Mathematical Programming manuscript No.
(will be inserted by the editor)

Defeng Sun · Jie Sun · Liwei Zhang

The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming

Received: 23 January 2006 / Accepted: 23 March 2007

Abstract We analyze the rate of local convergence of the augmented Lagrangian method in nonlinear semidefinite optimization. The presence of the positive semidefinite cone constraint requires extensive tools such as the singular value decomposition of matrices, an implicit function theorem for semismooth functions, and variational analysis on the projection operator in the symmetric matrix space. Without requiring strict complementarity, we prove that, under the constraint nondegeneracy condition and the strong second order sufficient condition, the rate of convergence is linear and the ratio constant is proportional to $1/c$, where c is the penalty parameter that exceeds a threshold $\overline{c} > 0$.

Keywords Nonlinear semidefinite programming · rate of convergence · the augmented Lagrangian method · variational analysis

Mathematics Subject Classification (2000) 90C22 · 65K05 · 49J52

The research of Defeng Sun is partly supported by the Academic Research Fund from the National University of Singapore. The research of Jie Sun and Liwei Zhang is partly supported by Singapore-MIT Alliance and by Grants RP314000-042/057-112 of the National University of Singapore. The research of Liwei Zhang is also supported by the National Natural Science Foundation of China under project grant No. 10471015 and by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, China.

Defeng Sun
Department of Mathematics, National University of Singapore, Republic of Singapore. E-mail: matsundf@nus.edu.sg

Jie Sun
Department of Decision Sciences, National University of Singapore, Republic of Singapore. E-mail: jsun@nus.edu.sg

Liwei Zhang
Department of Applied Mathematics, Dalian University of Technology, Dalian 116024, China. E-mail: lwzhang@dlut.edu.cn

1 Introduction

The nonconvex semidefinite programming problem has wide applications in system control, structural design, and other fields. It has recently become a focal point in optimization research. For example, in the recent release of the library COMPl$_{\rm e}$ib [24], a total of 168 test examples for nonlinear semidefinite programs, control system design, and related problems are collected. Among very few algorithms for this problem, the augmented Lagrangian method appears to perform well [26]. It naturally calls for a suitable theoretical explanation for this phenomenon. Note that algorithms for nonlinear semidefinite programs may display quite distinctive features from conventional nonlinear programming programs (see [14] for such an example). In its general setting, the augmented Lagrangian method can be used to solve the following optimization problem

$$\text{(OP)} \qquad \min \ f(x) \quad \text{s.t.} \quad h(x) = 0, \ g(x) \in K\,,$$

where $f : X \mapsto \Re$, $h : X \mapsto \Re^m$, and $g : X \mapsto Y$ are twice continuously differentiable functions, X and Y are two finite-dimensional real Hilbert spaces equipped with a scalar product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$, and K is a closed convex cone in Y. For any given $\overline{x} \in X$ and $\varepsilon > 0$, let the open ball be $\mathbf{B}_\varepsilon(\overline{x}) := \{x \in X \,|\, \|x - \overline{x}\| < \varepsilon\}$. Suppose that X' and Y' are two finite-dimensional real Hilbert spaces and that $F : X \times X' \mapsto Y'$. If F is Fréchet-differentiable at $(x, x') \in X \times X'$, then we use $\mathcal{J}F(x, x')$ (respectively, $\mathcal{J}_x F(x, x')$) to denote the Fréchet-derivative of F at (x, x') (respectively, the partial Fréchet-derivative of F at (x, x') with respect to x) and $\nabla F(x, x') := \mathcal{J}F(x, x')^*$, the adjoint of $\mathcal{J}F(x, x')$ (respectively, $\nabla_x F(x, x') := \mathcal{J}_x F(x, x')^*$, the adjoint of $\mathcal{J}_x F(x, x')$). Moreover, if F is twice Fréchet-differentiable at $(x, x') \in X \times X'$, we define

$$\mathcal{J}^2 F(x, x') := \mathcal{J}(\mathcal{J}F)(x, x')\,, \quad \mathcal{J}^2_{xx} F(x, x') := \mathcal{J}_x(\mathcal{J}_x F)(x, x'),$$

$$\nabla^2 F(x, x') := \mathcal{J}(\nabla F)(x, x')\,, \quad \text{and} \quad \nabla^2_{xx} F(x, x') := \mathcal{J}_x(\nabla_x F)(x, x')\,.$$

