Noname manuscript No.
(will be inserted by the editor)

Local convergence analysis of augmented Lagrangian method for nonlinear semidefinite programming

Shiwei Wang · Chao Ding

This version: October 31, 2021

Abstract The augmented Lagrangian method (ALM) has gained tremendous popularity for its elegant theory and impressive numerical performance since it was proposed by Hestenes and Powell in 1969. It has been widely used in numerous efficient solvers to improve numerical performance to solve many problems. In this paper, 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.

Keywords Nonlinear semidefinite programming · The augmented Lagrangian method · Local convergence rate · Semi-isolated calmness · Uniform quadratic growth · Uniform second order expansion

Mathematics Subject Classification (2010) 90C22 · 65K05 · 49J52

1 Introduction

The non-convex semidefinite programming problem is attracting more attention for its wide applications in machine learning, structural design, and other fields. In the well-known library COMPleib [31], there are about 168 test examples for nonlinear semidefinite programs, control system design, and related problems. In this paper, we consider the nonlinear semidefinite programming (NLSDP) problem in the following form:

$$\begin{aligned} \min_{x\in\mathcal{X}} \ & f(x) \\ \text{s.t.} \ & h(x)=0, \\ & G(x)\in\mathcal{S}^n_+, \end{aligned} \tag{1}$$

S.W. Wang
Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, P.R. China, School of Mathematical Sciences, University of Chinese Academy of Science, Beijing, P.R. China.
E-mail: wangshiwei182@mails.ucas.ac.cn

C. Ding
Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, P.R. China. The research of this author was supported by the National Natural Science Foundation of China under projects No. 12071464, No. 11671387, No. 11531014 and No. 11688101 and the Beijing Natural Science Foundation (Z190002).
E-mail: dingchao@amss.ac.cn

where $f : \mathcal{X} \to \Re$, $h : \mathcal{X} \to \mathcal{Y}$ and $G : \mathcal{X} \to \mathcal{S}^n$ are locally Lipschitz and twice continuously differentiable, $\mathcal{X}$, $\mathcal{Y}$ are given Euclidean spaces, $\mathcal{S}^n$ is the linear space of all $n \times n$ real symmetric matrices equipped with the usual Frobenius inner product and its induced norm. For notational simplicity, we use $\langle \cdot, \cdot \rangle$ to denote inner product of every Euclidean spaces and $\| \cdot \|$ to denote its induced norm. The exact meaning of these notations can be deduced from the context. $\mathcal{S}^n_+$ ($\mathcal{S}^n_-$) is used to represent the n-dimensional positive (negative) semidefinite cone. For a general constraint optimization problem, if the constraint set $\mathcal{K}$ is some polyhedron, we say the problem is polyhedral; otherwise, we say the problem is non-polyhedral. Especially, when $\mathcal{K} = \{0\} \times \Re^n_+$, it is the well-known nonlinear programming (NLP). It is easy to see that NLSDP is a non-polyhedral problem as it possesses the positive semidefinite (SDP) cone constraint.

In this paper, we will mainly focus on the local convergence analysis of augmented Lagrangian method (ALM) for (1). The Lagrangian function of problem (1) is defined by

$$L(x, y, \Gamma) := f(x) + \langle y, h(x) \rangle + \langle \Gamma, G(x) \rangle, \quad (x, y, \Gamma) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{S}^n. \tag{2}$$

For any $(y, \Gamma) \in \mathcal{Y} \times \mathcal{S}^n$, denote the first-order and second-order derivatives of $L(\cdot, y, \Gamma)$ at $x \in \mathcal{X}$ by $\nabla_x L(x, y, \Gamma)$ and $\nabla^2_{xx} L(x, y, \Gamma)$, respectively. Augmented Lagrangian function is firstly introduced by Arrow and Solow in 1958 [1] to study a differential equation method for solving equality constrained optimization problems. It is identified by Rockafellar in 1970 [44] and firstly studied in detail by Buys in his doctoral dissertation in 1973 [6] to inequality constraints. The augmented Lagrangian function of (1) takes the following form (cf. [50, Section 11.K] and [54])

$$\mathscr{L}(x, \lambda, \rho) := f(x) + \frac{\rho}{2} \operatorname{dist}^2(\Phi(x) + \frac{\lambda}{\rho}, \mathcal{K}) - \frac{\|\lambda\|^2}{2\rho}, \tag{3}$$

where $\lambda := (y, \Gamma)$, $\Phi(x) := (h(x), G(x))$, $\mathcal{K} := \{0\} \times \mathcal{S}^n_+$. And for a given z, $\operatorname{dist}(z, \mathcal{K})$ is the distance from point z to $\mathcal{K}$. For a given initial point $(x^0, \lambda^0) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{S}^n$ and a constant $\rho^0 > 0$, the $(k+1)$-th iteration of ALM for NLSDP (1) is

$$\begin{cases} x^{k+1} \approx \arg\min\{\mathscr{L}(x, \lambda^k, \rho^k)\}, \\ \lambda^{k+1} = \rho^k \left[\Phi(x^{k+1}) + \frac{\lambda^k}{\rho^k} - \Pi_{\mathcal{K}}(\Phi(x^{k+1}) + \frac{\lambda^k}{\rho^k})\right] \end{cases}$$

with ρ^{k+1} updated by certain rule.

The augmented Lagrangian method was firstly proposed by Hestenes [23] and Powell [41] for solving the equality constrained problem and was generalized by Rockafellar [45] to NLP. It rapidly grew into popularity for its mathematical elegance and impressive numerical performance in various areas, like statistical optimization (e.g., Lasso [59]), machine learning and game theory. It has also been implemented in many powerful large scale solvers like SDPNAL+ [61,63], QSDPNAL [32], SuitedLasso [33] 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 [41] 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 [46] and Tretykov [60] 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 [45].

For the convex 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 [47]. As stated in [47, 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 [48, 47, 39].

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 [47], 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 [42] but difficult to be verified for non-polyhedron. In 2019, Cui et al. [10] 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 [47] in giving a practical stopping criterion for ALM subproblem under the Robinson constraint qualification (RCQ) (for the improvement of implementable stopping criteria, see also [18]) 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 [4] 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. [8], Contesse-Becker [9], and Ito and Kunisch [25] 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 [19] 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 [22] established the local convergence for piecewise linear quadratic composite optimization problems under merely SOSC, which inspires us to study whether their results can be extended to NLSDP (1). Their success relies on the validity of upper Lipschitz continuous of KKT solution mapping when SOSC is satisfied, see [16, 26, 30, 37, 22]. However, this does not hold for non-polyhedral case as mentioned in [10] by using [5, Example 5.54]. For comprehensive surveys about the augmented Lagrangian method for nonlinear programming, please see [4, 20, 49]. Recently, Rockafellar [52] shaded lights on how to derive ALM convergence rate for non-convex "fully amenable" [50, 10F] problems through its connection with PPA for the dual problem and the variational sufficiency studied in [51, Page 6]. He successfully obtained the ALM primal R-linear convergence from the ALM dual Q-linear convergence for generalized NLP by assuming variational sufficiency. However, NLSDP does not belong to the fully amenable kind. Also, as mentioned in [52], variational sufficiency may fail even under SOSC.

For non-convex non-polyhedral problem, Sun et al. [58] proved the convergence rate of NLSDP under strongly SOSC [55] together with nondegeneracy (cf. [58] or [5]). In 2019, Kanzow and Steck [29] justified the primal-dual linear convergence of ALM under SOSC and strong 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) in [21] 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.

In this paper, under some mild conditions, we establish a locally (asymptotic Q-superlinear) Q-linear convergence of the primal-dual sequences generated by ALM without assuming the uniqueness of Lagrangian multipliers for NLSDP under SOSC and the semi-isolated calmness of the KKT solution mapping. Furthermore, for NLSDP, we provide a sufficient condition for the semi-isolated calmness of the KKT solution mapping. The remaining parts of this paper are organized as follows. In the next section, we introduce some preliminary knowledge in semidefinite cone and variational analysis. In Section 3, we study the uniform second-order expansion for the Moreau envelop of the indicator function of the SDP cone, which is essential to the main convergence result. In Section 4, we obtain the main result on the local (asymptotic Q-superlinear) Q-linear convergence of ALM for NLSDP with the help of uniform second-order growth condition of augmented Lagrangian function. Section 5 is devoted to the sufficient condition of semi-isolated calmness of the set of KKT points. Also in this section, we illustrate by two examples that the conditions proposed in our main result can be satisfied. We conclude our paper and make some comments in the final section.

2 Preliminaries

This section lists some preliminaries on the positive semidefinite cone and variational analysis, which will be used in this paper. Detailed discussions on these subjects can be found in [7,36,50].

Let $A \in \mathcal{S}^n$ be given. We use $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_n(A)$ to denote the eigenvalues of A (all real and counting multiplicity) arranging in nonincreasing order and use $\lambda(A)$ to denote the vector of the ordered eigenvalues of A. Let $\Lambda(A) := \mathrm{Diag}(\lambda(A))$. Also, we use $v_1(A) > \cdots > v_d(A)$ to denote the different eigenvalues. Consider the eigenvalue decomposition of A, 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

$$A = \begin{bmatrix} P_\alpha \ P_\beta \ P_\gamma \end{bmatrix} \begin{bmatrix} \Lambda(A)_{\alpha\alpha} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Lambda(A)_{\gamma\gamma} \end{bmatrix} \begin{bmatrix} P_\alpha^T \\ P_\beta^T \\ P_\gamma^T \end{bmatrix}. \tag{4}$$

where $\alpha := \{i : \lambda_i(A) > 0\}$, $\beta := \{i : \lambda_i(A) = 0\}$ and $\gamma := \{i : \lambda_i(A) < 0\}$. We use $\Pi_{\mathcal{S}^n_+}(A)$ to represent the projection from A to $\mathcal{S}^n_+$. From [56, Theorem 4.7] we know that the metric projection operator $\Pi_{\mathcal{S}^n_+}(\cdot)$ is directionally differentiable at any $A \in \mathcal{S}^n$ and the directional derivative of $\Pi_{\mathcal{S}^n_+}(\cdot)$ at A along direction $H \in \mathcal{S}^n$ is given by

$$\Pi'_{\mathcal{S}^n_+}(A; H) = P \begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & \Sigma_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} \\ \widetilde{H}_{\alpha\beta}^T & \Pi_{\mathcal{S}^{|\beta|}_+}(\widetilde{H}_{\beta\beta}) & 0 \\ \Sigma_{\alpha\gamma}^T \circ \widetilde{H}_{\alpha\gamma}^T & 0 & 0 \end{bmatrix} P^T, \tag{5}$$

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

$$\Sigma_{ij} := \frac{\max\{\lambda_i(A), 0\} - \max\{\lambda_j(A), 0\}}{\lambda_i(A) - \lambda_j(A)}, \quad i, j = 1, \ldots, n, \tag{6}$$

where $0/0$ is defined to be 1.

For a given Euclidean space $\mathcal{X}$. Let C be any subset in $\mathcal{X}$, the Bouligand tangent/contingent cone of C at x is a closed cone defined by

$$T_C(x) := \left\{ d \in \mathcal{X} \mid \exists\, t^k \downarrow 0 \text{ and } d^k \to d \text{ with } x + t^k d^k \in C \text{ for all } k \right\}.$$

The regular/Fréchet normal cone of C at x is defined by

$$\widehat{N}_C(x) := \left\{ v \in \mathcal{X} \mid \langle v, x' - x \rangle \leq o(\|x' - x\|)\ \forall\, x' \in C \right\}.$$

The limiting/Mordukhovich normal cone is defined by

$$N_C(x) := \left\{ \lim_{k\to\infty} v^k \mid v^k \in \widehat{N}_C(x^k),\ x^k \to x,\ x^k \in C \text{ for all } k \right\}.$$

When C is convex, the regular normal cone and limiting normal cone coincide with the normal cone in the sense of convex analysis [43], i.e., $N_C(x) = \widehat{N}_C(x) = \{v \in \mathcal{X} : \langle v, x' - x\rangle \le 0,\ \forall\, x' \in C\}$.

For any $x \in C$, the critical cone associated with $y \in N_C(x)$ of C is defined in [17, Page 98], i.e., $\mathcal{C}_C(x,y) = T_C(x) \cap (y)^{\perp}$. It is well known (see e.g., [55, (19)]) that the critical cone of SDP at a given $Y \in N_{\mathcal{S}^n_+}(X)$ can be completely described as

$$\mathcal{C}_{\mathcal{S}^n_+}(X,Y) := \{U \in \mathcal{S}^n \mid P_\beta^T U P_\beta \in \mathcal{S}^{|\beta|}_+,\ P_\beta^T U P_\gamma = 0,\ P_\gamma^T U P_\gamma = 0\}, \tag{7}$$

where $X + Y$ has the eigenvalue decomposition in (4).

Given a proper, lower semi-continuous function $f : \mathcal{X} \to (-\infty, +\infty]$ and a constant $\rho > 0$, the Moreau envelop of function f at $y \in \mathcal{X}$ is defined by

$$e_\rho f(y) := \inf_{x\in\mathcal{X}} \left\{ f(x) + \frac{1}{2\rho}\|x - y\|^2 \right\}. \tag{8}$$

In particular, when $\rho = 1$, we denote $e_1 f(y)$ as $ef(y)$ for simplicity. Especially, when f is the indicator function of $\mathcal{S}^n_+$ with $\rho = 1$ at $A \in \mathcal{S}^n$, we denote it as $e\delta_{\mathcal{S}^n_+}(A)$.

The following definition of second-order subderivative is taken from [50, Definition 13.3].

Definition 1 For $f : \mathcal{X} \to (-\infty, +\infty]$, any $\bar{x} \in \mathcal{X}$ with $f(\bar{x})$ finite and any $\bar{y} \in \mathcal{X}$, the second subderivative of f at $\bar{x}$ for $\bar{y}$ is defined as

$$\mathrm{d}^2 f(\bar{x}, \bar{y})(w) = \liminf_{t\downarrow 0, w'\to w} \frac{f(\bar{x} + tw') - f(\bar{x}) - t\langle \bar{y}, w\rangle}{\frac{1}{2}t^2}.$$

The second semiderivative of f at $\bar{x}$ is

$$\mathrm{d}^2 f(\bar{x})(w) = \liminf_{t\downarrow 0, w'\to w} \frac{f(\bar{x} + tw') - f(\bar{x}) - t\mathrm{d}f(\bar{x})(w)}{\frac{1}{2}t^2},$$

where $\mathrm{d}f(\bar{x})(w)$ is the subderivative of f at $\bar{x}$ defined in [50, Definition 8.1].

Especially, for augmented Lagrangian function of NLSDP (3), we use $\mathrm{d}^2_x\mathscr{L}$ to denote the partial second semiderivative on x. The following definition is an extension of the second order expansion of functions defined in [40, Definition 1.1].

Definition 2 Consider a function $f : \mathcal{X} \to \Re$ 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 exist a finite, continuous and positively homogeneous of degree 2 function g such that

$$f(\bar{x} + th) = f(\bar{x}) + t\langle \nabla f(\bar{x}), h\rangle + \frac{t^2}{2} g(h) + o(t^2\|h\|^2), \quad t \in \Re,\ h \in \mathcal{X}.$$

(ii) There exists a constant $r > 0$ such that $o(t^2\|h\|^2)$ is uniform for all $x \in \mathbb{B}_r(\bar{x})$ with certain conditions, i.e., for all $\varepsilon > 0$, there exist positive constants δ, r such that for all $x \in \mathbb{B}_r(\bar{x})$ with certain conditions and all $\|th\| \le \delta$, we have

$$\frac{f(x + th) - f(x) - t\langle \nabla f(x), h\rangle - \frac{t^2}{2} g(h)}{\|t^2 h^2\|} \le \varepsilon,$$

where δ is uniform for all $x \in \mathbb{B}_r(\bar{x})$ with certain conditions.

3 The uniform second order expansion of $e\delta_{\mathcal{S}^n_+}(A)$

In this section, we shall establish the uniform expansion of the Moreau envelop $e\delta_{\mathcal{S}^n_+}(\cdot)$ of the indicator function of $\mathcal{S}^n_+$, 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 $\mathcal{X}$, $\mathcal{Z}$, a positive constant r and $\overline{A} \in \mathcal{X}$. Considering a mapping $\Delta : \mathcal{X} \times \mathcal{X} \to \mathcal{Z}$. We say $\Delta(A, H) = O(\|H\|)$ with $O(\|H\|)$ uniform for all $A \in \mathbb{B}_r(\overline{A})$ with certain conditions if there exist positive constants l, δ, r such that for all $A \in \mathbb{B}_r(\overline{A})$ and all $\|H\| \le \delta$, we have

$$\frac{\|\Delta(A,H)\|}{\|H\|} \le l,$$

where δ and l are uniform for all $A \in \mathbb{B}_r(\overline{A})$ with certain conditions. We call δ the uniform radius and l 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 [57] and essentially proved in the derivation of [56, Lemma 4.12].

Lemma 1 *Given a fixed $\overline{A} \in \mathcal{S}^n$. Let $0 < r < \min_{i<j}\{v_i(\overline{A}) - v_j(\overline{A})\}/3$. For any $H \in \mathcal{S}^n$ and $A \in \mathbb{B}_r(\overline{A})$, let U be an orthogonal matrix such that*

$$U^T\big(\Lambda(A) + H\big)U = \Lambda\big(\Lambda(A) + H\big). \tag{9}$$

Then, for any $H \to 0$, we have

$$\begin{cases} U_{\bar\alpha_k\bar\alpha_l} = O(\|H\|), & k,l = 1,\cdots,\bar d, k \ne l \\ U_{\bar\alpha_k\bar\alpha_k}U^T_{\bar\alpha_k\bar\alpha_k} = I_{|\bar\alpha_k|} + O(\|H\|^2), & k = 1,\cdots,\bar d \end{cases} \tag{10}$$

where $\bar\alpha_k := \alpha_k(\overline{A}) = \{i \mid \lambda_i(\overline{A}) = v_k(\overline{A})\}$, $k = 1,\dots,\bar d$. Furthermore, for each $k \in \{1,\dots,\bar d\}$, there exists $Q_k \in \mathcal{O}^{|\bar\alpha_k|}$ such that

$$U_{\bar\alpha_k\bar\alpha_k} = Q_k + O(\|H\|^2) \tag{11}$$

and

$$Q_k^T H_{\bar\alpha_k\bar\alpha_k} Q_k = \Lambda_{\bar\alpha_k\bar\alpha_k}(\Lambda(X) + H) - Q_k^T\Lambda(A)_{\bar\alpha_k\bar\alpha_k}Q_k + O(\|H\|^2). \tag{12}$$

It is worth to note that the $O(\|H\|)$ and $O(\|H\|^2)$ above are uniform for all $A \in \mathbb{B}_r(\overline{A})$.

Proof. See Appendix A.

By applying Lemma 1, we can obtain the following result.

Lemma 2 *Given $\overline{A} \in \mathcal{S}^n$ and let $0 < r < \min_{i<j}\{v_i(\overline{A}) - v_j(\overline{A})\}/3$. For any $H \in \mathcal{S}^n$ and $A \in \mathbb{B}_r(\overline{A})$, let U be an orthogonal matrix such that*

$$U^T\big(A + H\big)U = \Lambda(A + H). \tag{13}$$

For all $l \in \{1,\dots,\bar d\}$, there exist $Q_l \in \mathcal{O}^{|\bar\alpha^l|}$ (depends on H) such that for all $H \to 0$,

$$(P^TU)_{\bar\alpha_k\bar\alpha_l} = \Theta_{kl} \circ (\widetilde{H}_{\bar\alpha_k\bar\alpha_l}Q_l) + O(\|H\|^2), \quad k \ne l$$

where $O(\|H\|^2)$ is uniform for all $A \in \mathbb{B}_r(\overline{A})$, $(\Theta_{kl})_{ij} = 1/\big((\Lambda(A)_{\bar\alpha_l\bar\alpha_l})_{ii} - (\Lambda(A)_{\bar\alpha_k\bar\alpha_k})_{jj}\big)$ and $\widetilde{H} = P^THP$, $P \in \mathcal{O}^n(A)$.

Proof. See Appendix C.

For two matrix $A, \overline{A} \in \mathcal{S}^n$, notation $\pi(A) = \pi(\overline{A})$ means these two matrix possess the same index sets of different eigenvalues, i.e, A and $\overline{A}$ both have $\bar{d}$ different eigenvalues with $\alpha_l(A) = \alpha_l(\overline{A})$ for all $l = 1, \ldots, \bar{d}$. Applying Lemma 1, we can get the uniform 1-order B-differentiability of projection function for SDP case, which is an enhancement of [14, Proposition 2.6].

Proposition 1 *Given a fixed $\overline{A} \in \mathcal{S}^n$ 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 $A \in \mathbb{B}_r(\overline{A})$ with $\pi(\overline{A}) = \pi(A)$, i.e., for $\mathcal{S}^n \ni H \to 0$,*

$$\varPi_{\mathcal{S}^n_+}(A + H) - \varPi_{\mathcal{S}^n_+}(A) - \varPi'_{\mathcal{S}^n_+}(A; H) = O(\|H\|^2) \tag{14}$$

and $O(\|H\|^2)$ is uniform for all $A \in \mathbb{B}_r(\overline{A})$ with $\pi(\overline{A}) = \pi(A)$.

Proof. See Appendix B.

