\overline{r} = \frac{r'}{1 + 2\sqrt{2}\,\overline{\sigma}\kappa_2} \quad \text{with} \quad r' = \min\{\widehat{r}, \frac{\widehat{r}}{2\tau}, \frac{r_1}{\sqrt{2}\,\overline{\sigma}\kappa_2 + 1}, r_2, r_4, r_3\}. \tag{4-65}$$

Pick $q \in (0,1)$ and $\overline{\rho} = \max\{\rho_3, \frac{2\sqrt{2}\kappa_1\kappa_2^2\overline{\sigma}^2}{q}, 4\kappa_1\kappa_2\overline{\sigma}\}$. By induction, we want to show that if $\bar{\zeta} \in \mathrm{ri}\,\mathcal{M}(\bar{x})$, for any starting point $(x^0, \zeta^0) \in \mathcal{B}_{\overline{r}}(\bar{x}, \bar{\zeta})$ the sequence generated by the algorithm with $\rho^k \geq \overline{\rho}$ and $\epsilon_k = o(R(x^k, \zeta^k))$, we have for all $k = 0, 1, \ldots$ the following relationships

$$(x^k, \zeta^k) \in \mathcal{B}_{r'}(\bar{x}, \bar{\zeta}), \tag{4-66}$$

$$\|L'_x(x^{k+1}, \zeta^{k+1})\| \leq \epsilon_k, \tag{4-67}$$

$$\|x^{k+1} - x^k\| + \|\zeta^{k+1} - \zeta^k\| \leq \overline{\sigma} R_k \tag{4-68}$$

hold. By induction, we firstly assume $k = 0$. Since $(x^0, \zeta^0) \in \mathcal{B}_{\overline{r}}(\bar{x}, \bar{\zeta})$ and $\overline{r} \le r'$, we have (4-66) holds. By Proposition 4.11 and $\zeta^0 \in \mathcal{B}_{\widehat{r}/2\tau}(\bar{\zeta})$, we can find $\widehat{x}^1 \in \operatorname{int}\mathcal{B}_{\widehat{r}}(\bar{x})$ satisfying $(\mathcal{L}_{\rho^0})'_x(\widehat{x}^1, \zeta^0) = 0$. From Remark 4.3, we know that we can find x^1 sufficiently close to $\widehat{x}^1$ such that x^1 satisfies the two relationships in Remark 4.3. Define further $\zeta^1 = \rho^0[\Phi(x^1) + \zeta^0/\rho^0 - \Pi_{\mathcal{K}}(\Phi(x^1) + \zeta^0/\rho^0)]$ and we have

$$\|L'_x(x^1, \zeta^1)\| = \|(\mathcal{L}_{\rho^0})'_x(x^1, \zeta^0)\| \le \epsilon_k.$$

It follows from Proposition 4.12 and Remark 4.3 that

$$\|x^1 - x^0\| + \|\zeta^1 - \zeta^0\| \le \overline{\sigma} R_0.$$

Thus, (x^1, ζ^1) is well defined and satisfies (4-67) and (4-68) for $k = 0$. Then we assume (x^k, ζ^k), $k = 0, 1, \ldots, s+1$ are well defined and (4-66)-(4-68) hold for $k = 0, 1, \ldots, s$. We now verify the existence of (x^{s+2}, ζ^{s+2}) and (4-66)-(4-68) satisfies for $k = s+1$. We first show that $(x^{s+1}, \zeta^{s+1}) \in \mathcal{B}_{r'}(\bar{x}, \bar{\zeta})$. Fix an integer k with $0 \le k \le s$. Since $(x^k, \zeta^k) \in \mathcal{B}_{r'}(\bar{x}, \bar{\zeta})$, it follows from (4-42) that

$$\begin{aligned} R_k &\le \kappa_2(\|x^k - \bar{x}\| + \operatorname{dist}(\zeta^k, \mathcal{M}(\bar{x}))) \\ &\le \kappa_2(\|x^k - \bar{x}\| + \|\zeta^k - \bar{\zeta}\|) \le \sqrt{2}\kappa_2\|(x^k, \zeta^k) - (\bar{x}, \bar{\zeta})\| \le \sqrt{2}\kappa r'. \end{aligned} \tag{4-69}$$

Thus we have

$$\begin{aligned} \|(x^{k+1}, \zeta^{k+1}) - (\bar{x}, \bar{\zeta})\| &\le \|x^{k+1} - x^k\| + \|\zeta^{k+1} - \zeta^k\| + \|(x^k, \zeta^k) - (\bar{x}, \bar{\zeta})\| \\ &\le \overline{\sigma} R_k + r' \le (\sqrt{2}\,\overline{\sigma}\kappa_2 + 1)r' \le r_1, \end{aligned}$$

which implies that $(x^{k+1}, \zeta^{k+1}) \in \mathcal{B}_{r_1}(\bar{x}, \bar{\zeta})$. From Definition 3.32 combined with [97, Theorem 3.1], we obtain

$$\begin{aligned} &\|x^{k+1} - \bar{x}\| + \operatorname{dist}(\zeta^{k+1}, \mathcal{M}(\bar{x})) \\ &\le \kappa_1 R_{k+1} = \kappa_1(\|L'_x(x^{k+1}, \zeta^{k+1})\| + \|\Phi(x^{k+1}) - \Pi_{\mathcal{K}}(\Phi(x^{k+1}) + \zeta^{k+1})\|) \\ &\le \kappa_1\epsilon_k + \kappa_1\|\Phi(x^{k+1}) - \Pi_{\mathcal{K}}(\Phi(x^{k+1}) + \zeta^{k+1})\|. \end{aligned}$$

Let $p^{k+1} = \Pi_{\mathcal{K}}(\Phi(x^{k+1}) + (\rho^k)^{-1}\zeta^k)$. From Algorithm 2 we have $\Phi(x^{k+1}) - p^{k+1} = (\rho^k)^{-1}(\zeta^{k+1} - \zeta^k)$. It follows from $\zeta^{k+1} = \rho^k\Pi_{-\mathcal{K}}(\Phi(x^{k+1}) + (\rho^k)^{-1}\zeta^k) = \nabla e_{1/\rho_k}\delta_{\mathcal{K}}(\Phi(x^{k+1}) + (\rho^k)^{-1}\zeta^k)$ and $\nabla e_r g(x) = (rI + (\partial g)^{-1})^{-1}(x)$ (see [1, Theorem 2.26]) that $\zeta^{k+1} \in N_{\mathcal{K}}(p^{k+1})$. By the nonexpansivness of $y \mapsto y - \Pi_{\mathcal{K}}(y + \zeta^{k+1})$, we obtained

$$\begin{aligned} &\|\Phi(x^{k+1}) - \Pi_{\mathcal{K}}(\Phi(x^{k+1}) + \zeta^{k+1})\| \\ &= \|\Phi(x^{k+1}) - \Pi_{\mathcal{K}}(\Phi(x^{k+1}) + \zeta^{k+1})\| - \|p^{k+1} - \Pi_{\mathcal{K}}(p^{k+1} + \zeta^{k+1})\| \\ &\le \|\Phi(x^{k+1}) - \Pi_{\mathcal{K}}(\Phi(x^{k+1}) + \zeta^{k+1}) - (p^{k+1} - \Pi_{\mathcal{K}}(p^{k+1} + \zeta^{k+1}))\| \\ &\le \|\Phi(x^{k+1}) - p^{k+1}\|. \end{aligned}$$

Thus we have

$$\|x^{k+1}-\bar{x}\|+\mathrm{dist}(\zeta^k,\mathcal{M}(\bar{x}))\le\kappa_1\epsilon_k+(\rho^k)^{-1}\kappa_1\|\zeta^{k+1}-\zeta^k\|,$$

which can be further calculated as

$$\begin{aligned}\|x^{k+1}-\bar{x}\|+\mathrm{dist}(\zeta^{k+1},\mathcal{M}(\bar{x}))&\overset{(4\text{-}68)}{\le}\kappa_1\epsilon_k+\frac{\overline{\sigma}\kappa_1}{\rho_k}R_k\overset{(4\text{-}64)}{\le}\frac{1}{4\kappa_2}R_k+\frac{1}{4\kappa_2}R_k\\&\overset{(4\text{-}42)}{\le}\frac{1}{2}\big(\|x^k-\bar{x}\|+\mathrm{dist}(\zeta^k,\mathcal{M}(\bar{x}))\big).\end{aligned}\tag{4-70}$$

It follows that

$$\|x^{k+1}-\bar{x}\|+\mathrm{dist}(\zeta^{k+1},\mathcal{M}(\bar{x}))\le\frac{1}{2^{k+1}}\big(\|x^0-\bar{x}\|+\mathrm{dist}(\zeta^0,\mathcal{M}(\bar{x}))\big).\tag{4-71}$$

Then we have

$$\begin{aligned}&\|(x^{s+1},\zeta^{s+1})-(x^0,\zeta^0)\|\le\sum_{k=0}^{s}\|(x^{k+1},\zeta^{k+1})-(x^k,\zeta^k)\|\overset{(4\text{-}68)}{\le}\overline{\sigma}\sum_{k=0}^{s}R_k\\&\overset{(4\text{-}69)}{\le}\overline{\sigma}\kappa_2\sum_{k=0}^{s}\big(\|x^k-\bar{x}\|+\mathrm{dist}(\zeta^k,\mathcal{M}(\bar{x}))\big)\\&\overset{(4\text{-}71)}{\le}\overline{\sigma}\kappa_2\sum_{k=0}^{s}\frac{1}{2^k}\big(\|x^0-\bar{x}\|+\mathrm{dist}(\zeta^0,\mathcal{M}(\bar{x}))\big)\\&\le 2\overline{\sigma}\kappa_2\big(\|x^0-\bar{x}\|+\mathrm{dist}(\zeta^0,\mathcal{M}(\bar{x}))\big)\le 2\overline{\sigma}\kappa_2\big(\|x^0-\bar{x}\|+\|\zeta^0-\bar{\zeta}\|\big).\end{aligned}$$

Thus we arrive at the estimate

$$\begin{aligned}&\|(x^{s+1},\zeta^{s+1})-(\bar{x},\bar{\zeta})\|\le\|(x^{s+1},\zeta^{s+1})-(x^0,\zeta^0)\|+\|(x^0,\zeta^0)-(\bar{x},\bar{\zeta})\|\\&\le 2\overline{\sigma}\kappa_2(\|x^0-\bar{x}\|+\|\zeta^0-\bar{\zeta}\|)+\|(x^0,\zeta^0)-(\bar{x},\bar{\zeta})\|\\&\le(2\sqrt{2}\overline{\sigma}\kappa_2+1)\|(x^0,\zeta^0)-(\bar{x},\bar{\zeta})\|\le(2\sqrt{2}\overline{\sigma}\kappa_2+1)\overline{r}=r',\end{aligned}$$

where the last inequality comes from $(x^0,\zeta^0)\in\mathcal{B}_{\overline{r}}(\bar{x},\bar{\zeta})$. Then we have verified $(x^{s+1},\zeta^{s+1})\in\mathcal{B}_{r'}(\bar{x},\bar{\zeta})$. By (4-65), we get $\zeta^{s+1}\in\mathcal{B}_{\overline{r}/2\tau}(\bar{\zeta})$, and hence Proposition 4.11 ensures the optimal solution $\widehat{x}^{s+2}$ such that $\widehat{x}^{s+2}\in\mathrm{int}\,\mathcal{B}_{\overline{r}}(\bar{x})$. Thus we have $(\mathcal{L}_{\rho^{s+1}})'_x(\widehat{x}^{s+2},\zeta^{s+1})=0$. Still from Remark 4.3, we can find x^{s+2} sufficiently close to $\widehat{x}^{s+2}$ such that x^{s+2} satisfies the two relationships in Remark 4.3 and we observe that

$$\|L'_x(x^{s+2},\zeta^{s+2})\|=\|(\mathcal{L}_{\rho^{s+1}})'_x(x^{s+2},\zeta^{s+1})\|\le\epsilon_{s+1}.$$

By Proposition 4.12 and Remark 4.3, we have

$$\|x^{s+2}-x^{s+1}\|+\|\zeta^{s+2}-\zeta^{s+1}\|\le\overline{\sigma}R_{s+1}.$$

Then we have finished verifying (4-66)-(4-68) for $k=s+1$. If $\bar{\zeta}\in\mathrm{rbd}\,\mathcal{M}(\bar{x})$ and Assumption 1 holds, we want to prove for any starting point $(x^0,\zeta^0)\in\mathcal{B}_{\overline{r}}(\bar{x},\bar{\zeta})$ the sequence generated by the algorithm with $\rho^k \geq \overline{\rho}$, $\epsilon_k = o(R(x^k, \zeta^k))$ and $R(\bar{x}, \zeta^k) > 0$ for all k, relationships (4-66)-(4-68) also hold for all $k = 0, 1, \ldots$. The proof is exactly the same as the $\bar{\zeta} \in \mathrm{ri}\,\mathcal{M}(\bar{x})$ case and the only difference lies in the Proposition 4.12 used above should be alternated by Corollary 4.13.

Then we prove the convergence of the sequence. Use the same argument as in the proofs of (4-71), we have

$$\|(x^{k+l}, \zeta^{k+l}) - (x^k, \zeta^k)\| \leq 2\overline{\sigma}\kappa_2\big(\|x^k - \bar{x}\| + \mathrm{dist}(\zeta^k, \mathcal{M}(\bar{x}))\big) \quad \text{for all} \quad k, l \in \mathbb{N}. \tag{4-72}$$

It follows from (4-71) that the righthand side of (4-72) goes to 0 as $k \to \infty$, which implies $\{(x^k, \zeta^k)\}$ is Cauchy. Assume $\{(x^k, \zeta^k)\}$ convergences to $(\bar{x}, \widetilde{\zeta})$, where $\widetilde{\zeta} \in \mathcal{M}(\bar{x})$. Let $l \to \infty$ in (4-72). Then we have

$$\|(x^k, \zeta^k) - (\bar{x}, \widetilde{\zeta})\| \leq 2\overline{\sigma}\kappa_2\big(\|x^k - \bar{x}\| + \mathrm{dist}(\zeta^k, \mathcal{M}(\bar{x}))\big),$$

which together with (4-70) verifies

$$\begin{aligned}
&\|(x^{k+1}, \zeta^{k+1}) - (\bar{x}, \widetilde{\zeta})\| \leq 2\overline{\sigma}\kappa_2\big(\|x^{k+1} - \bar{x}\| + \mathrm{dist}(\zeta^{k+1}, \mathcal{M}(\bar{x}))\big) \\
&\leq 2\overline{\sigma}\kappa_2\kappa_1(R_k^{-1}\epsilon_k + (\rho^k)^{-1}\overline{\sigma})R_k \\
&\leq 2\overline{\sigma}\kappa_2^2\kappa_1(R_k^{-1}\epsilon_k + (\rho^k)^{-1}\overline{\sigma})\big(\|x^k - \bar{x}\| + \mathrm{dist}(\zeta^k, \mathcal{M}(\bar{x}))\big) \\
&\leq 2\sqrt{2}\overline{\sigma}\kappa_1\kappa_2^2(R_k^{-1}\epsilon_k + (\rho^k)^{-1}\overline{\sigma})\|(x^k, \zeta^k) - (\bar{x}, \widehat{\zeta})\|.
\end{aligned}$$

Combining this with $\rho^k \geq \overline{\rho}$ and $\epsilon_k = o(R_k)$ result in

$$\limsup_{k\to\infty} \frac{\|(x^{k+1}, \zeta^{k+1}) - (\bar{x}, \widetilde{\zeta})\|}{\|(x^k, \zeta^k) - (\bar{x}, \widetilde{\zeta})\|} \leq \limsup_{k\to\infty} 2\sqrt{2}\overline{\sigma}\kappa_1\kappa_2^2(R_k^{-1}\epsilon_k + (\rho^k)^{-1}\overline{\sigma}) \leq q.$$

It follows from $q \in (0, 1)$ that the convergence rate is linear. Then we have completed the whole proof. □

Remark 4.4. If $\rho^k \to \infty$, we obtain the asymptotic Q-superlinear convergence rate of KKT pair from (4-63) as $\tau^k \to 0$. Regarding to the update of ρ^k, in [43, 97], they apply a practical rule in Algorithm 2, step 4 to update ρ^k, i.e., defining the auxiliary function by

$$V(x, \zeta, \rho) := \|(\mathcal{L}_\rho)'_x(x, \zeta)\| + \|\Phi(x) - \Pi_{\mathcal{K}}(\Phi(x) + \rho^{-1}\zeta)\|.$$

Suppose $c \in (0, 1)$. If $k = 0$ or $V(x^{k+1}, \zeta^k, \rho^k) \leq cV(x^k, \zeta^{k-1}, \rho^{k-1})$ holds, set $\rho^{k+1} := \rho^k$; otherwise, set $\rho^{k+1} := \varsigma\rho^k$. By taking advantage of Theorem 4.14 and similar manners as in [43], we are able to obtain the boundedness of $\{\rho^k\}_{k=1}^\infty$.

Remark 4.5. In Theorem 4.14, if $\bar{\zeta} \in \mathrm{rbd}\,\mathcal{M}(\bar{x})$ and Assumption 1 holds, we have shown that the primal-dual sequence $\{(x^k, \zeta^k)\}_{k\geq 0}$ generated by Algorithm 2 converge to $(\bar{x}, \widetilde{\zeta})$ for some $\widetilde{\zeta} \in \mathcal{M}(\bar{x})$ if $\zeta^k \notin \mathcal{M}(\bar{x})$ for sufficiently large k. It remains unknown from our approach whether the primal sequence $\{x^k\}_{k\geq 0}$ converges to $\bar{x}$ linearly if the dual sequence $\{\zeta^k\}_{k\geq 0}$ terminates finitely. This is a future work we are working on.

For NLSDP, the characterization of semi-isolated calmness is given in Section 3.4. It follows from the discussion under Theorem 3.34 that for NLSDP, if $\mathcal{M}(\bar{x})$ is a singleton, the result obtained in Theorem 4.14 reduces to [51, Theorem 4.2] for NLSDP.

To end this chapter, we propose two examples to illustrate that the conditions required in Theorem 4.14 can be satisfied indeed.