A feasible point $x \in X$ to (OP) is called a stationary point if there exist $\zeta \in \Re^m$ and $\xi \in Y$ such that the following Karush-Kuhn-Tucker (KKT) condition is satisfied at (x, ζ, ξ):

$$\nabla_x L_0(x, \zeta, \xi) = 0\,, \quad h(x) = 0\,, \quad g(x) \in K\,, \quad \xi \in K^*, \ \langle g(x), \xi \rangle = 0\,, \qquad (1)$$

where the Lagrangian function $L_0 : X \times \Re^m \times Y \mapsto \Re$ is defined as

$$L_0(x, \zeta, \xi) := f(x) + \langle \zeta, h(x) \rangle - \langle \xi, g(x) \rangle$$

and K^* is the dual cone of K, i.e.,

$$K^* := \{v \in Y \,|\, \langle v, z \rangle \geq 0 \ \ \forall\, z \in K\}\,.$$

Any point $(x, \zeta, \xi) \in X \times \Re^m \times Y$ satisfying (1) is called a KKT point and the corresponding point (ζ, ξ) is called a Lagrange multiplier at x. Let $\mathcal{M}(x)$ (maybe empty) be the set of all Lagrangian multipliers at x.

Let $c > 0$ be a parameter. The augmented Lagrangian function with the penalty parameter c for problem (OP) is defined as (cf. [40, Section 11.K] or [43])

$$L_c(x,\zeta,\xi) := f(x) + \langle \zeta, h(x)\rangle + \frac{c}{2}\|h(x)\|^2 + \frac{1}{2c}\left[\|\Pi_{K^*}(\xi - cg(x))\|^2 - \|\xi\|^2\right], \tag{2}$$

where $(x,\zeta,\xi) \in X \times \Re^m \times Y$ and $\Pi_{K^*}(\cdot)$ denotes the metric projection operator onto the set K^*. By observing (cf. [51])

$$\Pi_{K^*}(y) = \Pi_K(-y) + y \quad \text{and} \quad \langle \Pi_K(-y) + y, \Pi_K(-y)\rangle = 0 \quad \forall\, y \in Y,$$

we have for any $(x,\zeta,\xi) \in X \times \Re^m \times Y$ that

$$\lim_{c\downarrow 0} L_c(x,\zeta,\xi) = L_0(x,\zeta,\xi) - \lim_{c\downarrow 0} \frac{1}{2c}\|\Pi_K(cg(x)-\xi)\|^2 = \begin{cases} L_0(x,\zeta,\xi) & \text{if } \xi \in K^*, \\ -\infty & \text{otherwise.} \end{cases}$$

If there is no inequality constraint, problem (OP) specializes to

$$\min\ f(x) \quad \text{s.t. } h(x) = 0. \tag{3}$$

The corresponding augmented Lagrangian function is

$$L_c(x,\zeta) = f(x) + \langle \zeta, h(x)\rangle + \frac{c}{2}\|h(x)\|^2, \quad (x,\zeta) \in X \times \Re^m,$$

which was introduced by Arrow and Solow [2] in the study of a differential equation method for solving (3). The augmented Lagrangian method was initiated by Hestenes [19] and Powell [31] for solving the equality constrained problem (3) and was generalized by Rockafellar [34] to the following nonlinear programming problem

$$\text{(NLP)} \qquad \min\ f(x) \quad \text{s.t. } h(x) = 0\,,\ g(x) \geq 0\,,$$

where $f : \Re^n \mapsto \Re$, $h : \Re^n \mapsto \Re^m$, and $g : \Re^n \mapsto \Re^p$ are twice continuously differentiable. Problem (NLP) is a special case of (OP) with $X := \Re^n$, $Y := \Re^p$, and $K := \Re^p_+$.