With the help of the properties of the second subderivative, 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 3 *Given $(\varPhi(x), \lambda) \in \operatorname{gph} N_{\mathcal{K}}$. Denote $A = G(x) + \varGamma$ and A possesses the eigenvalue decomposition in (4). We have that for all $H \in \mathcal{S}^n$, the Moreau envelop of $\mathrm{d}^2\delta_{\mathcal{K}}(\varPhi(x), \lambda)(\cdot)$ can be calculated in the following form, i.e., for all $(b, B) \in \mathcal{Y} \times \mathcal{S}^n$,*

$$\begin{aligned}
&e_{1/(2\rho)}\big(\mathrm{d}^2\delta_{\mathcal{K}}(\varPhi(x), \lambda)\big)(b, B) = e_{1/(2\rho)}\big(\mathrm{d}^2\delta_{\mathcal{S}^n_+}(G(x), \varGamma)\big)(B) + \rho\|b\|^2 \\
=&2\rho \sum_{i\in\beta,\, j\in\gamma} \widetilde{B}_{ij}^2 + \rho \sum_{i,\, j\in\gamma} \widetilde{B}_{ij}^2 + \rho \mathrm{dist}^2(\widetilde{B}_{\beta\beta}, \mathcal{S}_+^{|\beta_0|}) \\
&+ 2\rho \sum_{i\in\alpha,\, j\in\gamma} \Big(\frac{\lambda_j(A)/\lambda_i(A)}{-\lambda_j(A)/\lambda_i(A) + \rho_3}\Big)^2 \widetilde{B}_{ij}^2 - 2 \sum_{i\in\alpha,\, j\in\gamma} \frac{\lambda_j(A)}{\lambda_i(A)} \Big(\frac{\rho\widetilde{B}_{ij}}{-\lambda_j(A)/\lambda_i(A) + \rho_3}\Big)^2 + \rho\|b\|^2,
\end{aligned}$$

where $\widetilde{B} = P^T B P$ with $P \in \mathcal{O}^n(A)$.

Proof. From [35, Theorem 6.2], we know that

$$\mathrm{d}^2\delta_{\mathcal{K}}(\varPhi(x), \lambda)(z) = -\varUpsilon_{G(x)}\big(\varGamma, Z\big) + \delta_{\mathcal{C}_{\mathcal{K}}}(\varPhi(x), \lambda)(z),$$

where $z = (s, Z) \in \mathcal{Y} \times \mathcal{S}^n$ and $\varUpsilon_{G(x)}\big(\varGamma, Z\big) = 2\langle \varGamma, Z G(x)^\dagger Z\rangle$ is the σ-term. Combining this with the definition of Moreau envelop in (8), we have

$$\begin{aligned}
&e_{1/(2\rho)}\big(\mathrm{d}^2\delta_{\mathcal{K}}(\varPhi(x), \lambda)\big)(b, B) = \inf_z\{\rho\|z - (b, B)\|^2 + \mathrm{d}^2\delta_{\mathcal{K}}(\varPhi(x), \lambda)(z)\} \\
=& \inf_{z\in\mathcal{C}_{\mathcal{K}}(\varPhi(x),\lambda)} \{-\varUpsilon_{G(x)}\big(\varGamma, Z\big) + \rho\|z - (b, B)\|^2\} \\
=& \inf_{Z\in\mathcal{C}_{\mathcal{S}^n_+}(G(x),\varGamma)} \{-\varUpsilon_{G(x)}\big(\varGamma, Z\big) + \rho\|Z - B\|^2\} + \rho\|b\|^2.
\end{aligned} \tag{15}$$

From [55, (28)], we have

$$-\varUpsilon_{G(x)}\big(\varGamma, Z\big) = -2 \sum_{i\in\alpha,\, j\in\gamma} \frac{\lambda_j(A)}{\lambda_i(A)} (P^T Z P)_{ij}^2.$$

From [55, (19)], we also have

$$\mathcal{C}_{\mathcal{S}^n_+}(G(x), \varGamma) = \Big\{Z \in \mathcal{S}^n : P_\beta^T Z P_\beta \in \mathcal{S}_+^{|\beta|}, P_\beta^T Z P_\gamma = 0, P_\gamma^T Z P_\gamma = 0\Big\}.$$

Let $\widetilde{Z}=P^TZP$, $\widetilde{B}=P^TBP$. Thus

$$\begin{aligned}&\inf_{Z\in\mathcal{C}_{\mathcal{S}^n_+}(G(x),\Gamma)}\{-\Upsilon_{G(x)}\big(\Gamma,Z\big)+\rho\|Z-B\|^2\}\\&=\inf_{Z\in\mathcal{C}_{\mathcal{S}^n_+}(G(x),\Gamma)}\{-2\sum_{i\in\alpha,\,j\in\gamma}\frac{\lambda_j(A)}{\lambda_i(A)}(P^TZP)^2_{ij}+\rho\|Z-B\|^2\}\\&=\inf_{Z\in\mathcal{C}_{\mathcal{S}^n_+}(G(x),\Gamma)}\{-2\sum_{i\in\alpha,\,j\in\gamma}\frac{\lambda_j(A)}{\lambda_i(A)}\widetilde{Z}^2_{ij}+\rho\sum_{i,j}|\widetilde{Z}_{ij}-\widetilde{B}_{ij}|^2\}.\end{aligned}\tag{16}$$

For all $i\in\alpha$, $j\in\alpha\cup\beta$ and $i\in\beta$, $j\in\alpha$, to obtain the minmum of (16), 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 $i,j\in\beta$, $\widetilde{Z}_{\beta\beta}=\Pi_{\mathcal{S}^{|\beta|}_+}(\widetilde{B}_{\beta\beta})$. Otherwise, $\widetilde{Z}_{ij}=0$. It follows from (16) that

$$\begin{aligned}&\inf_{Z\in\mathcal{C}_{\mathcal{S}^n_+}(G(x),\Gamma)}\{-\Upsilon_{G(x)}\big(\Gamma,Z\big)+\rho\|Z-B\|^2\}\\=&2\rho\sum_{i\in\beta,\,j\in\gamma}\widetilde{B}^2_{ij}+\rho\sum_{i,\,j\in\gamma}\widetilde{B}^2_{ij}+\rho\,\mathrm{dist}^2(\widetilde{B}_{\beta\beta},\mathcal{S}^{|\beta|}_+)\\&+2\rho\sum_{i\in\alpha,\,j\in\gamma}\Big(\frac{\lambda_j(A)/\lambda_i(A)}{-\lambda_j(A)/\lambda_i(A)+\rho}\Big)^2\widetilde{B}^2_{ij}-2\sum_{i\in\alpha,\,j\in\gamma}\frac{\lambda_j(A)}{\lambda_i(A)}\Big(\frac{\rho\widetilde{B}_{ij}}{-\lambda_j(A)/\lambda_i(A)+\rho}\Big)^2\end{aligned}\tag{17}$$

Combining (15) and (17) we have completed the proof. □

The following result illustrates the uniform second-order expansion for $e\delta_{\mathcal{S}^n_+}(\cdot)$, which will be used in the derivation of Proposition 3. It is worth to note that the second-order expansion for it is firstly studied in [40, Theorem 3.5]. By taking advantage of Lemmas 1-3, we can provide a direct proof here.

Proposition 2 *Given $(G(\bar{x}),\overline{\Gamma})\in\mathrm{gph}\,N_{\mathcal{S}^n_+}$. Denote $\overline{A}=G(\bar{x})+\overline{\Gamma}$ and $\overline{A}$ possesses the eigenvalue decomposition in* (4)*. Let $0<r<\min_{i<j}\{v_i(\overline{A})-v_j(\overline{A})\}/3$. For any $A:=G(\bar{x})+\Gamma\in\mathbb{B}_r(\overline{A})$ with $\Gamma\in N_{\mathcal{S}^n_+}(G(\bar{x}))$ and $\pi(A)=\pi(\overline{A})$, we have for all $H\to 0$,*

$$e\delta_{\mathcal{S}^n_+}(A+H)-e\delta_{\mathcal{S}^n_+}(A)=\langle\Pi_{\mathcal{S}^n_-}(A),H\rangle+\frac{1}{2}e\big(\mathrm{d}^2\delta_{\mathcal{S}^n_+}(G(\bar{x}),\Gamma)\big)(H)+O(\|H\|^3),$$

where $O(\|H\|^3)$ is uniform for all $A\in\mathbb{B}_r(\overline{A})$ with $\Gamma\in N_{\mathcal{S}^n_+}(G(\bar{x}))$ and $\pi(A)=\pi(\overline{A})$, $\mathrm{d}^2\delta_{\mathcal{S}^n_+}(G(\bar{x}),\Gamma)$ is defined in Definition 1.

Proof. From [50, Theorem 2.26], we have $\nabla e\delta_{\mathcal{S}^n_+}(A)=\Pi_{\mathcal{S}^n_-}(A)$. Denote $Q(A)=\Pi_{\mathcal{S}^n_-}(A)$. It follows that

$$\begin{aligned}&e\delta_{\mathcal{S}^n_+}(A+H)-e\delta_{\mathcal{S}^n_+}(A)=\frac{1}{2}\langle Q(A+H)-Q(A),Q(A+H)+Q(A)\rangle\\=&\frac{1}{2}\langle H,Q(A)+Q(A+H)-Q(A)\rangle+\frac{1}{2}\langle H,Q(A)\rangle\\&-\frac{1}{2}\langle\Pi_{\mathcal{S}^n_+}(A+H)-\Pi_{\mathcal{S}^n_+}(A),Q(A+H)\rangle-\frac{1}{2}\langle\Pi_{\mathcal{S}^n_+}(A+H)-\Pi_{\mathcal{S}^n_+}(A),Q(A)\rangle\\=&\langle H,Q(A)\rangle+\frac{1}{2}\langle H,Q'(A,H)+O(\|H\|^2)\rangle\\&-\frac{1}{2}\langle\Pi_{\mathcal{S}^n_+}(A+H)-\Pi_{\mathcal{S}^n_+}(A),Q(A+H)\rangle-\frac{1}{2}\langle\Pi_{\mathcal{S}^n_+}(A+H)-\Pi_{\mathcal{S}^n_+}(A),Q(A)\rangle\qquad(18)\\=&\langle H,Q(A)\rangle+\frac{1}{2}\langle H,Q'(A,H)\rangle+O(\|H\|^3)\\&-\frac{1}{2}\langle\Pi_{\mathcal{S}^n_+}(A+H)-\Pi_{\mathcal{S}^n_+}(A),Q(A+H)\rangle-\frac{1}{2}\langle\Pi_{\mathcal{S}^n_+}(A+H)-\Pi_{\mathcal{S}^n_+}(A),Q(A)\rangle\qquad(19)\end{aligned}$$

where (18) comes from Proposition 1. We denote $\lambda_i := \lambda_i(A)$, $\Lambda := \Lambda(A)$ and $\Xi := \Lambda(A+H)$ for short. It follows from $H = \Pi'_{\mathcal{S}^n_+}(A,H) + \Pi'_{\mathcal{S}^n_-}(A,H)$ and (5) that

$$\begin{aligned}
&\langle H, Q'(A,H)\rangle = \langle \Pi'_{\mathcal{S}^n_+}(A,H), \Pi'_{\mathcal{S}^n_-}(A,H)\rangle + \|\Pi'_{\mathcal{S}^n_-}(A,H)\|^2 \\
&= 2\sum_{i\in\alpha, j\in\gamma} \frac{\lambda_j^2 - \lambda_i\lambda_j}{(\lambda_i-\lambda_j)^2}\widetilde{H}_{ij}^2 + \sum_{i,j\in\gamma}\widetilde{H}_{ij}^2 + 2\sum_{i\in\beta, j\in\gamma}\widetilde{H}_{ij}^2 + \|\Pi_{\mathcal{S}^{|\beta|}_-}(\widetilde{H}_{\beta\beta})\|^2 \\
&= \inf_z\{\|z-H\|^2 + \mathrm{d}^2\delta_{\mathcal{K}}(G(\bar{x}),\Gamma)(z)\},
\end{aligned} \tag{20}$$

where the last equality comes from Lemma 3. It is easy to see that

$$\begin{aligned}
&-\langle \Pi_{\mathcal{S}^n_+}(A+H) - \Pi_{\mathcal{S}^n_+}(A), Q(A+H)\rangle = \langle \Pi_{\mathcal{S}^n_+}(A), Q(A+H)\rangle \\
&= \mathrm{tr}(\Lambda_{\alpha\alpha}(P^TU)_{\alpha\gamma}\Xi_{\gamma\gamma}(P^TU)^T_{\alpha\gamma}) + \mathrm{tr}(\Lambda_{\alpha\alpha}(P^TU)_{\alpha\beta}\Xi_{\beta\beta}(P^TU)^T_{\alpha\beta}).
\end{aligned} \tag{21}$$

It follows from Lemma 2 and the Lipschitz continuity of $\lambda(\cdot)$ that

$$\begin{aligned}
(P^TU)_{\alpha\gamma}\Xi_{\gamma\gamma}(P^TU)^T_{\alpha\gamma} &= \left[\Theta_{\gamma\alpha}\circ(\widetilde{H}_{\alpha\gamma}Q_\gamma)\right]\Xi_{\gamma\gamma}\left[\Theta^T_{\gamma\alpha}\circ(Q^T_\gamma\widetilde{H}^T_{\alpha\gamma})\right] + O(\|H\|^3) \\
&= \left[\Theta_{\gamma\alpha}\circ(\widetilde{H}_{\alpha\gamma}Q_\gamma)\right]\Lambda_{\gamma\gamma}\left[\Theta^T_{\gamma\alpha}\circ(Q^T_\gamma\widetilde{H}^T_{\alpha\gamma})\right] + O(\|H\|^3).
\end{aligned}$$

Thus we have

$$\mathrm{tr}(\Lambda_{\alpha\alpha}(P^TU)_{\alpha\gamma}\Xi_{\gamma\gamma}(P^TU)^T_{\alpha\gamma}) = \sum_{i\in\alpha, j\in\gamma}\frac{\lambda_i\lambda_j}{(\lambda_i-\lambda_j)^2}\|(\widetilde{H}_{\alpha\gamma}Q_\gamma)_{ij}\|^2 + O(\|H\|^3),$$

where $O(\|H\|^3)$ is uniform for all $A \in \mathbb{B}_r(\overline{A})$ with $\Gamma \in N_{\mathcal{S}^n_+}(G(\bar{x}))$ and $\pi(A) = \pi(\overline{A})$. Also, it is easy to see that $\mathrm{tr}(\Lambda_{\alpha\alpha}(P^TU)_{\alpha\beta}\Xi_{\beta\beta}(P^TU)^T_{\alpha\beta}) = O(\|H\|^3)$. Taking this into (21), we have

$$-\langle \Pi_{\mathcal{S}^n_+}(A+H) - \Pi_{\mathcal{S}^n_+}(A), Q(A+H)\rangle = \sum_{i\in\alpha, j\in\gamma}\frac{\lambda_i\lambda_j}{(\lambda_i-\lambda_j)^2}\|(\widetilde{H}_{\alpha\gamma}Q_\gamma)_{ij}\|^2 + O(\|H\|^3).$$

Similarly, we can compute $\langle \Pi_{\mathcal{S}^n_+}(A+H) - \Pi_{\mathcal{S}^n_+}(A), Q(A)\rangle$ in the exactly same way, i.e.,

$$\begin{aligned}
\langle \Pi_{\mathcal{S}^n_+}(A+H) - \Pi_{\mathcal{S}^n_+}(A), Q(A)\rangle &= \mathrm{tr}(\Lambda_{\gamma\gamma}(P^TU)_{\gamma\alpha}\Xi_{\alpha\alpha}(P^TU)^T_{\gamma\alpha}) + \mathrm{tr}(\Lambda_{\gamma\gamma}(P^TU)_{\gamma\beta}\Xi_{\beta\beta}(P^TU)^T_{\gamma\beta}) \\
&= \sum_{i\in\alpha, j\in\gamma}\frac{\lambda_i\lambda_j}{(\lambda_i-\lambda_j)^2}\|(\widetilde{H}_{\gamma\alpha}Q_\alpha)_{ij}\|^2 + O(\|H\|^3).
\end{aligned} \tag{22}$$

Combining (19), (20), (21) and (22) together, we have obtained the result. □

4 ALM convergence for NLSDP

In this section, we shall study the convergence of the inexact augmented Lagrangian method (ALM) for NLSDP (1) without requiring the uniqueness of Lagrangian multipliers. Recalling the augmented Lagrangian function of NLSDP defined in (3). We define the residual function by

$$R(x,\lambda) = \|\nabla_x L(x,\lambda)\| + \|\Phi(x) - \Pi_{\mathcal{K}}(\Phi(x)+\lambda)\|. \tag{23}$$

The augmented Lagrangian method is stated below.

Algorithm 1 (Augmented Lagrangian method)

Require: Let $(x^0, \lambda^0) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{S}^n$, $\rho^0 > 0$, $\varsigma > 1$, $\xi \in (0,1)$, $\{\epsilon_k\}_{k\geq 0}$ with $\epsilon_k > 0$ for all k and $\epsilon_k \to 0$ and set $k := 0$.

1: If (x^k, λ^k) satisfies a suitable termination criterion: STOP.

2: Compute x^{k+1} such that

$$\|\nabla_x \mathscr{L}(\cdot, \lambda^k, \rho^k)\| \leq \epsilon_k. \tag{24}$$

3: Update the vector of multipliers to

$$\lambda^{k+1} := \rho^k \left[\Phi(x^{k+1}) + \frac{\lambda^k}{\rho^k} - \Pi_{\mathcal{K}}(\Phi(x^{k+1}) + \frac{\lambda^k}{\rho^k})\right]. \tag{25}$$

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.

Before we go any further, it is necessary to introduce the definition of semi-isolated calmness. For given perturbation parameters $(a_1, a_2, b) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{S}^n$, the corresponding canonical perturbation counterpart of (1) is given by

$$\begin{aligned} \min_{x\in\mathcal{X}} \ & f(x) - \langle a_1, x\rangle \\ \text{s.t. } & h(x) - a_2 = 0, \\ & G(x) - b \in \mathcal{S}^n_+. \end{aligned} \tag{26}$$

Denote $S_{KKT}(a_1, a_2, b)$ the solution set of the KKT optimality condition for problem (26), i.e.,

$$S_{KKT}(a_1,a_2,b) = \left\{ (x,y,\Gamma) \in \mathcal{X}\times\mathcal{Y}\times\mathcal{S}^n : \begin{array}{l} \nabla_x L(x,y,\Gamma) - a_1 = 0, \\ h(x) - a_2 = 0, \\ \mathcal{S}^n_+ \ni (G(x) - b) \perp \Gamma \in \mathcal{S}^n_-. \end{array} \right\}. \tag{27}$$

For any KKT pair (x, y, Γ) that satisfies the KKT condition with $(\bar a_1, \bar a_2, \bar b) = (0,0,0)$, we call x the stationary point. For a given feasible solution $\bar x \in \mathcal{X}$ of (1) with $(\bar a_1, \bar a_2, \bar b) = (0,0,0)$, define $\mathcal{M}(\bar x)$ the set of all multiples $(y, \Gamma) \in \mathcal{Y} \times \mathcal{S}^n$ satisfying the KKT condition (27), i.e.,

$$\mathcal{M}(\bar x) = \{(y,\Gamma) \in \mathcal{Y}\times\mathcal{S}^n \mid (\bar x, y, \Gamma) \in S_{\mathrm{KKT}}(0,0,0)\}. \tag{28}$$

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 [37, Theorem 4.1], which is an extension of [27, 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 [37, Theorem 4.1], though this does not hold for the non-polyhedral case as stated in [38, Theorem 5.6].

Definition 3 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 $\mathbb{U}$ of $(0,0,0)$ and $\mathbb{V}$ of $(\bar x, \bar y, \overline{\Gamma})$ such that for all $(a_1, a_2, b) \in \mathbb{U}$,

$$\|x - \bar x\| + \mathrm{dist}((y,\Gamma), \mathcal{M}(\bar x)) \leq \kappa \|(a_1,a_2,b)\| \quad \forall\, (x,y,\Gamma) \in S_{\mathrm{KKT}}(a_1,a_2,b) \cap \mathbb{V}.$$

Let $\bar x$ be a stationary point of NLSDP (1). Assume $\mathcal{M}(\bar x)$ is nonempty with $(\bar y, \overline{\Gamma}) \in \mathcal{M}(\bar x)$, the critical cone of problem (1) is adopted from [55, (37)]

$$\mathcal{C}(\bar x) = \left\{ d \in \mathcal{X} : \nabla h(\bar x) d = 0, \nabla G(\bar x) d \in \mathcal{C}_{\mathcal{S}^n_+}(G(\bar x), \overline{\Gamma}) \right\},$$

where $\mathcal{C}_{\mathcal{S}^n_+}(G(\bar x), \overline{\Gamma})$ is the critical cone defined in (7). We also need the definition of second order sufficient condition for (1), which can be found from, e.g., [22, equation (2.11)].

Definition 4 Let $\bar{x}$ be a stationary point of NLSDP (1). 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

$$\langle \nabla^2_{xx} L(\bar{x}, \bar{y}, \overline{\Gamma})d, d\rangle - \Upsilon_{G(\bar{x})}\big(\overline{\Gamma}, \nabla G(\bar{x})d\big) > 0, \quad \forall\, 0 \neq d \in \mathcal{C}(\bar{x}).$$

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

4.1 Properties of augmented Lagrangian function and the solution of subproblem (24)

To establish the ALM convergence of NLSDP, we firstly need the quadratic growth condition for augmented Lagrangian function as presented in [58] and [22]. The (uniform) positive definite condition is critical in establishing the (uniform) quadratic growth condition. In [22, 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 [58, Proposition 4], although they supposed nondegeneracy and strongly SOSC. It follows from [35, Theorem 8.4] that for any given $\lambda \in \mathcal{M}(\bar{x})$, function $x \mapsto \mathscr{L}(x, \lambda, \rho)$ is twice semidifferentiable with respect to x at $\bar{x}$. Then we obtain the following lemma.