Example 4.1.

$$\begin{array}{ll} \min & \frac{1}{2}x^3 \\ \text{s.t} & -x^2 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \in \mathbb{S}^3_+, \quad \Leftarrow \Gamma \end{array}$$

It is easy to see that the optimal solution is $\bar{x} = 0$, its corresponding multiplier set is $\mathcal{M}(\bar{x}) = \{\Gamma \mid \Gamma \in \mathbb{S}^3_-\}$. Pick $\overline{\Gamma} = \mathrm{Diag}(0, -1, -2)$. Then for all $\Gamma \in \mathcal{B}_{\min\{1/3, r_1\}}(\overline{\Gamma}) \backslash \mathcal{M}(\bar{x})$ with r_1 taken from Definition 3.32 with $\mathcal{V} = \mathcal{B}_{r_1}(\bar{x}, \bar{\lambda})$, we know that

$$\Pi_{\mathcal{M}(\bar{x})}(\Gamma) = Q\mathrm{Diag}(\min\{0, \Gamma_1\}, \Gamma_2, \Gamma_3)Q^T,$$

where $Q \in \mathcal{O}^3(\Gamma)$. Let $\widehat{\Gamma} = Q\mathrm{Diag}(0, \Gamma_2, \Gamma_3)Q^T$. It follows from Definition 3.32 and [97, Theorem 3.1] that

$$\|\Pi_{\mathcal{M}(\bar{x})}(\Gamma) - \widehat{\Gamma}\| \leq \mathrm{dist}(\Gamma, \mathcal{M}(\bar{x})) = O(R(x, \Gamma)).$$

Thus Assumption 1 is satisfied. It is easy to calculate $L''_{xx}(\bar{x}, \overline{\Gamma}) = 4 > 0$, which implies SOSC holds at $(\bar{x}, \overline{\Gamma})$. Furthermore, it is obvious that bounded linear regular holds. Then we get the semi-isolated calmness of S_{KKT} at $\big(0, (\bar{x}, \overline{\Gamma})\big)$ holds by Theorem 3.34. Thus, if we have an oracle of finding a starting point in $\mathcal{B}_{\bar{r}}(\bar{x}, \overline{\Gamma})$ (cf. Theorem 4.14), the sequence $\{(x^k, \Gamma^k)\}$ generated by ALM converges to $(\bar{x}, \widetilde{\Gamma})$ for some $\widetilde{\Gamma} \in \mathcal{M}(\bar{x})$ while $\widetilde{\Gamma}$ does not have to be $\overline{\Gamma}$ or $\widehat{\Gamma}$.

We also provide another nontrivial NLSDP example, which is modified from the example proposed in the arxiv version of [36, Example 2] for different purpose.

Example 4.2. Consider the following example

$$\begin{array}{lll} \min & \frac{1}{2}x^2 + 2t & \\ \text{s.t} & tA - x^2 I_2 \in \mathbb{S}^2_+, & \Leftarrow \Gamma \\ & t \geq 0, & \Leftarrow y \end{array}$$

where $A = \begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix}$. This problem possesses the unique optimal solution $(\bar{t}, \bar{x}) = (0, 0)$. The corresponding multiplier is

$$\mathcal{M}(\bar{t}, \bar{x}) = \{(\Gamma, y) \in \mathbb{S}^2_- \times \mathbb{R} \mid \langle A, -\Gamma \rangle \leq 2\}.$$

We can pick $\bar{y} = 0$ and $\overline{\Gamma} = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$. For all $\Gamma \in \mathcal{B}_{\min\{r_1, 1/(2\sqrt{10})\}}(\overline{\Gamma})$ with r_1 taken from Definition 3.32 with $\mathcal{V} = \mathcal{B}_{r_1}(\bar{x}, \bar{\lambda})$, we know that $\langle A, -\Gamma \rangle < 2$, which implies that $\Pi_{\mathcal{M}(\bar{t},\bar{x})}(\Gamma) = \Pi_{\mathcal{S}^2_-}(\Gamma)$. Suppose $\Gamma = Q\mathrm{Diag}(\Gamma_1, \Gamma_2)Q^T$. Let $\widehat{\Gamma} = Q\mathrm{Diag}(0, \Gamma_2)Q^T$. Then we have

$$\|\Pi_{\mathcal{M}(\bar{x})}(\Gamma) - \widehat{\Gamma}\| \leq \mathrm{dist}(\Gamma, \mathcal{M}(\bar{x})) = O(R(x, \Gamma)),$$

which verifies Assumption 1. It is easy to see that $\begin{bmatrix} -1/2 & 0 \\ 0 & -1/2 \end{bmatrix} \in \mathcal{G}_1(\bar{t}, \bar{x}) \cap \mathrm{ri}\, \mathcal{G}_2(\bar{t}, \bar{x})$, which implies the validity of boundedly linear regularity by [96, Corollary 3]. It is easy to check that SOSC holds since $L''(\bar{t}, \bar{x}, \overline{\Gamma}, \bar{y}) = \begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix}$ and for all $w = (w_1, w_2) \in \mathcal{C}(\bar{t}, \bar{x})$, we have $w_1 = 0$. Thus we have SOSC holds at $(\bar{t}, \bar{x}, \overline{\Gamma}, \bar{y})$. It follows that semi-isolated calmness of S_{KKT} at $\big(0, (\bar{t}, \bar{x}, \overline{\Gamma}, \bar{y})\big)$ holds by Theorem 3.34. In this case, if we have an oracle of finding a starting point in $\mathcal{B}_{\bar{r}}(\bar{t}, \bar{x}, \overline{\Gamma}, \bar{y})$ (cf. Theorem 4.14), the sequence $\{(t^k, x^k, \Gamma^k, y^k)\}$ generated by ALM converges to $(\bar{t}, \bar{x}, \widetilde{\Gamma}, \widetilde{y})$ for some $(\widetilde{\Gamma}, \widetilde{y}) \in \mathcal{M}(\bar{t}, \bar{x})$. Here $(\widetilde{\Gamma}, \widetilde{y})$ does not have to be $(\overline{\Gamma}, \bar{y})$ or $(\widehat{\Gamma}, \widehat{y})$.

Chapter 5 Convergence analysis of ALM for NLSDP under strong variational sufficiency

In the above chapter, we mainly relax the uniqueness of multiplier under the convergence rate analysis of ALM for NLSDP. However, the requirement of Assumption 1 is very hard to be checked. Moreover, there are two problems left to be handled. Firstly, as [36] shows, usually, the dual Q-linear convergence rate together with the KKT residual R-linear rate is enough in practical solvers. Therefore it is meaningful to study whether we can get a convergence result of that kind instead of a primal-dual type under a weaker condition. Secondly, although we consider solving the ALM subproblem inexactly, we have not put forward a practical relative error criterion for it.

To overcome these difficulties, we turn our spot to a newly proposed property named strong variational sufficiency, which turns out to be of great use in the convergence analysis of multiplier methods. However, what this property implies for non-polyhedral problems remains a puzzle. In this chapter, we prove the equivalence between the strong variational sufficiency and the strong second order sufficient condition (SOSC) for nonlinear semidefinite programming (NLSDP), without requiring the uniqueness of multiplier or any other constraint qualifications. Based on this equivalence, the local convergence property of the augmented Lagrangian method (ALM) for NLSDP can be established under strong SOSC in the absence of constraint qualifications. Moreover, under the strong SOSC, we can apply the semi-smooth Newton method to solve the ALM subproblems of NLSDP as the positive definiteness of the generalized Hessian of augmented Lagrangian function is satisfied.

It is worth to note that the equality constraint $h(x) = 0$ with h being twice continuously differentiable is omitted in this chapter for simplicity as it is a polyhedral constraint and we can add it into the indicator function of the cartesian product with $\mathbb{S}^n_+$ and $\{0\}$.

To see how ALM (1-13) works for nonconvex problems without constraint qualifications, we recall Example 4.1. The optimal solution of (4.1) is $\bar{x} = 0$ with the corresponding multiplier set $\mathcal{M}(\bar{x}) := \{Y \mid Y \in \mathbb{S}^3_-\}$. Due to the unboundedness of $\mathcal{M}(\bar{x})$, we know from [104, Theorem 4.1] that the Robinson constraint qualification [105] does not hold at $\bar{x}$. Pick a particular multiplier

$$\overline{Y} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \in \mathcal{M}(\bar{x}).$$

It is clear that $L''_{xx}(\bar{x}, \overline{Y}) = 4 > 0$, which implies strong SOSC (Definition 3.20) holds at $(\bar{x}, \overline{Y})$. We directly apply the ALM (Algorithm 2) to problem (4.1), and we find that the corresponding ALM subproblem in (1-13) can be solved exactly. Then, it can be observed from Figure 5-1 that for fixed ρ^k and $\widetilde{\rho}^k$, $\mathrm{dist}(Y^k, \mathcal{M}(\bar{x}))$, the distance between the k-th iteration Y^k and $\mathcal{M}(\bar{x})$, converges to zero linearly.

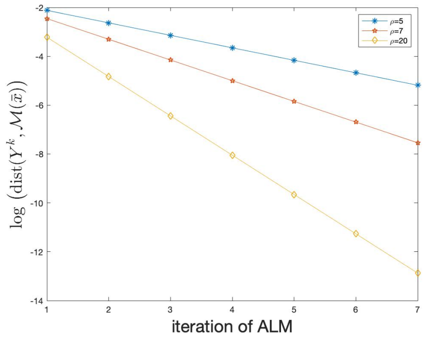


图 5-1 ALM for solving (4.1) with different penalty parameters $\rho^k = \rho$.

5.1 Necessary background

To characterize strong variational sufficient condition mentioned in Definition 1.1, the tools of the second subderivative and generalized quadratic form are needed.

Definition 5.1. *[1, Definition 13.6] Let $\bar{x}$ be a point where the function $f : \mathbb{X} \to [-\infty, +\infty]$ is finite. f is twice epi-differentiable at $\bar{x}$ for v if the functions*

$$\Delta_t^2 f(\bar{x} \mid v)(u) = \frac{f(\bar{x} + tu) - f(\bar{x}) - t\langle v, u \rangle}{\frac{1}{2}t^2}$$

epi-converge to $d^2 f(\bar{x} \mid v)$ as $t \downarrow 0$, where $d^2 f(\bar{x} \mid v)$ is the second subderivative of f at $\bar{x}$ for v defined as

$$d^2 f(\bar{x} \mid v)(w) = \liminf_{t \downarrow 0, u \to w} \Delta_t^2 f(\bar{x} \mid v)(u).$$

We know from [100, Theorem 3.6] that $\delta_{\mathbb{S}^n_+}$ is twice epi-differentiable at X for Y with $Y \in \mathcal{N}_{\mathbb{S}^n_+}(X)$.

Definition 5.2. *[55] A generalized linear mapping $\mathcal{R}$ from $\mathbb{X}$ to $\mathbb{Y}$, is a set-valued mapping for which $\mathrm{gph}\,\mathcal{R}$ is a subspace of $\mathbb{X} \times \mathbb{Y}$. This means that $\mathrm{dom}\mathcal{R}$ is a subspace $\mathbb{Z}$ of $\mathbb{X}$, $\mathcal{R}(0)$ is a subspace $\mathbb{Z}'$ of $\mathbb{Y}$, and there is an ordinary linear mapping $\mathcal{R}_0 : \mathbb{Z} \to \mathbb{Y}$ such that $\mathcal{R}(x) = \mathcal{R}_0(x) + \mathbb{Z}'$ for $x \in \mathbb{Z}$. We call a function $q : \mathbb{X} \to (-\infty, +\infty]$ is a generalized quadratic form on $\mathbb{X}$ if $q(0) = 0$ and the subgradient mapping $\partial q : \mathbb{X} \rightrightarrows \mathbb{X}$ is generalized linear. A function g on $\mathbb{X}$ will be called generalized twice differentiable at x for a subgradient y if it is twice epi-differentiable at x for y with the second-order subderivative $d^2 g(x \mid y)$ being a generalized quadratic form.*

The following definition of the quadratic bundle is taken from [55], which is essential in characterizing the strong variational sufficiency.

Definition 5.3. *For general optimization problem* (1-10)*, suppose* $(\bar{x}, \overline{Y})$ *is a KKT pair. The quadratic bundle of* θ *is defined as*

$$\mathrm{quad}\,\theta(G(\bar{x}) \mid \overline{Y}) = \begin{cases} \textit{the collection of generalized quadratic forms } q \textit{ for which} \\ \exists (X^k, Y^k) \to (G(\bar{x}), \overline{Y}) \textit{ with } \theta \textit{ generalized twice differentiable} \\ \textit{at } X^k \textit{ for } Y^k \textit{ and such that the generalized quadratic} \\ \textit{forms } q_k = \frac{1}{2} d^2\theta(X^k \mid Y^k) \textit{ converge epigraphically to } q. \end{cases}$$

As mentioned in [55], the variational sufficient condition guarantees the local optimality for (1-14). The following result is taken from [55, Theorem 5], which is useful for the subsequent analysis.

Proposition 5.4. *For general optimization problem* (1-10)*, strong variational sufficient condition for local optimality with respect to* $(\bar{x}, \overline{Y})$ *is equivalent to that every* $q \in \mathrm{quad}\,\theta(G(\bar{x}) \mid \overline{Y})$ *has*

$$\frac{1}{2}\langle L''_{xx}(\bar{x}, \overline{Y})d, d\rangle + q(G'(\bar{x})d) > 0 \quad \text{when } d \neq 0. \tag{5-1}$$

5.2 The characterization of strong variational sufficient condition for NLSDP

In this section, we will study the relationship between strong variational sufficient condition and the well-known strong SOSC (2-51) for NLSDP by combining (5-1) and [100, Theorem 3.3] together. Firstly, we use nonlinear second order cone programming (NLSOC) as an example to illustrate our approach. The general NLSOC can be written as

$$\begin{aligned} \min_{x \in \mathbb{X}} \quad & f(x) \\ \text{s.t.} \quad & g(x) \in \mathcal{K}, \end{aligned} \tag{5-2}$$

where $f : \mathbb{X} \to \mathbb{R}, g : \mathbb{X} \to \mathbb{R}^m$ are both twice continuously differentiable, $\mathcal{K}$ is the second order cone defined as $\mathcal{K} = \{(u_1, \ldots, u_m) \in \mathbb{R}^m \mid u_1 \geq \|(u_2, \ldots, u_m)\|\} = \{u \mid h(u) \leq 0\}$ with $h(u) = -u_1 + \|(u_2, \ldots, u_m)\|$. The ordinary Lagrangian function of problem (5-2) is defined by

$$L(x, y) := f(x) + \langle y, g(x)\rangle, \quad (x, y) \in \mathbb{X} \times \mathbb{R}^m. \tag{5-3}$$

Given a stationary point $\bar{x}$. Let

$$\mathcal{M}_{\mathrm{soc}}(\bar{x}) := \left\{ y \in \mathbb{R}^m \;\middle|\; \begin{array}{l} L'_x(\bar{x}, y) = 0, \\ \mathcal{K} \ni g(\bar{x}) \perp y \in \mathcal{K}^\circ \end{array} \right\}.$$

be the set of all multipliers $y \in \mathbb{R}^m$ satisfying the KKT condition for (5-2), where $\mathcal{K}^\circ$ is the polar cone of $\mathcal{K}$ defined in [59, Section 14]. The strong SOSC at $(\bar{x}, \bar{y})$ is defined as

$$\langle L''_{xx}(\bar{x}, \bar{y})d, d\rangle - \Upsilon_{g(\bar{x})}\big(\bar{y}, g'(\bar{x})d\big) > 0 \quad \forall\, 0 \neq g'(\bar{x})d \in \operatorname{aff} \mathcal{C}_{\mathcal{K}}(g(\bar{x}), \bar{y}), \tag{5-4}$$

where the explicit form of $\Upsilon_{g(\bar{x})}\big(\bar{y}, g'(\bar{x})d\big)$ is given in [106, Theorem 29].

As illustrated in [55, Example 3], the explicit form of the quadratic bundle $\operatorname{quad} \delta_{\mathcal{K}}(g(\bar{x}) \mid \bar{y})$ for certain reference KKT point $(\bar{x}, \bar{y})$ can be obtained by using corresponding results for polyhedral problems. For example, if $g(\bar{x}) \in \operatorname{int} \mathcal{K}$, we have $\bar{y} = 0$. It can be checked directly by using [100, Theorem 3.3] and [106, (36),(43)] that the quadratic bundle $\operatorname{quad} \delta_{\mathcal{K}}(g(\bar{x}) \mid \bar{y})$ consists of solely $q \equiv 0$. If $g(\bar{x}) \in \operatorname{bd} \mathcal{K} \backslash \{0\}$, for any nonzero $\bar{y} \in \mathcal{N}_{\mathcal{K}}(g(\bar{x}))$, the quadratic bundle $\operatorname{quad} \delta_{\mathcal{K}}(g(\bar{x}) \mid \bar{y})$ consists of the generalized quadratic form

$$q = \frac{1}{2} d^2 \delta_{\mathcal{K}}(g(\bar{x}) \mid \bar{y}) \quad \text{with} \quad q(w) = \begin{cases} \frac{1}{2} w \cdot h''(g(\bar{x}))w & \text{if } h'(g(\bar{x}))w = 0, \\ \infty & \text{otherwise}. \end{cases} \tag{5-5}$$

If $g(\bar{x}) = 0$ and $\bar{y} \in \operatorname{int}(-\mathcal{K})$, the quadratic bundle consists of $q = \delta_{\{0\}}$. If $g(\bar{x}) = 0$ and $\bar{y} \in \operatorname{bd}(-\mathcal{K})\backslash\{0\}$, both $\delta_{\{0\}}$ and (5-5) constitute the quadratic bundle. It can be checked directly that for NLSOC, strong variational sufficient condition is equivalent to strong SOSC when the reference point $(\bar{x}, \bar{y})$ lies in one of the above circumstances.

Thus, only two circumstances for NLSOC remain to be discussed. The first one is $g(\bar{x}) = 0$ and $\bar{y} = 0$. Since $\mathcal{K}$ is C^2−cone reducible, we know from [100, Theorem 3.3], [106, Theorem 29] that for any $(g^k, y^k) \to (g(\bar{x}), \bar{y})$ and $w \in \mathbb{R}^m$,

$$d^2 \delta_{\mathcal{K}}(g^k \mid y^k)(w) = w^T \mathcal{H}(g^k, y^k) w + \delta_{\mathcal{C}_{\mathcal{K}}(g^k, y^k)}(w) \geq 0, \tag{5-6}$$

where $\mathcal{H}(g^k, y^k) = \frac{y_1^k}{g_1^k}(g^k)'^T(1\ 0^T; 0\ \ -I_{m-1})(g^k)'$ if $g^k \in \operatorname{bd} \mathcal{K}\backslash\{0\}$ and $\mathcal{H}(g^k, y^k) = 0$ otherwise. If we pick $g^k \in \operatorname{int} \mathcal{K} \to g(\bar{x})$ and $y^k = 0$ for all k, it is easy to see from [106, Theorem 29, (35), (36)] that

$$d^2 \delta_{\mathcal{K}}(g^k \mid y^k)(w) = 0 + \delta_{\mathcal{C}_{\mathcal{K}}(g^k, y^k)}(w),$$

where $\mathcal{C}_{\mathcal{K}}(g^k, y^k) = \mathbb{R}^m$. Thus we have proved $0 \in \operatorname{quad} \delta_{\mathcal{K}}(g(\bar{x}) \mid \bar{y})$. Combining this with (5-6), it follows that in this case, (5-1) is equivalent to strong SOSC.

The second case is $g(\bar{x}) \in \operatorname{bd}\mathcal{K}\backslash\{0\}$ and $\bar{y} = 0$. Let $g^k \in \operatorname{int} \mathcal{K} \to g(\bar{x})$ and $y^k = 0$ for each k. We know that

$$d^2 \delta_{\mathcal{K}}(g^k \mid y^k)(w) = 0 + \delta_{\mathcal{C}_{\mathcal{K}}(g^k, y^k)}(w),$$

where $\mathcal{C}_{\mathcal{K}}(g^k, y^k) = \mathbb{R}^m$. As $k \to \infty$, we know that $\lim_{k \to \infty} d^2 \delta_{\mathcal{K}}(g^k \mid y^k)(w) = 0$. Using [106, (35), (36), (43)] again, we also have (5-1) is equivalent to strong SOSC.