For the equality constrained optimization problem (3), Powell sketched a proof in [31] to show that if the linear independence constraint qualification and the second-order sufficient condition are satisfied, then the augmented Lagrangian method can converge locally at a linear rate without having $c \to \infty$. For convex programming, Rockafellar [34] established a saddle point theorem in terms of $L_c(\cdot)$ and Rockafellar [35] and Tretyakov [48] proved the global convergence of the augmented Lagrangian method for any $c > 0$.

The augmented Lagrangian method for solving (OP) can be stated as follows. Let $c_0 > 0$ be given. Let $(\zeta^0, \xi^0) \in \Re^m \times K^*$ be the initial estimated Lagrange multiplier. At the kth iteration, determine x^k by minimizing $L_{c_k}(x,\zeta^k,\xi^k)$, compute (ζ^{k+1},ξ^{k+1}) by

$$\begin{cases} \zeta^{k+1} := \zeta^k + c_k h(x^k)\,, \\ \xi^{k+1} := \Pi_{K^*}(\xi^k - c_k g(x^k))\,, \end{cases}$$

and update c_{k+1} by

$$c_{k+1} := c_k \quad \text{or} \quad c_{k+1} := \kappa c_k$$

according to certain rules, where $\kappa > 1$ is a preselected positive number. If the sequence of parameters $\{c_k\}$ is chosen to satisfy $c_k \to +\infty$, then the global convergence of the augmented Lagrangian method can be similarly discussed to the penalty function method [4]. If c_k has a finite limit, then there exists a positive integer N_0 such that $c_k \equiv c$ for $k \geq N_0$ and some positive number c. In this paper, instead of considering global convergence properties, we consider the local convergence properties of the augmented Lagrangian method for (OP) when the second case occurs; namely the case in which $c_k \equiv c$ for all sufficient large k. For simplicity in our analysis, for k sufficiently large, we choose x^k as an exact local solution of $L_c(\cdot, \zeta^k, \xi^k)$.

In [3] (also see [4, Section 2.2]), Bertsekas established an important result on the linear rate of convergence of the augmented Lagrangian method for the equality constrained problem (3), in which the ratio constant is proportional to $1/c$. The significance of Bertsekas's result resides in the fact that theoretically, subject to numerical stability, we can select a large c to accelerate the convergence, which partially explains why the practical performance of this method has been good. In [4, Chapter 3], assuming the strict complementarity condition, Bertsekas also discussed similar results for nonlinear programming (NLP). On the other hand, without assuming the strict complementarity condition, many authors (e.g., Conn et al. [11], Contesse-Becker[12], and Ito and Kunisch [21]) derived linear convergence rate for the augmented Lagrangian method. For more on the augmented Lagrangian method for nonlinear programming, see the two monographs [4,17] and the survey paper [39].

The main objective of this paper is to study, without assuming the strict complementarity, the rate of convergence of the augmented Lagrangian method for solving the nonlinear semidefinite programming problem

$$\text{(NLSDP)} \qquad \min \ f(x) \quad \text{s.t.} \quad h(x) = 0\,, \ g(x) \in \mathcal{S}^p_+\,,$$

where $\mathcal{S}^p_+$ is the cone of all positive semidefinite matrices in $\mathcal{S}^p$, the linear space of all p by p symmetric matrices in $\Re^{p\times p}$. The difficulty for achieving this objective lies in the facts that the positive semidefinite cone $\mathcal{S}^p_+$ is non-polyhedral for $p > 1$ and very few established tools exist for dealing with the augmented Lagrangian method in such a general setting. A work of similar nature (but of different target) is Pennanen's local convergence analysis of proximal point methods for the inclusion problem