Lemma 4 *Let $\bar{x} \in \mathcal{X}$ be a stationary point to the NLSDP* (1) *and $\bar{\lambda} \in \mathcal{M}(\bar{x})$* (28)*. Then the following conditions are equivalent:*

(a) the SOSC holds at $(\bar{x}, \bar{\lambda})$ (see Definition 4);
(b) there are positive constants ρ_3, ε', l' such that for all $\lambda \in \mathcal{M}(\bar{x}) \cap \mathbb{B}_{\varepsilon'}(\bar{\lambda})$ and all $\rho \geq \rho_3$,

$$\mathrm{d}^2_x \mathscr{L}\big(\bar{x}, \lambda, \rho\big)(w) \geq l'\|w\|^2 \quad \forall\, w \in \Re^n \backslash \{0\}, \tag{29}$$

where $\mathrm{d}^2_x \mathscr{L}$ is the second semiderivative defined by Definition 1.

Proof. "(b)⇒(a)": It follows from [35, Theorem 8.4], directly.

"(a)⇒(b)": First, it follows from [50, Proposition 13.5] that the second semiderivative is lower semicontinuous and positive homogenous of degree 2. Thus, by [35, Theorem 8.4 (ii)], we know that the SOSC is equivalent to the existence of $l' > 0$ such that

$$\mathrm{d}^2_x \mathscr{L}\big(\bar{x}, \bar{\lambda}, \rho_3\big)(w) \geq 2l' \quad \forall\, w \in \mathbb{S}, \tag{30}$$

where $\mathbb{S}$ is the unite sphere of $\mathcal{X}$.

Next, we shall show that there exists an $\varepsilon' > 0$ such that for any $\lambda \in \mathcal{M}(\bar{x}) \cap \mathbb{B}_{\varepsilon'}(\bar{\lambda})$,

$$\mathrm{d}^2_x \mathscr{L}\big(\bar{x}, \lambda, \rho_3\big)(w) \geq l' \quad \forall\, w \in \mathbb{S}.$$

From [35, Theorem 8.3 (i)], we know that for any $\lambda \in \mathcal{M}(\bar{x})$,

$$\begin{aligned}\mathrm{d}^2_x \mathscr{L}(\bar{x}, \lambda, \rho_3)(w) =& \langle \nabla^2_{xx} L(\bar{x}, \lambda)w, w\rangle \\ &+ \inf_z \{\rho_3\|z - \nabla \Phi(\bar{x})w\|^2 + \mathrm{d}^2\delta_{\mathcal{K}}(\Phi(\bar{x}), \lambda)(z)\}.\end{aligned} \tag{31}$$

Choose $\varepsilon'' \in (0, l'/(2\|\nabla^2 \Phi(\bar{x})\|))$ if $\|\nabla^2 \Phi(\bar{x})\| \neq 0$ and $\varepsilon'' > 0$ otherwise. For each $w \in \mathbb{S}$, it is clear that for any $\lambda \in \mathbb{B}_{\varepsilon''}(\bar{\lambda})$,

$$\begin{aligned}\langle \nabla^2_{xx} L(\bar{x}, \lambda)w, w\rangle &= \langle \nabla^2_{xx} L(\bar{x}, \bar{\lambda})w, w\rangle + \langle \lambda - \bar{\lambda}, \nabla^2 \Phi(\bar{x})(w, w)\rangle \\ &\geq \langle \nabla^2_{xx} L(\bar{x}, \bar{\lambda})w, w\rangle - \|\nabla^2 \Phi(\bar{x})\| \cdot \|\lambda - \bar{\lambda}\| \\ &\geq \langle \nabla^2_{xx} L(\bar{x}, \bar{\lambda})w, w\rangle - \frac{l'}{2}.\end{aligned} \tag{32}$$

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 (4). Let $0 < r < \min_{i<j}\{v_i(\overline{A}) - v_j(\overline{A})\}/3$. For all $\Gamma \in \mathbb{B}_r(\overline{\Gamma})$, denote $\alpha \cup \beta_+ := \{i \mid \lambda(A) > 0\}$, $\beta_0 := \{i \mid \lambda(A) = 0\}$ and $\gamma \cup \beta_- := \{i \mid \lambda(A) < 0\}$. It is easy to see $\beta = \beta_0 \cup \beta_-$ as $\beta_+ = \varnothing$. Let $\widetilde{Z} = P^T Z P$, $\widetilde{B} = P^T(\nabla G(\bar{x})w)P$. It follows from Lemma 3 that

$$\begin{aligned}
&\inf_z\{\rho_3\|z - \nabla\Phi(\bar{x})w\|^2 + \mathrm{d}^2\delta_{\mathcal{K}}(\Phi(\bar{x}),\lambda)(z)\}\\
=&\inf_{Z\in\mathcal{C}_{\mathcal{S}^n_+}(G(\bar{x}),\Gamma)}\{-\Upsilon_{G(\bar{x})}(\Gamma, Z) + \rho_3\|Z - \nabla G(\bar{x})w\|^2\} + \rho_3\|\nabla h(\bar{x})w\|^2\\
=&2\rho_3\sum_{i\in\beta_0,\, j\in\gamma\cup\beta_-}\widetilde{B}_{ij}^2 + \rho_3\sum_{i,\, j\in\gamma\cup\beta_-}\widetilde{B}_{ij}^2 + \rho_3\mathrm{dist}^2(\widetilde{B}_{\beta_0\beta_0}, \mathcal{S}_+^{|\beta_0|})\\
&+ 2\rho_3\sum_{i\in\alpha,\, j\in\gamma\cup\beta_-}\Big(\frac{\lambda_j(A)/\lambda_i(A)}{-\lambda_j(A)/\lambda_i(A) + \rho_3}\Big)^2\widetilde{B}_{ij}^2\\
&- 2\sum_{i\in\alpha,\, j\in\gamma\cup\beta_-}\frac{\lambda_j(A)}{\lambda_i(A)}\Big(\frac{\rho_3\widetilde{B}_{ij}}{-\lambda_j(A)/\lambda_i(A) + \rho_3}\Big)^2 + \rho_3\|\nabla h(\bar{x})w\|^2.
\end{aligned}\tag{33}$$

From [13, 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

$$P = \overline{P}Q + O(\|\Gamma - \overline{\Gamma}\|),$$

where $\bar{d}$ is the number of the different eigenvalues of $\overline{A}$ and $Q_l \in \mathcal{O}^{|\bar{\alpha}_l|}$. We can take Γ such that $\|\Gamma - \overline{\Gamma}\| \leq \min\{r, \sqrt{2l'}/2\}$. Let $\widehat{B} = \overline{P}^T(\nabla G(\bar{x})w)\overline{P}$. Then we have

$$\begin{aligned}
&2\rho_3\sum_{i\in\beta_0,\, j\in\gamma\cup\beta_-}\widetilde{B}_{ij}^2 + \rho_3\sum_{i,\, j\in\gamma\cup\beta_-}\widetilde{B}_{ij}^2 + \rho_3\mathrm{dist}^2(\widetilde{B}_{\beta_0\beta_0}, \mathcal{S}_+^{|\beta_0|})\\
=&2\rho_3\sum_{i\in\beta,\, j\in\gamma}\widetilde{B}_{ij}^2 + \rho_3\sum_{i,\, j\in\gamma}\widetilde{B}_{ij}^2 + \rho_3\mathrm{dist}^2(\widetilde{B}_{\beta_0\beta_0}, \mathcal{S}_+^{|\beta_0|}) + 2\rho_3\sum_{i\in\beta_0,\, j\in\beta_-}\widetilde{B}_{ij}^2 + \rho_3\sum_{i,\, j\in\beta_-}\widetilde{B}_{ij}^2\\
=&2\rho_3\sum_{i\in\beta,\, j\in\gamma}\widehat{B}_{ij}^2 + \rho_3\sum_{i,\, j\in\gamma}\widehat{B}_{ij}^2 + \rho_3\mathrm{dist}^2(\widetilde{B}_{\beta\beta}, \mathcal{C}_{\mathcal{S}_+^{|\beta|}}) + O(\|\Gamma - \overline{\Gamma}\|)\\
\geq&\rho_3\mathrm{dist}^2(B_{\beta\beta}, \mathcal{S}_+^{|\beta|}) + 2\rho_3\sum_{i\in\beta,\, j\in\gamma}\widehat{B}_{ij}^2 + \rho_3\sum_{i\in\gamma,\, j\in\gamma}\widehat{B}_{ij}^2 + O(\|\Gamma - \overline{\Gamma}\|),
\end{aligned}\tag{34}$$

where the last inequality comes from

$$\mathcal{S}_+^{|\beta|} \supseteq \mathcal{C}_{\mathcal{S}_+^{|\beta|}} := \{W \in \mathcal{S}_+^{|\beta|} \mid \overline{P}_{\beta_0}^T W \overline{P}_{\beta_0} \in \mathcal{S}_+^{|\beta|},\ \overline{P}_{\beta_0}^T W \overline{P}_{\beta_-} = 0,\ \overline{P}_{\beta_-}^T W \overline{P}_{\beta_-} = 0\}.$$

By the Lipschitz continuity of $\lambda(\cdot)$, we have

$$\begin{aligned}
&2\rho_3\sum_{i\in\alpha,\, j\in\gamma\cup\beta_-}\Big(\frac{\lambda_j(A)/\lambda_i(A)}{-\lambda_j(A)/\lambda_i(A) + \rho_3}\Big)^2\widetilde{B}_{ij}^2 - 2\sum_{i\in\alpha,\, j\in\gamma\cup\beta_-}\frac{\lambda_j(A)}{\lambda_i(A)}\Big(\frac{\rho_3\widetilde{B}_{ij}}{-\lambda_j(A)/\lambda_i(A) + \rho_3}\Big)^2\\
\geq&2\rho_3\sum_{i\in\alpha,\, j\in\gamma}\Big(\frac{\lambda_j(\Gamma)}{-\lambda_j(\Gamma) + \rho_3\lambda_i(G(\bar{x}))}\Big)^2\widetilde{B}_{ij}^2 - 2\sum_{i\in\alpha,\, j\in\gamma}\frac{\lambda_j(\Gamma)}{\lambda_i(G(\bar{x}))}\Big(\frac{\rho_3\widetilde{B}_{ij}\lambda_i(G(\bar{x}))}{-\lambda_j(\Gamma) + \rho_3\lambda_i(G(\bar{x}))}\Big)^2\\
=&2\rho_3\sum_{i\in\alpha,\, j\in\gamma}\Big(\frac{\lambda_j(\overline{\Gamma})}{-\lambda_j(\overline{\Gamma}) + \rho_3\lambda_i(G(\bar{x}))}\Big)^2\widetilde{B}_{ij}^2 - 2\sum_{i\in\alpha,\, j\in\gamma}\frac{\lambda_j(\overline{\Gamma})}{\lambda_i(G(\bar{x}))}\Big(\frac{\rho_3\widetilde{B}_{ij}\lambda_i(G(\bar{x}))}{-\lambda_j(\overline{\Gamma}) + \rho_3\lambda_i(G(\bar{x}))}\Big)^2 + O(\|\Gamma - \overline{\Gamma}\|)\\
=&2\rho_3\sum_{i\in\alpha,\, j\in\gamma}\Big(\frac{\lambda_j(\overline{\Gamma})}{-\lambda_j(\overline{\Gamma}) + \rho_3\lambda_i(G(\bar{x}))}\Big)^2\widehat{B}_{ij}^2 - 2\sum_{i\in\alpha,\, j\in\gamma}\frac{\lambda_j(\overline{\Gamma})}{\lambda_i(G(\bar{x}))}\Big(\frac{\rho_3\widehat{B}_{ij}\lambda_i(G(\bar{x}))}{-\lambda_j(\overline{\Gamma}) + \rho_3\lambda_i(G(\bar{x}))}\Big)^2 + O(\|\Gamma - \overline{\Gamma}\|).
\end{aligned}\tag{35}$$

It follows from Lemma 3 again that if we combine (34) and (35) together, the right hand side is exactly

$$\inf_{Z\in\mathcal{C}_{\mathcal{S}^n_+}(G(\bar{x}),\overline{\Gamma})}\{-\Upsilon_{G(\bar{x})}\big(\overline{\Gamma},Z\big)+\rho_3\|Z-\nabla G(\bar{x})w\|^2\}+O(\|\Gamma-\overline{\Gamma}\|).$$

Combining this fact with (33), we have

$$\begin{aligned}
&\inf_z\{\rho_3\|z-\nabla\Phi(\bar{x})w\|^2+\mathrm{d}^2\delta_{\mathcal{K}}(\Phi(\bar{x}),\lambda)(z)\}\\
&=\inf_{Z\in\mathcal{C}_{\mathcal{S}^n_+}(G(\bar{x}),\Gamma)}\{-\Upsilon_{G(\bar{x})}\big(\Gamma,Z\big)+\rho_3\|Z-\nabla G(\bar{x})w\|^2\}+\rho_3\|\nabla h(\bar{x})w\|^2\\
&\geq\inf_{Z\in\mathcal{C}_{\mathcal{S}^n_+}(G(\bar{x}),\overline{\Gamma})}\{-\Upsilon_{G(\bar{x})}\big(\overline{\Gamma},Z\big)+\rho_3\|Z-\nabla G(\bar{x})w\|^2\}+\rho_3\|\nabla h(\bar{x})w\|^2+O(\|\Gamma-\overline{\Gamma}\|)\\
&\geq\inf_z\{\rho_3\|z-\nabla\Phi(\bar{x})w\|^2+\mathrm{d}^2\delta_{\mathcal{K}}(\Phi(\bar{x}),\bar{\lambda})(z)\}-l' \qquad (36)
\end{aligned}$$

Let $\varepsilon'=\min\{\varepsilon'',\sqrt{2l'}/2,r\}$. By (30), (31), (32) and (36), we have verified that for any $\lambda\in\mathcal{M}(\bar{x})\cap\mathbb{B}_{\varepsilon'}(\bar{\lambda})$,

$$\mathrm{d}_x^2\mathscr{L}\big(\bar{x},\lambda,\rho_3\big)(w)\geq\mathrm{d}_x^2\mathscr{L}\big(\bar{x},\bar{\lambda},\rho_3\big)(w)-l'\geq l'\quad\forall\,w\in\mathbb{S}.$$

Using the positive homogeneity of the second semiderivative yields (29) for $\rho=\rho_3$ and for all $\lambda\in\mathcal{M}(\bar{x})\cap\mathcal{B}_{\varepsilon'}(\bar{\lambda})$. Recall that the function

$$\rho\mapsto e_{1/2\rho}\big(\mathrm{d}^2\delta_{\mathcal{K}}(\Phi(\bar{x}),\lambda)\big)(\nabla\Phi(\bar{x})w)$$

is nondecreasing. Therefor the function $\rho\mapsto\mathrm{d}_x^2\mathscr{L}\big((\bar{x},\bar{\lambda},\rho),0\big)(w)$ is also nondecreasing. This yields (29) for all $\lambda\in\mathcal{M}(\bar{x})\cap\mathcal{B}_{\varepsilon'}(\bar{\lambda})$ 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 2.

Proposition 3 *Let $\bar{x}\in\mathcal{X}$ be a stationary point to the NLSDP* (1) *and $\bar{\lambda}\in\mathcal{M}(\bar{x})$. 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 $\lambda\in\mathcal{M}(\bar{x})\cap\mathbb{B}_r(\bar{\lambda})$ with $\pi(\Gamma)=\pi(\overline{\Gamma})$ and any $\rho>0$, we have*

$$\frac{f(\bar{x})-\mathscr{L}(x,\lambda,\rho)}{\|x-\bar{x}\|^2}=-\frac{1}{2}\mathrm{d}_x^2\mathscr{L}(\bar{x},\lambda,\rho)(\frac{x-\bar{x}}{\|x-\bar{x}\|})+O(\|x-\bar{x}\|), \qquad (37)$$

where $O(\|x-\bar{x}\|)$ is uniform for all $\lambda\in\mathcal{M}(\bar{x})\cap\mathbb{B}_r(\bar{\lambda})$ with $\pi(\Gamma)=\pi(\overline{\Gamma})$.

Proof. From [22, Proposition 3.2], we know that for all $\lambda\in\mathcal{M}(\bar{x})$ and any $\rho>0$, $f(\bar{x})=\mathscr{L}(\bar{x},\lambda,\rho)$. It follows that

$$\begin{aligned}
\mathscr{L}(x,\lambda,\rho)-f(\bar{x})=&f(x)-f(\bar{x})+\langle y,h(x)\rangle+\frac{\rho}{2}\|h(x)\|^2-\big(\langle y,h(\bar{x})\rangle+\frac{\rho}{2}\|h(\bar{x})\|^2\big)\\
&+\rho[e\delta_{\mathcal{S}^n_+}(G(x)+\rho^{-1}\Gamma)-e\delta_{\mathcal{S}^n_+}(G(\bar{x})+\rho^{-1}\Gamma)]. \qquad (38)
\end{aligned}$$

It can be checked directly from Proposition 2 that

$$\begin{aligned}
&e\delta_{\mathcal{S}^n_+}(G(x)+\rho^{-1}\Gamma)-e\delta_{\mathcal{S}^n_+}(G(\bar{x})+\rho^{-1}\Gamma)-\langle\nabla_x e\delta_{\mathcal{S}^n_+}(G(\bar{x})+\rho^{-1}\Gamma),x-\bar{x}\rangle\\
=&\langle\Pi_{\mathcal{S}^n_-}(G(\bar{x})+\rho^{-1}\Gamma),G(x)-G(\bar{x})\rangle+\frac{1}{2}e\big(\mathrm{d}^2\delta_{\mathcal{S}^n_+}(G(\bar{x}),\Gamma)\big)\big(G(x)-G(\bar{x})\big)\\
&-\langle\nabla(e\delta_{\mathcal{S}^n_+})(G(\bar{x})+\rho^{-1}\Gamma),\nabla G(\bar{x})(x-\bar{x})\rangle+O(\|G(x)-G(\bar{x})\|^3)\\
=&\langle\Pi_{\mathcal{S}^n_-}(G(\bar{x})+\rho^{-1}\Gamma),\frac{1}{2}\nabla^2G(\bar{x})(x-\bar{x},x-\bar{x})\rangle+O(\|G(x)-G(\bar{x})\|^3)\\
&+O(\|x-\bar{x}\|^3)+\frac{1}{2}e\big(\mathrm{d}^2\delta_{\mathcal{S}^n_+}(G(\bar{x}),\Gamma)\big)\big(\nabla G(\bar{x})(x-\bar{x})+O(\|x-\bar{x}\|^2)\big). \qquad (39)
\end{aligned}$$

From the explicit form of $e\big(\mathrm{d}^2\delta_{\mathcal{S}^n_+}(G(\bar{x}),\Gamma)\big)(\cdot)$ in Lemma 3, we know that

$$\begin{aligned} &e\big(\mathrm{d}^2\delta_{\mathcal{S}^n_+}(G(\bar{x}),\Gamma)\big)\big(\nabla G(\bar{x})(x-\bar{x})+O(\|x-\bar{x}\|^2)\big) \\ &= e\big(\mathrm{d}^2\delta_{\mathcal{S}^n_+}(G(\bar{x}),\Gamma)\big)\big(\nabla G(\bar{x})(x-\bar{x})\big)+O(\|x-\bar{x}\|^3), \end{aligned} \tag{40}$$

where $O(\|x-\bar{x}\|^2)$ and $O(\|x-\bar{x}\|^3)$ in (39) and (40) are uniform for all $\lambda\in\mathcal{M}(\bar{x})\cap\mathbb{B}_r(\bar{\lambda})$ 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 (38), (39), (40) and [35, Theorem 8.3] with $\bar{\lambda}\in\mathcal{M}(\bar{x})$, we have attained (37). □

Combining Lemma 4 and Proposition 3 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 [58, Proposition 1] and extended to general C^2-reducible constrained optimization by Mohammadi et al. [35, Theorem 8.4] under weaker condition.

Theorem 1 *Let $\bar{x}\in\mathcal{X}$ be a stationary point to the NLSDP* (1) *and $\bar{\lambda}\in\mathcal{M}(\bar{x})$* (28). *Then we have the following two results:*

(a) If $\bar{\lambda}\in\operatorname{ri}\mathcal{M}(\bar{x})$ (the relative interior of $\mathcal{M}(\bar{x})$), the SOSC holds at $(\bar{x},\bar{\lambda})$ if and only if there are positive constants ρ_3, θ, ε, l such that for all $\rho\geq\rho_3$ and all $\lambda\in\mathcal{M}(\bar{x})\cap\mathbb{B}_\varepsilon(\bar{\lambda})$ the uniform quadratic growth condition

$$\mathscr{L}(x,\lambda,\rho)\geq f(\bar{x})+l\|x-\bar{x}\|^2 \quad \text{for all } x\in\mathbb{B}_\theta(\bar{x}) \tag{41}$$

is satisfied.

(b) If $\bar{\lambda}\in\operatorname{rbd}\mathcal{M}(\bar{x})$ (the relative boundary of $\mathcal{M}(\bar{x})$), the SOSC holds at $(\bar{x},\bar{\lambda})$ if and only if there are positive constants ρ_3, θ, ε, l such that (41) *holds uniformly for all $\rho\geq\rho_3$ and all $\lambda\in\mathcal{M}(\bar{x})\cap\mathbb{B}_\varepsilon(\bar{\lambda})$ with $\pi(\Gamma)=\pi(\overline{\Gamma})$.*

Proof. "$\Leftarrow$" can be obtained from [35, Theorem 8.4]. Then we are going to verify the opposite direction. It follows from Lemma 4 that there exist the positive constants l', ε' and ρ_3 for which (29) is satisfied for all $\lambda\in\mathcal{M}(\bar{x})\cap\mathbb{B}_{\varepsilon'}(\bar{\lambda})$ and all $\rho\geq\rho_3$. Using this and $f(\bar{x})=\mathscr{L}(\bar{x},\lambda,\rho_3)$ for any $\lambda\in\mathcal{M}(\bar{x})$, which can be obtained by the same proof of [22, Proposition 3.2 (a)], we deduce from (29) that for any given $\lambda\in\mathcal{M}(\bar{x})\cap\mathbb{B}_{\varepsilon'}(\bar{\lambda})$ there exists $\theta_\lambda>0$ for which we have

$$\mathscr{L}(x,\lambda,\rho_3)\geq f(\bar{x})+\frac{l'}{2}\|x-\bar{x}\|^2 \quad \text{for all } x\in\mathbb{B}_{\theta_\lambda}(\bar{x}), \tag{42}$$

where the constant l' can be chosen the same for all the multipliers $\lambda\in\mathcal{M}(\bar{x})\cap\mathbb{B}_{\varepsilon'}(\bar{\lambda})$. It can be obtained directly from the definition of the second subderivative. The radii of the balls centered at $\bar{x}$ in (42), however, depend on λ. If $\bar{\lambda}\in\operatorname{ri}\mathcal{M}(\bar{x})$, we argue that a common radius can be chosen for all the multipliers $\lambda\in\mathcal{M}(\bar{x})$ that are sufficiently close to $\bar{\lambda}$. Its proof is exactly the same as the proof of [21, Theorem 4.5]. We omit it here for simplicity.