Thus, we immediately obtain the following characterization of strong variational sufficiency for NLSOC, which is a supplement for [55, Example 3].

Proposition 5.5. *Let $\bar{x} \in \mathbb{X}$ be a stationary point to the NLSOC* (5-2) *and $\bar{y} \in \mathcal{M}(\bar{x})$. The strong variational sufficient condition* (5-1) *with respect to $(\bar{x}, \bar{y})$ holds if and only if the strong SOSC* (5-4) *holds at $(\bar{x}, \bar{y})$.*

It is worth noting that the above proposition is not surprising. If $g(\bar{x}) \neq 0$, we can obtain this result by regarding the NLSOC problem as a polyhedral problem as shown in [55, Example 3]. If $g(\bar{x}) = 0$, the $\sigma-$term happens to be 0, which makes the calculation of the quadratic bundle much simpler. When it comes to NLSDP problem (2-48), things are not so easy as we can neither regard SDP as a polyhedral problem nor have a simplification for its $\sigma-$term (2-52). However, the success of this approach to NLSOC gives us a hint that this approach may also work for NLSDP. Before we put forward our main result, we need the following proposition on the quadratic bundle defined by Definition 5.3 for $\delta_{\mathbb{S}^n_+}$. Let

$$\mathcal{M}(\bar{x}) := \left\{ Y \in \mathbb{R}^m \,\middle|\, \begin{array}{l} L'_x(\bar{x}, Y) = 0, \\ \mathbb{S}^n_+ \ni g(\bar{x}) \perp Y \in \mathbb{S}^n_- \end{array} \right\} \tag{5-7}$$

be the set of all multipliers $Y \in \mathbb{S}^n$ satisfying the KKT condition with respect to $\bar{x}$ for (2-48).

Proposition 5.6. *Let $\bar{x} \in \mathbb{X}$ be a stationary point to NLSDP* (2-48) *and $\overline{Y} \in \mathcal{M}(\bar{x})$, where $\mathcal{M}(\bar{x})$ is given in* (5-7)*. Let $A = G(\bar{x}) + \overline{Y}$, which possesses the decomposition* (2-42)*. Then, there exists $q \in \operatorname{quad} \delta_{\mathbb{S}^n_+}(G(\bar{x}) \mid \overline{Y})$ such that for all $H \in \mathbb{S}^n$,*

$$\begin{aligned} q(H) &= -\frac{1}{2}\Upsilon_{G(\bar{x})}(\overline{Y}, H) + \delta_{\operatorname{aff} \mathcal{C}_{\mathbb{S}^n_+}(G(\bar{x}), \overline{Y})}(H) \\ &= \sum_{i \in \alpha, j \in \gamma} \frac{-\lambda_j(A)}{\lambda_i(A)} (\widetilde{H}_{ij})^2 + \delta_{\operatorname{aff} \mathcal{C}_{\mathbb{S}^n_+}(G(\bar{x}), \overline{Y})}(H), \end{aligned} \tag{5-8}$$

where $\widetilde{H} = P^T H P$.

Proof. For each k, choose

$$X^k = P \begin{bmatrix} \Lambda(A)_{\alpha\alpha} & 0 & 0 \\ 0 & (z^k)_\beta & 0 \\ 0 & 0 & 0 \end{bmatrix} P^T \quad \text{and} \quad Y^k = \overline{Y},$$

where $z^k \downarrow 0$ (this notation means for each k, $z^k > 0$ and $z^k \to 0$ as $k \to \infty$) with each $(z^k)_i$ non-increasing on k. Let $A^k = X^k + Y^k$ for each k. Since $\mathbb{S}^n_+$ is C^2-cone reducible [5, Example 3.140], we know from [100, Theorem 3.3], (2-46) and (2-52) that for any

$H \in \mathbb{S}^n$,

$$\begin{aligned}\frac{1}{2}d^2\delta_{\mathbb{S}^n_+}(X^k \mid Y^k)(H) &= \sum_{i\in\alpha\cup\beta, j\in\gamma} \frac{-\lambda_j(A^k)}{\lambda_i(A^k)}\big(\widetilde{H}_{ij}\big)^2 + \delta_{\mathcal{C}_{\mathbb{S}^n_+}(X^k,Y^k)}(H) \\ &= \sum_{i\in\alpha, j\in\gamma} \frac{-\lambda_j(A)}{\lambda_i(A)}\big(\widetilde{H}_{ij}\big)^2 + \sum_{i\in\beta, j\in\gamma} \frac{-\lambda_j(A)}{\lambda_i(X^k)}\big(\widetilde{H}_{ij}\big)^2 \\ &\quad + \delta_{\mathcal{C}_{\mathbb{S}^n_+}(X^k,Y^k)}(H),\end{aligned}$$

where $\mathcal{C}_{\mathbb{S}^n_+}(X^k, Y^k) = \{H \in \mathbb{S}^n \mid \widetilde{H}_{\gamma\gamma} = 0\}$. It is worth to note that $d^2\delta_{\mathbb{S}^n_+}(X^k \mid Y^k)(H)$ is a generalized quadratic form by using Definition 5.2, as $d^2\delta_{\mathbb{S}^n_+}(X^k \mid Y^k)(0) = 0$ and $\partial d^2\delta_{\mathbb{S}^n_+}(X^k \mid Y^k)(H) = \mathcal{R}(H) + \mathcal{N}_{\mathcal{C}_{\mathbb{S}^n_+}(X^k,Y^k)}(H)$ with $\mathcal{N}_{\mathcal{C}_{\mathbb{S}^n_+}(X^k,Y^k)}(H)$ being a subspace and

$$\mathcal{R}(H)(\Delta H) = -4\sum_{i\in\alpha, j\in\gamma} \frac{\lambda_j(A)}{\lambda_i(A)}(\widetilde{H})_{ij} \cdot (P^T\Delta HP)_{ij} - 4\sum_{i\in\beta, j\in\gamma} \frac{\lambda_j(A)}{\lambda_i(A)}(\widetilde{H})_{ij} \cdot (P^T\Delta HP)_{ij}$$

being linear on H. This also implies that $\delta_{\mathbb{S}^n_+}$ is generalized twice differentiable at X^k for Y^k.

Let $f_1^k(H) = \sum_{i\in\alpha,j\in\gamma} \frac{-\lambda_j(A)}{\lambda_i(A)}\big(\widetilde{H}_{ij}\big)^2$, $f_2^k(H) = \delta_{\mathcal{C}_{\mathbb{S}^n_+}(X^k,Y^k)}(H)$ and $f_3^k(H) = \sum_{i\in\beta,j\in\gamma} \frac{-\lambda_j(A)}{\lambda_i(X^k)}\big(\widetilde{H}_{ij}\big)^2$. By the definition of continuous convergence mentioned in [1, page 250], we know that f_1^k converges continuously to f_1 with

$$f_1(H) = \sum_{i\in\alpha, j\in\gamma} \frac{-\lambda_j(A)}{\lambda_i(A)}\big(\widetilde{H}_{ij}\big)^2.$$

It follows from [1, Theorem 7.11] that f_1^k epi-converges to f_1. It can be checked easily that f_1^k also pointwise converges [1, page 239] to f_1. Since $\mathcal{C}_{\mathbb{S}^n_+}(X^k, Y^k)$ is a constant closed set, we know that f_2^k pointwise converges and epi-converges [1, Proposition 4.4] to f_2 with $f_2(H) = \delta_{\mathcal{C}_{\mathbb{S}^n_+}(X^k,Y^k)}(H)$. It follows from [1, Theorem 7.46(a)] that $f_1^k + f_2^k$ epi-convergences to $f_1 + f_2$. Also, it can be checked directly that $f_1^k + f_2^k$ pointwise convergences to f_1+f_2. By the construction of z^k, the sequence of $\{f_3^k\}$ is nondecreasing ($f_3^k \le f_3^{k+1}$). We know from [1, Proposition 7.4] that f_3^k epi-converges to $\sup_k\{\mathrm{cl} f_3^k\}$, where $\mathrm{cl} f_3^k$ is the closure of f_3^k. It is easy to see that $\sup_k\{\mathrm{cl} f_3^k\} = \sup_k\{f_3^k\} = \delta_{\mathcal{V}}$ with $\mathcal{V} = \{H \in \mathbb{S}^n \mid \widetilde{H}_{\beta\gamma} = 0\}$ and f_3^k pointwise converges to $\delta_{\mathcal{V}}$. Since $f_1^k + f_2^k$ pointwise and epi-converges to $f_1 + f_2$, by using [1, Theorem 7.46(a)] again, we obtain that $f_1^k + f_2^k + f_3^k$ epi-converges to $f_1 + f_2 + f_3$.

Combining the above discussion with (2-47), the generalized quadratic form $\frac{1}{2}d^2\delta_{\mathbb{S}^n_+}(X^k \mid Y^k)(H)$ converges epigraphically to generalized quadratic form

$$\sum_{i\in\alpha, j\in\gamma} \frac{-\lambda_j(A)}{\lambda_i(A)}(\widetilde{H}_{ij})^2 + \delta_{\mathrm{aff}\,\mathcal{C}_{\mathbb{S}^n_+}(G(\bar{x}),\overline{Y})}(H).$$

Thus we have verified this proposition. □

The following result is on the explicit characterization of the strong variational sufficiency of local optimality for NLSDP, which is the main result of this chapter.

Theorem 5.7. *Let* $\bar{x} \in \mathbb{X}$ *be a stationary point to the NLSDP* (2-48) *and* $\overline{Y} \in \mathcal{M}(\bar{x})$. *Then the following three conditions are equivalent*

(i) the strong variational sufficient condition with respect to $(\bar{x}, \overline{Y})$ *holds;*

(ii) the strong second order sufficient condition (SOSC) (2-51) *holds at* $(\bar{x}, \overline{Y})$*;*

(iii) there exist $\rho_0 > 0$ *and* $\underline{\eta} > 0$ *such that for any* $\rho \geq \rho_0$ *and any* $W \in \partial_B \Pi_{\mathbb{S}^n_-}(G(\bar{x}) + \rho^{-1}\overline{Y})$,

$$\langle d, \mathcal{A}_\rho(\overline{Y}, W)d\rangle \geq \underline{\eta}\|d\|^2 \quad \forall d \in \mathbb{X},$$

where $\mathcal{A}_\rho(\overline{Y}, W)$ *is defined by* (2-55).

Proof. "$(i) \Longrightarrow (ii)$": This direction can be obtained directly from Proposition 5.6 and Proposition 5.4 as we only need to substitute (5-8) into (5-1).

"$(ii) \Longrightarrow (iii)$": This proof sketch is similar to [50, Proposition 4] although they require the validity of nondegeneracy, which is superfluous here. It follows from Definition 2-51 that there exists $\eta_0 > 0$ such that

$$\langle L''_{xx}(\bar{x}, \overline{Y})d, d\rangle - \Upsilon_{G(\bar{x})}\big(\overline{Y}, G'(\bar{x})d\big) \geq \eta_0\|d\|^2 \quad \forall\ G'(\bar{x})d \in \operatorname{aff}\mathcal{C}_{\mathbb{S}^n_+}(G(\bar{x}), \overline{Y}).$$

By using [50, Lemma 7], there exist two positive numbers ρ_1 and $\underline{\eta} \in (0, \eta_0/2]$ such that for any $\rho' \geq \rho_1$,

$$\begin{aligned}&\langle L''_{xx}(\bar{x}, \overline{Y})d, d\rangle - \Upsilon_{G(\bar{x})}\big(\overline{Y}, G'(\bar{x})d\big)\\ &+ \rho'\|\overline{P}_\gamma^T(G'(\bar{x})d)\overline{P}_\gamma\|^2 + \rho'\|\overline{P}_\beta^T(G'(\bar{x})d)\overline{P}_\gamma\|^2 \geq 2\underline{\eta}\|d\|^2 \quad \forall d \in \mathbb{X}.\end{aligned}$$

Suppose a sufficient large $\rho_0 \geq \rho_1$. For any $\rho \geq \rho_0$ and $d \in \mathcal{X}$, we have

$$\begin{aligned}&-\Upsilon_{G(\bar{x})}(\overline{Y}, G'(\bar{x})d) - 2\rho\sum_{i\in\alpha, j\in\gamma}\frac{-\lambda_j}{\rho\lambda_i - \lambda_j}(\overline{P}^T(G'(\bar{x})d)\overline{P})_{ij}^2\\ &= 2\sum_{i\in\alpha, j\in\gamma}\frac{-\lambda_j}{\lambda_i}(\overline{P}^T(G'(\bar{x})d)\overline{P})_{ij}^2 - 2\rho\sum_{i\in\alpha, j\in\gamma}\frac{-\lambda_j}{\rho\lambda_i - \lambda_j}(\overline{P}^T(G'(\bar{x})d)\overline{P})_{ij}^2\\ &= 2\sum_{i\in\alpha, j\in\gamma}\frac{\lambda_j^2}{\lambda_i(\rho\lambda_i - \lambda_j)}(\overline{P}^T(G'(\bar{x})d)\overline{P})_{ij}^2 \leq \underline{\eta}\|d\|^2,\end{aligned}$$

where the last inequality can be obtained by sufficiently large ρ. Combining the above two relations together, we have for any $\rho \geq \rho_0$, $\rho' \geq \rho_0 \geq \rho_1$,

$$\begin{aligned}&\langle L''_{xx}(\bar{x}, \overline{Y})d, d\rangle + 2\rho\sum_{i\in\alpha, j\in\gamma}\frac{-\lambda_j}{\rho\lambda_i - \lambda_j}(\overline{P}^T(G'(\bar{x})d)\overline{P})_{ij}^2\\ &+ \rho'\|\overline{P}_\gamma^T(G'(\bar{x})d)\overline{P}_\gamma\|^2 + \rho'\|\overline{P}_\beta^T(G'(\bar{x})d)\overline{P}_\gamma\|^2 \geq \underline{\eta}\|d\|^2.\end{aligned}$$

Then it can be checked directly that for all $d \in \mathbb{X}$ and $\rho \geq \rho_0$, $\langle d, \mathcal{A}_\rho(\overline{Y}, W)d\rangle \geq \underline{\eta}\|d\|^2$.

"$(iii) \Longrightarrow (i)$": It follows directly from [55, Theorem 3] as $\mathcal{A}_\rho(\overline{Y}, W)$ coincides with [55, (3.6)]. □

Remark 5.1. Next, we will discuss the relationship between tilt stability and variationally strong convexity/strong variational sufficient condition. Given a function $\Xi : \mathbb{X} \to (-\infty, +\infty]$. As shown in [107], for a general optimization problem

$$\min_{z} \ \Xi(z),$$

the variationally strong convexity of Ξ at $\bar{z}$ for $0 \in \partial\Xi(\bar{z})$ with modulus $\sigma > 0$ implies that $\bar{z}$ is a tilt stable local minimizer of Ξ with modulus σ^{-1}. A point $\bar{z} \in \operatorname{dom}\Xi$ is a tilt stable local minimizer of Ξ if there is a neighborhood $\mathcal{Z}$ of $\bar{z}$ such that the mapping

$$T : v \mapsto \arg\min_{z\in\mathcal{Z}}\{\Xi(z) - \langle v, z\rangle\}$$

is single-valued and Lipschitz continuous on some neighborhood of 0 with $T(0) = \bar{z}$.

However, it is worth to note that tilt stability usually does not imply variationally strong convexity as explained in [56, Remark 2.8] and [107]. If Ξ is amenable [1, Chapter 13], the equivalence between the two properties is verified in [56, Proposition 2.9]. In particular for NLSDP, it can be checked directly that the Robinson constraint qualification (RCQ) [105], i.e.,

$$\mathcal{N}_{\mathbb{S}^n_+}(G(\bar{x})) \cap (G'(\bar{x})\mathbb{X})^{\perp} = 0,$$

is equivalent to the constraint qualification in [1, Definition 10.23]. Thus, the general objective function of NLSDP with $z = x$ and $\Xi = f + \delta_{\mathbb{S}^n_+}(G(\cdot))$ is amenable. It follows that for NLSDP variationally strong convexity is equivalent to tilt stability under RCQ. For more information on the characterization of variationally strong convexity, the readers may refer to [20, 55, 56].

In [55, (2.18)], Rockafellar defines the so-called augmented tilt stability. The augmented tilt stability of (1-10) is said to hold for a given $\rho > 0$ with $\mathcal{L}_\rho(x, Y)$ being locally convex-concave on the neighborhood $\mathcal{X} \times \mathcal{Y}$, if there is a neighborhood $\mathcal{V}$ of 0 such that the mapping

$$(v, Y) \in \mathcal{V} \times \mathcal{Y} \mapsto \arg\min_{x\in\mathcal{X}}\{\mathcal{L}_\rho(x, Y) - \langle v, x\rangle\} \tag{5-9}$$

is single-valued and Lipschitz continuous. As shown in [55, (1.5)], the mapping (5-9) can also be written as

$$(v, Y) \in \mathcal{V} \times \mathcal{Y} \mapsto \arg\min_{x\in\mathcal{X},u}\{\phi_\rho(x, u) - \langle Y, u\rangle - \langle v, x\rangle\}, \tag{5-10}$$

where ϕ_ρ is given in (1-15). From (5-10), it seems that augmented tilt stability is an application of tilt stability with $z = (x, u)$ and $\Xi = \phi_\rho$. In [55, Theorem 2], Rockafellar also shows augmented tilt stability is equivalent to the strong variational sufficient condition, i.e., the variationally strong convexity of ϕ_ρ, without constraint qualifications. By using Theorem 5.7, the strong SOSC is also equivalent to the augmented tilt stability.

Recently, Khanh et al. [56, Theorem 6.5] also provide a sufficient condition for (strong) variational sufficiency of strongly amenable problem in the form of (1-10). However, the strong variational sufficiency mentioned in [56] is different from the one used here. In [56], strong variational sufficiency is defined as the variationally strong convexity of the function $x \to f(x) + \theta(G(x))$, while this chapter focuses on that of the perturbed function $(x, u) \to f(x) + \theta(G(x) + u) + \frac{r}{2}\|u\|^2$. In their result, they require a condition named second order qualification condition [108, (3.15)] at $\bar{x}$, i.e.,

$$\ker G'(\bar{x})^* \cap \partial^2\theta(G(\bar{x}), \overline{Y})(0) = \{0\}, \tag{5-11}$$

where $\partial^2\theta(G(\bar{x}), \overline{Y})$ is defined in [108, Definition 2.1]. As illustrated in [56, Lemma 7.2], the second order qualification condition is equivalent to linear independent constraint qualification (LICQ) for nonlinear programming. By employing [80, Theorem 3.1], we shall establish the following characterization of the second order qualification condition for NLSDP.

Corollary 5.8. *Suppose $\bar{x}$ is a feasible point for NLSDP* (2-48) *and $\overline{Y} \in \mathcal{N}_{\mathbb{S}^n_+}(G(\bar{x}))$. Suppose $G(\bar{x}) + \overline{Y}$ possesses the eigenvalue decomposition* (2-42). *Then, the second order qualification condition* (5-11) *at $(G(\bar{x}), \overline{Y})$ is equivalent to the nondegeneracy condition at $(G(\bar{x}), \overline{Y})$, i.e.,*

$$\ker G'(\bar{x})^* \cap \operatorname{lin}\big(\mathcal{T}_{\mathbb{S}^n_+}(G(\bar{x}))\big)^{\perp} = \{0\}, \tag{5-12}$$

where $\operatorname{lin}\big(\mathcal{T}_{\mathbb{S}^n_+}(G(\bar{x}))\big)$ is the linearity space of $\mathcal{T}_{\mathbb{S}^n_+}(G(\bar{x}))$, i.e., the largest linear space in $\mathcal{T}_{\mathbb{S}^n_+}(G(\bar{x}))$.