$$0 \in \mathcal{T}(x)\,, \tag{4}$$

where $\mathcal{T}$ is a set-valued mapping from a Hilbert space X' to itself [30]. Based in part on Rockafellar's convergence analysis for the inclusion problem (4) with monotone operators [37,38], Pennanen [30] established local linear convergence results of the proximal point methods under the condition that $\mathcal{T}^{-1}$ has a Lipschitz localization property at a solution $\overline{x}$ to (4). One interesting part of Pennanen's results is that he used his theory to establish the local linear convergence of the proximal point method of multipliers (the regularized augmented Lagrangian method) for solving (NLP) without assuming the strict complementarity condition. This suggests that one may do the same for (NLSDP). However, by focusing on the optimization problem (OP) instead of the more general inclusion problem (4), we hope to gain more by using the rich symmetry structure uniquely possessed by this optimization problem. Indeed, we are not only able to prove that the ratio constant is proportional to $1/c$ with the penalty parameter c exceeding a threshold $\overline{c} > 0$, but also able to provide nice properties on the generalized Hessian of the dual function used in our analysis (cf. Proposition 5) for (NLSDP) that relate the augmented Lagrangian method to an approximate generalized Newton method.

The organization of this paper is as follows. In Section 2, we discuss several technical results used in our convergence analysis. In Section 3, we develop a general theory on the rate of convergence of the augmented Lagrangian method for a class of constrained optimization problems under two basic assumptions. Section 4 is devoted to applying the theory developed in Section 3 to nonlinear semidefinite programming. Finally, we give our conclusions in Section 5. To show how the removal of strict complementarity complicates the analysis, we provide a simple proof of the counterpart under strict complementarity as an appendix.

2 Preliminaries

To analyze the problem without the strict complementarity condition, we use tools from semismooth matrix functions. This section serves as a preparation for our analysis. We will cite and prove some results that are essential to our discussion.

Let X and Y be two finite-dimensional real Hilbert spaces. Let $\mathcal{O}$ be an open set in X and $\Phi : \mathcal{O} \subseteq X \to Y$ be a locally Lipschitz continuous function on the open set $\mathcal{O}$. By Rademacher's theorem, Φ is almost everywhere Fréchet-differentiable in $\mathcal{O}$. We denote by $\mathcal{D}_\Phi$ the set of Fréchet-differentiable points of Φ in $\mathcal{O}$. Then, the Bouligand-subdifferential of Φ at $x \in \mathcal{O}$, denoted $\partial_B \Phi(x)$, is

$$\partial_B \Phi(x) := \left\{ \lim_{k\to\infty} \mathcal{J}\Phi(x^k) \,|\, x^k \in \mathcal{D}_\Phi, x^k \to x \right\} .$$

Clarke's generalized Jacobian of Φ at x is the convex hull of $\partial_B \Phi(x)$ (see [10]), i.e.,

$$\partial \Phi(x) = \operatorname{conv} \{\partial_B \Phi(x)\} .$$

The following concept of semismoothness was first introduced by Mifflin [28] for functionals and was extended by Qi and Sun [32] to vector valued functions.

Definition 1 Let $\Phi : \mathcal{O} \subseteq X \mapsto Y$ be a locally Lipschitz continuous function on the open set $\mathcal{O}$. We say that Φ is semismooth at a point $x \in \mathcal{O}$ if

(i) Φ is directionally differentiable at x; and

(ii) for any $\Delta x \in X$ and $V \in \partial\Phi(x+\Delta x)$ with $\Delta x \to 0$,

$$\Phi(x+\Delta x) - \Phi(x) - V(\Delta x) = o(\|\Delta x\|)\,.$$

Furthermore, Φ is said to be strongly semismooth at $x \in \mathcal{O}$ if Φ is semismooth at x and for any $\Delta x \in X$ and $V \in \partial\Phi(x+\Delta x)$ with $\Delta x \to 0$,

$$\Phi(x+\Delta x) - \Phi(x) - V(\Delta x) = \mathrm{O}(\|\Delta x\|^2)\,.$$

By combining Clarke's implicit function theorem for locally Lipschitz continuous functions [10, Section 7.1] with [44, Theorem 1.1] and [23, Lemma 2], we can get the following lemma of implicit functions directly. Here and below,

$$\pi_x \partial H(\overline{x}, \overline{y}) = \text{ the projection of } \partial H(\overline{x}, \overline{y}) \text{ onto the space } X.$$