If $\bar{\lambda}\in\operatorname{rbd}\mathcal{M}(\bar{x})$, we prove that a common radius can be chosen for all $\lambda\in\mathcal{M}(\bar{x})\cap\mathbb{B}_\varepsilon(\bar{\lambda})$ with $\pi(\Gamma)=\pi(\overline{\Gamma})$. Following the proof of [22, Proposition 3.2], we know that $f(\bar{x})=\mathscr{L}(\bar{x},\lambda,\rho_3)$. It is easy to see that for all $x\in\mathbb{B}_{\theta_{\bar{\lambda}}}(\bar{x})$, we have

$$\frac{f(\bar{x})-\mathscr{L}(x,\bar{\lambda},\rho_3)}{\|x-\bar{x}\|^2}\leq-\frac{l'}{2}.$$

From Proposition 3 with $\lambda\in\mathcal{M}(\bar{x})$ and the positive homogenous of degree 2 of second semiderivative, we have

$$\frac{f(\bar{x})-\mathscr{L}(x,\lambda,\rho_3)}{\|x-\bar{x}\|^2}=-\frac{1}{2}\mathrm{d}_x^2\mathscr{L}(\bar{x},\lambda,\rho_3)(\frac{x-\bar{x}}{\|x-\bar{x}\|})+O(\|x-\bar{x}\|),$$

where $O(\|x-\bar{x}\|)$ is uniform for all $\lambda \in \mathcal{M}(\bar{x}) \cap \mathbb{B}_r(\bar{\lambda})$ with $\pi(\Gamma)=\pi(\overline{\Gamma})$, with its uniform radius ι and uniform constant c.

From the proof of the Lemma 4, we know for all $\lambda \in \mathcal{M}(\bar{x}) \cap \mathbb{B}_{\varepsilon'}(\bar{\lambda})$,

$$\mathrm{d}_x^2\mathscr{L}(\bar{x},\lambda,\rho_3)(\frac{x-\bar{x}}{\|x-\bar{x}\|}) \geq \mathrm{d}_x^2\mathscr{L}(\bar{x},\bar{\lambda},\rho_3)(\frac{x-\bar{x}}{\|x-\bar{x}\|}) + O(\|\Gamma-\overline{\Gamma}\|).$$

Suppose the uniform constant for $O(\|\Gamma-\overline{\Gamma}\|)$ is s. Let $\|x-\bar{x}\| \leq \min\{\theta_{\bar{\lambda}},\iota,l'/(16c)\} := \theta$ and $\|\lambda-\bar{\lambda}\| \leq \min\{r,\varepsilon',l'/(16s)\} := \varepsilon$ with $\lambda \in \mathcal{M}(\bar{x})$, $\pi(\Gamma)=\pi(\overline{\Gamma})$. It is easy to see that $\|\Gamma-\overline{\Gamma}\| \leq \|\lambda-\bar{\lambda}\| \leq \varepsilon$. Thus we have

$$\begin{aligned}
&\frac{f(\bar{x})-\mathscr{L}(x,\lambda,\rho_3)}{\|x-\bar{x}\|^2} = -\frac{1}{2}\mathrm{d}_x^2\mathscr{L}(\bar{x},\lambda,\rho_3)(\frac{x-\bar{x}}{\|x-\bar{x}\|}) + O(\|x-\bar{x}\|)\\
&\leq -\frac{1}{2}\mathrm{d}_x^2\mathscr{L}(\bar{x},\bar{\lambda},\rho_3)(\frac{x-\bar{x}}{\|x-\bar{x}\|}) + O(\|\Gamma-\overline{\Gamma}\|) + \frac{l'}{16}\\
&\leq -\frac{1}{2}\mathrm{d}_x^2\mathscr{L}(\bar{x},\bar{\lambda},\rho_3)(\frac{x-\bar{x}}{\|x-\bar{x}\|}) + \frac{l'}{8}\\
&= \frac{f(\bar{x})-\mathscr{L}(x,\bar{\lambda},\rho_3)}{\|x-\bar{x}\|^2} + O(\|x-\bar{x}\|) + \frac{l'}{8} \leq \frac{f(\bar{x})-\mathscr{L}(x,\bar{\lambda},\rho_3)}{\|x-\bar{x}\|^2} + \frac{3l'}{16}.
\end{aligned}$$

Taking the supremum of $x \in \mathbb{B}_\theta(\bar{x})$ on both sides and we have for all $\lambda \in \mathbb{B}_\varepsilon(\bar{\lambda}) \cap \mathcal{M}(\bar{x})$ with $\pi(\Gamma)=\pi(\overline{\Gamma})$,

$$\sup_{x\in\mathbb{B}_\theta(\bar{x})} \frac{f(\bar{x})-\mathscr{L}(x,\lambda,\rho_3)}{\|x-\bar{x}\|^2} \leq \sup_{x\in\mathbb{B}_\theta(\bar{x})} \frac{f(\bar{x})-\mathscr{L}(x,\bar{\lambda},\rho_3)}{\|x-\bar{x}\|^2} + \frac{3l'}{16} \leq -\frac{5l'}{16}.$$

It follows that for all $x \in \mathbb{B}_\theta(\bar{x})$ and $\lambda \in \mathbb{B}_\varepsilon(\bar{\lambda}) \cap \mathcal{M}(\bar{x})$ with $\pi(\Gamma)=\pi(\overline{\Gamma})$,

$$\mathscr{L}(x,\lambda,\rho_3) \geq f(\bar{x}) + \frac{5l'}{16}\|x-\bar{x}\|^2.$$

From [50, Exercise 11.56], we know that for all $\rho \geq \rho_3$, $\mathscr{L}(x,\lambda,\rho) \geq \mathscr{L}(x,\lambda,\rho_3)$. Setting $l = \frac{5l'}{16}$ and we have proved (41). □

The locally Lipschitz continuous property of local minimizer of augmented Lagrangian function for NLSDP is established in [58, Theorem 1] under nondegeneracy and strong SOSC. It is worth noting that [22, Proposition 5.2] have verified the uniformly isolated calmness of the local minimizers of the augmented Lagrangian 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 [58, Theorem 1].

Proposition 4 *Let $\bar{x} \in \mathcal{X}$ be a stationary point to the NLSDP* (1) *and $\bar{\lambda} \in \mathcal{M}(\bar{x})$* (28). *Suppose $(\bar{x},\bar{\lambda})$ satisfies SOSC. Then there are positive constants $\tau,\rho_3,\widehat{r} > 0$ such that for every $\rho \geq \rho_3$ and $\lambda \in \mathbb{B}_{\widehat{r}/2\tau}(\bar{\lambda})$, the set of the local minimizers of function $\mathcal{L}(x,\lambda,\rho)$ over $x \in \mathbb{B}_{\widehat{r}}(\bar{x})$, defined by $\mathcal{S}_\rho(\lambda)$, satisfies the uniform isolated calmness property, i.e.,*

$$\mathcal{S}_\rho(\lambda) \subseteq \{\bar{x}\} + \tau\|\lambda-\bar{\lambda}\|\mathbb{B}$$

and satisfies $\varnothing \neq \mathcal{S}_\rho(\lambda) \subseteq \operatorname{int}\mathbb{B}_{\widehat{r}}(\bar{x})$, where $\mathbb{B}$ is the unite ball in $\mathcal{X}$.

Proof. It follows from the continuity of $\mathscr{L}$ and the compactness of $\mathbb{B}_{\widehat{r}}(\bar{x})$ that $\mathcal{S}_\rho(\lambda) \neq \varnothing$ (cf. [53, Theorem 4.16]). From [50], we know that the Lagrangian function $\mathscr{L}(x, \lambda, \rho)$ is concave on λ. Combined with [35, 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 $x \in \mathbb{B}_{\widehat{r}}(\bar{x})$

$$\begin{aligned}\mathscr{L}(x, \lambda, \rho) &\geq \mathscr{L}(x, \bar{\lambda}, \rho) - \langle \nabla_\lambda \mathscr{L}(x, \lambda, \rho), \bar{\lambda} - \lambda \rangle \\ &= \mathscr{L}(x, \bar{\lambda}, \rho) - \langle \Phi(x) - \Pi_{\mathcal{K}}(\Phi(x) + \rho^{-1}\lambda), \bar{\lambda} - \lambda \rangle \\ &\geq f(\bar{x}) + l\|x - \bar{x}\|^2 - \langle \Phi(x) - \Pi_{\mathcal{K}}(\Phi(x) + \rho^{-1}\lambda), \bar{\lambda} - \lambda \rangle,\end{aligned}$$

where the l, ρ_3 and $\widehat{r}$ is the same as in [35, Theorem 8.4]. Let $u \in \mathcal{S}_\rho(\lambda) := \arg\min\{\mathscr{L}(x, \lambda, \rho) \mid x \in \mathbb{B}_{\widehat{r}}(\bar{x})\}$ and we have

$$\mathscr{L}(u, \lambda, \rho) \leq \mathscr{L}(\bar{x}, \lambda, \rho) = f(\bar{x}) + \frac{\rho}{2}\mathrm{dist}(\Phi(\bar{x}) + \rho^{-1}\lambda, \mathcal{K})^2 - \frac{1}{2}\rho^{-1}\|\lambda\|^2 \leq f(\bar{x}).$$

Define $\tau := \frac{\kappa_\Phi}{l} + \sqrt{\frac{\kappa_\Phi^2}{l^2} + \frac{1}{l\rho_3}}$ and fix $\lambda \in \mathbb{B}_{\widehat{r}/2\tau}(\bar{\lambda})$ and $\rho \geq \rho_3$, where κ_Φ is the Lipschitzian constant of Φ. Combining the above two inequalities together, we have

$$\|u - \bar{x}\|^2 \leq \frac{1}{l}\langle \Phi(u) - \Pi_{\mathcal{K}}(\Phi(u) + \rho^{-1}\lambda), \bar{\lambda} - \lambda \rangle. \tag{43}$$

Since $\bar{\lambda} \in N_{\mathcal{K}}(\Phi(\bar{x}))$, we have $\Phi(\bar{x}) = \Pi_{\mathcal{K}}(\Phi(\bar{x}) + \rho^{-1}\bar{\lambda})$. It follows that

$$\begin{aligned}&\|\Phi(u) - \Pi_{\mathcal{K}}(\Phi(u) + \rho^{-1}\lambda)\| \\ &= \|\Phi(u) - \Phi(\bar{x}) + \Pi_{\mathcal{K}}(\Phi(\bar{x}) + \rho^{-1}\bar{\lambda}) - \Pi_{\mathcal{K}}(\Phi(u) + \rho^{-1}\lambda)\| \\ &\leq 2\|\Phi(u) - \Phi(\bar{x})\| + \rho^{-1}\|\bar{\lambda} - \lambda\| \leq 2\kappa_\Phi\|u - \bar{x}\| + \rho^{-1}\|\bar{\lambda} - \lambda\|.\end{aligned}$$

Combining the above two inequalities together leads us to

$$\|u - \bar{x}\|^2 \leq \frac{1}{l}\big(2\kappa_\Phi\|u - \bar{x}\| + \rho^{-1}\|\bar{\lambda} - \lambda\|\big)\|\lambda - \bar{\lambda}\|.$$

The latter inequality can be written as the following form

$$l\|u - \bar{x}\|^2 - 2\kappa_\Phi\|\lambda - \bar{\lambda}\| \cdot \|u - \bar{x}\| - \rho^{-1}\|\lambda - \bar{\lambda}\|^2 \leq 0.$$

It follows that

$$\|u - \bar{x}\| \leq \Big(\frac{\kappa_\Phi}{l} + \sqrt{\frac{\kappa_\Phi^2}{l^2} + \frac{1}{l\rho}}\Big)\|\lambda - \bar{\lambda}\| \leq \tau\|\lambda - \bar{\lambda}\| \leq \tau\widehat{r}/2\tau < \widehat{r}.$$

Then we have completed the proof. □

Remark 1 Recalling the definition of residual function $R(x, \lambda)$ (23). It is easy to know that for KKT point $(\bar{x}, \bar{\lambda})$, there exist r_2 and κ_2 such that for all $(x, \lambda) \in \mathbb{B}_{r_2}(\bar{x}, \bar{\lambda})$,

$$R(x, \lambda) \leq \kappa_2\big(\|x - \bar{x}\| + \mathrm{dist}(\lambda, \mathcal{M}(\bar{x}))\big). \tag{44}$$

Its proof is in the same way as in [22, Proposition 5.4]. Moreover, by Theorem 1 and the proof of (43), we can prove in the same approach that if $\bar{\lambda} \in \mathrm{ri}\,\mathcal{M}(\bar{x})$, for all $\mu \in \mathcal{M}(\bar{x}) \cap \mathbb{B}_\varepsilon(\bar{\lambda})$, there exist $\rho_3 > 0$, $\theta > 0$ and $l > 0$ such that for all $\rho \geq \rho_3$, $x \in \mathbb{B}_\theta(\bar{x})$, $\lambda \in \mathcal{Y} \times \mathcal{S}^n$,

$$\|x - \bar{x}\|^2 \leq \frac{1}{l}\langle \Phi(x) - \Pi_{\mathcal{K}}(\Phi(x) + \rho^{-1}\lambda), \mu - \lambda \rangle. \tag{45}$$

Similarly, if $\bar{\lambda} \in \mathrm{rbd}\,\mathcal{M}(\bar{x})$, we have that there also exists $\varepsilon > 0$ such that for all $\mu = (y, \Gamma) \in \mathcal{M}(\bar{x}) \cap \mathbb{B}_\varepsilon(\bar{\lambda})$ with $\pi(\Gamma) = \pi(\overline{\Gamma})$, (45) holds for all $\rho \geq \rho_3$, $x \in \mathbb{B}_\theta(\bar{x})$, $\lambda \in \mathcal{Y} \times \mathcal{S}^n$.

Remark 2 It is worth noting that [51, Theorems 1 and 2] obtained augmented tilt stability under the variational sufficient condition. The relationship between augmented tilt stability and uniform isolated calmness of $\mathcal{S}_\rho(\lambda)$ remains unknown to us though it seems the former is stronger. However, as mentioned in [52], the variational sufficient condition used in [51] may fail when SOSC holds. The variational sufficient condition can be satisfied for fully amenable problems, which include NLP, nonlinear second-order cone programming (NLSOC) and exclude NLSDP obviously. This implies that the approach taken in this paper is different from that of [51].

4.2 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 [22, Theorem 5.5], which mainly focuses on the polyhedron case. We illustrate it in two different cases.

Proposition 5 *Let $\bar{x} \in \mathcal{X}$ be a stationary point to the NLSDP* (1) *and $\bar{\lambda} \in \mathcal{M}(\bar{x})$* (28)*. Suppose $(\bar{x}, \bar{\lambda})$ satisfies SOSC and the semi-isolated calmness (see Definition 3) holds for S_{KKT} at $\big((0,0),(\bar{x},\bar{\lambda})\big)$. If $\bar{\lambda} \in \operatorname{ri}\mathcal{M}(\bar{x})$, then there exists positive constants r_3, κ_3 and ρ_3 such that for all $\rho \geq \rho_3$, $(x,\lambda) \in \mathbb{B}_{r_3}(\bar{x},\bar{\lambda})$ with $R(x,\lambda) > 0$, and all the optimal solutions u to problem*

$$\min \mathscr{L}(w,\lambda,\rho) \quad \text{subject to} \quad w \in \mathbb{B}_{\widehat{r}}(\bar{x}) \tag{46}$$

with $\widehat{r}$ obtained in Proposition 4, the error bound estimate

$$\|u - x\| + \|\nabla e_{1/\rho}\delta_{\mathcal{K}}(\Phi(u) + \rho^{-1}\lambda) - \lambda\| \leq \kappa_3 R(x,\lambda) \tag{47}$$

holds. If $\bar{\lambda} \in \operatorname{rbd}\mathcal{M}(\bar{x})$, (47) *also holds for all $\rho \geq \rho_3$, $(x,\lambda) \in \mathbb{B}_{r_3}(\bar{x},\bar{\lambda})$ with $\pi(\Gamma_\pi) = \pi(\overline{\Gamma})$ and $R(x,\lambda) > 0$, where $\Pi_{\mathcal{M}(\bar{x})}(\lambda) = (y_\pi, \Gamma_\pi)$.*

Proof. By Proposition 4, we know that for every $\lambda \in \mathbb{B}_{\widehat{r}/2r}(\bar{\lambda})$ and every $\rho \geq \rho_3$ any optimal solution u to (46) satisfies the first-order optimality condition

$$\nabla_x \mathscr{L}(u,\lambda,\rho) = 0. \tag{48}$$

If $\bar{\lambda} \in \operatorname{ri}\mathcal{M}(\bar{x})$, assume by contradiction that the error bound estimate (47) fails, which implies there exists a sequence $\{(x^k,\lambda^k,\rho^k)\}_{k\in\mathbb{N}} \subset \mathcal{X}\times\mathcal{Y}\times\mathcal{S}^n\times[\rho_3,\infty)$ with $(x^k,\lambda^k) \to (\bar{x},\bar{\lambda})$ and $\rho^k \geq \rho_3$ as $k \to \infty$ such that

$$\|u^k - x^k\| + \|d^k - \lambda^k\| > kR_k \tag{49}$$

$$\text{with } d^k := \nabla e_{1/\rho_k}\delta_{\mathcal{K}}(\Phi(u^k) + \frac{\lambda^k}{\rho^k}) = \rho^k\big(\Phi(u^k) + \frac{\lambda^k}{\rho^k} - \Pi_{\mathcal{K}}(\Phi(u^k) + \frac{\lambda^k}{\rho^k})\big), \tag{50}$$

where u^k is an optimal solution to (46) for $(\lambda,\rho) = (\lambda^k,\rho^k)$ and $R_k := R(x^k,\lambda^k)$ for each $k \in \mathbb{N}$. If $\bar{\lambda} \in \operatorname{rbd}\mathcal{M}(\bar{x})$, we also assume by contradiction and the only difference lies in the supposed sequence λ^k satisfies $\pi(\Gamma_\pi^k) = \pi(\overline{\Gamma})$ in addition. Denote by s_k the left hand side of (49) and we have $R_k = o(s_k)$. It follows that

$$\nabla_x L(x^k,\lambda^k) + o(s_k) = 0 \quad \text{and} \quad \Phi(x^k) + o(s_k) = \Pi_{\mathcal{K}}(\Phi(x^k) + \lambda^k). \tag{51}$$

Passing to a subsequence if necessary, we can find $(\xi,\eta) \in \mathcal{X}\times\Re^e\times\mathcal{S}^n$ such that

$$\frac{u^k - x^k}{s_k} \to \xi \quad \text{and} \quad \frac{d^k - \lambda^k}{s_k} \to \eta := (\eta_0,\eta_1) \in \mathcal{Y}\times\mathcal{S}^n \quad \text{with} \quad (\xi,\eta) \neq 0 \tag{52}$$

We know from the definition of semi-isolated calmness and [28, Theorem 3.1] that for (x,λ) sufficiently close to $(\bar{x},\bar{\lambda})$,

$$\|x-\bar{x}\|+\operatorname{dist}(\lambda,\mathcal{M}(\bar{x}))\le \kappa R(x,\lambda).$$

Combining this with (44), we have $R_k\to 0$ since $(x^k,\lambda^k)\to(\bar{x},\bar{\lambda})$. Set $\mu^k:=\Pi_{\mathcal{M}(\bar{x})}(\lambda^k):=(\mu_1^k,\mu_2^k)\in\mathcal{Y}\times\mathcal{S}^n$. Thus we can assume without loss of generality that $x^k-\bar{x}=O(R_k)$ and $\lambda^k-\mu^k=O(R_k)$ for all $k\in\mathbb{N}$, which in turn results in

$$x^k-\bar{x}=o(s_k)\quad\text{and}\quad \lambda^k-\mu^k=o(s_k)\quad\text{as } k\to\infty. \tag{53}$$

The latter along with $\lambda^k\to\bar{\lambda}$ tells us that $\mu^k\to\bar{\lambda}$. So we get $\mu^k\in\mathcal{M}(\bar{x})\cap\mathbb{B}_\varepsilon(\bar{\lambda})$ for all k sufficiently large. Suppose $\bar{\lambda}\in\operatorname{ri}\mathcal{M}(\bar{x})$. It is clear from Remark 1 that if $\bar{\lambda}\in\operatorname{ri}\mathcal{M}(\bar{x})$, (45) holds for all $\mu\in\mathcal{M}(\bar{x})\cap\mathbb{B}_\varepsilon(\bar{\lambda})$, which implies for all $\rho^k\ge\rho_3$,

$$\|u^k-\bar{x}\|^2\le\frac{1}{\rho^k l}\langle d^k-\lambda^k,\mu^k-\lambda^k\rangle\le\frac{1}{\rho^k l}\|d^k-\lambda^k\|\cdot\|\mu^k-\lambda^k\|. \tag{54}$$