Proof. It follows from [11, (18)] that

$$\operatorname{lin}\big(\mathcal{T}_{\mathbb{S}^n_+}(G(\bar{x}))\big) = \{H \in \mathbb{S}^n \mid [P_\beta\ P_\gamma]^T H [P_\beta\ P_\gamma] = 0\},$$

which implies that $\operatorname{lin}\big(\mathcal{T}_{\mathbb{S}^n_+}(G(\bar{x}))\big)^{\perp} = \{H \in \mathbb{S}^n \mid P_\alpha^T H P_\alpha = 0,\ P_\alpha^T H P_{\beta\cup\gamma} = 0\}$. We know from [80, Theorem 3.1] that

$$\begin{aligned} \partial^2\delta_{\mathbb{S}^n_+}(G(\bar{x}), \overline{Y})(0) &= \{H \in \mathbb{S}^n \mid P_\alpha^T H P = 0,\ (P_\beta^T H P_\beta, 0) \in \mathcal{N}_{\operatorname{gph}\mathcal{N}_{\mathbb{S}^{|\beta|}_+}}(0,0)\} \\ &\subseteq \operatorname{lin}\big(\mathcal{T}_{\mathbb{S}^n_+}(G(\bar{x}))\big)^{\perp}. \end{aligned}$$

Thus, we immediately obtain the second order qualification condition from the nondegeneracy (5-12).

Conversely, we know from the explicit characterization of $\mathcal{N}_{\operatorname{gph}\mathcal{N}_{\mathbb{S}^{|\beta|}_+}}(0,0)$ obtained in [80, Theorem 3.1] that if $(H, 0) \in \mathcal{N}_{\operatorname{gph}\mathcal{N}_{\mathbb{S}^{|\beta|}_+}}(0,0)$, then H is free, which implies that

$$\partial^2\delta_{\mathbb{S}^n_+}(G(\bar{x}), \overline{Y})(0) = \operatorname{lin}\big(\mathcal{T}_{\mathbb{S}^n_+}(G(\bar{x}))\big)^{\perp}.$$

Thus, we obtain the nondegeneracy from the second order qualification condition (5-11), directly. □

In [56, Theorem 6.5 (ii)], the authors provide a sufficient condition for the strong variational sufficiency mentioned in [56, Definition 6.1] via coderivetive for strongly amenable problem in the form of (1-10). In this chapter, we present an explicit characterization of strong variational sufficiency for NLSDP without any constraint qualification. By combining [56, Theorem 6.2], Theorem 5.7 and [16, Theorem 5.6, Lemma 6.3] together, we know that under second order qualification condition, the two strong variational sufficiencies are the same. However, their relationship without the nondegeneracy condition remains unclear.

Moreover, by using [109, Example 2.2], it can be proved in a similar manner to [55, Theorem 3] that the variational sufficiency of local optimality for NLSDP (2-48) is equivalent to the existence of $\rho_0 > 0$ and a convex open neighborhood $\mathcal{V}$ of $(\bar{x}, \overline{Y})$ such that for any $\rho \geq \rho_0$, $(x, Y) \in \mathcal{V}$ and any $W \in \partial_B \Pi_{\mathbb{S}^n_-}(G(x) + \rho^{-1}Y)$,

$$\langle d, \mathcal{A}_\rho(Y, W)d \rangle \geq 0 \quad \forall d \in \mathbb{X}.$$

However, whether the variational sufficiency of local optimality is equivalent to certain second order optimality condition without any constraint qualification for NLSDP (2-48) is still a future work that we are working on.

5.3 Semi-smooth Newton-CG based ALM for nonconvex NLSDP

In this section, we apply the main result Theorem 5.7 to study the local convergence rate of the (extended) ALM (1-13) for solving NLSDP. The detail algorithm is stated in Algorithm 2. By [20, Theorems 1.1 and 1.2], the strong variational sufficient condition with respect to $(\bar{x}, \overline{Y})$ for local optimality holds if and only if there exist $\bar{\rho} > 0$ and a closed convex neighborhood $\mathcal{X} \times \mathcal{Y}$ of $(\bar{x}, \overline{Y})$ such that for all $\rho \geq \bar{\rho}$, $\mathcal{L}_\rho(x, Y)$ is strongly convex in $x \in \mathcal{X}$ with modulus $s > 0$ [1] for all $Y \in \mathcal{Y}$ and concave in $Y \in \mathcal{Y}$ for all $x \in \mathcal{X}$.

In the above algorithm, subproblem (5-13) is solved inexactly. Three increasing tightness stopping criteria for the updating of x^{k+1} are illustrated in [20, equation (1.15)]:

$$\left(2\widetilde{\rho}^k[\mathcal{L}_{\rho^k}(x^{k+1}, Y^k) - \inf_{\mathcal{X}} \mathcal{L}_{\rho^k}(\cdot, Y^k)]\right)^{1/2} \leq \begin{cases} (a)\ \epsilon_k, \\ (b)\ \epsilon_k \min\{1, \|\widetilde{\rho}^k(\mathcal{L}_{\rho^k})'_Y(x^{k+1}, Y^k)\|\}, \\ (c)\ \epsilon_k \min\{1, \|\widetilde{\rho}^k(\mathcal{L}_{\rho^k})'_Y(x^{k+1}, Y^k)\|^2\} \end{cases} \tag{5-14}$$

$$\text{with } \epsilon_k \in (0,1)\ \sum_{k=0}^{\infty} \epsilon_k = \sigma < \infty \text{ and } \rho^k \to \rho^\infty \leq \infty.$$

It is worth to note that (c) is first introduced in [110] to support linear convergence in partnership with strong variational sufficiency. It follows from [20, Theorem 3.1] that

[1]A function $\psi : \mathbb{X} \to \mathbb{R}$ is said to be strongly convex with modulus $s > 0$ if $\psi - \frac{s}{2}\|\cdot\|^2$ is convex.

Algorithm 2 Augmented Lagrangian method for solving (2-48)

Require: Let $(x^0, Y^0) \in \mathbb{X} \times \mathbb{S}^n$, $\rho^0 > \bar{\rho}$. Set $k := 0$.

1: If (x^k, Y^k) satisfies a suitable termination criterion: STOP.

2: Compute x^{k+1} such that

$$x^{k+1} \approx \bar{x}^{k+1} = \arg\min_{x \in \mathcal{X}} \mathcal{L}_{\rho^k}(x, Y^k). \tag{5-13}$$

3: Update multipliers to

$$Y^{k+1} := Y^k + \widetilde{\rho}^k \left[G(x^{k+1}) - \Pi_{\mathbb{S}^n_+}(G(x^{k+1}) + Y^k/\rho^k) \right],$$

where $\widetilde{\rho}^k = \rho^k - \bar{\rho}$.

4: Update nondecreasing positive sequence ρ^{k+1} according to certain rules.

5: Set $k \leftarrow k + 1$ and go to 1.

(5-14) can be replaced by

$$\sqrt{\widetilde{\rho}^k} \|(\mathcal{L}_{\rho^k})'_x(x^{k+1}, Y^k)\| \leq \begin{cases} (a)\ \epsilon'_k, \\ (b)\ \epsilon'_k \min\{1, \|\widetilde{\rho}^k (\mathcal{L}_{\rho^k})'_Y(x^{k+1}, Y^k)\|\}, \\ (c)\ \epsilon'_k \min\{1, \|\widetilde{\rho}^k (\mathcal{L}_{\rho^k})'_Y(x^{k+1}, Y^k)\|^2\}, \end{cases} \tag{5-15}$$

where $\epsilon'_k = \epsilon_k \sqrt{s}$, as the strong convexity of $\mathcal{L}_{\rho^k}(\cdot, Y^k)$ with modulus s guarantees $\mathcal{L}_{\rho^k}(x^{k+1}, Y^k) - \inf_{\mathcal{X}} \mathcal{L}_{\rho^k}(\cdot, Y^k) \leq \frac{1}{2s} \|(\mathcal{L}_{\rho^k})'_x(x^{k+1}, Y^k)\|^2$.

By applying the local duality, which comes from the strong variational sufficient condition through [55, Theorem 1], we suppose $S_{KKT}(0,0) \cap \mathcal{X} \times \mathcal{Y} \neq \emptyset$, where $\mathcal{X} \times \mathcal{Y}$ is the neighborhood mentioned at the beginning of this section. As mentioned in [20, page 9-10], we can define the associated local primal and dual problems to $\mathcal{X} \times \mathcal{Y}$ in the following sense. The associated local primal problem is

$$\min\ \widehat{f}(x) := \sup_{Y \in \mathcal{Y}} \mathcal{L}_{\bar{\rho}}(x, Y) \quad \text{over} \quad x \in \mathcal{X}. \tag{5-16}$$

The associated local dual problem is

$$\max\ \widehat{h}(Y) := \inf_{x \in \mathcal{X}} \mathcal{L}_{\bar{\rho}}(x, Y) \quad \text{over} \quad Y \in \mathcal{Y}. \tag{5-17}$$

In [20, Theorem 2.1], the author reveals the connection between problems (2-48), (5-16) and (5-17), which is of great use in the following discussion.

5.3.1 Convergence analysis of ALM

By taking advantage of Theorem 5.7 and the genius work [20], we can view the convergence analysis of ALM for NLSDP as a direct extension. Recalling the definition of bounded linear regularity given in Definition 3.33. A sufficient condition to guarantee the property of bounded linear regularity is established in [96, Corollary 3]. Denote

$$\mathcal{G}_1(\bar{x}) = \{Y \in \mathbb{S}^n \mid f'(\bar{x}) + G'(\bar{x})^* Y = 0\} \ \text{ and } \ \mathcal{G}_2(\bar{x}) = \{Y \in \mathbb{S}^n \mid Y \in \mathcal{N}_{\mathbb{S}^n_+}(G(\bar{x}))\}. \tag{5-18}$$

It is easy to see that $\mathcal{G}_1(\bar{x})$ is a polyhedron and $\mathcal{G}_2(\bar{x})$ is convex.

For a stationary point $\bar{x}$ of NLSDP (2-48), define

$$\xi(G'(\bar{x})) = \min\{\|G'(\bar{x})^*\eta\| : \eta \in \mathcal{G}_1(\bar{x})^{\perp},\ \|\eta\| = 1\}. \tag{5-19}$$

The following condition is adopted from [20] as it is essential in verifying the convergence rate.

Condition 1. There exist $b > 0$ and $\varepsilon > 0$ such that the local dual problem (5-17) satisfies $\widehat{h}(Y) \leq \max_{\mathcal{Y}} \widehat{h} - b\,\mathrm{dist}^2(Y, \mathcal{H})$ when $\|Y - \overline{Y}\| < \varepsilon$, where $\mathcal{Y}$ is the closed convex neighborhood of $\overline{Y}$ given at the beginning of Section 5.3, $\mathcal{H} := \arg\max_{\mathcal{Y}} \widehat{h}$.

The following result, which is originally proposed in [20, Theorem 4.2] for polyhedral case, provides the sufficiency of Condition 1. As pointed out by [20], for non-polyhedral cases such as NLSDP, similar results can be obtained under the bounded linear regularity assumption [96]. For the sack of complementary, we include the proof here. Moreover, it follows from [20, page 34] that Condition 1 trivially holds if $G'(\bar{x}) = 0$.

Proposition 5.9. *Let $\bar{x} \in \mathbb{X}$ be a stationary point to the NLSDP* (2-48) *and $\overline{Y} \in \mathcal{M}(\bar{x})$, where $\mathcal{M}(\bar{x})$ is given in* (5-7)*. Suppose $G'(\bar{x}) \neq 0$. If strong SOSC with respect to $(\bar{x}, \overline{Y})$ holds and the collection $\{\mathcal{G}_1(\bar{x}), \mathcal{G}_2(\bar{x})\}$ is boundedly linearly regular, where $\mathcal{G}_1(\bar{x})$ and $\mathcal{G}_2(\bar{x})$ are given in* (5-18)*, then we have $\xi(G'(\bar{x})) > 0$, where $\xi(G'(\bar{x}))$ is defined by* (5-19) *and Condition 1 holds for*

$$b = \frac{\kappa}{a_2 + a_1} \quad \textit{with} \quad a_2 = b_0^{-1} + 2\bar{\rho} \quad \textit{and} \quad a_1 = \frac{2\|L''_{xx}(\bar{x}, \overline{Y}) + \hat{\rho} I\|}{\xi(G'(\bar{x}))^2}, \tag{5-20}$$

where $\hat{\rho} = \lambda_{\max}(\bar{\rho} G'(\bar{x})^* G'(\bar{x})) + \varepsilon$ and b_0 is the quadratic parameter κ given in [98, Proposition 2.1].

Proof. Let $\mu = \max_{\mathcal{Y}} \widehat{h}$. It follows from [20, Theorem 2.1] that $\mathcal{M}(\bar{x}) \cap \mathrm{int}\,\mathcal{Y} = \mathcal{H}$, where $\mathcal{M}(\bar{x})$ is given in (3-92). Shrinking the neighborhood $\mathcal{U}$ of $\overline{Y}$ if necessary, we have for all $Y \in \mathcal{U}$, $\mathrm{dist}(Y, \mathcal{M}(\bar{x})) = \mathrm{dist}(Y, \mathcal{H})$. We want to identify b such that on a neighborhood of $\overline{Y}$,

$$\mu - \widehat{h}(Y) \geq b\,\mathrm{dist}^2(Y, \mathcal{M}(\bar{x})).$$

By using the definition of bounded linear regularity, we have there exists a constant $\kappa > 0$ such that

$$\mathrm{dist}(Y, \mathcal{M}(\bar{x})) \leq \kappa \max\{\mathrm{dist}(Y, \mathcal{G}_1(\bar{x})), \mathrm{dist}(Y, \mathcal{G}_2(\bar{x}))\} \quad \forall\, Y \in \mathcal{U}.$$

It follows that

$$\mathrm{dist}^2(Y, \mathcal{M}(\bar{x})) \leq \kappa^2\big(\mathrm{dist}^2(Y, \mathcal{G}_1(\bar{x})) + \mathrm{dist}^2(Y, \mathcal{G}_2(\bar{x}))\big) \quad \forall\, Y \in \mathcal{U}.$$

Then we only need to prove there exist some $a_1 > 0$ and $a_2 > 0$ such that

$$\mu - \widehat{h}(Y) \geq a_1^{-1}\,\mathrm{dist}^2(Y, \mathcal{G}_1(\bar{x})) \quad \text{and} \quad \mu - \widehat{h}(Y) \geq a_2^{-1}\,\mathrm{dist}^2(Y, \mathcal{G}_2(\bar{x})). \tag{5-21}$$

Firstly, we shall show that the first inequality in (5-21) holds. According to the definition of $\widehat{h}$, we obtain that

$$\mu - \widehat{h}(Y) = \max_{\bar{x}+w\in\mathcal{X}} \left\{ \bar{g}_{\bar{\rho}}^*\big(Y + \bar{\rho}\overline{G}(w)\big) - \big(\bar{f}(w) + \langle \overline{G}(w), Y\rangle + \frac{\bar{\rho}}{2}\|\overline{G}(w)\|^2\big)\right\}, \tag{5-22}$$

Since $\min \bar{g}^* = 0$, it follows from (5-22) that for $k(w, Y) = \bar{f}(w) + \langle Y, \overline{G}(w)\rangle + \frac{\bar{\rho}}{2}\|\overline{G}(w)\|^2$,

$$\mu - \widehat{h}(Y) \geq -\min_{\bar{x}+w\in\mathcal{W}} k(w, Y). \tag{5-23}$$

It has $k(0, Y) = 0$ and $\nabla_w k(0, Y) = \nabla f(\bar{x}) + \nabla G(\bar{x})^* Y = \nabla G(\bar{x})^*(Y - \overline{Y})$, where the last equality comes from $\nabla f(\bar{x}) + \nabla G(\bar{x})^*\overline{Y} = \nabla_x L(\bar{x}, \overline{Y}) = 0$. Also, we have $k''_{ww}(0, Y) = L''_{xx}(\bar{x}, Y) + \bar{\rho} G'(\bar{x})^* G'(\bar{x})$. By the twice differentiable continuity of $L(x, Y)$ on (x, Y), we know that there exists some ε, $\theta > 0$ such that for all $w \in \mathbb{X}$, $\langle w, (L''_{xx}(\bar{x}, \overline{Y}) + \varepsilon I - L''_{xx}(\bar{x}, Y))w\rangle \geq 0$ for all $\|Y - \overline{Y}\| \leq \theta$. Moreover, we can pick sufficient large ρ' such that for all $w \in \mathbb{X}$, $\langle w, (\rho' I - \bar{\rho} G'(\bar{x})^* G'(\bar{x}))w\rangle \geq 0$. It follows that $\forall \rho > \rho' + \varepsilon, \exists\, \mu > 0$ such that for mapping $H_\rho = L''_{xx}(\bar{x}, \overline{Y}) + \rho I$, for all $\|w\| \leq \mu$,

$$k(w, Y) \leq \langle G'(\bar{x})^*(Y - \overline{Y}), w\rangle + \frac{1}{2} w^T H_\rho w \quad \text{when } \|Y - \overline{Y}\| \leq \theta. \tag{5-24}$$

Combining (5-23) and (5-24) together, we know that when $\|Y - \overline{Y}\| \leq \theta$,

$$\begin{aligned} \mu - \widehat{h}(Y) &\geq -\min_{\|w\|\leq\mu} \left\{ \langle G'(\bar{x})^*(Y - \overline{Y}), w\rangle + \frac{1}{2} w^T H_\rho w \right\} \\ &\geq -\min_{\|w\|\leq\mu} \left\{ \langle G'(\bar{x})^*(Y - \overline{Y}), w\rangle + \frac{1}{2}\|H_\rho\|\|w\|^2 \right\}. \end{aligned}$$

It is easy to see the minimizer is $w = -\|H_\rho\|^{-1} G'(\bar{x})^*(Y - \overline{Y})$ when $\|H_\rho\|^{-1}\|G'(\bar{x})^*(Y - \overline{Y})\| \leq \mu$, which can be obtained when θ is sufficiently small. It follows that

$$\begin{aligned} -\min_{\|w\|\leq\mu} \left\{ \langle G'(\bar{x})^*(Y - \overline{Y}), w\rangle + \frac{1}{2}\|H_\rho\|\|w\|^2 \right\} &= \frac{1}{2\|H_\rho\|}\|G'(\bar{x})^*(Y - \overline{Y})\|^2 \\ &\geq \frac{1}{2\|H_\rho\|}\xi(G'(\bar{x}))^2 \mathrm{dist}^2(Y, \mathcal{G}_1(\bar{x})), \end{aligned}$$

where the last inequality comes from $\|Y - \overline{Y}\| = \|\Pi_{\mathcal{G}_1(\bar{x})^\perp}(Y)\| = \mathrm{dist}(Y, \mathcal{G}_1(\bar{x}))$. Pick $\hat{\rho} = \rho' + \varepsilon'$ with $\varepsilon' > \varepsilon$. Then we have established the first inequality in (5-21) with a_1 as in (5-20).