Lemma 1 *Suppose that $H : X \times Y \mapsto X$ is a locally Lipschitz continuous function in an open neighborhood of $(\overline{x}, \overline{y}) \in X \times Y$ with $H(\overline{x}, \overline{y}) = 0$. If every element in $\pi_x \partial H(\overline{x}, \overline{y})$ is nonsingular, then there exist an open neighborhood $\mathcal{O}_Y$ of $\overline{y}$ and a locally Lipschitz continuous function $x(\cdot) : \mathcal{O}_Y \mapsto X$ satisfying $x(\overline{y}) = \overline{x}$ such that for every $y \in \mathcal{O}_Y$,*

$$H(x(y), y) = 0\,.$$

Furthermore, if H is (strongly) semismooth at every point in the open neighborhood of $(\overline{x}, \overline{y})$, then $x(\cdot)$ is (strongly) semismooth at every point in $\mathcal{O}_Y$.

The following two lemmas on the Bouligand-subdifferential of composite functions are useful in determining $\pi_x \partial_B(\nabla_x L_c)(\cdot)$. The first one is proved in [45, Lemma 2.1] and the second one needs a proof, which will be given here.

Lemma 2 *Let $F : X \mapsto Y$ be a continuously differentiable function on an open neighborhood $\mathcal{O}$ of $\bar{x} \in X$ and $\Phi : \mathcal{O}_Y \subseteq Y \mapsto X'$ be a locally Lipschitz continuous function on an open set $\mathcal{O}_Y$ containing $\bar{y} := F(\bar{x})$, where X' is a finite-dimensional real vector space. Suppose that Φ is directionally differentiable at every point in $\mathcal{O}_Y$ and that $\mathcal{J}F(\bar{x}) : X \to Y$ is onto. Then it holds that*

$$\partial_B(\Phi{*}F)(\bar{x}) = \partial_B\Phi(\bar{y})\mathcal{J}F(\bar{x})\,,$$

where "$*$" stands for the composite operation.

Lemma 3 *Let $F : X \mapsto Y$ be a continuously differentiable function on an open neighborhood $\mathcal{O}$ and $\Phi : X \to X'$ be a locally Lipschitz continuous function on $\mathcal{O}$, where X' is a finite-dimensional real vector space. Suppose that Φ is semismooth at every point in $\mathcal{O}$. Let $\Psi : X \mapsto Y'$ be defined as*

$$\Psi(x) := F(x)\Phi(x) \equiv F(x)\cdot\Phi(x)\,, \quad x \in X\,,$$

where Y' is a finite-dimensional real vector space and "$\cdot$" is a bi-linear operator from $Y \times X'$ to Y'. Then for every $x \in \mathcal{O}$ and $\Delta x \in X$,

$$\partial_B\Psi(x)(\Delta x) = \mathcal{J}F(x)(\Delta x)\Phi(x) + F(x)\partial_B\Phi(x)(\Delta x)\,. \tag{5}$$

Proof. Let $x \in \mathcal{O}$ and $\Delta x \in X$ be two arbitrary but fixed points. By using the fact that if Φ is Fréchet differentiable at $y \in \mathcal{O}$, then Ψ is also Fréchet differentiable at y we obtain

$$\partial_B \Psi(x)(\Delta x) \supseteq \mathcal{J}F(x)(\Delta x)\Phi(x) + F(x)\partial_B \Phi(x)(\Delta x)\,.$$

Conversely, let $W \in \partial_B \Psi(x)$. Then there exists a sequence of Fréchet differentiable points $\{x^k\} \subseteq \mathcal{O}$ converging to x such that $W = \lim_{k\to\infty} \mathcal{J}\Psi(x^k)$. Since Φ is assumed to be semismooth at each x^k, we have

$$\Phi'(x^k; \Delta x) \in \partial_B \Phi(x^k)(\Delta x).$$

Thus, for any $k \geq 1$,

$$\begin{aligned}\mathcal{J}\Psi(x^k)(\Delta x) &= \mathcal{J}F(x^k)(\Delta x)\Phi(x^k) + F(x^k)\Phi'(x^k; \Delta x)\\ &\in \mathcal{J}F(x^k)(\Delta x)\Phi(x^k) + F(x^k)\partial_B \Phi(x^k)(\Delta x)\,,\end{aligned}$$

which, together with the upper semicontinuity of $\partial_B \Phi(\cdot)$, implies

$$W(\Delta x) = \lim_{k\to\infty} \mathcal{J}\Psi(x^k)(\Delta x) \in \mathcal{J}F(x)(\Delta x)\Phi(x) + F(x)\partial_B \Phi(x)(\Delta x)\,.$$