If $\bar{\lambda}\in\operatorname{rbd}\mathcal{M}(\bar{x})$, still from Remark 1, we know that (54) also holds for all $\rho^k\ge\rho_3$ when $\pi(\mu_2^k)=\pi(\bar{\Gamma})$. For proof simplicity, we only write down the $\bar{\lambda}\in\operatorname{ri}\mathcal{M}(\bar{x})$ case here as an example while the other case can be obtained similarly. It is easy to see from (49) that

$$\begin{aligned}
&\|d^k-\lambda^k\|=\|\rho^k\big(\Phi(u^k)+\frac{\lambda^k}{\rho^k}-\Pi_{\mathcal{K}}(\Phi(u^k)+\frac{\lambda^k}{\rho^k})\big)-\lambda^k\|=\rho^k\|\Phi(u^k)-\Pi_{\mathcal{K}}(\Phi(u^k)+\frac{\lambda^k}{\rho^k})\|\\
&\le\rho^k\big(\operatorname{dist}(\Phi(u^k),\mathcal{K})+\|\Pi_{\mathcal{K}}(\Phi(u^k))-\Pi_{\mathcal{K}}(\Phi(u^k)+\frac{\lambda^k}{\rho^k})\|\big)\\
&\le\rho^k\big(\operatorname{dist}(\Phi(u^k),\mathcal{K})+\|\frac{\lambda^k}{\rho^k}\|\big)\le\rho^k\big(\|\Phi(\bar{x})-\Phi(u^k)\|+\operatorname{dist}(\Phi(\bar{x}),\mathcal{K})+\|\frac{\lambda^k}{\rho^k}\|\big).
\end{aligned} \tag{55}$$

Combining (54), (55) and the boundedness of u^k from Proposition 4 together, we have

$$\|u^k-\bar{x}\|^2\le\frac{1}{l}\big(\|\Phi(\bar{x})-\Phi(u^k)\|+\|\frac{\lambda^k}{\rho^k}\|\big)\cdot\|\mu^k-\lambda^k\|\to 0\quad\text{as } k\to 0$$

Multiplying (54) by s_k^2, we have

$$\frac{\|u^k-\bar{x}\|^2}{s_k^2}\le\frac{1}{\rho^k l}\frac{\|d^k-\lambda^k\|}{s_k}\cdot\frac{\|\mu^k-\lambda^k\|}{s_k}\to 0\quad\text{as}\quad k\to\infty \tag{56}$$

along with (52) and (53). Thus we get $u^k-\bar{x}=o(s_k)$. Combining this with (53) clearly shows that

$$\xi=\lim_{k\to\infty}\frac{u^k-x^k}{s_k}=\lim_{k\to\infty}\frac{u^k-\bar{x}}{s_k}-\lim_{k\to\infty}\frac{x^k-\bar{x}}{s_k}=0-0=0. \tag{57}$$

If either the sequence $\{\rho^k\}_{k\in\mathbb{N}}$ or $\{\rho^k/s_k\}_{k\in\mathbb{N}}$ is bounded, it follows from $u^k-\bar{x}=o(s_k)$ that $\frac{\rho^k}{s_k}\|u^k-\bar{x}\|\to 0$ as $k\to\infty$. It is easy to know that

$$\begin{aligned}
\nabla_\lambda\mathscr{L}(u^k,\lambda^k,\rho^k)&=(\rho^k)^{-1}\big(\nabla e_{1/\rho^k}\delta_{\mathcal{K}}(\Phi(u^k+(\rho^k)^{-1}\lambda^k)-\lambda^k)\big)\\
&=\Phi(u^k)-\Pi_{\mathcal{K}}(\Phi(u^k)+(\rho^k)^{-1}\lambda^k)
\end{aligned}$$

and

$$\begin{aligned}
&\|\Phi(u^k)-\Pi_{\mathcal{K}}(\Phi(u^k)+(\rho^k)^{-1}\lambda^k)\|\\
&=\|\Phi(u^k)-\Phi(\bar{x})+\Pi_{\mathcal{K}}(\Phi(\bar{x})+(\rho^k)^{-1}\mu^k)-\Pi_{\mathcal{K}}(\Phi(u^k)+(\rho^k)^{-1}\lambda^k)\|\\
&\le 2\|\Phi(u^k)-\Phi(\bar{x})\|+(\rho^k)^{-1}\|\mu^k-\lambda^k\|\le 2\kappa_\Phi\|u^k-\bar{x}\|+(\rho^k)^{-1}\|\mu^k-\lambda^k\|,
\end{aligned}$$

where κ_Φ is the Lipschitzian modulus of Φ. Combining the above two equations together brings us to

$$\begin{aligned}\frac{\|d^k-\lambda^k\|}{s_k} &= \frac{\rho^k}{s_k}\|\Phi(u^k)-\Pi_{\mathcal{K}}(\Phi(u^k)+(\rho^k)^{-1}\lambda^k)\| \\ &\le 2\kappa_\Phi\frac{\rho^k}{s_k}\|u^k-\bar{x}\|+\frac{\|\mu^k-\lambda^k\|}{s_k}\to 0 \quad \text{as} \quad k\to\infty,\end{aligned}$$

which by (53) yields $\eta=0$. This is a contradiction with (52) since we showed that $(\xi,\eta)=0$.

Assume now that both sequences $\{\rho^k\}_{k\in\mathbb{N}}$ and $\{\rho^k/s_k\}_{k\in\mathbb{N}}$ are unbounded. We can assume by passing to a subsequence if necessary that

$$\rho^k\to\infty \quad \text{and} \quad \frac{\rho^k}{s_k}\to\infty \quad \text{as} \quad k\to\infty. \tag{58}$$

Since u^k is an optimal solution to (46) associated with (λ^k,ρ^k), we deduce from (48) that

$$\nabla_x\mathscr{L}(u^k,\lambda^k,\rho^k)=0.$$

Along with (51) we have

$$\begin{aligned}0&=\nabla_x\mathscr{L}(u^k,\lambda^k,\rho^k)-\nabla_x L(x^k,\lambda^k)+o(s_k)\\ &=\nabla f(u^k)-\nabla f(x^k)+\nabla\Phi(u^k)^*\nabla e_{1/\rho_k}\delta_{\mathcal{K}}(\Phi(u^k)+\frac{\lambda^k}{\rho^k})-\nabla\Phi(x^k)^*\lambda^k+o(s^k)\\ &=\big(\nabla\Phi(u^k)^*-\nabla\Phi(x^k)^*\big)\lambda^k+\nabla\Phi(u^k)^*(d^k-\lambda^k)+o(s_k)\\ &=\nabla\Phi(u^k)^*(d^k-\lambda^k)+o(s_k),\end{aligned}$$

where the last two equalities result from the Lipschitz continuity of ∇f and $\nabla\Phi$ around $\bar{x}$ and from the fact that $u^k-x^k=o(s_k)$. Dividing both sides by s_k and then letting $k\to\infty$, we have

$$\nabla\Phi(\bar{x})^*\eta=0.$$

Then we show that $\eta=0$. Denote $E=\{z\in\mathcal{Y}\times\mathcal{S}^n \mid \nabla\Phi(\bar{x})^*(\bar{x})z=0\}$ and it is easy to know that $E^\perp=\operatorname{Range}\nabla\Phi(\bar{x})$. So $\nabla\Phi(\bar{x})(u^k-\bar{x})\in E^\perp$. From (56), we have

$$\frac{\rho^k}{s_k}\|u^k-\bar{x}\|^2\le\frac{1}{l}\frac{\|d^k-\lambda^k\|}{s_k}\cdot\|\mu^k-\lambda^k\|\to 0 \quad \text{as} \quad k\to\infty.$$

Since Π_E is a linear mapping due to E being a subspace, we arrive at

$$\begin{aligned}\Pi_E\big(\frac{\rho^k}{s_k}(\Phi(u^k)-\Phi(\bar{x}))\big)&=\Pi_E\big(\frac{\rho^k}{s_k}\nabla\Phi(\bar{x})(u^k-\bar{x})+O(\frac{\rho^k}{s_k}\|u^k-\bar{x}\|^2)\big)\\ &=\Pi_E\big(O(\frac{\rho^k}{s_k}\|u^k-\bar{x}\|^2)\big)\to 0 \quad \text{as} \quad k\to\infty.\end{aligned} \tag{59}$$

Fix $k\in\mathbb{N}$ and set $z^k:=\Phi(u^k)+(\rho^k)^{-1}(\lambda^k-d^k)$. It follows from $d^k=\nabla e_{1/\rho_k}\delta_{\mathcal{K}}(\Phi(u^k)+(\rho^k)^{-1}\lambda^k)$ and $\nabla e_r g(x)=(rI+(\partial g)^{-1})^{-1}(x)$ (see [50, Theorem 2.26]) that $d^k\in\partial\delta_{\mathcal{K}}(z^k)$. Using (52) and (58) allows us to arrive at

$$z^k=\big(0,G(u^k)-\frac{(d_1)^k-\Gamma^k}{s_k}\cdot\frac{s_k}{\rho^k}\big)\to\big(0,G(\bar{x})\big)=:\bar{z} \quad \text{as} \quad k\to\infty.$$

Denote $d^k=(d_0^k,d_1^k)$ and $z^k=(0,z_1^k)$. By [11, Lemma 4 and (10)], we have $\lambda(d_1^k)\in\partial\delta_{\widehat{\mathcal{S}}}(\lambda(z_1^k))$ and $\widehat{\mathcal{S}}=\{v\in\Re^n : \max_{1\le i\le n}\{\langle -e^i,v\rangle\}\le 0\}$. Remembering $\mu^k\in\mathcal{M}(\bar{x})$, we obtain $\mu_1^k\in N_{\mathcal{S}_+^n}(\bar{z}_1)$, which implies $\lambda(\mu_1^k) \in \partial\delta_{\widehat{\mathcal{S}}}(\lambda(\bar{z}_1))$. Taking a subsequence if necessary, it is easy to see that $\lambda(z_1^k)$ and $\lambda(\bar{z}_1)$ are in the following two sets

$$\lambda(z_1^k) \in \mathcal{C}^k := \{v \in \Re^n \mid v_{\alpha\cup\beta_+} > 0, v_{(\alpha\cup\beta_+)^c} = 0\}$$

and

$$\lambda(\bar{z}_1) \in \overline{\mathcal{C}} := \{v \in \Re^n \mid v_\alpha > 0, v_{\alpha^c} = 0\},$$

where $\alpha \cup \beta_+$ is the index set of the positive components of z_1^k and α is the index set of the positive components of $\bar{z}_1$. Suppose $\overline{P}^k \in \mathcal{O}^n(\bar{z}_1) \cap \mathcal{O}^n(\mu_1^k)$ and $P^k \in \mathcal{O}^n(z_1^k) \cap \mathcal{O}^n(d_1^k)$. Thus we have

$$z^k - \bar{z} \in \{0\} \times \Theta^k := \Omega^k.$$

where $\Theta^k := \{P^k \mathrm{Diag}(w)(P^k)^T \in \mathcal{S}^n \mid w \in \mathcal{C}^k\} - \{\overline{P}^k \mathrm{Diag}(w)(\overline{P}^k)^T \in \mathcal{S}^n \mid w \in \overline{\mathcal{C}}\}$. It follows from the difinition of z^k and $\bar{z}$ that

$$\Phi(u^k) - \Phi(\bar{x}) - \frac{d^k - \lambda^k}{\rho^k} = z^k - \bar{z} \in \Omega^k.$$

Since $z_1^k \to \bar{z}_1$ and taking a subsequence if necessary, we know that $\{P^k\}_{k=1}^\infty$ converges to an orthogonal matrix $\widehat{P}^1$ and $\{\overline{P}^k\}_{k=1}^\infty$ converges to an orthogonal matrix $\widehat{P}^2$. From [13, Proposition 2.5], we can assume

$$\widehat{P}^1 = \overline{P} Q^1 = \left[\overline{P}_{\nu^1} Q_1^1 \; \overline{P}_{\nu^2} Q_2^1 \cdots \; \overline{P}_{\nu^p} Q_p^1\right]$$

and

$$\widehat{P}^2 = \overline{P} Q^2 = \left[\overline{P}_{\nu^1} Q_1^2 \; \overline{P}_{\nu^2} Q_2^2 \cdots \; \overline{P}_{\nu^p} Q_p^2\right],$$

where $\overline{P} \in \mathcal{O}^n(\bar{z}_1)$, $Q_l^s \in \mathcal{O}^{|\nu^l|}$, $s = 1, 2$ and $\nu^l = \{i \mid \lambda_i(\bar{z}_1) = v_l(\bar{z}_1)\}, l = 1, \ldots, p$.

Multipling by ρ_k/s_k and using the linearity of Π_E allow us to get

$$\Pi_E\big(\frac{\rho_k}{s_k}(\Phi(u^k) - \Phi(\bar{x}))\big) - \Pi_E\big(\frac{d^k - \lambda^k}{s^k}\big) \in \Pi_E(\Omega^k),$$

which together with (52) and (59) yields $-\Pi_E(\eta) \in \Pi_E(\overline{\Omega})$, where $\overline{\Omega} = \{0\} \times (\Theta_1 - \Theta_2)$, $\Theta_1 = \{\widehat{P}^1 \mathrm{Diag}(w)(\widehat{P}^1)^T \in \mathcal{S}^n \mid w \in \mathcal{C}^k\}$ and $\Theta_2 = \{\widehat{P}^2 \mathrm{Diag}(w)(\widehat{P}^2)^T \in \mathcal{S}^n \mid w \in \overline{\mathcal{C}}\}$. Since $\eta \in E$, the latter confirms that

$$-\eta = -\Pi_E(\eta) = \Pi_E(\zeta) \quad \text{for some} \quad \zeta := (0, \zeta_1) \in \overline{\Omega}.$$

By the representation of $\partial\delta_{\widehat{\mathcal{S}}}(\lambda(\bar{z}_1))$ in [11, (11)], we conclude that

$$\lambda(\mu_1^k) \in \mathrm{cone}\{-e_i \mid i \in \bar{\iota}_1\} := \overline{\mathcal{A}} \quad \text{where} \quad \bar{\iota}_1 = \alpha^c$$

and

$$\lambda(d_1^k) \in \mathrm{cone}\{-e_i \mid i \in \iota_1^k\} := \mathcal{A}^k \quad \text{where} \quad \iota_1^k = (\alpha \cup \beta_+)^c.$$

Thus we have $\dfrac{\mu^k - d^k}{s^k} \in \mathcal{Y} \times \Sigma^k$, where $\Sigma^k := \{\overline{P}^k \mathrm{Diag}(w)(\overline{P}^k)^T \in \mathcal{S}^n \mid w \in \overline{\mathcal{A}}\} - \{P^k \mathrm{Diag}(w)(P^k)^T \in \mathcal{S}^n \mid w \in \mathcal{A}^k\}$. Moreover, it is easy to verify

$$\frac{\mu^k - d^k}{s^k} = \frac{\mu^k - \lambda^k}{s^k} + \frac{\lambda^k - d^k}{s^k} \to -\eta \quad \text{as} \quad k \to \infty.$$

It follows that $-\eta \in \mathcal{Y} \times (\Sigma_2 - \Sigma_1)$, where $\Sigma_1 = \{\widehat{P}^1 \mathrm{Diag}(w)(\widehat{P}^1)^T \in \mathcal{S}^n \mid w \in \mathcal{A}^k\}$ and $\Sigma_2 = \{\widehat{P}^2 \mathrm{Diag}(w)(\widehat{P}^2)^T \in \mathcal{S}^n \mid w \in \overline{\mathcal{A}}\}$. It is easy to see that

$$\overline{P}^T(-\eta_1)\overline{P} := Q^2 \begin{bmatrix} 0_\alpha & 0 & 0 \\ 0 & -_{\beta_+} & 0 \\ 0 & 0 & -_{(\alpha\cup\beta_+)^c} \end{bmatrix} (Q^2)^T - Q^1 \begin{bmatrix} 0_\alpha & 0 & 0 \\ 0 & 0_{\beta_+} & 0 \\ 0 & 0 & -_{(\alpha\cup\beta_+)^c} \end{bmatrix} (Q^1)^T$$

and

$$\overline{P}^T(\zeta_1)\overline{P} := Q^1 \begin{bmatrix} +_\alpha & 0 & 0 \\ 0 & +_{\beta_+} & 0 \\ 0 & 0 & 0_{(\alpha\cup\beta_+)^c} \end{bmatrix} (Q^1)^T - Q^2 \begin{bmatrix} +_\alpha & 0 & 0 \\ 0 & 0_{\beta_+} & 0 \\ 0 & 0 & 0_{(\alpha\cup\beta_+)^c} \end{bmatrix} (Q^2)^T,$$

where $-_{\beta_+}$ denoted the $|\beta_+|$-dimensional diagonal matrix whose diagonal components are all negative, the other notations are similarly. It follows that

$$\begin{aligned} \|\eta\|^2 &= \langle \eta, \eta \rangle = \langle \eta, \Pi_E(\eta) \rangle = -\langle \eta, \Pi_E(\zeta) \rangle = -\langle \eta, \Pi_E(\zeta) + \Pi_{E^\perp}(\zeta) \rangle = \langle -\eta, \zeta \rangle \\ &= \langle -\eta_1, \zeta_1 \rangle \le 0. \end{aligned}$$

Then we have verified that $(\xi, \eta) = (0, 0)$, which contradicts with (52). Thus we have completed the proof. □

Before we put forward the main result of this paper, we need to propose the following assumption, which is needed in the proof of our main result. In Section 5, we will give 2 example (Example 1, 2) to show the validity of this assumption.

Assumption 1 *For all $\lambda = (y, \Gamma) \notin \mathcal{M}(\bar{x})$ sufficiently close to $\bar{\lambda} \in \mathrm{rbd}\, \mathcal{M}(\bar{x})$ and x sufficiently close to $\bar{x}$, there also exists $\widehat{\lambda} \in \mathcal{M}(\bar{x})$ with $\pi(\widehat{\Gamma}) = \pi(\overline{\Gamma})$ such that*

$$\|\Pi_{\mathcal{M}(\bar{x})}(\lambda) - \widehat{\lambda}\| = O(R(x, \lambda)).$$

By adding Assumption 1 to the conditions in Proposition 5, we can get a further result of Proposition 5 for the case when $\bar{\lambda} \in \mathrm{rbd}\, \mathcal{M}(\bar{x})$.

Corollary 1 *Besides the conditions in Proposition 5, if the Assumption 1 also holds, we have when $\bar{\lambda} \in \mathrm{rbd}\, \mathcal{M}(\bar{x})$, (47) also holds for all $\rho \ge \rho_3$ and $(x, \lambda) \in \mathbb{B}_{r_3}(\bar{x}, \bar{\lambda})$ with $R(x, \lambda) > 0$ and $R(\bar{x}, \lambda) > 0$.*

Proof. The proof is exactly the same as the proof of Proposition 5. Suppose the contradiction sequence (x^k, λ^k) also satisfies $\lambda^k \notin \mathcal{M}(\bar{x})$. The only difference lies in (53) and (54), 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 (53) and (54) can be alternated by $\widehat{\mu}^k$. Then we have completed the proof. □

Remark 3 Given $(x^k, \lambda^k) \in \mathbb{B}_{r_3}(\bar{x}, \bar{\lambda})$ and $\rho^k \ge \rho_3$ with r_3 and ρ_3 taken from Proposition 5. Suppose x^{k+1} is the optimal solution to (46). Increasing κ_3 if necessary. Similarly as in [22, Remark 5.6], we know that for $\widetilde{x}^{k+1}$ sufficiently close to x^{k+1}, we also have

$$\|\nabla_x \mathscr{L}(\widetilde{x}^{k+1}, \lambda^k, \rho^k)\| \le \epsilon_k$$

and

$$\|\widetilde{x}^{k+1} - x^k\| + \|\nabla e_{1/\rho} \delta_{\mathcal{K}}(\Phi(\widetilde{x}^{k+1}) + (\rho^k)^{-1}\lambda^k) - \lambda^k\| \le \kappa_3 R(x^k, \lambda^k).$$

Next we are going to propose the main result of this paper, which is inspired by [22, Theorem 5.6]. It illustrates the local linear convergence of ALM for NLSDP without requiring the uniqueness of multipliers by applying Proposition 5 and (45).

Theorem 2 *Let $\bar{x} \in \mathcal{X}$ be a stationary point to the NLSDP (1) and $\bar{\lambda} \in \mathcal{M}(\bar{x})$ (28). Suppose $(\bar{x}, \bar{\lambda})$ satisfies SOSC and the semi-isolated calmness (see Definition 3) holds for S_{KKT} at $\big((0,0), (\bar{x}, \bar{\lambda})\big)$.*

(i) If $\bar{\lambda} \in \operatorname{ri}\mathcal{M}(\bar{x})$*, then there exist positive constants* $\overline{r}$*,* $\overline{\zeta}$*,* $\overline{\varrho}$ *such that for any starting point* $(x^0,\lambda^0) \in \mathbb{B}_{\overline{r}}(\bar{x},\bar{\lambda})$ *the primal-dual sequence* $\{(x^k,\lambda^k)\}_{k\geq 0}$ *generated by Algorithm 1 with* $\rho^k \geq \overline{\varrho}$ *and* $\epsilon_k = o(R(x^k,\lambda^k))$ *for all* k *satisfies the estimate*

$$\|x^{k+1}-x^k\| + \|\lambda^{k+1}-\lambda^k\| \leq \overline{\zeta} R(x^k,\lambda^k). \tag{60}$$

(ii) If $\bar{\lambda} \in \operatorname{rbd}\mathcal{M}(\bar{x})$ *and Assumption 1 holds, then there exist positive constants* $\overline{r}$*,* $\overline{\zeta}$*,* $\overline{\varrho}$ *such that for any starting point* $(x^0,\lambda^0) \in \mathbb{B}_{\overline{r}}(\bar{x},\bar{\lambda})$ *the primal-dual sequence* $\{(x^k,\lambda^k)\}_{k\geq 0}$ *generated by Algorithm 1 with* $\rho^k \geq \overline{\varrho}$ *and* $\epsilon_k = o(R(x^k,\lambda^k))$ *and* $\lambda^k \notin \mathcal{M}(\bar{x})$ *for all* k *satisfies the estimate* (60).