By a similar manner as the proof of [20, Theorem 4.2], we obtain the second inequality in (5-21) for $a_2 = b_0^{-1} + 2\bar{\rho}$. We omit it here for simplicity. Thus we have completed the proof. □

Next, we shall present the local convergence result of ALM for NLSDP. We say a sequence $\{z^k\} > 0$ converges Q-linearly to 0 at a rate c if $\limsup_{k\to\infty} \frac{z^{k+1}}{z^k} \leq c < \infty$. When

$c = 0$, we say $\{z^k\}$ converges Q-superlinearly to 0. Moreover, a sequence $\{y^k\} > 0$ converges R-linearly to 0 at a rate c if $y^k \leq z^k$ with $\{z^k\} > 0$ converges Q-linearly to 0 at that rate. The following closeness condition relative to the closed convex set $\mathcal{M}(\bar{x})$ is taken from [20, Theorem 2.2]. Recall that $\mathcal{X}$, $\mathcal{Y}$ are the closed convex neighborhood of $\bar{x}$, $\overline{Y}$ mentioned in the beginning of Section 5.3.

Condition 2. We say the initial point Y^0 and $\sigma > 0$ in (5-14) satisfies the following closedness condition relative to the closed convex set $\mathcal{M}(\bar{x})$ (5-7) if there exists $\eta > \mathrm{dist}(Y^0, \mathcal{M}(\bar{x})) + \sigma$ such that

$$\left\{Y \mid \|Y - Y^0\| \leq 3\eta\right\} \subset \mathcal{Y}.$$

It is easy to see that this condition indicates Y^0 to be sufficiently close to $\mathcal{M}(\bar{x})$ and $\overline{Y}$. Also, the computations of subproblems at each iteration need to be sufficiently accurate. If subproblems are solved exactly, a sufficient condition for the above one is that there exists $\eta > 0$ such that $\mathcal{B}_\eta(Y^0) \subseteq \mathrm{int}\,\mathcal{Y}$ and $\mathrm{dist}(Y^0, \mathcal{M}(\bar{x})) \leq \eta/3$.

By using Theorem 5.7 and [20, Theorem 2.2, 2.3, 3.1, 3.2], we immediately get the following local convergence result of ALM for solving NLSDP.

Theorem 5.10. *Let $\bar{x} \in \mathcal{X}$ be a stationary point to the NLSDP* (2-48) *and $\overline{Y} \in \mathcal{M}(\bar{x})$. Suppose the strong SOSC* (2-51) *holds at $(\bar{x}, \overline{Y})$. Let the initial point Y^0 and σ in* (5-14) *satisfy Condition 2. Suppose the set $\{x \mid -(\mathcal{L}_{\bar{\rho}})'_x(x, Y) \in \mathcal{N}_{\mathcal{X}}(x)\}$ is nonempty and bounded when $Y \in \mathrm{int}\,\mathcal{Y}$.*

(i) Under stopping criterion (5-15 a), we have the sequence $\{Y^k\}$ converges within $\mathrm{int}\,\mathcal{Y}$ to a particular $\widehat{Y} \in \mathrm{int}\,\mathcal{Y}$. Moreover, both x^k and $\bar{x}^k$ converge to $\bar{x}$.

(ii) Stopping criterion in (i) is strengthened into (5-15 b) and suppose $\{\mathcal{G}_1(\bar{x}), \mathcal{G}_2(\bar{x})\}$ is also boundedly linearly regular, where $\mathcal{G}_1(\bar{x})$ and $\mathcal{G}_2(\bar{x})$ are given in (5-18). *Then we have $\mathrm{dist}(Y^k, \mathcal{M}(\bar{x})) \to 0$ with*

$$\mathrm{dist}(Y^{k+1}, \mathcal{M}(\bar{x})) \leq \frac{1}{\sqrt{1 + b^2(\rho^\infty)^2}}\mathrm{dist}(Y^k, \mathcal{M}(\bar{x}))$$

and $\bar{x}^k \to \bar{x}$ with

$$\|\bar{x}^k - \bar{x}\| \leq \frac{1}{s}\mathrm{dist}(Y^k, \mathcal{M}(\bar{x})),$$

where $\bar{x}^k$ is the exact solution of subproblems in (5-13)*, b is given in Condition 1 and s is the strong convexity modulus of $\mathcal{L}_\rho(x, Y)$ on x with $\rho \geq \bar{\rho}$.*

(iii) If stopping criterion in (ii) is strengthened into (5-15 c), we have $Y^k \to \widehat{Y}$ with

$$\|Y^{k+1} - \widehat{Y}\| \leq \frac{1}{\sqrt{1 + b^2(\rho^\infty)^2}}\|Y^k - \widehat{Y}\|.$$

Moreover, if the stopping criterion is further supplemented by

$$\|(\mathcal{L}_{\rho^k})'_x(x^{k+1}, Y^k)\| \leq c\|Y^{k+1} - Y^k\| \text{ for some fixed } c,$$

we have $x^k \to \bar{x}$ *with*

$$\|x^k - \bar{x}\| \le p\|Y^k - \widehat{Y}\|$$

for some $p > 0$.

As illustrated in [20, Theorem 2.3], the condition "set $\{x \mid -(\mathcal{L}_{\bar{\rho}})'_x(x,Y) \in \mathcal{N}_{\mathcal{X}}(x)\}$ is nonempty and bounded when $Y \in \operatorname{int}\mathcal{Y}$" can be reduced to "the existence of $Y \in \operatorname{int}\mathcal{Y}$ such that $\{x \mid -(\mathcal{L}_{\bar{\rho}})'_x(x,Y) \in \mathcal{N}_{\mathcal{X}}(x)\}$ being nonempty and bounded". This condition is trivially satisfied as $\mathcal{X}$ is a neighborhood of $\bar{x}$ and $(\bar{x},\overline{Y})$ belongs to the set. It is worth to note that as illustrated in [20, Theorem 2.3], the result in Theorem 5.10 (i) only requires variational sufficiency. Under variational sufficiency, the convergence of x^k to $\bar{x}$ can not be obtained. Meanwhile, under the stopping criterion (5-15 b), we may not be able to obtain from Theorem 5.10 (ii) the convergence of the primal iteration sequence $\{x^k\}$, since the exact solution $\bar{x}^k$ of subproblems in (5-13) is unknown in practice. Next, we shall show that the KKT residual of NLSDP (2-48) also converges R-linearly, which means that the KKT residual can be used as a verifiable stopping criterion for ALM. Its proof sketch is inspired by [36, Theorem 2].

Proposition 5.11. *Suppose the conditions in Theorem 5.10 (ii) hold. Define the following residual function*

$$R(x,Y) := \|L'_x(x,Y)\| + \|G(x) - \Pi_{\mathbb{S}^n_+}(G(x)+Y)\|. \tag{5-25}$$

Then, for k sufficiently large, if $\sqrt{\bar{\rho}^k}\epsilon_k < 1$, *we have there exists* $a > 0$ *such that*

$$R(x^{k+1},Y^{k+1}) \le c^k \operatorname{dist}(Y^k, \mathcal{M}(\bar{x}))$$

with $c^k = (\epsilon'_k\sqrt{\bar{\rho}^k} + (1+\bar{\rho}+a\bar{\rho})(\bar{\rho}^k)^{-1})(1-\sqrt{\bar{\rho}^k}\epsilon_k)^{-1}$.

Proof. Let $\widehat{Y}^{k+1} = Y^k + \rho^k(G(x^{k+1}) - \Pi_{\mathbb{S}^n_+}(G(x^{k+1}) + (\rho^k)^{-1}Y^k))$ and $z^{k+1} = \Pi_{\mathbb{S}^n_+}(G(x^{k+1}) + (\rho^k)^{-1}Y^k)$. We have $G(x^{k+1}) - z^{k+1} = (\bar{\rho}^k)^{-1}(Y^{k+1} - Y^k)$. It follows directly from (5-15 b) that for each k,

$$\|L'_x(x^{k+1},\widehat{Y}^{k+1})\| = \|(\mathcal{L}_{\rho^k})'_x(x^{k+1},Y^k)\| \le \epsilon'_k\sqrt{\bar{\rho}^k}\|Y^{k+1} - Y^k\|.$$

Then there exists $a > 0$ such that

$$\begin{aligned}
&\|L'_x(x^{k+1},Y^{k+1})\| = \|L'_x(x^{k+1},\widehat{Y}^{k+1}) + G'(x^{k+1})^*(Y^{k+1} - \widehat{Y}^{k+1})\| \\
&\le \|L'_x(x^{k+1},\widehat{Y}^{k+1})\| + \|G'(x^{k+1})^*(Y^{k+1} - \widehat{Y}^{k+1})\| \le \|L'_x(x^{k+1},\widehat{Y}^{k+1})\| + a\|Y^{k+1} - \widehat{Y}^{k+1}\| \\
&= \|L'_x(x^{k+1},\widehat{Y}^{k+1})\| + a\bar{\rho}\|G(x^{k+1}) - \Pi_{\mathbb{S}^n_+}(G(x^{k+1}) + (\rho^k)^{-1}Y^k)\| \\
&= \|L'_x(x^{k+1},\widehat{Y}^{k+1})\| + a\bar{\rho}(\bar{\rho}^k)^{-1}\|Y^{k+1} - Y^k\| \le (\epsilon'_k\sqrt{\bar{\rho}^k} + a\bar{\rho}(\bar{\rho}^k)^{-1})\|Y^{k+1} - Y^k\|,
\end{aligned} \tag{5-26}$$

where the second inequality follows from the twice differentiable continuity and the boundedness of x^k obtained from Theorem 5.10(a). It can be verified directly from [1, Theorem 2.26] that $\widehat{Y}^{k+1} \in \partial\delta_{\mathbb{S}^n_+}(z^{k+1})$ and

$$\begin{aligned}
&\|G(x^{k+1}) - \Pi_{\mathbb{S}^n_+}(G(x^{k+1}) + \widehat{Y}^{k+1})\| \\
&= \|G(x^{k+1}) - \Pi_{\mathbb{S}^n_+}(G(x^{k+1}) + \widehat{Y}^{k+1})\| - \|z^{k+1} - \Pi_{\mathbb{S}^n_+}(\widehat{Y}^{k+1} + z^{k+1})\| \\
&\le \|G(x^{k+1}) - \Pi_{\mathbb{S}^n_+}(G(x^{k+1}) + \widehat{Y}^{k+1}) - (z^{k+1} - \Pi_{\mathbb{S}^n_+}(\widehat{Y}^{k+1} + z^{k+1}))\| \\
&\le \|G(x^{k+1}) - z^{k+1}\| = (\widetilde{\rho}^k)^{-1}\|Y^{k+1} - Y^k\|.
\end{aligned}$$

It then follows that

$$\begin{aligned}
&\|G(x^{k+1}) - \Pi_{\mathbb{S}^n_+}(G(x^{k+1}) + Y^{k+1})\| \\
&\le \|G(x^{k+1}) - \Pi_{\mathbb{S}^n_+}(G(x^{k+1}) + \widehat{Y}^{k+1})\| + \|\Pi_{\mathbb{S}^n_+}(G(x^{k+1}) + \widehat{Y}^{k+1}) - \Pi_{\mathbb{S}^n_+}(G(x^{k+1}) + Y^{k+1})\| \\
&\le (\widetilde{\rho}^k)^{-1}\|Y^{k+1} - Y^k\| + \|\widehat{Y}^{k+1} - Y^{k+1}\| \\
&= (\widetilde{\rho}^k)^{-1}\|Y^{k+1} - Y^k\| + \bar{\rho}\|G(x^{k+1}) - \Pi_{\mathbb{S}^n_+}(G(x^{k+1}) + (\rho^k)^{-1}Y^k)\| \\
&= (\widetilde{\rho}^k)^{-1}\|Y^{k+1} - Y^k\| + \bar{\rho}(\widetilde{\rho}^k)^{-1}\|Y^{k+1} - Y^k\| = (1 + \bar{\rho})(\widetilde{\rho}^k)^{-1}\|Y^{k+1} - Y^k\|. \quad (5\text{-}27)
\end{aligned}$$

Combining (5-26) and (5-27) together, we obtain

$$R(x^{k+1}, Y^{k+1}) \le (\epsilon_k'\sqrt{\widetilde{\rho}^k} + (1 + \bar{\rho} + a\bar{\rho})(\widetilde{\rho}^k)^{-1})\|Y^{k+1} - Y^k\|. \quad (5\text{-}28)$$

Then we will prove $\|Y^{k+1} - Y^k\| \le \mathrm{dist}(Y^k, \mathcal{M}(\bar{x}))$. Let $P_k(Y^k) = \arg\max\{\widehat{h}(Y) - \frac{1}{2\widetilde{\rho}^k}\|Y - Y^k\|^2\}$. Then, we have

$$\begin{aligned}
&\|Y^{k+1} - Y^k\| \le \|Y^{k+1} - P_k(Y^k)\| + \|P_k(Y^k) - Y^k\| \\
&\le (2\widetilde{\rho}^k(\mathcal{L}_{\rho^k}(x^{k+1}, Y^k) - \inf_{\mathcal{X}} \mathcal{L}_{\rho^k}(\cdot, Y^k)))^{1/2} + \|P_k(Y^k) - Y^k\| \\
&\le \sqrt{\widetilde{\rho}^k}\epsilon_k\|Y^{k+1} - Y^k\| + \|P_k(Y^k) - Y^k\|, \quad (5\text{-}29)
\end{aligned}$$

where the second inequality follows from [20, (2.19)] and the last one follows from (5-15 b) and the paragraph under (5-15). By using [36, Proposition 1(b), the proof of Lemma 3] and [20, Theorem 2.1], we have

$$\|P_k(Y^k) - Y^k\| \le \mathrm{dist}(Y^k, \mathcal{M}_D), \quad (5\text{-}30)$$

where $\mathcal{M}_D$ denotes the solution set of (5-17). It follows from [20, Theorem 2.1] that $\mathrm{dist}(Y^k, \mathcal{M}_D) = \mathrm{dist}(Y^k, \mathcal{M}(\bar{x}) \cap \mathcal{Y}) = \mathrm{dist}(Y^k, \mathcal{M}(\bar{x}))$ since for k sufficiently large, $Y^k \in \mathrm{int}\,\mathcal{Y}$. Based on (5-28)-(5-30), we obtain that for k sufficiently large,

$$R(x^{k+1}, Y^{k+1}) \le (\epsilon_k'\sqrt{\widetilde{\rho}^k} + (1 + \bar{\rho} + a\bar{\rho})(\widetilde{\rho}^k)^{-1})(1 - \sqrt{\widetilde{\rho}^k}\epsilon_k)^{-1}\mathrm{dist}(Y^k, \mathcal{M}(\bar{x})).$$

This completes the proof. □

Remark 5.2. We compare our results with existing ALM convergence results for non-convex non-polyhedral problems. [51] justified the primal-dual linear convergence of

ALM under SOSC and strong Robinson constraint qualification (SRCQ) for $\mathcal{C}^2$-cone reducible constrained problems, which include NLSDP and NLSOC. [50] proved the convergence rate of NLSDP under strong SOSC together with nondegeneracy. Results in the former chapter do provide the linear rate under SOSC and semi-isolated calmness of the KKT pair without requiring the multiplier to be unique. However, some other assumptions are also needed. In this chapter, to obtain the Q-linear convergence for multiplier and R-linear convergence for primal variable, we assume strong SOSC (Definition 3.20), which is much stronger than the aforementioned SOSC. However, we do not assume any restriction on the dual variable.

In [20, Example 5.3], the author also studied the ALM convergence for second order cone programming when $G(\bar{x}) \neq 0$. Moreover, [52] gives the primal-dual linear convergence for NLSOC when the multiplier is unique while they only require SOSC instead of strong SOSC. By using Proposition 5.5, without any constraint qualification, the local convergence results of ALM for NLSOC can be obtained immediately.

5.3.2 Solving ALM subproblem: semi-smooth Newton-CG method

Although the convergence properties of ALM have been established, it is equally important to study how to solve the subproblem (5-13). To guarantee the convergence rate of ALM, we need to employ the stopping criterion (5-15 b or c). To meet the requirement of this stopping criterion, we consider using the semi-smooth Newton-CG method to solve the subproblem as it possesses the quadratic convergence rate under suitable conditions.

It follows from the proposed characterization of strong variational sufficiency (Theorem 5.7) that for NLSDP (2-48) if the strong SOSC (2-51) holds then there exists a neighborhood $\mathcal{B}_r(\bar{x}, \overline{Y})$ of $(\bar{x}, \overline{Y})$ such that for all $(x, Y) \in \mathcal{B}_r(\bar{x}, \overline{Y})$, every elements in $\pi_x \partial_B \big((\mathcal{L}_\rho)'_x\big)(x, Y)$ is positive definite. Thus, the validity of semi-smooth Newton-CG method to solve the subproblems (5-13) is ensured. The algorithm to solve the $(k+1)$-th subproblem is stated below (see Algorithm 3). The convergence analysis framework for Algorithm 3 is well-known [111, Theorem 3.2] (see also [26, Theorem 3.4, 3.5]). We omit the detailed proof here for simplicity.

Proposition 5.12. *Suppose the strong SOSC* (2-51) *holds at* $(\bar{x}, \overline{Y}) \in S_{KKT}(0,0)$. *Then, Algorithm 3 is well-defined and any accumulation point* $\hat{x}$ *of* $\{x_j\}$ *generated by Algorithm 3 is the optimal solution to the subproblem* (5-13). *Furthermore, suppose that at each step* j *CG terminates, i.e.,*

$$\|(V_j + \varepsilon_j I)d_j + (\mathcal{L}_{\rho^k})'_x(x_j, Y^k)\| \leq \nu_j.$$

Then the whole sequence $\{x_j\}$ *converges to* $\hat{x}$ *and*

$$\|x_{j+1} - \hat{x}\| = O(\|x_j - \hat{x}\|^{1+\tau}).$$

Algorithm 3 Semi-smooth Newton-CG method for solving (5-13)

Require: Set the initial point $x_0 = x^k$, where x^k is obtained from the k-th iteration of Algorithm 2. Let $\mu \in (0, 1/2)$, $\tau \in (0, 1]$, $\tau_1, \tau_2, \bar{\nu} \in (0, 1)$ and $\theta \in (0, 1)$. Set $j := 0$.