Moreover, for each case, the sequence is convergent to $(\bar{x},\widehat{\lambda})$ *for some* $\widehat{\lambda} \in \mathcal{M}(\bar{x})$ *and its rate of convergence is linear, i.e., for* k *sufficiently large,*

$$\|(x^{k+1},\lambda^{k+1}) - (\bar{x},\widehat{\lambda})\| \leq \tau^k \|(x^k,\lambda^k) - (\bar{x},\widehat{\lambda})\|, \tag{61}$$

where $\tau^k = 2\sqrt{2\zeta}\kappa_1\kappa_2^2(R_k^{-1}\epsilon_k + (\rho^k)^{-1}\overline{\zeta})$.

Proof. Consider $R_k := R(x^k,\lambda^k)$. If $R_k = 0$ for some k, then the pair (x^k,λ^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 with $\mathbb{V} = \mathbb{B}_{r_1}(\bar{x},\bar{\lambda})$ and $\kappa = \kappa_1$, κ_2 and r_2 from (44), ρ_3, κ_3 and r_3 from Proposition 5 or Corollary 1, τ and $\widehat{r}$ from Proposition 4. By the definition of ϵ_k, we can find $r_4 > 0$ such that

$$\epsilon(x,\lambda) \leq \frac{1}{4\kappa_1\kappa_2} R(x,\lambda) \quad \text{whenever} \quad (x,\lambda) \in \mathbb{B}_{r_4}(\bar{x},\bar{\lambda}). \tag{62}$$

Define $\overline{\zeta} = \kappa_3$ and

$$\overline{r} = \frac{r'}{1+2\sqrt{2}\,\overline{\zeta}\kappa_2} \quad \text{with} \quad r' = \min\{\widehat{r}, \frac{\widehat{r}}{2\tau}, \frac{r_1}{\sqrt{2}\,\overline{\zeta}\kappa_2+1}, r_2, r_4, r_3\}. \tag{63}$$

Pick $q \in (0,1)$ and $\overline{\rho} = \max\{\rho_3, \frac{2\sqrt{2}\kappa_1\kappa_2^2\overline{\zeta}^2}{q}, 4\kappa_1\kappa_2\overline{\zeta}\}$. By induction, we want to show that if $\bar{\lambda} \in \operatorname{ri}\mathcal{M}(\bar{x})$, for any starting point $(x^0,\lambda^0) \in \mathbb{B}_{\overline{r}}(\bar{x},\bar{\lambda})$ the sequence generated by the algorithm with $\rho^k \geq \overline{\rho}$ and $\epsilon_k = o(R(x^k,\lambda^k))$, we have for all $k = 0,1,\ldots$ the following relationships

$$(x^k,\lambda^k) \in \mathbb{B}_{r'}(\bar{x},\bar{\lambda}), \tag{64}$$

$$\|\nabla_x L(x^{k+1},\lambda^{k+1})\| \leq \epsilon_k, \tag{65}$$

$$\|x^{k+1}-x^k\| + \|\lambda^{k+1}-\lambda^k\| \leq \overline{\zeta} R_k \tag{66}$$

hold. By induction, we firstly assume $k=0$. Since $(x^0,\lambda^0) \in \mathbb{B}_{\overline{r}}(\bar{x},\bar{\lambda})$ and $\overline{r} \leq r'$, we have (64) holds. By Proposition 4 and $\lambda^0 \in \mathbb{B}_{\widehat{r}/2\tau}(\bar{\lambda})$, we can find $\widehat{x}^1 \in \operatorname{int}\mathbb{B}_{\widehat{r}}(\bar{x})$ satisfying $\nabla_x \mathscr{L}(\widehat{x}^1,\lambda^0,\rho^0) = 0$. From Remark 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 3. Define further $\lambda^1 = \rho^0[\Phi(x^1) + \lambda^0/\rho^0 - \Pi_{\mathcal{K}}(\Phi(x^1) + \lambda^0/\rho^0)]$ and we have

$$\|\nabla_x L(x^1,\lambda^1)\| = \|\nabla_x \mathscr{L}(x^1,\lambda^0,\rho^0)\| \leq \epsilon_k.$$

It follows from Proposition 5 and Remark 3 that

$$\|x^1-x^0\| + \|\lambda^1-\lambda^0\| \leq \overline{\zeta} R_0.$$

Thus, (x^1,λ^1) is well defined and satisfies (65) and (66) for $k=0$. Then we assume (x^k,λ^k), $k = 0,1,\ldots,s+1$ are well defined and (64)-(66) hold for $k = 0,1,\ldots,s$. We now verify the existence of (x^{s+2},λ^{s+2}) and (64)-(66) satisfies for $k=s+1$. We first show that $(x^{s+1},\lambda^{s+1})\in\mathbb{B}_{r'}(\bar{x},\bar{\lambda})$. Fix an integer k with $0\le k\le s$. Since $(x^k,\lambda^k)\in\mathbb{B}_{r'}(\bar{x},\bar{\lambda})$, it follows from (44) that

$$\begin{aligned} R_k &\le \kappa_2(\|x^k-\bar{x}\|+\operatorname{dist}(\lambda^k,\mathcal{M}(\bar{x}))) \\ &\le \kappa_2(\|x^k-\bar{x}\|+\|\lambda^k-\bar{\lambda}\|)\le\sqrt{2}\kappa_2\|(x^k,\lambda^k)-(\bar{x},\bar{\lambda})\|\le\sqrt{2}\kappa r'. \end{aligned} \tag{67}$$

Thus we have

$$\begin{aligned} \|(x^{k+1},\lambda^{k+1})-(\bar{x},\bar{\lambda})\| &\le \|x^{k+1}-x^k\|+\|\lambda^{k+1}-\lambda^k\|+\|(x^k,\lambda^k)-(\bar{x},\bar{\lambda})\| \\ &\le \bar{\zeta}R_k+r'\le(\sqrt{2}\,\bar{\zeta}\kappa_2+1)r'\le r_1, \end{aligned}$$

which implies that $(x^{k+1},\lambda^{k+1})\in\mathbb{B}_{r_1}(\bar{x},\bar{\lambda})$. From Definition 3 combined with [28, Theorem 3.1], we obtain

$$\begin{aligned} &\|x^{k+1}-\bar{x}\|+\operatorname{dist}(\lambda^{k+1},\mathcal{M}(\bar{x})) \\ &\le \kappa_1R_{k+1}=\kappa_1(\|\nabla_xL(x^{k+1},\lambda^{k+1})\|+\|\Phi(x^{k+1})-\Pi_{\mathcal{K}}(\Phi(x^{k+1})+\lambda^{k+1})\|) \\ &\le \kappa_1\epsilon_k+\kappa_1\|\Phi(x^{k+1})-\Pi_{\mathcal{K}}(\Phi(x^{k+1})+\lambda^{k+1})\|. \end{aligned}$$

Let $s^{k+1}=\Pi_{\mathcal{K}}(\Phi(x^{k+1})+(\rho^k)^{-1}\lambda^k)$. From Algorithm 1 we have $\Phi(x^{k+1})-s^{k+1}=(\rho^k)^{-1}(\lambda^{k+1}-\lambda^k)$. It follows from $\lambda^{k+1}=\nabla e_{1/\rho_k}\delta_{\mathcal{K}}(\Phi(x^{k+1})+(\rho^k)^{-1}\lambda^k)$ and $\nabla e_rg(x)=(rI+(\partial g)^{-1})^{-1}(x)$ (see [50, Theorem 2.26]) that $\lambda^{k+1}\in N_{\mathcal{K}}(s^{k+1})$. By the nonexpansivness of $y\mapsto y-\Pi_{\mathcal{K}}(y+\lambda^{k+1})$, we obtained

$$\begin{aligned} &\|\Phi(x^{k+1})-\Pi_{\mathcal{K}}(\Phi(x^{k+1})+\lambda^{k+1})\| \\ &=\|\Phi(x^{k+1})-\Pi_{\mathcal{K}}(\Phi(x^{k+1})+\lambda^{k+1})\|-\|s^{k+1}-\Pi_{\mathcal{K}}(s^{k+1}+\lambda^{k+1})\| \\ &\le\|\Phi(x^{k+1})-\Pi_{\mathcal{K}}(\Phi(x^{k+1})+\lambda^{k+1})-(s^{k+1}-\Pi_{\mathcal{K}}(s^{k+1}+\lambda^{k+1}))\| \\ &\le\|\Phi(x^{k+1})-s^{k+1}\|. \end{aligned}$$

Thus we have

$$\|x^{k+1}-\bar{x}\|+\operatorname{dist}(\lambda^k,\mathcal{M}(\bar{x}))\le\kappa_1\epsilon_k+(\rho^k)^{-1}\kappa_1\|\lambda^{k+1}-\lambda^k\|,$$

which can be further calculated as

$$\begin{aligned} \|x^{k+1}-\bar{x}\|+\operatorname{dist}(\lambda^{k+1},\mathcal{M}(\bar{x})) &\overset{(66)}{\le} \kappa_1\epsilon_k+\frac{\bar{\zeta}\kappa_1}{\rho_k}R_k\overset{(62)}{\le}\frac{1}{4\kappa_2}R_k+\frac{1}{4\kappa_2}R_k \\ &\overset{(44)}{\le}\frac{1}{2}\big(\|x^k-\bar{x}\|+\operatorname{dist}(\lambda^k,\mathcal{M}(\bar{x}))\big). \end{aligned} \tag{68}$$

It follows that

$$\|x^{k+1}-\bar{x}\|+\operatorname{dist}(\lambda^{k+1},\mathcal{M}(\bar{x}))\le\frac{1}{2^{k+1}}\big(\|x^0-\bar{x}\|+\operatorname{dist}(\lambda^0,\mathcal{M}(\bar{x}))\big). \tag{69}$$

Then we have

$$\begin{aligned} &\|(x^{s+1},\lambda^{s+1})-(x^0,\lambda^0)\|\le\sum_{k=0}^{s}\|(x^{k+1},\lambda^{k+1})-(x^k,\lambda^k)\|\overset{(66)}{\le}\bar{\zeta}\sum_{k=0}^{s}R_k \\ &\overset{(67)}{\le}\bar{\zeta}\kappa_2\sum_{k=0}^{s}\big(\|x^k-\bar{x}\|+\operatorname{dist}(\lambda^k,\mathcal{M}(\bar{x}))\big) \\ &\overset{(69)}{\le}\bar{\zeta}\kappa_2\sum_{k=0}^{s}\frac{1}{2^k}\big(\|x^0-\bar{x}\|+\operatorname{dist}(\lambda^0,\mathcal{M}(\bar{x}))\big) \\ &\le 2\bar{\zeta}\kappa_2\big(\|x^0-\bar{x}\|+\operatorname{dist}(\lambda^0,\mathcal{M}(\bar{x}))\big)\le 2\bar{\zeta}\kappa_2\big(\|x^0-\bar{x}\|+\|\lambda^0-\bar{\lambda}\|\big). \end{aligned}$$

Thus we arrive at the estimate

$$\begin{aligned}&\|(x^{s+1},\lambda^{s+1})-(\bar{x},\bar{\lambda})\| \le \|(x^{s+1},\lambda^{s+1})-(x^0,\lambda^0)\| + \|(x^0,\lambda^0)-(\bar{x},\bar{\lambda})\|\\ &\le 2\bar{\zeta}\kappa_2(\|x^0-\bar{x}\|+\|\lambda^0-\bar{\lambda}\|) + \|(x^0,\lambda^0)-(\bar{x},\bar{\lambda})\|\\ &\le (2\sqrt{2}\bar{\zeta}\kappa_2+1)\|(x^0,\lambda^0)-(\bar{x},\bar{\lambda})\| \le (2\sqrt{2}\bar{\zeta}\kappa_2+1)\overline{r} = r',\end{aligned}$$

where the last inequality comes from $(x^0,\lambda^0)\in\mathbb{B}_{\overline{r}}(\bar{x},\bar{\lambda})$. Then we have verified $(x^{s+1},\lambda^{s+1})\in\mathbb{B}_{r'}(\bar{x},\bar{\lambda})$. By (63), we get $\lambda^{s+1}\in\mathbb{B}_{\widehat{r}/2\tau}(\bar{\lambda})$, and hence Proposition 4 ensures the optimal solution $\widehat{x}^{s+2}$ such that $\widehat{x}^{s+2}\in\operatorname{int}\mathbb{B}_{\widehat{r}}(\bar{x})$. Thus we have $\nabla_x\mathscr{L}(\widehat{x}^{s+2},\lambda^{s+1},\rho^{s+1})=0$. Still from Remark 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 3 and we observe that

$$\|\nabla_x L(x^{s+2},\lambda^{s+2})\| = \|\nabla_x\mathscr{L}(x^{s+2},\lambda^{s+1},\rho^{s+1})\| \le \epsilon_{s+1}.$$

By Proposition 5 and Remark 3, we have

$$\|x^{s+2}-x^{s+1}\| + \|\lambda^{s+2}-\lambda^{s+1}\| \le \bar{\zeta}R_{s+1}.$$

Then we have finished verifying (64)-(66) for $k=s+1$. If $\bar{\lambda}\in\operatorname{rbd}\mathcal{M}(\bar{x})$ and Assumption 1 holds, we want to prove for any starting point $(x^0,\lambda^0)\in\mathbb{B}_{\overline{r}}(\bar{x},\bar{\lambda})$ the sequence generated by the algorithm with $\rho^k\ge\overline{\rho}$, $\epsilon_k=o(R(x^k,\lambda^k))$ and $R(\bar{x},\lambda^k)>0$ for all k, relationships (64)-(66) also hold for all $k=0,1,\ldots$. The proof is exactly the same as the $\bar{\lambda}\in\operatorname{ri}\mathcal{M}(\bar{x})$ case and the only difference lies in the Proposition 5 used above should be alternated by Corollary 1.

Then we prove the convergence of the sequence. Use the same argument as in the proofs of (69), we have

$$\|(x^{k+l},\lambda^{k+l})-(x^k,\lambda^k)\| \le 2\bar{\zeta}\kappa_2\big(\|x^k-\bar{x}\|+\operatorname{dist}(\lambda^k,\mathcal{M}(\bar{x}))\big) \quad \text{for all} \quad k,l\in\mathbb{N}. \tag{70}$$

It follows from (69) that the righthand side of (70) goes to 0 as $k\to\infty$, which implies $\{(x^k,\lambda^k)\}$ is Cauchy. Assume $\{(x^k,\lambda^k)\}$ convergences to $(\bar{x},\widehat{\lambda})$, where $\widehat{\lambda}\in\mathcal{M}(\bar{x})$. Let $l\to\infty$ in (70). Then we have

$$\|(x^k,\lambda^k)-(\bar{x},\widehat{\lambda})\| \le 2\bar{\zeta}\kappa_2\big(\|x^k-\bar{x}\|+\operatorname{dist}(\lambda^k,\mathcal{M}(\bar{x}))\big),$$

which together with (68) verifies

$$\begin{aligned}&\|(x^{k+1},\lambda^{k+1})-(\bar{x},\widehat{\lambda})\| \le 2\bar{\zeta}\kappa_2\big(\|x^{k+1}-\bar{x}\|+\operatorname{dist}(\lambda^{k+1},\mathcal{M}(\bar{x}))\big)\\ &\le 2\bar{\zeta}\kappa_2\kappa_1(R_k^{-1}\epsilon_k+(\rho^k)^{-1}\bar{\zeta})R_k\\ &\le 2\bar{\zeta}\kappa_2^2\kappa_1(R_k^{-1}\epsilon_k+(\rho^k)^{-1}\bar{\zeta})\big(\|x^k-\bar{x}\|+\operatorname{dist}(\lambda^k,\mathcal{M}(\bar{x}))\big)\\ &\le 2\sqrt{2}\bar{\zeta}\kappa_1\kappa_2^2(R_k^{-1}\epsilon_k+(\rho^k)^{-1}\bar{\zeta})\|(x^k,\lambda^k)-(\bar{x},\widehat{\lambda})\|.\end{aligned}$$

Combining this with $\rho^k\ge\overline{\rho}$ and $\epsilon_k=o(R_k)$ result in

$$\limsup_{k\to\infty}\frac{\|(x^{k+1},\lambda^{k+1})-(\bar{x},\widehat{\lambda})\|}{\|(x^k,\lambda^k)-(\bar{x},\widehat{\lambda})\|} \le \limsup_{k\to\infty} 2\sqrt{2}\bar{\zeta}\kappa_1\kappa_2^2(R_k^{-1}\epsilon_k+(\rho^k)^{-1}\bar{\zeta}) \le q.$$

It follows from $q\in(0,1)$ that the convergence rate is linear. Then we have completed the whole proof. □

Remark 4 If $\rho^k \to \infty$, we obtain the asymptotic Q-superlinear convergence rate of KKT pair from (61) as $\tau^k \to 0$. Regarding to the update of ρ^k, in [28,22], they apply a practical rule in Algorithm 1, step 4 to update ρ^k, i.e., defining the auxiliary function by

$$V(x,\lambda,\rho) := \|\nabla_x \mathscr{L}(x,\lambda,\rho)\| + \|\Phi(x) - \Pi_{\mathcal{K}}(\Phi(x) + \rho^{-1}\lambda)\|.$$

If $k = 0$ or $V(x^{k+1},\lambda^k,\rho^k) \le \xi V(x^k,\lambda^{k-1},\rho^{k-1})$ holds, set $\rho^{k+1} := \rho^k$; otherwise, set $\rho^{k+1} := \varsigma\rho^k$. By taking advantage of Theorem 2 and similar manners as in [22], we are able to obtain the boundedness of $\{\rho^k\}$.

In addiction, 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. As mention in [15, Theorem 24], the isolated calmness of S_{KKT} is equivalent to SOSC and SRCQ [5, (4.125)] under RCQ [5, Definition 2.86]. Thus, the result obtained in Theorem 2 reduces to [29, Theorem 4.2] for NLSDP.

Remark 5 In Theorem 2, if $\bar{\lambda} \in \operatorname{rbd}\mathcal{M}(\bar{x})$ and Assumption 1 holds, we have shown that the primal-dual sequence $\{(x^k,\lambda^k)\}_{k\ge 0}$ generated by Algorithm 1 converge to $(\bar{x},\widehat{\lambda})$ for some $\widehat{\lambda} \in \mathcal{M}(\bar{x})$ if $\lambda^k \notin \mathcal{M}(\bar{x})$ for sufficiently large k. It remains unknown from our approach whether the primal sequence $\{x^k\}_{k\ge0}$ converges to $\bar{x}$ linearly if the dual sequence $\{\lambda^k\}_{k\ge0}$ terminates finitely. This is a future work we are working on.

5 A sufficient condition for the semi-isolated calmness of S_{KKT} (27)

In this section, we give a sufficient condition for the semi-isolated calmness of KKT pair. In order to reach the goal, we need the definition of bounded linear regularity of a collection of closed convex sets, which can be found in, e.g., [2, Definition 5.6].

Definition 5 Let $D_1, D_2, \ldots, D_m \subseteq \mathcal{X}$ be closed convex sets for some positive integer m. Suppose that $D := D_1 \cap D_2 \cap \ldots \cap D_m$ is non-empty. The collection $\{D_1, D_2, \ldots, D_m\}$ is said to be boundedly linearly regular if for every bounded set $\mathcal{B} \subseteq \mathcal{X}$, there exists a constant $\kappa > 0$ such that

$$\operatorname{dist}(x,D) \leqslant \kappa \max\left\{\operatorname{dist}(x,D_1),\ldots,\operatorname{dist}(x,D_m)\right\}, \forall x \in \mathcal{B}.$$

A sufficient condition to guarantee the property of bounded linear regularity is established in [3, Corollary 3]. Denote $\mathcal{G}_1(\bar{x}) = \{(y,\Gamma) \in \mathcal{Y}\times\mathcal{S}^n \mid \nabla f(\bar{x}) + \nabla h(\bar{x})^* y + \nabla G(\bar{x})^*\Gamma = 0\}$ and $\mathcal{G}_2(\bar{x}) = \{(y,\Gamma) \in \mathcal{Y}\times\mathcal{S}^n \mid \Gamma \in \mathcal{N}_{\mathcal{S}^n_+}(G(\bar{x}))\}$. It is easy to see that $\mathcal{G}_1(\bar{x})$ is a polyhedron and $\mathcal{G}_2(\bar{x})$ is convex. Along with [28, Theorem 3.1], the following result gives a sufficient condition for semi-isolated calmness. Its proof is inspired from [38, Theorem 5.9].