1: Given a maximum number of CG iterations $t_j > 0$ and compute

$$\nu_j = \min\{\bar{\nu}, \|(\mathcal{L}_{\rho^k})'_x(x_j, Y^k)\|^{1+\tau}\}.$$

2: Choose $W_j \in \partial\Pi_{\mathbb{S}^n_-}(G(x_j) + (\rho^k)^{-1}Y^k)$. Let $V_j = L''_{xx}(x_j, \Pi_{\mathbb{S}^n_-}(G(x_j) + (\rho^k)^{-1}Y^k) + \rho^k G'(x_j)^* W_j G'(x_j)$ and $\varepsilon_j = \tau_1 \min\{\tau_2, \|\nabla_x \mathcal{L}_{\rho^k}(x_j, Y^k)\|\}$. Apply the CG algorithm (CG(ν_j, t_j)) mentioned in [26, Algorithm 1] to find an approximate solution $d_j \in \mathbb{X}$ to

$$(V_j + \varepsilon_j I)d_j = -(\mathcal{L}_{\rho^k})'_x(x_j, Y^k)$$

such that

$$\|(V_j + \varepsilon_j I)d_j + (\mathcal{L}_{\rho^k})'_x(x_j, Y^k)\| \leq \nu_j.$$

3: Set $\zeta_j = \theta^{m_j}$, where m_j is the first non-negative integral number such that

$$\mathcal{L}_{\rho^k}(x_j + \theta^{m_j} d_j, Y^k) \leq \mathcal{L}_{\rho^k}(x_j, Y^k) + \mu\theta^{m_j}\langle(\mathcal{L}_{\rho^k})'_x(x_j, Y^k), d^j\rangle.$$

4: Set $x_{j+1} = x_j + \zeta_j d_j$ and $j = j + 1$.

Remark 5.3. It is worth noting that for convex NLSDP, the convergence result of ALM can be established under weaker conditions such that the optimal $\bar{x}$ can be non-unique (see [84, Theorem 20]). This implies that when the problem is reduced to convex case, Theorem 5.10 is much weaker than [84, Theorem 20] as it requires strong SOSC, which implies the local uniqueness of $\bar{x}$. However, the semi-smooth Newton algorithm may fail to solve the subproblem in the absence of strong SOSC since it is equivalent to the positive definiteness of the generalized Hessian of Newton equation of the subproblem (5-13) (Theorem 5.10). Thus, by Theorem 5.7, the strong SOSC seems to be not only sufficient to the local fast linear convergence rate of ALM, but also necessary for the invertibility of generalized Hessian of augmented Lagrangian function for NLSDP, which is crucial for the semi-smooth Newton CG method for solving the ALM subproblem (5-13).

5.4 Numerical experiments

Consider the following optimization problem

$$\begin{array}{ll} \min\limits_{X \in \mathbb{S}^n} & \frac{1}{2}\langle X, Q \circ X\rangle \\ \text{s.t.} & X \in \mathbb{S}^n_+ \\ & B \circ X = 0, \end{array} \tag{5-31}$$

where "$\circ$" denotes the Hadamard product of matrices, i.e., for any U and $V \in \mathbb{R}^{p\times q}$, $(U\circ V)_{ij} = U_{ij}V_{ij}$,

$$B = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 \\ 1 & \cdots & 1 & 1 \end{bmatrix} \quad \text{and} \quad Q = \begin{bmatrix} q & 1 & \cdots & 1 & 0 \\ 1 & q & \cdots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \cdots & q & 0 \\ 0 & 0 & \cdots & 0 & -1 \end{bmatrix}$$

with a given $q \geq n-1$. The Lagrangian function of (5-31) is given by

$$L(X,Y,Z) = \frac{1}{2}\langle X, Q\circ X\rangle + \langle X, Y\rangle + \langle B\circ X, Z\rangle.$$

It can be checked directly that $\overline{X} = 0$ is a local optimal solution with the multiplier $\overline{Y} = \mathrm{Diag}(0,\ldots,0,-1)$ and $\overline{Z} = \mathrm{Diag}(0,\ldots,0,1)$. In fact, the corresponding multiplier set of $\overline{X}$ is given by

$$\mathcal{M}(\overline{X}) = \{(Y,Z)\in\mathbb{S}^n\times\mathbb{S}^n \mid Y + B\circ Z = 0, Y\in\mathbb{S}^n_-\}.$$

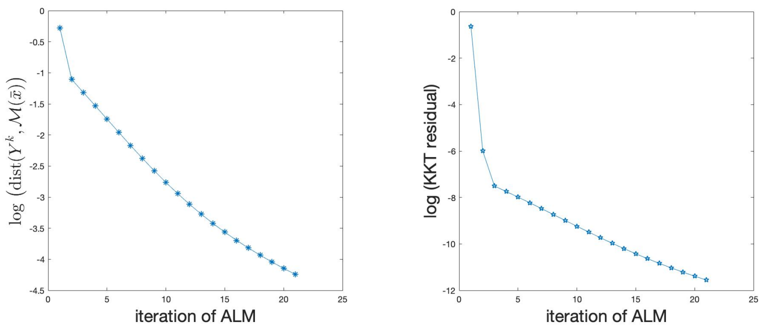
As the problem (4.1) mentioned in the beginning of this chapter, it follows from [104, Theorem 4.1] that the Robinson constraint qualification [105] does not hold at $\overline{X}$, due to the unboundedness of $\mathcal{M}(\bar{x})$. It is clear that $L''_{XX}(\overline{X},\overline{Y},\overline{Z}) = Q$, which is positive definite over $d \in \{d\in\mathbb{S}^n \mid B\circ d = 0, d\in \mathrm{aff}\,\mathcal{C}_{\mathbb{S}^n_+}(\overline{X},\overline{Y})\}$, implies the validity of strong SOSC (2-51). Also, boundedly linear regularity is satisfied at $(\overline{X},\overline{Y},\overline{Z})$ since $(Y,Z)\in\mathcal{G}_1(\overline{X})\cap\mathcal{G}_2(\overline{X})$ with $Y = \mathrm{Diag}(-1,\ldots,-1)$, $Z = \mathrm{Diag}(0,\ldots,0,1)$.

Next, we shall apply Algorithm 2 to solve problem (5-31) with different dimensions. The subproblem (5-13) is solved by Algorithm 3 where the exactness in (5-15) is chosen as $\epsilon'_k = 0.01\times(1/1.05)^{k-1}$ and the stopping criterion (5-15 b) is employed. The algorithm is stopped when the KKT residual $R(x^k, Y^k)$ defined in (5-25) is less than `1e-5`. The codes are implemented in Matlab (R2018a), and the numerical experiments are run under a 64-bit MacOS on an Intel Cores i5 2.3GHz CPU with 8GB memory. The following table (Table **??**) shows the numerical results of different dimensions of (5-31). Noting that the distance from (Y^k, Z^k) to $\mathcal{M}(\overline{X})$ is difficult to compute, we

n	q	iteration	KKT residual	cpu(s)
3	2	8	8.27e-06	0.22
100	200	11	8.98e-06	3.11
1000	1500	21	9.57e-06	1083.34

表 5-1 Numerical results of semi-smooth Newton-CG based ALM for problem (5-31)**.**

use the following alternative $\mathrm{dist}((Y^k,Z^k),\mathcal{M}(\overline{X})) = O(\|Y^k + B\circ Z^k\| + \|\Pi_{\mathbb{S}^n_+}(Y^k)\|)$ as boundedly linear regularity is satisfied. The detail iterative performance of ALM for solving problem (5-31) with $n = 1000$ and $q = 1500$ is also reported in Figure



(a) (b)

图 5-2 Semi-smooth Newton-CG based ALM for problem (5-31) **with** $n = 1000$ **and** $q = 1500$**.**

5-2. It can be seen from Figure 5-2 (a) that although the problem is nonconvex and the multiplier set is not a singleton, ALM also possesses a linear rate of convergence of $\mathrm{dist}((Y, Z), \mathcal{M}(\overline{X}))$ for the validity of strong variational sufficiency. Meanwhile, the right one shows that the KKT residual also converges to 0 as the algorithm proceeds. Moreover, by strong variational sufficiency, we know from Theorem 5.7 that the Hessian matrix V_j in Algorithm 3 Step 2 is always positive definite when k is sufficiently large. Indeed, the minimum eigenvalue of the Hessian matrix V_j is positive (e.g., $\lambda_{\min}(V_j) \approx 1.01$ for the case $n = 3$ and $q = 2$; $\lambda_{\min}(V_j) \approx 1.26$ for the case $n = 100$ and $q = 200$). It is also worth to note that the distance from (Y^k, Z^k) to $(\overline{Y}, \overline{Z})$ does not converge to 0, which meets our theory as the limit point $(\widehat{Y}, \widehat{Z})$ in Theorem 5.10 may be different from the reference point $(\overline{Y}, \overline{Z})$.

It is well-known that the condition number of V_j in Algorithm 3 Step 2 is proportional to ρ^k. In fact, the estimation of the condition number of V_j can be obtained in a similar manner of [50, Lemma 10] as the smallest eigenvalue lies in a constant interval while the largest one lies in an interval which is propositional to ρ^k. Thus, in practice, we usually set an upper bound for ρ^k, although ρ^k can be infinity in theory.

It can be seen from Theorem 5.10 that the convergence analysis of Algorithm 2 requires a good starting point of the multiplier while no restriction to the initial primal variable x^0 is needed. To verify the convergence theory (Theorem 5.10), in the numerical experiments of (5-31) presented in Table 5-1, we choose X^0 randomly. (Y^0, Z^0) is chosen to be $(\overline{Y}, \overline{Z}) + \eta(P_1, P_2)$, where P_1, P_2 are symmetry matrices that are uniform randomly generated. Typically, η is chosen to be small, e.g., 0.1.

Next, we will discuss how to generate the initial point more practical. Condition 2 indicates that (Y^0, Z^0) should be sufficiently close to $(\overline{Y}, \overline{Z})$, at which strong SOSC is satisfied. How to find (Y^0, Z^0) satisfying Condition 2 for nonconvex optimization problem is very challenging. A natural idea is to apply first order methods to find a satisfactory initial point. Moreover, we know from Theorem 5.7 that a good starting point may be the one at which the positive definiteness of generalized Hessian is satisfied.

In practice, when the generalized Hessian is not positive definite, we can apply APG instead of the semismooth Newton to solve the ALM subproblem.

Chapter 6 Conclusions and future work

In this thesis, we focus on a class of optimization problem named composition matrix optimization problems, which involves a bunch of practical non-polyhedral optimization problems, e.g., NLSDP. Theoretically, we study several variational properties of it. We give an explicit characterization of the coderivative for the graph of subdifferential of composite piecewise affine function problems. And applying the characterization to depict the Lipschitzian full stability of CMatOP. As a by-product, we also give the equivalence between strong regularity and Lipschitzian full stability with nondegeneracy. Moreover, the equivalence between isolated calmness and SOSC with SRCQ for CMatOP is also established. A sufficient condition has also been provided for semi-isolated calmness. The above variational properties are all very critical in perturbation analysis and algorithm convergence analysis.

Practically, we explore the convergence behaviors of ALM for CMatOP, particularly for NLSDP. We have shown that the augmented Lagrangian method convergences linearly for NLSDP under certain conditions without requiring the uniqueness of the Lagrangian multiplier. During the establishment of ALM convergence, we obtain the uniform second expansion of the Moreau envelope of SDP and give several sufficient conditions for the semi-isolated calmness of S_{KKT}.

To overcome the weakness of the requirement of the abstract Assumption 1, we derive the equivalence between the strong variational sufficiency and the strong SOSC for NLSDP and NLSOC without requiring the uniqueness of multiplier or any other constraint qualification. By using the equivalence result, the local convergence property of ALM for NLSDP can be established under merely strong SOSC instead of plus it with any constraint qualification. As a direct application, we are able to show that the positive definiteness of the generalized Hessian of augmented Lagrangian function, which is critical in the use of semi-smooth Newton method for NLSDP, is satisfied under strong SOSC.

However, there are still many issues that deserve to be explored further. Firstly, how to explore the characterization of Lipschitizan full stability under weaker conditions, e.g., RCQ for CMatOPs is an interesting one. Also, it is nature to consider whether tilt stability is equivalent to some second order optimality condition. As for the ongoing works of Chapter 4, we are still working on providing sufficient (necessary) conditions to Assumption 1. Lastly, we still do not have a satisfactory characterization for variational sufficiency. Moreover, for the implementation of ALM for solving nonconvex cases, it is still very challenging for finding a good starting point, e.g., the one close enough to a KKT pair of a local optimal solution which satisfies the strong SOSC condition.

References

[1] Rockafellar R T, Wets R J B. Variational analysis [M]. Springer, Berlin, 1998.

[2] Candes E J, Tao T. The power of convex relaxation: near-optimal matrix completion [J]. IEEE Transactions on Information Theory, 2009, 56: 2053-2080.

[3] Recht B, M. Fazel P A P. Guaranteed minimum rank solutions to affine matrix equations via nuclear norm minimization [J]. SIAM Review, 2010, 52: 471-501.

[4] Cui Y, Ding C. Nonsmooth composite matrix optimization: Strong regularity, constraint nondegeneracy and beyond [J]. arXiv, 2019, 1907.13253.

[5] Bonnans J F, Shapiro A. Perturbation analysis of optimization problems [M]. Springer, New York, 2000.

[6] Francisco F, Pang J S. Finite-dimensional variational inequalities and complementarity problems [M]. Springer, New York, 2003.

[7] Klatte D, Kummer B. Nonsmooth equations in optimization: regularity, calculus, methods, and applications [M]. Kluwer Academic Publishers, 2002.

[8] Bonnans J, Cominetti R, Shapiro A. Sensitivity analysis of optimization problems under second order regular constraints [J]. Mathematics of Operations Research, 1998, 23: 806-831.

[9] Bonnans J F, Shapiro A. Optimization problems with perturbations: A guided tour [J]. SIAM review, 1998, 40: 228-264.

[10] Bonnans J, Cominetti R, Shapiro A. Second order optimality conditions based on parabolic second order tangent sets [J]. SIAM Journal on Optimization, 1999, 9: 466-492.

[11] Sun D F. The strong second order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications [J]. Mathematics of Operations Research, 2006, 31: 761-776.

[12] Chan Z X, Sun D F. Constraint nondegeneracy, strong regularity, and nonsingularity in semidefinite programming [J]. SIAM Journal on Numerical Analysis, 2008, 19: 370-396.

[13] Robinson S M. Strongly regular generalized equations [J]. Mathematics of Operations Research, 1980, 5: 43-62.

[14] Levy A B, Poliquin R A, Rockafellar R T. Stability of locally optimal solutions [J]. SIAM Journal on Optimization, 2000, 10: 580-604.

[15] Mordukhovich B S, Rockafellar R T, Sarabi M E. Characterizations of full stability in constrained optimization [J]. SIAM Journal on Optimization, 2013, 23: 1810-1849.

[16] Mordukhovich B S, Nghia T T A, Rockafellar R T. Full stability in finite-dimensional optimization [J]. Mathematics of Operations Research, 2015, 40: 226-252.

[17] Dontchev A L, Rockafellar R T. Characterizations of strong regularity for variational inequalities over polyhedral convex sets [J]. SIAM Journal on Optimization, 1996, 6: 1087-1105.

[18] Mordukhovich B S, Outrata J V, Sarabi M E. Full stability of locally optimal solutions in second-order cone programming [J]. SIAM Journal on Optimization, 2014, 24: 1581-1613.

[19] Shapiro A, Sun J. Some properties of the augmented lagrangian in cone constrained optimization [J]. Mathematics of Operations Research, 2004, 29: 479-491.

[20] Rockafellar R T. Convergence of augmented lagrangian methods in extensions beyond nonlinear programming [J]. Mathematical Programming, 2022, to appear.

[21] Hestenes M R. Multiplier and gradient methods [J]. Journal of Optimization Theory and Applications, 1969, 4: 303-320.

[22] Powell M J D. A method for nonlinear constraints in minimization problems [J]. In Fletcher, R. (ed.) Optimization, 1969, Academic, New York: 283-298.

[23] Rockafellar R T. A dual approach to solving nonlinear programming problems by unconstrained optimization [J]. Mathematical Programming, 1973, 5: 354-373.

[24] Tibshirani R. Regression shrinkage and selection via the lasso [J]. Journal of the Royal Statistical Society. Series B. Statistical Methodology, 1996, 58: 267-288.

[25] Yang L Q, Sun D F, Toh K C. Sdpnal+: a majorized semismooth newton-cg augmented lagrangian method for semidefinite programming with nonnegative constraints [J]. Mathematical Programming Computation, 2015, 7: 331-366.

[26] Zhao X Y, Sun D F, Toh K C. A newton-cg augmented lagrangian method for semidefinite programming [J]. SIAM Journal on Optimization, 2010, 20: 1737-1765.

[27] Li X D, Sun D F, Toh K C. Qsdpnal: A two-phase augmented lagrangian method for convex quadratic semidefinite programming [J]. Mathematical Programming Computation, 2018, 10: 703-743.

[28] Li X D, Sun D F, Toh K C. A highly efficient semismooth newton augmented lagrangian method for solving lasso problems [J]. SIAM Journal on Optimization, 2018, 28: 433-458.

[29] Rockafellar R T. The multiplier method of hestenes and powell applied to convex programming [J]. Journal of Optimization Theory and Applications, 1973, 12: 555-562.

[30] Tretyakov N V. A method of penalty estimates for convex programming problems [J]. Ékonomika i Matematicheskie Metody, 1973, 9: 525-540.

[31] Rockafellar R T. Augmented lagrangians and applications of the proximal point algorithm in convex programming [J]. Mathematics of Operations Research, 1976, 1: 97-116.

[32] Rockafellar R T. Monotone operators and the proximal point algorithm [J]. SIAM Journal on Control and Optimization, 1976, 14: 877-898.

[33] Pennanen T. Local convergence of the proximal point algorithm and multiplier methods without monotonicity [J]. Mathematics of Operations Research, 2002, 27: 170-191.

[34] Luque F J. Asymptotic convergence analysis of the proximal point algorithm [J]. SIAM Journal on Optimization, 1984, 22: 277-293.

[35] Robinson S M. Some continuity properties of polyhedral multifunctions [J]. SIAM Journal on Control and Optimization, 1981, 28: 206-214.

[36] Cui Y, Sun D F, Toh K C. On the r-superlinear convergence of the kkt residuals generated by the augmented lagrangian method for convex composite conic programming [J]. Mathematical Programming, 2019, 178: 381-415.

[37] Eckstein J, Silva P J S. A practical relative error criterion for augmented lagrangians [J]. Mathematical Programming, 2013, 141: 319-348.

[38] Bertsekas D. Constrained optimization and lagrange multipliers methods [M]. Academic Press, New York, 1982.

[39] Conn A R, Gould N I M, Toint P. A globally convergent augmented lagrangian algorithm for optimization with general constraints and simple bounds [J]. SIAM Journal on Numerical Analysis, 1991, 28: 545-572.

[40] Contesse-Becker L. Extended convergence results for the method of multipliers for non-strictly binding inequality constraints [J]. Journal of Optimization Theory and Applications, 1993, 79: 273-310.

[41] Ito K, Kunisch K. The augmented lagrangian method for equality and inequality constraints in hilbert spaces [J]. Mathematical Programming, 1990, 46: 341-360.

[42] Fernandez D, Solodov M V. Local convergence of exact and inexact augmented lagrangian methods under the second-order sufficient optimality condition [J]. SIAM Journal on Optimization, 2012, 22: 384-407.

[43] Hang N T V, Sarabi M E. Local convergence analysis of augmented lagrangian methods for piecewise linear-quadratic composite optimization problems [J]. SIAM Journal on Optimization, 2021, 31: 2665-2694.

[44] Dontchev A L, Rockafellar R T. Characterizations of lipschitz stability in nonlinear programming [J]. Mathematical Programming With Data Perturbations, 1997, Marcel Dekker, New York: 65-82.

[45] Izmailov A F, Kurennoy A S, Solodov M V. A note on upper lipschitz stability, error bounds, and critical multipliers for lipschitz-continuous kkt systems [J]. Mathematical Programming, 2013, 142: 591-604.

[46] Klatte D. Upper lipschitz behavior of solutions to perturbed $c^{1,1}$ programs [J]. Mathematical Programming, 2000, 88: 285-311.

[47] Mordukhovich B, Sarabi E. Critical multipliers in variational systems via second-order generalized differentiation [J]. Mathematical Programming, 2018, 169: 605-648.

[48] Golshtein E G, Tretyakov N V. Modified lagrangians and monotone maps in optimization [M]. Wiley, New York, 1989.

[49] Rockafellar R T. Lagrange multipliers and optimality [J]. SIAM Review, 1993, 35: 183-238.

[50] Sun D F, Sun J, Zhangn L W. The rate of convergence of the augmented lagrangian method for nonlinear semidefinite programming [J]. Mathematical Programming, 2008, 114: 349-391.

[51] Kanzow C, Steck D. Improved local convergence results for augmented lagrangian methods in c^2-cone reducible constrained optimization [J]. Mathematical Programming, 2019, 177: 425-438.

[52] Hang N T V, Mordukhovich B, Sarabi E. Augmented lagrangian method for second-order cone programs under second-order sufficiency [J]. Journal of Global Optimization, 2022, 82: 51-81.

[53] Rockafellar R T. Progressive decoupling of linkages in optimization and variational inequalities with elicitable convexity or monotonicity [J]. Set-valued and Variational Analysis, 2019, 27: 863-893.

[54] Rockafellar R T. Progressive decoupling of linkages in monotone variational inequalities and convex optimization [J]. Yokohama Publishers, 2019: 271-291.

[55] Rockafellar R T. Augmented lagrangians and hidden convexity in sufficient conditions for local optimality [J]. Mathematical Programming, 2022, to appear.

[56] Khanh P D, Mordukhovich B S, Phat V T. Variational convexity of functions and variational sufficiency in optimization [J]. arXiv: 2208.14399, 2022.

[57] Mordukhovich B S. Variational analysis and generalized differentiation, i: Basic theory [M]. Grundlehren Series (Fundamental Principles of Mathematical Sciences), Vol.330, Springer, New York, 2006.

[58] Clarke F H, Ledyaev Y S, Stern R J, et al. Nonsmooth analysis and control theory [M]. Springer, New York, 1998.

[59] Rockafellar R T. Convex analysis [M]. Vol. 36. Princeton university press, 1970.

[60] Clarke F H. Optimization and nonsmooth analysis [M]. John Wiley and Sons, New York, 1983.

[61] Dontchev A L, Rockafellar R T. Implicit functions and solution mappings: A view from variational analysis, 2nd edition [M]. Springer, Dordrecht, 2014.

[62] Ding C. An introduction to a class of matrix optimization problems [J]. Thesis, 2014, National University of Singapore.

[63] Ding C, Sun D F, Toh K C. Spectral operators of matrices [J]. Mathematical Programming, 2018, 168: 509-531.

[64] Francisco F, Pang J S. Finite-dimensional variational inequalities and complementarity problems [M]. Springer Science and Business Media, 2007.

[65] Ding C, Sun D F, Toh K C. Spectral operators of matrices: semismoothness and characterizations of the generalized jacobian [J]. SIAM Journal on Optimization, 2020, 30: 630-659.

[66] Mordukhovich B S, Nghia T T A. Second-order variational analysis and characterizations of tilt-stable optimal solutions in infinite-dimensional spaces [J]. Nonlinear Analysis, 2013, 86: 159-180.

[67] Fan K. On a theorem of weyl concerning eigenvalues of affine transformations [J]. Proceedings of the National Academy of Sciences of the United States of America, 1949, 35: 652-655.

[68] Lancaster P. On eigenvalues of matrices dependent on a parameter [J]. Numerische Mathematik, 1964, 6: 377-387.

[69] Torki M. Second-order directional derivatives of all eigenvalues of a symmetric matrix [J]. Nonlinear Anal-Theor, 2001, 46: 1133-1150.

[70] Lewis A S. Convex analysis on the hermitian matrices [J]. SIAM Journal on Optimization, 1996, 6: 164-177.

[71] Lewis A S. Derivatives of spectral functions [J]. Mathematics of Operations Research, 1996, 21: 576-588.

[72] Mordukhovich B S, Sarabi M E. Generalized differentiation of piecewise affine functions in second-order variational analysis [J]. Nonlinear Analysis-Theory, Methods and Applications, 2016, 132: 240-273.

[73] Davis C. All convex invariant functions of hermitian matrices [J]. Archiv der Mathematik, 1957, 8: 276-278.

[74] Lewis A S. The convex analysis of unitarily invariant matrix functions [J]. Journal of Convex Analysis, 1995, 2: 173-183.

[75] Ding C, Sun D F, Sun J, et al. Spectral operators of matrices [J]. Mathematical Programming, 2018, 168: 509-531.

[76] Ben-Tal A, Zowe J. Necessary and sufficient optimality conditions for a class of nonsmooth minimization problems [J]. Mathematical Programming, 1982, 24: 70-91.

[77] Wu B, Ding C, Sun D F, et al. On the moreau-yosida regularization of the vector k-norm related functions [J]. SIAM Journal on Optimization, 2014, 24: 766-794.

[78] Overton M L, Womersley R S. Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices [J]. Mathematical Programming, 1993, 62: 321-357.

[79] Sun D F, Sun J. Semismooth matrix valued functions [J]. Mathematics of Operations Research, 2002, 27: 150-169.

[80] Ding C, Sun D F, Ye J. First order optimality conditions for mathematical programs with semidefinite cone complementarity constraints [J]. Mathematical Programming, 2014, 147: 539-579.

[81] Mordukhovich B S, Sarabi M E. Critical multipliers in variational systems via second-order generalized differentiation [J]. Mathematical Programming, 2018, 169: 605-648.

[82] Sun D F, Sun J. Strong semismoothness of eigenvalues of symmetric matrices and its applications in inverse eigenvalue problems [J]. SIAM Journal on Numerical Analysis, 2003, 40: 2352-236.

[83] Robinson S M. Local structure of feasible sets in nonlinear programming, part ii: Nondegeneracy [J]. Mathematical Programming Study, 1984, 22: 217-230.

[84] Cui Y, Ding C, Zhao X Y. Quadratic growth conditions for convex matrix optimization problems associated with spectral functions [J]. SIAM Journal on Optimization, 2017, 27: 2332-2355.

[85] Mordukhovich B S, Nam N M, Nhi N T Y. Partial second-order subdifferentials in variational analysis and optimization [J]. Numerical Functional Analysis and Optimization, 2014, 35: 1113-1151.

[86] Robinson S M. Stability theory for systems of inequality constraints, part i: linear systems [J]. SIAM Journal on Numerical Analysis, 1975, 12: 754-769.

[87] Aubin J P. Full stability in finite-dimensional optimization [J]. Mathematics of Operations Research, 1994, 9: 87-111.

[88] Ye J, Yen X. Necessary optimality conditions for optimization problems with variational inequality constraints [J]. Mathematics of Operations Research, 1997, 22: 977-997.

[89] Ye J. Constraint qualification and necessary optimality conditions for optimization problems with variational inequality constraints [J]. SIAM Journal on Optimization, 2000, 10: 943-962.

[90] Gfrerer H, Klatte D. Lipschitz and hölder stability of optimization problems and generalized equations [J]. Mathematical Programming, 2016, 158: 35-75.

[91] Ding C, Sun D F, Zhang L W. Characterization of the robust isolated calmness for a class of conic programming problems [J]. SIAM Journal on Optimization, 2017, 27: 67-90.

[92] Han D R, Sun D F, Zhang L W. Linear rate convergence of the alternating direction method of multipliers for convex composite quadratic and semi-definite programming [J]. arxiv:1508.02134, 2015: 1-40.

[93] Izmailov A F, Solodov M V. Newton-type methods for optimization and variational problems [M]. Springer, New York, 2014.

[94] Mordukhovich B S, Sarabi M E. Criticality of lagrange multipliers in variational systems [J]. SIAM Journal on Optimization, 2019, 29: 425-438.

[95] Bauschke H H, Borwein J M. On projection algorithms for solving convex feasibility problems [J]. SIAM Review, 1996, 38: 367-426.

[96] Bauschke H H, Borwein J M, Li W. Strong conical hull intersection property, bounded linear regularity, jameson ' s property (g), and error bounds in convex optimization [J]. Mathematical Programming, 1999, 86: 135-160.

[97] Kanzow C, Steck D. On error bounds and multiplier methods for variational problems in banach spaces [J]. SIAM Journal on Control and Optimization, 2018, 56: 1716-1738.

[98] Cui Y, Sun D F, Toh K C. On the asymptotic superlinear convergence of the augmented lagrangian method for semidefinite programming with multiple solutions [J]. arXiv:1610.00875, 2016.

[99] Poliquin R A, Rockafellar R. Generalized hessian properties of regularized nonsmooth functions [J]. SIAM Journal on Optimization, 1996, 6: 1121-1137.

[100] Mohammadi A, Mordukhovich B S, Sarabi M E. Parabolic regularity in geometric variational analysis [J]. Transactions of the American Mathematical Society, 2021, 374: 1711-1763.

[101] Rudin W. Principles of mathematical analysis citation third edition [M]. McGraw-Hill, 1976.

[102] Zarantonello E H. Projections on convex sets in hilbert space and spectral theory: Part i. projections on convex sets: Part ii. spectral theory [J]. Contributions to Nonlinear Functional Analysis, 1971: 237-424.

[103] Robinson S M. Generalized equations and their solutions, part ii: Applications to nonlinear programming [J]. Mathematical Programming Study, 1982, 19: 200-221.

[104] Zowe J, Kurcyusz S. Regularity and stability for the mathematical programming problem in banach spaces [J]. Applied Mathematics and Optimization, 1979, 5: 49-62.

[105] Robinson S M. Stability theory for systems of inequality constraints, part ii: differentiable nonlinear systems [J]. SIAM Journal on Numerical Analysis, 1976, 13: 497-513.

[106] Bonnans J F, C H R. Perturbation analysis of second-order cone programming problems [J]. Mathematical Programming, 2005, 104: 205-227.

[107] Rockafellar R T. Variational convexity and the local monotonicity of subgradient mappings [J]. Vietnam Journal of Mathematics, 2019, 47: 547-561.

[108] Mordukhovich B S, Rockafellar R T. Second-order subdifferential calculus with applications to tilt stability in optimization [J]. SIAM Journal on Optimization, 2012, 22: 953-986.

[109] Urruty J B H, Strodiot J J, Nguyen V H. Generalized hessian matrix and second-order optimality conditions for problems with $c^{1,1}$ data [J]. Applied Mathematics and Optimization, 1984, 11: 43-56.

[110] Rockafellar R T. Advances in convergence and scope of the proximal point algorithm [J]. Journal of Nonlinear and Convex Analysis, 2021, 22: 2347-2375.

[111] Qi L, Sun J. A nonsmooth version of newton's method [J]. Mathematical Programming, 1993, 58: 353-367.

[112] Kong W W, Melo J G, Monteiro R D. Iteration-complexity of a proximal augmented lagrangian method for solving nonconvex composite optimization problems with nonlinear convex constraints [J]. Mathematics of Operations Research, 2022, https://doi.org/10.1287/moor.2022.1301.

Appendix 1 Proof of Theorem 3.13 $(i) \Rightarrow (ii)$

Consider

$$\begin{aligned} \min_x \quad & f(x, \bar{p}) + \theta(g_1(x, \bar{p})) \\ \text{s.t.} \quad & h(x, \bar{p}) = 0, \\ & g_2(x, \bar{p}) \in \mathcal{K}. \end{aligned} \tag{A1-1}$$

Problem (A1-1) can be written in the following constrained form:

$$\begin{aligned} \min_x \quad & f(x, \bar{p}) + t \\ \text{s.t.} \quad & (g_1(x, \bar{p}), t) \in \operatorname{epi}\theta, \\ & G_2(x, \bar{p}) \in \mathcal{S}, \end{aligned} \tag{A1-2}$$

where $G_2(x, \bar{p}) = (h(x, \bar{p}), g_2(x, \bar{p}))$ and $\mathcal{S} = \{0\} \times \mathcal{K}$. Firstly, we claim that the strong regularity of problem (A1-1) is equivalent to strong regularity of problem (A1-2). The KKT system for (A1-1) is

$$\begin{cases} f_x'(x, \bar{p}) + (g_1)_x'(x, \bar{p})^* Y + (G_2)_x'(x, \bar{p})^* W = 0, \\ Y \in \partial\theta(g_1(x, \bar{p})) \quad \Leftrightarrow \quad g_1(x, \bar{p}) \in \partial\theta^*(Y), \\ W \in \mathcal{N}_{\mathcal{S}}(G_2(x, \bar{p})). \end{cases}$$

Also, KKT system for (A1-2) is

$$\begin{cases} f_x'(x, \bar{p}) + (g_1)_x'(x, \bar{p})^* Y + (G_2)_x'(x, \bar{p})^* W = 0, \\ 1 + \tau = 0, \\ W \in \mathcal{N}_{\mathcal{S}}(G_2(x, \bar{p})), \\ \begin{pmatrix} Y \\ \tau \end{pmatrix} \in \mathcal{N}_{\operatorname{epi}\theta} \begin{pmatrix} g_1(x, \bar{p}) \\ t \end{pmatrix} \quad \Leftrightarrow \quad \begin{pmatrix} g_1(x, \bar{p}) \\ t \end{pmatrix} \in \mathcal{N}_{(\operatorname{epi}\theta)^\circ} \begin{pmatrix} Y \\ \tau \end{pmatrix} \end{cases}$$

which is equivalent to

$$\begin{cases} f_x'(x, \bar{p}) + (g_1)_x'(x, \bar{p})^* Y + (G_2)_x'(x, \bar{p})^* W = 0, \\ W \in \mathcal{N}_{\mathcal{S}}(G_2(x, \bar{p})) \quad \Leftrightarrow \quad G_2(x, \bar{p}) \in \mathcal{N}_{\mathcal{S}^\circ}(W) \\ \begin{pmatrix} Y \\ -1 \end{pmatrix} \in \mathcal{N}_{\operatorname{epi}\theta} \begin{pmatrix} g_1(x, \bar{p}) \\ t \end{pmatrix} \quad \Leftrightarrow \quad \begin{pmatrix} g_1(x, \bar{p}) \\ t \end{pmatrix} \in \mathcal{N}_{(\operatorname{epi}\theta)^\circ} \begin{pmatrix} Y \\ -1 \end{pmatrix}. \end{cases}$$

If (A1-2) is strongly regular at its local minimizer $(\bar{x}, \bar{t}, \overline{W}, \overline{Y})$, we have the solution of

$$\begin{cases} f_x'(x, \bar{p}) + (g_1)_x'(x, \bar{p})^* Y + (G_2)_x'(x, \bar{p})^* W = \varepsilon_1, \\ W \in \mathcal{N}_{\mathcal{S}}(G_2(x, \bar{p}) + \varepsilon_2), \\ \begin{pmatrix} Y \\ -1 \end{pmatrix} \in \mathcal{N}_{\operatorname{epi}\theta} \begin{pmatrix} g_1(x, \bar{p}) + \varepsilon_3 \\ t + \varepsilon_4 \end{pmatrix} \end{cases} \tag{A1-3}$$

is unique and Lipschitz continuous. Denote its solution as $x(\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4)$, $Y(\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4)$, $t(\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4)$ and $W(\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4)$. From [1, Theorem 8.9], its also the solution of the

$$\begin{cases} f_x'(x,\bar{p}) + (g_1)_x'(x,\bar{p})^*Y + (G_2)_x'(x,\bar{p})^*W = \varepsilon_1, \\ W \in \mathcal{N}_{\mathcal{S}}(G_2(x,\bar{p}) + \varepsilon_2), \\ Y \in \partial\theta(g_1(x,\bar{p}) + \varepsilon_3). \end{cases} \tag{A1-4}$$

Thus the uniqueness and Lipschitz continuity holds. It follows that (A1-1) is strongly regular. Conversely, suppose (A1-1) is strongly regular. For generalized equation (A1-3), we get its solution $x(\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4)$, $W(\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4)$, $Y(\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4)$ and $t(\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4) = \theta(g_1(x,\bar{p}) + \varepsilon_3) - \varepsilon_4$ from [1, Theorem 8.9]. Since $x(\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4)$, $W(\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4)$, $Y(\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4)$ are also the solution of (A1-4), we know they are unique and Lipschitz continous around $(\bar{x},\overline{Y},\overline{W})$. Since $t(\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4) = \theta(g_1(x,\bar{p}) + \varepsilon_3) - \varepsilon_4$, it is easy to see its uniqueness and Lipschitz continuity. Thus we have (A1-2) is strongly regular at $(\bar{x},\bar{t},\overline{W},\overline{Y})$. It follows that the strong regularity of (A1-1) and (A1-2) are equivalent to each other. Then we are ready to prove Theorem 3.13 $(i) \Rightarrow (ii)$.

Proof of Theorem 3.13 $(i) \Rightarrow (ii)$**:**

For problem (3-36), suppose that strong regularity holds at $(\bar{x},\overline{Y},\bar{y},\overline{Z})$, $\bar{x}$ is the local minimizer for problem (3-36). By the former discussion, we know that for problem

$$\begin{aligned} \min_{x\in\mathbb{X},t\in\mathbb{R}} \quad & f(x,\bar{p}) + t \\ \text{s.t.} \quad & (g_1(x,\bar{p}),t) \in \operatorname{epi}\theta, \\ & h(x,\bar{p}) = 0, \\ & g_2(x,\bar{p}) \in \mathcal{K}, \end{aligned}$$

strong regularity also holds at $(\bar{x},\bar{t},\overline{Y},\bar{y},\overline{Z})$ with $\bar{x}$ being a local minimizer, where $\bar{t} = \theta(g_1(\bar{x},\bar{p}))$. It follows from [5, Theorem 5.24] that nondegeneracy and uniform quadratic growth condition are satisfied at $(\bar{x},\bar{t})$. Uniform quadratic growth condition ensures us to find $l > 0$ and neighborhoods $\mathcal{P}\times\mathcal{V}$ of $(\bar{p},0)$, $\mathcal{X}\times\mathcal{T}$ of $(\bar{x},\bar{t})$ such that for any $(p,v)\in\mathcal{P}\times\mathcal{V}$, there is a unique stationary point $(\bar{x}(p,v),\bar{t}(p,v))\in\mathcal{X}\times\mathcal{T}$ of problem $\mathcal{P}(p,v)$ that satisfies

$$f(x,p)+t-\langle x,v\rangle \geq f(\bar{x}(p,v),p)+\bar{t}(p,v)-\langle\bar{x}(p,v),v\rangle+l||(x,t)-(\bar{x}(p,v),\bar{t}(p,v))||^2,$$

for all $(x,t)\in\mathcal{X}\times\mathcal{T}$ that satisfies $(g_1(x,p),t)\in\operatorname{epi}\theta$, $h(x,p)=0$ and $g_2(x,p)\in\mathcal{K}$.