Theorem 3 *Let $\bar{x} \in \mathcal{X}$ be a stationary point to the NLSDP* (26) *with* $(a_1,a_2,b) = (0,0,0)$ *and* $(\bar{y},\overline{\Gamma}) \in \mathcal{M}(\bar{x})$. *Suppose SOSC holds at* $(\bar{x},\bar{y},\overline{\Gamma})$ *and*

$$\mathcal{G}_1(\bar{x}) \cap \operatorname{ri}\mathcal{G}_2(\bar{x}) \neq \varnothing. \tag{71}$$

Then there exist a constant $\kappa_1 > 0$, neighborhoods $\mathbb{V} := \mathbb{B}_{r_1}(\bar{x},\bar{y},\overline{\Gamma})$ of $(\bar{x},\bar{y},\overline{\Gamma})$ and $\mathbb{U}$ of $(0,0,0)$ *such that for any $(a_1,a_2,b) \in \mathbb{U}$,*

$$\|x-\bar{x}\| + \operatorname{dist}((y,\Gamma),\mathcal{M}(\bar{x})) \le \kappa_1\|(a_1,a_2,b)\| \quad \forall\,(x,y,\Gamma) \in S_{\mathrm{KKT}}(a_1,a_2,b)\cap\mathbb{V}.$$

Proof. For the given $\bar{x}$, define $S(a,b) := \{(y,\Gamma) \in \mathcal{Y}\times\mathcal{S}^n \mid a = \nabla_x L(\bar{x},y,\Gamma),\ b + G(\bar{x}) \in N^{-1}_{\mathcal{S}^n_+}(\Gamma)\}$. It is clear that $S(0,0) \equiv \mathcal{M}(\bar{x})$. From [3, Corollary 3] and (71), we know boundedly linearly regular holds for the collection of sets $\{\mathcal{G}_1(\bar{x}),\mathcal{G}_2(\bar{x})\}$. From the definition of boundedly linearly regular, we know that for all $(y,\Gamma) \in \mathbb{B}_r(\bar{y},\overline{\Gamma})$ with a given $r > 0$, there exists a constant $\hat{\kappa}$ such that

$$\begin{aligned}\operatorname{dist}\big((y,\Gamma),S(0,0)\big) &= \operatorname{dist}\big((y,\Gamma),\mathcal{G}_1(\bar{x})\cap\mathcal{G}_2(\bar{x})\big)\\ &\le \hat{\kappa}\big(\operatorname{dist}((y,\Gamma),\mathcal{G}_1(\bar{x})) + \operatorname{dist}((y,\Gamma),\mathcal{G}_2(\bar{x}))\big).\end{aligned}$$

By Hoffman's error bound [24], we have

$$\operatorname{dist}\big((y,\Gamma),S(0,0)\big)\le\hat\kappa\big(\|\nabla_x L(\bar x,y,\Gamma)\|+\operatorname{dist}(\Gamma,N_{\mathcal{S}^n_+}(G(\bar x)))\big).$$

It follows from the metric subregularity of $N_{\mathcal{S}^n_-}$ with $N^{-1}_{\mathcal{S}^n_+}=N_{\mathcal{S}^n_-}$ (or use [12, Proposition 14] directly) that there exists $\kappa'>0$ such that

$$\begin{aligned}\operatorname{dist}\big((y,\Gamma),S(0,0)\big)&\le\hat\kappa\big(\|\nabla_x L(\bar x,y,\Gamma)\|+\kappa'\operatorname{dist}(G(\bar x),N^{-1}_{\mathcal{S}^n_+}(\Gamma))\big)\\&\le\tilde\kappa\big(\|\nabla_x L(\bar x,y,\Gamma)\|+\operatorname{dist}(G(\bar x),N^{-1}_{\mathcal{S}^n_+}(\Gamma))\big),\end{aligned}\tag{72}$$

where $\tilde\kappa=\max\{\hat\kappa,\hat\kappa\kappa'\}$.

Claim. There are numbers $\varepsilon'>0,\ell'\ge 0$ and neighborhood $\mathbb{V}$ of $0\in\mathcal{X}\times\mathcal{Y}\times\mathcal{S}^n$ such that for any $(a_1,a_2,b)\in\mathbb{U}$ and any $(x,y,\Gamma)\in S_{\rm KKT}(a_1,a_2,b)\cap\mathbb{B}_{\varepsilon'}(\bar x,\bar y,\overline{\Gamma})$ we have the estimate

$$\|x-\bar x\|\le\ell'(\|a_1\|+\|a_2\|+\|b\|).\tag{73}$$

To prove this claim, suppose on the contrary that (73) fails, i.e., for any $k\in\mathbb{N}$ there are $((a_1)_k,(a_2)_k,b_k)\in\mathbb{B}_{1/k}(0)$ and $(x_k,y_k,\Gamma_k)\in S_{\rm KKT}\left((a_1)_k,(a_2)_k,b_k\right)\cap\mathbb{B}_{1/k}(\bar x,\bar y,\overline{\Gamma})$ satisfying

$$\frac{\|x_k-\bar x\|}{\|(a_1)_k\|+\|(a_2)_k\|+\|b_k\|}\to\infty\text{ as }k\to\infty,$$

which yields $(a_1)_k=o(\|x_k-\bar x\|)$, $(a_2)_k=o(\|x_k-\bar x\|)$ and $b_k=o(\|x_k-\bar x\|)$. Letting $X_k=G(x_k)+b_k$, it follows that $(X_k,\Gamma_k)\in\operatorname{gph}N_{\mathcal{S}^n_+}$. From (72), we know that for sufficiently large k, the estimate

$$\operatorname{dist}\big((y_k,\Gamma_k),S(0,0)\big)\le\tilde\kappa\big(\|\nabla_x L(\bar x,y_k,\Gamma_k)\|+\operatorname{dist}(G(\bar x),\mathcal{N}^{-1}_{\mathcal{S}^n_+}(\Gamma_k))\big)$$

holds. It follows that

$$\begin{aligned}&\operatorname{dist}\big((y_k,\Gamma_k),S(0,0)\big)\le\tilde\kappa\|\nabla_x L(\bar x,y_k,\Gamma_k)\|+\tilde\kappa\|G(\bar x)-G(x_k)-b_k\|\\&\le\tilde\kappa(\|\lambda_k\|\cdot\|\nabla\Phi(x_k)-\nabla\Phi(\bar x)\|+\|\nabla\Phi(x_k)^*\lambda_k+\nabla f(x_k)\|\\&\quad+\|\nabla f(x_k)-\nabla f(\bar x)\|+\|G(\bar x)-G(x_k)\|+\|b_k\|)\\&\le\tilde\kappa(l\|\lambda_k\|\cdot\|x_k-\bar x\|+\|(a_1)_k\|+2l\|x_k-\bar x\|+\|b_k\|),\end{aligned}\tag{74}$$

where l is the max Lipschitz constant for G, ∇f and $\nabla\Phi$ at $\bar x$. Thus there is $(y'_k,\Gamma'_k)\in\mathcal{M}(\bar x)=S(0,0)$ such that the sequence $\frac{(y_k,\Gamma_k)-(y'_k,\Gamma'_k)}{\|x_k-\bar x\|}$ is bounded. Then we have a convergent subsequence

$$\eta_k:=\frac{(y_k,\Gamma_k)-(y'_k,\Gamma'_k)}{\|x_k-\bar x\|}\to\eta\text{ as }k\to\infty\text{ with some }\eta\in\Re^e\times\mathcal{S}^n.\tag{75}$$

Taking a subsequence if necessary, we have

$$\xi_k:=\frac{x_k-\bar x}{\|x_k-\bar x\|}\to\xi\text{ as }k\to\infty.$$

Denote $s_k:=\|x_k-\bar x\|$ and hence deduce from $(x_k,y_k,\Gamma_k)\in S_{\rm KKT}\left((a_1)_k,(a_2)_k,b_k\right)$ that

$$o(s_k)=\nabla_x L(x_k,y_k,\Gamma_k),\quad h(x_k)=(a_2)_k\quad\text{and}\quad\Gamma_k\in N_{\mathcal{S}^n_+}(X_k).$$

Combining this with (75) lead us to

$$o(s_k)=(a_1)_k=\nabla_x L(x_k,y_k,\Gamma_k)=\nabla_x L(x_k,\bar y,\overline{\Gamma})-\nabla_x L(\bar x,\bar y,\overline{\Gamma})+\nabla\Phi(x_k)^*(\lambda_k-\bar\lambda).$$

Recalling [5, Page 241], the reduced problem takes this form:

$$\min_{x\in\mathcal{X}} f(x) \quad \text{subject to} \quad h(x)=0,\ \Xi(G(x))\in\mathcal{C},$$

where $\mathcal{C}$ is the reduced cone of $\mathcal{S}_+^n$ and Ξ is the corresponding reduced function. We denote the Lagrangian function of reduced problem as $\mathcal{L}(x,y,\mu)=f(x)+\langle y,h(x)\rangle+\langle\mu,\Xi(G(x))\rangle$. It is easy to see that $\nabla_x L(x,\bar y,\overline{\Gamma})=\nabla_x\mathcal{L}(x,\bar y,\bar\mu)$, where $\nabla\Xi(G(\bar x))^*\bar\mu=\overline{\Gamma}$. Suppose $\Phi(x)=(h(x),G(x))\in\mathcal{Y}\times\mathcal{S}_+^n$. Since $(\bar y,\overline{\Gamma})$, $(y_k',\Gamma_k')\in\mathcal{M}(\bar x)$, we know that $\nabla_x L(\bar x,\bar y,\overline{\Gamma})=\nabla_x L(\bar x,y_k',\Gamma_k')$. It follows that $\nabla\Phi(\bar x)^*\big((y_k',\Gamma_k')-(\bar y,\overline{\Gamma})\big)=0$. Then we have

$$\begin{aligned}
o(s_k)=(a_1)_k &= \nabla_x\mathcal{L}(x_k,\bar y,\bar\mu)-\nabla\mathcal{L}_x(\bar x,\bar y,\bar\mu)+\nabla\Phi(x_k)^*\big((y_k,\Gamma_k)-(\bar y,\overline{\Gamma})\big)\\
&= \nabla_{xx}^2\mathcal{L}(\bar x,\bar y,\bar\mu)(x_k-\bar x)+\nabla\Phi(\bar x)^*\big((y_k,\Gamma_k)-(\bar y,\overline{\Gamma})\big)+o(s_k)\\
&= \nabla_{xx}^2\mathcal{L}(\bar x,\bar y,\bar\mu)(x_k-\bar x)+\nabla\Phi(\bar x)^*\big((y_k,\Gamma_k)-(y_k',\Gamma_k')\big)+o(s_k),
\end{aligned}$$

which in turn yields the equality

$$\nabla_{xx}^2\mathcal{L}(\bar x,\bar y,\bar\mu)\xi+\nabla\Phi(\bar x)^*\eta=0.$$

Recalling that $(y_k',\Gamma_k')\in N_{\{0\}\times\mathcal{S}_+^n}\big(h(\bar x),G(\bar x)\big)$ and $(y_k,\Gamma_k)\in\mathcal{N}_{\{0\}\times\mathcal{S}_+^n}\big(h(x_k)-(a_2)_k,X_k\big)$. It follows from the monotonicity of normal cone mappings to convex sets that

$$0\le\left\langle\frac{\Gamma_k-\Gamma_k'}{s_k},\frac{X_k-G(\bar x)}{s_k}\right\rangle+\langle\frac{y_k-y_k'}{s_k},\frac{h(x_k)-(a_2)_k-h(\bar x)}{s_k}\rangle.$$

This therefore implies that

$$\langle\eta,\nabla\Phi(\bar x)\xi\rangle\ge 0.$$

Thus we have

$$0=\langle 0,\xi\rangle=\langle\nabla_{xx}^2\mathcal{L}(\bar x,\bar y,\bar\mu)\xi,\xi\rangle+\langle\eta,\nabla\Phi(\bar x)\xi\rangle\ge\langle\nabla_{xx}^2\mathcal{L}(\bar x,\bar y,\bar\mu)\xi,\xi\rangle. \tag{76}$$

It is worth to note that SOSC corresponds to the reduced second-order condition

$$\left\langle\nabla_{xx}^2\mathcal{L}(\bar x,\bar y,\bar\mu)\xi,\xi\right\rangle>0 \quad \forall\, 0\ne\xi\in\mathcal{C}(\bar x).$$

Then we prove $\nabla G(\bar x)\xi\in\mathcal{C}_{\mathcal{S}_+^n}(G(\bar x),\overline{\Gamma})$ and $\nabla h(\bar x)\xi=0$, which implies $\xi\in\mathcal{C}(\bar x)$ from [5, (3.271), Proposition 3.10] and thus leads to a contradiction.

Since $\mathcal{S}_+^n$ is a closed convex cone, it follows from $(X_k,\Gamma_k)\in\operatorname{gph}N_{\mathcal{S}_+^n}$ that $X_k\in\mathcal{S}_+^n$ and $\langle X_k,\Gamma_k\rangle=0$. Thus, we have

$$0=\langle X_k,\Gamma_k\rangle=\langle G(x_k)+b_k,\Gamma_k\rangle.$$

Since $\langle G(\bar x),\Gamma_k\rangle\le 0$, we know the above equation can be written as

$$0\le\langle G(x_k)+b_k-G(\bar x),\Gamma_k\rangle=\langle o(s_k)+\nabla G(\bar x)(x_k-\bar x),\Gamma_k\rangle,$$

and hence $\langle\nabla G(\bar x)\xi,\overline{\Gamma}\rangle\ge 0$. We have that $G(\bar x)+s_k(\nabla G(\bar x)\xi_k+o(s_k)/s_k)=X^k\in\mathcal{S}_+^n$, which implies $\nabla G(\bar x)\xi\in T_{\mathcal{S}_+^n}(G(\bar x))$. It follows that $\langle\nabla G(\bar x)\xi,\overline{\Gamma}\rangle=0$ as $\overline{\Gamma}\in N_{\mathcal{S}_+^n}(G(\bar x))$. Also, we have $(a_2)_k=h(x_k)-h(\bar x)=\nabla h(\bar x)(x_k-\bar x)+o(s_k)$, it follows that $0=\nabla h(\bar x)\xi$. Combining this, SOSC and (76) together, we have obtained the contradiction and thus completed the proof of this claim.

Taking the obtained $\mathbb{U}$ from the above claim. Following the same proof as (74), we know that there exist a constant $\kappa''>0$ and a neighborhood $\mathbb{B}_{\varepsilon''}(\bar x,\bar y,\overline{\Gamma})$ of $(\bar x,\bar y,\overline{\Gamma})$ such that for any $(a_1,a_2,b)\in\mathbb{U}$ and $(x,y,\Gamma)\in S_{\mathrm{KKT}}(a_1,a_2,b)\cap\mathbb{B}_{\varepsilon''}(\bar x,\bar y,\overline{\Gamma})$,

$$\operatorname{dist}((y,\Gamma),\mathcal{M}(\bar x))\le\kappa''(l\|\lambda\|\cdot\|x-\bar x\|+\|a_1\|+2l\|x-\bar x\|+\|b\|).$$

Combining with the claim and let $r_1 = \min\{r, \varepsilon', \varepsilon''\}$, we have completed the whole proof. $\square$

We can also study the validity of semi-isolated calmness directly from [38, 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 $(a, b) = (0, 0)$, where $S(a, b) = \{(y, \Gamma) \in \mathcal{Y} \times \mathcal{S}^n_+ \mid \nabla_x L(\bar{x}, y, \Gamma) = a,\ \Gamma \in \mathcal{N}_{\mathcal{S}^n_+}(G(\bar{x}) + b)\}$. It is easy to see that the strong regularity/isolated calmness/Aubin of S implies the calmness of S. (The definition of calmness, isolated calmness and Aubin can be seen in e.g., [62,50] and [17]).

To end this paper, we propose two examples to illustrate that the conditions required in Theorem 2 can be satisfied indeed.

Example 1

$$\begin{aligned} &\min\ \tfrac{1}{2}x^3 \\ &\text{s.t}\ \ -x^2 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \in \mathcal{S}^3_+, \quad \Leftarrow \Gamma \end{aligned}$$

It is easy to see that the optimal solution is $\bar{x} = 0$, its corresponding multiplier set is $\mathcal{M}(\bar{x}) = \{\Gamma \mid \Gamma \in \mathcal{S}^3_-\}$. Pick $\overline{\Gamma} = \mathrm{Diag}(0, -1, -2)$. Then for all $\Gamma \in \mathbb{B}_{\min\{1/3, r_1\}}(\overline{\Gamma}) \backslash \mathcal{M}(\bar{x})$ with r_1 taken from Definition 3 with $\mathbb{V} = \mathbb{B}_{r_1}(\bar{x}, \bar{\lambda})$, 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 and [28, Theorem 3.1] that

$$\|\Pi_{\mathcal{M}(\bar{x})}(\Gamma) - \widehat{\Gamma}\| = \mathrm{dist}(\Gamma, \mathcal{M}(\bar{x})) = O(R(x, \Gamma)).$$

Thus Assumption 1 is satisfied. It is easy to calculate $\nabla^2_{xx} L(\bar{x}, \overline{\Gamma}) = 2 > 0$, 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.

We also provide another nontrivial NLSDP example, which is modified from the example proposed in the arxiv version of [10, Example 2] for different purpose.

Example 2 Consider the following example

$$\begin{aligned} &\min\ \tfrac{1}{2}x^2 + 2t \\ &\text{s.t}\ \ tA - x^2 I_2 \in \mathcal{S}^2_+, \quad \Leftarrow \Gamma \\ &\qquad t \geq 0, \qquad\qquad \Leftarrow y \end{aligned}$$

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

$$\mathcal{M}(\bar{t}, \bar{x}) = \{(\Gamma, y) \in \mathcal{S}^2_- \times \Re \mid \langle A, -\Gamma \rangle \leq 2\}.$$

We can pick $\bar{y} = 0$ and $\overline{\Gamma} = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$. For all $\Gamma \in \mathbb{B}_{\min\{r_1, 1/(2\sqrt{10})\}}(\overline{\Gamma})$ with r_1 taken from Definition 3 with $\mathbb{V} = \mathbb{B}_{r_1}(\bar{x}, \bar{\lambda})$, 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 $\Gamma = Q\mathrm{Diag}(\Gamma_1, \Gamma_2)Q^T$ with $Q \in \mathcal{O}^2(\Gamma)$. Let $\widehat{\Gamma} = Q\mathrm{Diag}(0, \Gamma_2)Q^T$. Then we have

$$\|\Pi_{\mathcal{M}(\bar{x})}(\Gamma) - \widehat{\Gamma}\| = \mathrm{dist}(\Gamma, \mathcal{M}(\bar{x})) = O(R(x, \Gamma)),$$

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 [3, Corollary 3]. It is easy to check that SOSC holds since $\nabla^2 L(\bar{t}, \bar{x}, \overline{\Gamma}, \bar{y}) = \begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix}$ 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.

6 Conclusions

In this paper, 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}. In the future, we will mainly focus on the following three ongoing works. Firstly, as [10] 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. Inspired by work [52], which extends the convex framework of ALM convergence to the non-convex case by variational sufficiency, we see hope in extending it to non-convex non-polyhedral problems. Secondly, although we consider solving the ALM subproblem inexactly, we have not put forward a practical relative error criterion for it. This is a future work we focus on and [18,10] may provide some inspirations. Thirdly, we are also working on providing sufficient (necessary) conditions to Assumption 1.

Appendix A Proof of Lemma 1

For each $k \in \{1,\dots,\bar{d}\}$, let $\Lambda_{\bar{\alpha}_k\bar{\alpha}_k} = \mathrm{Diag}(\lambda_{\bar{\alpha}_k}(A))$ and $\Xi_{\bar{\alpha}_k\bar{\alpha}_k} = \mathrm{Diag}(\lambda_{\bar{\alpha}_k}(A+H))$. We first show that (10) holds. If $\bar{d}=1$, i.e., $\lambda_1(\overline{A}) = \cdots = \lambda_n(\overline{A})$, the first equation in (10) trivially holds. Next we assume that $\bar{d} \geq 2$. From (9), we have for any $\mathcal{S}^n \ni H \to 0$,

$$\begin{bmatrix} \Lambda_{\bar{\alpha}_1\bar{\alpha}_1} & 0 & \dots & 0 \\ 0 & \Lambda_{\bar{\alpha}_2\bar{\alpha}_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Lambda_{\bar{\alpha}_{\bar{d}}\bar{\alpha}_{\bar{d}}} \end{bmatrix} U + HU = U \begin{bmatrix} \Xi_{\bar{\alpha}_1\bar{\alpha}_1} & 0 & \dots & 0 \\ 0 & \Xi_{\bar{\alpha}_2\bar{\alpha}_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Xi_{\bar{\alpha}_{\bar{d}}\bar{\alpha}_{\bar{d}}} \end{bmatrix}.$$

It follows from $A \in \mathbb{B}_r(\overline{A})$ that $\lambda_i(A) \neq \lambda_j(A)$ whenever $i \in \bar{\alpha}_l$, $j \in \bar{\alpha}_k$ with $k \neq l$. It is easy to see that for all $i \in \bar{\alpha}_k$, $j \in \bar{\alpha}_l$ with $k \neq l$, $U_{ij} = \frac{(HU)_{ij}}{\Lambda_{ii}-\Xi_{jj}}$. Then we have for all $\|H\| \leq \delta := r/6$,

$$\frac{\|U_{\bar{\alpha}_k\bar{\alpha}_l}\|}{\|H\|} \leq \sum_{i\in\bar{\alpha}_k, j\in\bar{\alpha}_l} \frac{1}{(\Lambda_{ii}-\Xi_{jj})^2} \leq \sum_{i\in\bar{\alpha}_k, j\in\bar{\alpha}_l} \frac{1}{(|v_i(\overline{A})-v_j(\overline{A})|-2r-2\delta)^2} := l,$$

where δ and l are independent of A. Hence we obtain that

$$U_{\bar{\alpha}_k\bar{\alpha}_l} = O(\|H\|) \quad \forall\, 1 \leq k \neq l \leq \bar{d},$$

where $O(\|H\|)$ is uniform for all $A \in \mathbb{B}_r(\overline{A})$. By using the fact that U is orthogonal, we obtain directly that the second equation in (10) holds. In order to prove (11), we consider the SVD of each $U_{\bar{\alpha}_k\bar{\alpha}_k}$, $k=1,\dots,\bar{d}$. Fix $k \in \{1,\dots,\bar{d}\}$. Let W and V be in $\mathcal{O}^{|\bar{\alpha}_k|}$ such that $U_{\bar{\alpha}_k\bar{\alpha}_k} = W\Sigma V^T$, where Σ is a nonnegative diagonal matrix. From (10), we obtain that for all $A \in \mathbb{B}_r(\overline{A})$,

$$W\Sigma^2 W^T = I_{|\bar{\alpha}_k|} + O(\|H\|^2),$$

which is equivalent to

$$\Sigma^2 = W^T W + O(\|H\|^2) = I_{|\bar{\alpha}_k|} + O(\|H\|^2).$$

Since Σ is a nonnegative diagonal matrix, we may conclude that

$$\Sigma = \mathrm{Diag}(1+O(\|H\|^2)\dots 1+O(\|H\|^2)).$$

Therefore, from $U_{\bar{\alpha}_k\bar{\alpha}_k} = W\Sigma V^T$, we have $U_{\bar{\alpha}_k\bar{\alpha}_k} = WV^T + O(\|H\|^2)$. Since $WV^T \in \mathcal{O}^{|\bar{\alpha}_k|}$, we know that for all $A \in \mathbb{B}_r(\overline{A})$, (11) holds. Next, we shall show (12) holds. For each $k \in \{1,\ldots,\bar{d}\}$ by comparing the k-th diagonal block of both sides of (9), we obtain that

$$U_{\bar{\alpha}_k}^T(\Lambda(A)+H)U_{\bar{\alpha}_k} = \Xi_{\bar{\alpha}_k\bar{\alpha}_k}. \tag{77}$$

Fix $k \in \{1,\ldots,\bar{d}\}$. From (10) and (77), we know that

$$\begin{aligned} U_{\bar{\alpha}_k}^T\Lambda(A)U_{\bar{\alpha}_k} &= \begin{bmatrix} O(\|H\|) \; U_{\bar{\alpha}_k\bar{\alpha}_k} \; O(\|H\|) \end{bmatrix} \begin{bmatrix} \Lambda_1(A) & 0 & 0 \\ 0 & \Lambda(A)_{\bar{\alpha}_k\bar{\alpha}_k} & 0 \\ 0 & 0 & \Lambda_2(A) \end{bmatrix} \begin{bmatrix} O(\|H\|) \\ U_{\bar{\alpha}_k\bar{\alpha}_k} \\ O(\|H\|) \end{bmatrix} \\ &= O(\|H\|^2)\Lambda_1(A) + U_{\bar{\alpha}_k\bar{\alpha}_k}^T\Lambda(A)_{\bar{\alpha}_k\bar{\alpha}_k}U_{\bar{\alpha}_k\bar{\alpha}_k} + O(\|H\|^2)\Lambda_2(A). \end{aligned}$$

It follows that

$$\Xi_{\bar{\alpha}_k\bar{\alpha}_k} - \big(O(\|H\|^2)\Lambda_1(A) + U_{\bar{\alpha}_k\bar{\alpha}_k}^T\Lambda(A)_{\bar{\alpha}_k\bar{\alpha}_k}U_{\bar{\alpha}_k\bar{\alpha}_k} + O(\|H\|^2)\Lambda_2(A)\big) = U_{\bar{\alpha}_k\bar{\alpha}_k}^T H_{\bar{\alpha}_k\bar{\alpha}_k} U_{\bar{\alpha}_k\bar{\alpha}_k} + O(\|H\|^2).$$

Since $U_{\bar{\alpha}_k\bar{\alpha}_k} = Q_k + O(\|H\|^2)$ and $\|\Lambda(A)\| \le \|\Lambda(\overline{A})\| + r$, we obtain that

$$Q_k^T H_{\bar{\alpha}_k\bar{\alpha}_k} Q_k = \Xi_{\bar{\alpha}_k\bar{\alpha}_k} - Q_k^T\Lambda(A)_{\bar{\alpha}_k\bar{\alpha}_k}Q_k + O(\|H\|^2).$$

Hence (12) holds with the uniform $O(\|H\|^2)$ for all $A \in \mathbb{B}_r(\overline{A})$. The proof is completed.