Let $\Theta(p,v,\tau) := \{(x,t) \mid v \in f(x,p), \tau = -1\}$. Picking any $(p,v,\tau,u,w) \in \operatorname{gph}\Theta\cap(\mathcal{P}\times\mathcal{V}\times[-1]\times\mathcal{X}\times T)$, we have $u = \bar{x}(p,v)$, $w = \theta(g_1(\bar{x}(p,v),p))$. It follows that for all $(x,t)\in\mathcal{X}\times\mathcal{T}$ that satisfies $(g_1(x,p),t)\in\operatorname{epi}\theta$, $h(x,p)=0$ and $g_2(x,p)\in\mathcal{K}$, we have

$$f(x,p)+t-\langle x,v\rangle \geq f(u,p)+w-\langle u,v\rangle+l||(x,t)-(u,w)||^2.$$

Then we have for all $x \in \mathcal{X}$,

$$
\begin{aligned}
& f(x,p) + \theta(g_1(x,p)) + \delta_{\{0\}\times\mathcal{K}}(h(x,p), g_2(x,p)) \\
& \geq f(u,p) + \theta(g_1(u,p)) + \delta_{\{0\}\times\mathcal{K}}(h(u,p), g_2(u,p)) + \langle x-u, v\rangle + l||x-u||^2,
\end{aligned}
$$

which implies that the uniform second-order growth condition holds. Then employing Lemma 3.10 and [16, Theorem 4.1], we arrive at (ii). □

Appendix 2 A claim of when $\overline{\eta}_1 \subseteq \eta_1^k$

Theorem A.1. *Given* $(\overline{X}, \overline{Y}) \in \operatorname{gph} \partial\theta_1$. *If*

$$\begin{cases} \sum_{j \in \overline{\eta}_1} c_j a^j = 0 \textit{ implies } \sum_{j \in \overline{\eta}_1} c_j = 0, \\ \sum_{j \in \bar{\iota}_1 \backslash \overline{\eta}_1} c_j a^j = 0 \textit{ implies } \sum_{j \in \bar{\iota}_1 \backslash \overline{\eta}_1} c_j = 0, \end{cases} \tag{A1-1}$$

then we have for any $(X^k, Y^k) \to (\overline{X}, \overline{Y})$ *with* $(X^k, Y^k) \in \operatorname{gph} \partial\theta$, $\overline{\eta}_1 \subseteq \eta_1^k$ *for each* k.

Proof: Let $\lambda(Y^k) := \mathrm{y}^k, \lambda(\overline{Y}) = \mathrm{y}$. Since $\overline{\eta}_1 \subseteq \iota_1^k \subseteq \bar{\iota}_1$ and $(X^k, Y^k) \in \operatorname{gph} \partial\theta_1$, by using (2-13), we can write y^k in the following form

$$\mathrm{y}^k = \sum_{j \in \overline{\eta}_1} h_j^k a^j + \sum_{j \in \iota_1^k \backslash \overline{\eta}_1} h_j^k a^j,$$

where $\sum_{j \in \iota_1^k} h_j^k = 1$, $h_j^k \geq 0$. Consider the linear dependent relation that there exist some $g_j \neq 0$ with $j \in \bar{\iota}_1$ such that

$$\begin{cases} \sum_{j \in \bar{\iota}_1} g_j a^j = 0 \\ \sum_{j \in \bar{\iota}_1} g_j = 0 \end{cases},$$

which is equivalent to

$$\begin{cases} \sum_{j \in \overline{\eta}_1} g_j a^j = -\sum_{j \in \bar{\iota}_1 \backslash \overline{\eta}_1} g_j a^j \\ \sum_{j \in \overline{\eta}_1} g_j = -\sum_{j \in \bar{\iota}_1 \backslash \overline{\eta}_1} g_j \end{cases}. \tag{A1-2}$$

Assume the basis of the solution subspace of equation (A1-2) is $\{\mathrm{g}_1, \ldots, \mathrm{g}_r\}$. For each k, modifying y^k as

$$\begin{aligned} \mathrm{y}^k &= \sum_{j \in \overline{\eta}_1} (h_j^k + \beta_k^* (\mathrm{g}_1)_j) a^j + \sum_{j \in \bar{\iota}_1 \backslash \overline{\eta}_1} (h_j^k - \beta_k^* (\mathrm{g}_1)_j) a^j \\ &= \sum_{j \in \overline{\eta}_1} \overline{h}_j^k a^j + \sum_{j \in \bar{\iota}_1 \backslash \overline{\eta}_1} \overline{h}_j^k a^j \end{aligned} \tag{A1-3}$$

where $\beta_k^* = \min_{j \in \bar{\iota}_1 \backslash \overline{\eta}_1} \{\frac{h_j^k}{(\mathrm{g}_1)_j}\}$. It follows that $\sum_{j \in \bar{\iota}_1} \overline{h}_j^k = 1$ and $\overline{h}_j^k \geq 0$ for all $j \in \bar{\iota}_1 \backslash \overline{\eta}_1$. Similarly, we can do the same operation for $\{\mathrm{g}_2, \ldots, \mathrm{g}_r\}$, we omit it here for simplicity. Here, we still apply (A1-3) to denote the equation we obtained after such a series of operations.

For any linear dependent relation that satisfies there exist some $d_j \neq 0$ with $j \in \bar{\iota}_1$ such that

$$\begin{cases} \sum_{j \in \overline{\eta}_1} d_j a^j = \sum_{j \in \bar{\iota}_1 \backslash \overline{\eta}_1} d_j a^j \\ \sum_{j \in \overline{\eta}_1} d_j \neq \sum_{j \in \bar{\iota}_1 \backslash \overline{\eta}_1} d_j \end{cases}, \tag{A1-4}$$

assume the basis of the solution subspace of equation (A1-4) is $\{\mathrm{d}_1, \ldots, \mathrm{d}_s\}$. Modifying y^k in (A1-3) as

$$\mathrm{y}^k = \sum_{j\in\overline{\eta}_1}(\overline{h}_j^k + \beta_k^{**}(\mathrm{d}_1)_j)a^j + \sum_{j\in\overline{\iota}_1\backslash\overline{\eta}_1}(\overline{h}_j^k - \beta_k^{**}(\mathrm{d}_1)_j)a^j,$$

where $\beta_k^{**} := \min_{j\in\overline{\iota}_1\backslash\overline{\eta}_1}\{\frac{\overline{h}_j^k}{(\mathrm{d}_1)_j}\}$. It is easy to see that for all $j \in \overline{\iota}_1\backslash\overline{\eta}_1$, $\overline{h}_j^k - \beta_k^{**}(\mathrm{d}_1)_j \geq 0$. Similarly, we can do the same operation for $\{\mathrm{d}_2, \ldots, \mathrm{d}_s\}$, we still omit it here for simplicity and apply the above equation to denote the equation we obtained after such a series of operations. Since the linear relationships (A1-2) and (A1-4) go through all linear dependent relations, we have there is no left linear dependent relationships between $\sum_{j\in\overline{\eta}_1}(\overline{h}_j^k + \beta_k^{**}(\mathrm{d}_1)_j)a^j$ and $\sum_{j\in\overline{\iota}_1\backslash\overline{\eta}_1}(\overline{h}_j^k - \beta_k^{**}(\mathrm{d}_1)_j)a^j$. As $k \to \infty$, we have

$$\sum_{j\in\overline{\eta}_1}(\overline{h}_j^k + \beta_k^{**}(\mathrm{d}_1)_j)a^j \to \mathrm{y} = \sum_{j\in\overline{\eta}_1} h_j a^j \quad \text{and} \quad \sum_{j\in\overline{\iota}_1\backslash\overline{\eta}_1}(\overline{h}_j^k - \beta_k^{**}(\mathrm{d}_1)_j)a^j \to 0.$$

From the condition in this theorem, we have as $k \to \infty$,

$$\sum_{j\in\overline{\eta}_1}(\overline{h}_j^k + \beta_k^{**}(\mathrm{d}_1)_j - h_j) \to 0 \quad \text{and} \quad \sum_{j\in\overline{\iota}_1\backslash\overline{\eta}_1}(\overline{h}_j^k - \beta_k^{**}(\mathrm{d}_1)_j) \to 0.$$

Combining the above two relations together, we have $\beta_k^{**}(\sum_{j\in\overline{\eta}_1}(\mathrm{d}_1)_j - \sum_{j\in\overline{\iota}_1\backslash\overline{\eta}_1}(\mathrm{d}_1)_j) + 1 - 1 \to 0$, which implies $\beta_k^{**} \to 0$ as $k \to \infty$. It follows that

$$\lim_{k\to\infty}\sum_{j\in\overline{\eta}_1}\overline{h}_j^k a^j = \lim_{k\to\infty}\sum_{j\in\overline{\eta}_1}(\overline{h}_j^k + \beta_k^{**}(\mathrm{d}_1)_j)a^j = \mathrm{y}$$

and

$$\lim_{k\to\infty}\sum_{j\in\overline{\iota}_1\backslash\overline{\eta}_1}\overline{h}_j^k a^j = \lim_{k\to\infty}\sum_{j\in\overline{\iota}_1\backslash\overline{\eta}_1}(\overline{h}_j^k - \beta_k^{**}(\mathrm{d}_1)_j)a^j = 0.$$

Now for all $j \in \overline{\eta}_1$, we modify $\overline{h}_j^k$ to make $\overline{h}_j^k \to h_j$ holds. Again from the condition in this theorem, we consider the following equation.

$$\begin{cases} \sum_{j\in\overline{\eta}_1} c_j a^j = 0 \\ \sum_{j\in\overline{\eta}_1} c_j = 0 \end{cases}. \tag{A1-5}$$

Let $t = |\overline{\eta}_1|$. Denote the solution subspace of (A1-5) as $\mathrm{span}\{\mathrm{c}^1, \ldots, \mathrm{c}^s\}$. Then for each k, we want to find $\{u_j^k\}_{j=1}^s$ such that

$$\begin{pmatrix} \overline{h}_1^k \\ \vdots \\ \overline{h}_t^k \end{pmatrix} - u_1^k \begin{pmatrix} \mathrm{c}_1^1 \\ \vdots \\ \mathrm{c}_t^1 \end{pmatrix} - \cdots - u_1^k \begin{pmatrix} \mathrm{c}_1^s \\ \vdots \\ \mathrm{c}_t^s \end{pmatrix} \to \begin{pmatrix} h_1 \\ \vdots \\ h_t \end{pmatrix},$$

i.e.,

$$\lim_{k\to\infty} C\begin{pmatrix} u_1^k \\ \vdots \\ u_s^k \end{pmatrix} = \lim_{k\to\infty}\begin{pmatrix} \mathrm{c}_1^1 & \cdots & \mathrm{c}_1^s \\ \vdots & & \vdots \\ \mathrm{c}_t^1 & \cdots & \mathrm{c}_t^s \end{pmatrix}\begin{pmatrix} u_1^k \\ \vdots \\ u_s^k \end{pmatrix} = \lim_{k\to\infty}\begin{pmatrix} \overline{h}_1^k - h_1 \\ \vdots \\ \overline{h}_t^k - h_t \end{pmatrix} = \lim_{k\to\infty}\begin{pmatrix} \delta_1^k \\ \vdots \\ \delta_t^k \end{pmatrix}.$$

If the limit of the right hand side of the above equation does not exist, we need to take a subsequence. As for the matrix C, we have the following facts:

(1) Exchange any two columns is ok.

(2) Every column can be divided by a constant.

(3) $l_j + rl_i$ is ok, where l_j represents the j-th column of the matrix.

So we can make $\mathrm{c}_1^1 = 1$. i.e.,

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ * & * & \cdots & * \end{pmatrix}\begin{pmatrix} (u_1)_1^k \\ \vdots \\ (u_1)_s^k \end{pmatrix} = \begin{pmatrix} \delta_1^k \\ \vdots \\ \delta_t^k \end{pmatrix}.$$

Then we do the same thing for $l_2, l_3, \cdots$, we get

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ * & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ * & * & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ * & * & \cdots & * \end{pmatrix}\begin{pmatrix} (u_{s-1})_1^k \\ \vdots \\ (u_{s-1})_s^k \end{pmatrix} = \begin{pmatrix} \delta_1^k \\ \vdots \\ \delta_t^k \end{pmatrix}, \tag{A1-6}$$

where $(u_{s-1})^k$ denotes the solution u^k of the equation which is operated for $s-1$ times. By solving (A1-6), we have $(u_{s-1})_1^k = \delta_1^k$, $(u_{s-1})_2 = \delta_2^k - \cdots$, $(u_{s-1})_3 = \delta_3^k - \cdots - \cdots$, $\cdots$ Then we get the upper $t-1$ rows of (A1-6) hold and the last row still holds by the relation $\sum_{j\in\overline{\eta}_1} c_j^i = 0$. Thus we have modified $\overline{h}_j^k$ such that $\overline{h}_j^k \to h_j$, which implies $\overline{\eta}_1 \subseteq \eta_1^k$. □

Then we prove the "⊆" in Proposition 3.12 (i) holds if $\overline{\eta}_1 \subseteq \eta_1^k$ and $\overline{\eta}_2 \subseteq \eta_2^k$. Firstly, we prove when $\overline{\eta}_1 \subseteq \eta_1^k$, condition $\operatorname{dom}\mathcal{D}^*\partial\theta(\overline{X}_1,\overline{Y})\big((g_1)'_x(\bar{x},\bar{p})(\cdot)\big) \subseteq \operatorname{aff}\mathcal{C}_\theta(\overline{X}_1,\overline{Y})$ holds. To show it, pick any $\mathrm{w} \in \operatorname{dom}\mathcal{D}^*\partial\theta(\overline{X}_1,\overline{Y})\big((g_1)'_x(\bar{x},\bar{p})(\cdot)\big)$ and find $T \in \mathcal{D}^*\partial\theta(\overline{X}_1,\overline{Y})\big((g_1)'_x(\bar{x},\bar{p})\mathrm{w}\big)$. Let $\overline{A} = \overline{X}_1 + \overline{Y}$, $\overline{P} \in \mathcal{O}^m(\overline{A})$ and $W = (g_1)'_x(\bar{x},\bar{p})\mathrm{w}$. We only need to show the three conditions in [4, Proposition 5] hold for W. Then we confirm the three conditions one by one.

(i) in [4, Proposition 5]: Since $\overline{D}_{ij} = \frac{\lambda_i(\overline{X}_1)-\lambda_j(\overline{X}_1)}{\lambda_i(\overline{A})-\lambda_j(\overline{A})}$, we know that if $i \in \chi_q^p$, $j \in \chi_{q'}^p$ with $q \neq q'$, we have $\overline{D}_{ij} = 0$. It follows from the first equation in (3-32) that for all $i \in \chi_q^p$, $j \in \chi_{q'}^p$ with $q \neq q'$, $W_{ij} = 0$. Thus (i) is right.

(iii) in [4, Proposition 5]: For each $p \in \{1,\ldots,d_1\}$ and $q \in \{1,\ldots,u^p\}$, if $q \in \mathcal{E}^p$, then there exist $i, j \in \chi_q^p$ such that $(a^w)_i \neq (a^w)_j$ for some $w \in \overline{\eta}_1$. Consider the $m \times m$ permutation matrix $Q^{i,j}$ satisfying

$$(Q^{i,j}a^w)_z = \begin{cases} (a^w)_j, & z = i \\ (a^w)_i, & z = j \\ (a^w)_z, & \text{otherwise.} \end{cases} \tag{A1-7}$$

Since $\lambda_i(\overline{X}_1) = \lambda_j(\overline{X}_1)$ and $\lambda_i(\overline{Y}) = \lambda_j(\overline{Y})$, it is clear that $Q^{i,j}\lambda(\overline{X}_1) = \lambda(\overline{X}_1)$ and $Q^{i,j}\lambda(\overline{Y}) = \lambda(\overline{Y})$. It then follows from [4, Corollary 1] that there exists $w' \in \overline{\eta}_1$ such that $a^{w'} = Q^{i,j}a^w$. Therefore, we derive from the last two relationship in (3-32) that

$$\langle \kappa(\mathscr{D}^\pi(-\widetilde{W})), a^w - a^{w'} \rangle = (\kappa(\mathscr{D}^\pi(-\widetilde{W}))_i - \kappa(\mathscr{D}^\pi(-\widetilde{W}))_j)((a^w)_i - (a^{w'})_j) = 0,$$

which implies that

$$\kappa(\mathscr{D}^\pi(-\widetilde{W}))_i = \kappa(\mathscr{D}^\pi(-\widetilde{W}))_j.$$

For any $i' \in \chi_p^q$ with $i' \neq i$ and $i' \neq j$, if $(a^w)_{i'} \neq (a^w)_i$, by replacing i by i' and j by i in the above argument, we obtain that

$$\kappa(\mathscr{D}^\pi(-\widetilde{W}))_{i'} = \kappa(\mathscr{D}^\pi(-\widetilde{W}))_i = \kappa(\mathscr{D}^\pi(-\widetilde{W}))_j;$$

otherwise if $(a^w)_{i'} = (a^w)_i$, then by replacing i by i' in the above argument, we can also obtain the above equality. Consequently, we know that for any $q \in \mathcal{E}^p$, there exists some $u_p^q \in \mathbb{R}$ such that for any $i \in \chi_p^q$

$$\operatorname{diag}(-\widetilde{W})_i = \kappa(\mathscr{D}^\pi(-\widetilde{W}))_i = u_p^q$$

So $\operatorname{diag}(-\widehat{W})_i = u_p^q$, which shows the property (iii).

equation (24) in [4]: We observe that for each $p \in \{1, \ldots, d_1\}$, if $q \notin \mathcal{E}^p$, then for any $w \in \overline{\eta}_1$, there exists a scalar $\tilde{u}_q^p$ such that

$$(a^w)_i = (a^w)_j = \tilde{u}_q^p, \quad \forall i, j \in \chi_q^p,$$

which yields

$$\langle \operatorname{diag}(-\widehat{W}), a^w \rangle = \langle \kappa(\mathscr{D}^\pi(-\widetilde{W})), a^w \rangle, \quad \forall w \in \overline{\eta}_1. \tag{A1-8}$$

So $\langle \kappa(\mathscr{D}^\pi(-\widetilde{W})), a^w - a^{w'} \rangle = \langle \operatorname{diag}(-\widehat{W}), a^w - a^{w'} \rangle$, if $w, w' \in \overline{\eta}_1$. Together with $\overline{\eta}_1 \subseteq \eta_1^k$ and (3-32), we have $\langle \kappa(\mathscr{D}^\pi(-\widetilde{W})), a^w - a^{w'} \rangle = \langle \operatorname{diag}(-\widehat{W}), a^w - a^{w'} \rangle = 0$, which implies equation (24) in [4] holds.

Similarly, we can prove when $\overline{\eta}_2 \subseteq \eta_2^k$, condition $\operatorname{dom} \mathcal{D}^*\mathcal{N}_{\{0\}\times\mathcal{K}}\big(G_2(\bar{x}, \bar{p}), \bar{y}, \overline{Z}\big)\big(d(\cdot)\big) \subseteq \operatorname{aff} \mathcal{C}_{\mathcal{K}}(\overline{X}_2, \overline{Z}) \cap \ker h_x'(\bar{x}, \bar{p})$ holds. Combining the above discussion together, we know the "$\subseteq$“ in Proposition 3.12 (i) holds.