Appendix B Proof of Proposition 1

Firstly, we show (14) 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

$$\Lambda(A) + H = U\Lambda(Z)U^T. \tag{78}$$

Let $\delta > 0$ be any fixed number such that $0 \le \delta \le \frac{\lambda_{|\alpha|}(\overline{A})-r}{2}$ if $\alpha \neq \varnothing$ and be any fixed positive number otherwise. Then, define the following continuous scalar function

$$f(t) := \begin{cases} t & \text{if} \quad t > \delta \\ 2t-\delta & \text{if} \quad \frac{\delta}{2} < t < \delta \\ 0 & \text{if} \quad t < \frac{\delta}{2}. \end{cases}$$

Therefore, we have

$$\big\{\lambda_1(A),\ldots,\lambda_{|\alpha|}(A)\big\} \in (\delta,+\infty) \quad \text{and} \quad \big\{\lambda_{|\alpha|+1}(A),\ldots,\lambda_n(A)\big\} \in (-\infty,\frac{\delta}{2}).$$

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 $(-\infty,\frac{\delta}{2}) \cup (\delta,+\infty)$, It is well-known that for H sufficiently close to zero,

$$F(A+H) - F(A) - F'(A)H = O(\|H\|^2) \tag{79}$$

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 $A \in \mathbb{B}_r(\overline{A})$ 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

$$R'(A;H) = \Pi'_{\mathcal{S}^n_+}(A;H) - F'(A)H = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \Pi_{\mathcal{S}^{|\beta|}_+}(H_{\beta\beta}) & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

By the Lipschitz continuity of $\lambda(\cdot)$, we know that for H sufficiently close to zero,

$$\big\{\lambda_1(Z),\ldots,\lambda_{|\alpha|}(Z)\big\} \in (\delta,+\infty), \quad \big\{\lambda_{|\alpha|+1}(Z),\ldots,\lambda_{|\beta|}(Z)\big\} \in (-\infty,\frac{\delta}{2})$$

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,

$$R(A+H) = \Pi_{\mathcal{S}^n_+}(A+H) - F(A+H) = U \begin{bmatrix} 0 & 0 & 0 \\ 0 & \Pi_{\mathcal{S}^{|\beta|}_+}(\Lambda(Z)_{\beta\beta}) & 0 \\ 0 & 0 & 0 \end{bmatrix} U^T.$$

Since $U \in \mathcal{O}^n(Z)$, we know from Lemma 1 that for any $\mathcal{S}^n \ni H \to 0$, there exists an orthogonal matrix $Q \in \mathcal{O}^{|\beta|}$ such that

$$U_\beta = \begin{bmatrix} O(\|H\|) \\ U_{\beta\beta} \\ O(\|H\|) \end{bmatrix} \quad \text{and} \quad U_{\beta\beta} = Q + O(\|H\|^2), \tag{80}$$

Therefore, by noting that $\Pi_{\mathcal{S}^{|\beta|}_+}(\Lambda(Z)_{\beta\beta}) = O(\|H\|)$ and $O(\|H\|)$ is uniform for $A \in \mathbb{B}_r(\overline{A})$ with $\pi(\overline{A}) = \pi(A)$, we obtain from the above discussion that

$$R(A+H) - R(A) - R'(A;H) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & Q\Pi_{\mathcal{S}^{|\beta|}_+}(\Lambda(Z)_{\beta\beta})Q^T - \Pi_{\mathcal{S}^{|\beta|}_+}(H_{\beta\beta}) & 0 \\ 0 & 0 & 0 \end{bmatrix} + O(\|H\|^2)$$

By (78) and (80), 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 $A \in \mathbb{B}_r(\overline{A})$ with $\pi(\overline{A}) = \pi(A)$. Combining this with the globally Lipschitz continuity of $\Pi_{\mathcal{S}^{|\beta|}_+}(\cdot)$ and $\Pi_{\mathcal{S}^{|\beta|}_+}(Q\Lambda(Z)_{\beta\beta}Q^T) = Q\Pi_{\mathcal{S}^{|\beta|}_+}(\Lambda(Z)_{\beta\beta})Q^T$, we obtain that

$$Q\Pi_{\mathcal{S}^{|\beta|}_+}(\Lambda(Z)_{\beta\beta})Q^T - \Pi_{\mathcal{S}^{|\beta|}_+}(H_{\beta\beta}) = O(\|H\|^2)\,.$$

Therefore,

$$R(A+H) - R(A) - R'(A;H) = O(\|H\|^2)\,. \tag{81}$$

By combining (79) and (81), we know that for any $\mathcal{S}^n \ni H \to 0$,

$$\Pi_{\mathcal{S}^n_+}(\Lambda(A)+H) - \Pi_{\mathcal{S}^n_+}(\Lambda(A)) - \Pi'_{\mathcal{S}^n_+}(\Lambda(A);H) = O(\|H\|^2) \tag{82}$$

and $O(\|H\|^2)$ is uniform for $A \in \mathbb{B}_r(\overline{A})$ with $\pi(\overline{A}) = \pi(A)$. Next, we consider the case that $A = P^T\Lambda(A)P$. Re-write (78) as

$$\Lambda(A) + P^T HP = P^T U\Lambda(Z)U^T P\,.$$

Let $\widetilde{H} := P^T HP$. Then, we have $\Pi_{\mathcal{S}^n_+}(A+H) = P\,\Pi_{\mathcal{S}^n_+}(\Lambda(A)+\widetilde{H})P^T$. Therefore, since $P \in \mathcal{O}^n$, we know from (82) and (5) that for any $\mathcal{S}^n \ni H \to 0$, (14) holds.

Appendix C Proof of Lemma 2

We first consider the case where A is diagonal. For notational simplicity, let $\Lambda = \Lambda(A)$ and $\Xi = \Lambda(A+H)$. From (13), we have $AU + HU = U\Xi$, which implies

$$\Lambda_{\bar{\alpha}_k\bar{\alpha}_k}U_{\bar{\alpha}_k\bar{\alpha}_l} + (HU)_{\bar{\alpha}_k\bar{\alpha}_l} = U_{\bar{\alpha}_k\bar{\alpha}_l}\Xi_{\bar{\alpha}_l\bar{\alpha}_l}\,.$$

It follows that

$$\Lambda_{\bar{\alpha}_k\bar{\alpha}_k}U_{\bar{\alpha}_k\bar{\alpha}_l} + \sum_{j=1}^{\bar{d}} H_{\bar{\alpha}_k\bar{\alpha}_j}U_{\bar{\alpha}_j\bar{\alpha}_l} = U_{\bar{\alpha}_k\bar{\alpha}_l}\Xi_{\bar{\alpha}_l\bar{\alpha}_l}\,.$$

This, together with Lemma 1 shows that

$$U_{\bar{\alpha}_k\bar{\alpha}_l} = \Sigma_{kl}\circ\sum_{j=1}^{\bar{d}} H_{\bar{\alpha}_k\bar{\alpha}_j}U_{\bar{\alpha}_j\bar{\alpha}_l} = \Sigma_{kl}\circ H_{\bar{\alpha}_k\bar{\alpha}_l}Q_l + O(\|H\|^2) \tag{83}$$

where $(\Sigma_{kl})_{ij} = 1/((\Xi_{\bar{\alpha}_l\bar{\alpha}_l})_i - (\Lambda_{\bar{\alpha}_k\bar{\alpha}_k})_j)$. It is easy to see that $1/\big((\Xi_{\bar{\alpha}_l\bar{\alpha}_l})_i - (\Lambda_{\bar{\alpha}_k\bar{\alpha}_k})_j\big) = 1/\big((\Lambda_{\bar{\alpha}_l\bar{\alpha}_l})_i - (\Lambda_{\bar{\alpha}_k\bar{\alpha}_k})_j\big) + O(\|H\|)$. Combining this with (83), we have

$$U_{\bar{\alpha}_k\bar{\alpha}_l} = \Theta_{kl}\circ H_{\bar{\alpha}_k\bar{\alpha}_l}Q_l + O(\|H\|^2),\ \text{with } O(\|H\|^2) \text{ uniform for all } A \in \mathbb{B}_r(\overline{A}).$$

Next we consider $A = P^T\Lambda(A)P^T$. Re-write (13) as

$$\Lambda(A) + P^T HP = P^T U\Lambda(A+H)U^T P\,.$$

Let $\widetilde{H} := P^T HP$. Since P is an orthogonal matrix, the following proof is the same as the diagonal case. Thus we have completed the proof.

References

1. K. J. Arrow, R. M. Solow, Gradient methods for constrained maxima with weakened assumptions. In: Arrow, K.J., Hurwicz, L., Uzawa, H., (eds.) Studies in Linear and Nonlinear Programming. Stanford University Press, Stanford, (1958) 165–176 .
2. H. H. Bauschke, J. M. Borwein, On projection algorithms for solving convex feasibility problems, SIAM Review, 38 (1996) 367–426.
3. H. H. Bauschke, J. M. Borwein, W. Li, Strong conical hull intersection property, bounded linear regularity, Jameson's property (G), and error bounds in convex optimization, Mathematical Programming, 86 (1999) 135–160.
4. D. Bertsekas, Constrained Optimization and Lagrange Multipliers Methods, Academic Press, New York, 1982.
5. J. F. Bonnans , A. Shapiro, Perturbation analysis of optimization problems, Springer, New York, 2000.
6. J. D. Buys, Dual Algorithms for Constrained Optimization Problems, Doctoral dissertation, University of Leiden, Leiden, the Netherlands, 1972.

7. F. H. Clarke, Optimization and Nonsmooth Analysis, Wiley-Interscience, New York, 1983.
8. A. R. Conn, N. I. M. Gould, Ph. Toint, A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds, SIAM Journal on Numerical Analysis, 28 (1991) 545-572.
9. L. Contesse-Becker, Extended convergence results for the method of multipliers for non-strictly binding inequality constraints, Journal of Optimization Theory and Applications, 79 (1993) 273-310.
10. Y. Cui, D. F. Sun, K. C. Toh, On the R-superlinear convergence of the KKT residuals generated by the augmented Lagrangian method for convex composite conic programming, Mathematical Programming, 178 (2019) 381-415. (arXiv:1610.00875, 2016)
11. Y. Cui, C. Ding, Nonsmooth composite matrix optimization: strong regularity, constraint nondegeneracy and beyond, July 2019, arXiv: 1907.13253.
12. Y. Cui, C. Ding, X. Y. Zhao, Quadratic growth conditions for convex matrix optimization problems associated with spectral functions, SIAM Journal on Optimization, 27 (2017) 2332-2355.
13. C. Ding, An Introduction to a Class of Matrix Optimization Problems. PhD thesis, National University of Singapore. (2012).
14. C. Ding, D. F. Sun, J. J. Ye, First order optimality conditions for mathematical programs with semidefinite cone complementarity constraints, Mathematical Programming, 147 (2014) 539-579.
15. C. Ding, D. F. Sun, L. W. Zhang, Characterization of the robust isolated calmness for a class of conic programming problems, SIAM Journal on Optimization, 27 (2017) 67-90.
16. A. L. Dontchev, R. T. Rockafellar, Characterizations of Lipschitz stability in nonlinear programming, Mathematical Programming With Data Perturbations, Marcel Dekker, New York, (1997) 65-82.
17. A. L. Dontchev, R. T. Rockafellar, Implicit Functions and Solution Mappings, Springer, Heidelberg, 2014.
18. J. Eckstein, P. J. S. Silva, A practical relative error criterion for augmented Lagrangians, Mathematical Programming, 141 (2013) 319-348.
19. D. Fernandez, M. V. Solodov, Local convergence of exact and inexact augmented Lagrangian methods under the second-order sufficient optimality condition, SIAM Journal on Optimization, 22 (2012) 384-407.
20. E. G. Golshtein, N. V. Tretyakov, Modified Lagrangians and Monotone Maps in Optimization, Wiley, New York (1989)
21. N. T. V. Hang, B. Mordukhovich, E. Sarabi, Augmented lagrangian method for second-order cone programs under second-order sufficiency, Journal of Global Optimization, (2021) 1-31.
22. N. T. V. Hang, M. E. Sarabi, Local convergence analysis of augmented lagrangian methods for piecewise linear-quadratic composite optimization problems, arXiv preprint arXiv:2010.11379 (2020).
23. M. R. Hestenes, Multiplier and gradient methods, Journal of Optimization Theory and Applications, 4 (1969) 303-320.
24. A. J. Hoffman, On approximate solutions of systems of linear inequalities, Journal of Research of the National Bureau of Standards, 49 (1952) 263-265.
25. K. Ito, K. Kunisch, The augmented Lagrangian method for equality and inequality constraints in Hilbert spaces, Mathematical Programming, 46 (1990) 341-360.
26. A. F. Izmailov, A. S. Kurennoy, M. V. Solodov, A note on upper Lipschitz stability, error bounds, and critical multipliers for Lipschitz-continuous KKT systems, Mathematical Programming, 142 (2013) 591-604.
27. A. F. Izmailov, M. V. Solodov, Newton-Type Methods for Optimization and Variational Problems, Springer, New York, 2014.
28. C. Kanzow, D. Steck, On error bounds and multiplier methods for variational problems in Banach spaces, SIAM Journal on Control and Optimization, 56 (2018) 1716-1738.
29. C. Kanzow, D. Steck, Improved local convergence results for augmented Lagrangian methods in C^2-cone reducible constrained optimization, Mathematical Programming, 177 (2019) 425-438.
30. D. Klatte, Upper Lipschitz behavior of solutions to perturbed $C^{1,1}$ programs, Mathematical Programming, 88 (2000) 285-311.
31. F. Leibfritz, COMPleib 1.1: COnstraint Matrix-optimization Problem Library—a collection of test examples for nonlinear semidefinite programs, control system design and related problems. Technical Report, Department of Mathematics, University of Trier, Germany (2005).
32. X. D. Li, D. F. Sun, K. C. Toh, QSDPNAL: A two-phase augmented Lagrangian method for convex quadratic semidefinite programming, Mathematical Programming Computation, 10 (2018) 703-743.
33. X. D. Li, D. F. Sun, K. C. Toh, A highly efficient semismooth Newton augmented Lagrangian method for solving Lasso problems, SIAM Journal on Optimization 28 (2018) 433-458.
34. F. J. Luque, Asymptotic convergence analysis of the proximal point algorithm, SIAM Journal on Optimization, 22 (1984) 277-293.
35. A. Mohammadi, B. S. Mordukhovich and M. E. Sarabi, Parabolic Regularity in Geometric Variational Analysis, Transactions of the American Mathematical Society, 374 (2021) 1711-1763.
36. B. S. Mordukhovich, Variational Analysis and Generalized Differentiation, Vol. 1: Basic Theory, Vol. 2: Applications, Springer, Berlin, 2006.
37. B. Mordukhovich, E. Sarabi, Critical multipliers in variational systems via second-order generalized differentiation, Mathematical Programming, 169 (2018) 605-648.
38. B. Mordukhovich, E. Sarabi, Criticality of Lagrange multipliers in variational systems, SIAM Journal on Optimization, 29 (2019) 425-438.

39. T. Pennanen, Local convergence of the proximal point algorithm and multiplier methods without monotonicity, Mathematics of Operations Research, 27 (2002) 170-191.
40. R. A. Poliquin, R.T. Rockafellar, Generalized hessian properties of regularized nonsmooth functions, SIAM Journal on Optimization, 6 (1996) 1121-1137.
41. M. J. D. Powell, A method for nonlinear constraints in minimization problems, In Fletcher, R. (ed.) Optimization, pp. 283–298. Academic, New York (1969)
42. S. M. Robinson, Some continuity properties of polyhedral multifunctions, SIAM Journal on Control and Optimization, 28 (1981) 206-214.
43. R. T. Rockafellar, Convex analysis, Vol. 36. Princeton university press, 1970.
44. R. T. Rockafellar, New applications of duality in convex programming, Proceedings of the 4th Conf. on Probability, Brasov, Rumania, 1971, pp. 37-81; written version of a talk also given at the 7th International Symposium on Mathematical Programming in the Hague, 1970.
45. R. T. Rockafellar, A dual approach to solving nonlinear programming problems by unconstrained optimization, Mathematical Programming, 5 (1973) 354-373.
46. R. T. Rockafellar, The multiplier method of Hestenes and Powell applied to convex programming, Journal of Optimization Theory and Applications, 12 (1973) 555-562.
47. R. T. Rockafellar, Augmented Lagrangians and applications of the proximal point algorithm in convex programming, Mathematics of Operations Research, 1 (1976) 97-116.
48. R. T. Rockafellar, Monotone operators and the proximal point algorithm, SIAM Journal on Control and Optimization, 14 (1976) 877-898.
49. R. T. Rockafellar, Lagrange multipliers and optimality, SIAM Review, 35 (1993) 183-238.
50. R. T. Rockafellar, R. J-B. Wets, Variational analysis, Springer, Berlin, 1998.
51. R. T. Rockafellar, Augmented Lagrangians and hidden convexity in sufficient conditions for local optimality, Mathematical Programming, to appear.
52. R. T. Rockafellar, Convergence of Augmented Lagrangian Methods in Extensions Beyond Nonlinear Programming, Preprint, `https://sites.math.washington.edu/~rtr/papers/rtr258-ExtendedALM.pdf`, 2021.
53. W. Rudin, Principles of Mathematical Analysis citation third edition, McGraw-Hill, 1976.
54. A. Shapiro, J. Sun, Some properties of the augmented Lagrangian in cone constrained optimization, Mathematics of Operations Research, 29 (2004) 479-491.
55. D. F. Sun, The strong second order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications, Mathematics of Operations Research, 31 (2006) 761-776.
56. D. F. Sun, J. Sun, Semismooth matrix valued functions, Mathematics of Operations Research, 27 (2002) 150-169.
57. D. F. Sun, J. Sun, Strong Semismoothness of Eigenvalues of Symmetric Matrices and Its Application to Inverse Eigenvalue Problems. SIAM Journal on Numerical Analysis, 40 (2003) 2352-2367.
58. D. F. Sun, J. Sun, L. W. Zhang, The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming, Mathematical Programming, 114 (2008) 349-391.
59. R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society. Series B. Statistical Methodology, 58 (1996) 267-288.
60. N. V. Tretyakov, A method of penalty estimates for convex programming problems, Ékonomika i Matematicheskie Metody, 9 (1973) 525-540.
61. L. Q. Yang, D. F. Sun, and K. C. Toh, SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints, Mathematical Programming Computation, 7 (2015) 331-366.
62. J. J. Ye, X. Y. Ye, Necessary optimality conditions for optimization problems with variational inequality constraints, Mathematics of Operations Research, 22 (1997) 977-997.
63. X. Y. Zhao, D. F. Sun, K. C. Toh, A Newton-CG augmented Lagrangian method for semidefinite programming, SIAM Journal on Optimization, 20 (2010) 1737-1765.