Consequently, (5) holds. □

Let K be a closed convex set in Y. It is well known [51] that the metric projector $\Pi_K(\cdot)$ is Lipschitz continuous with the Lipschitz constant 1. Then for any $y \in Y$, $\partial \Pi_K(y)$ is well defined. Below is a lemma on the general properties of $\partial \Pi_K(\cdot)$.

Lemma 4 [27, Proposition 1] *Let $K \subseteq Y$ be a closed convex set. Then, for any $y \in Y$ and $V \in \partial \Pi_K(y)$, it holds that*

(i) V is self-adjoint.
(ii) $\langle d, Vd\rangle \geq 0 \quad \forall\, d \in Y$.
(iii) $\langle Vd, d - Vd\rangle \geq 0 \quad \forall\, d \in Y$.

For discussions on nonlinear semidefinite programming we need more properties about the Bouligand-subdifferential of the metric projector $\Pi_{\mathcal{S}^p_+}(\cdot)$ over $\mathcal{S}^p_+$ under the Frobenius inner product in $\mathcal{S}^p$. We write $C \succeq 0$ to mean that C is a symmetric positive semidefinite matrix. Let $\overline{Z} \in \mathcal{S}^p$ and $\overline{Z}_+ := \Pi_{\mathcal{S}^p_+}(\overline{Z})$. Suppose that $\overline{Z}$ has the following spectral decomposition

$$\overline{Z} = \overline{P}\Lambda\overline{P}^T, \tag{6}$$

where Λ is the diagonal matrix of eigenvalues of $\overline{Z}$ and $\overline{P}$ is a corresponding orthogonal matrix of the orthonormal eigenvectors. Then

$$\overline{Z}_+ = \overline{P}\Lambda_+\overline{P}^T,$$

where Λ_+ is the diagonal matrix whose diagonal entries are the nonnegative parts of the respective diagonal entries of Λ [20, 49]. Define three index sets of positive, zero, and negative eigenvalues of $\overline{Z}$, respectively, as

$$\alpha := \{i \,|\, \lambda_i > 0\},\ \beta := \{i \,|\, \lambda_i = 0\},\ \gamma := \{i \,|\, \lambda_i < 0\}.$$

Write

$$\Lambda = \begin{bmatrix} \Lambda_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Lambda_\gamma \end{bmatrix} \quad \text{and } \overline{P} = [\,\overline{P}_\alpha \; \overline{P}_\beta \; \overline{P}_\gamma\,]$$

with $\overline{P}_\alpha \in \Re^{p\times|\alpha|}$, $\overline{P}_\beta \in \Re^{p\times|\beta|}$, and $\overline{P}_\gamma \in \Re^{p\times|\gamma|}$. Let Θ be any matrix in $\mathcal{S}^p$ with entries

$$\begin{cases} \Theta_{ij} = \dfrac{\max\{\lambda_i, 0\} + \max\{\lambda_j, 0\}}{|\,\lambda_i\,| + |\,\lambda_j\,|} & \text{if } (i,j) \notin \beta\times\beta\,, \\ \Theta_{ij} \in [0,1] & \text{if } (i,j) \in \beta\times\beta\,. \end{cases} \tag{7}$$

The projection operator $\Pi_{\mathcal{S}^p_+}(\cdot)$ is directionally differentiable everywhere in $\mathcal{S}^p$ [5,6] and is a strongly semismooth matrix-valued function [46]. For any $H \in \mathcal{S}^p$, we have

$$\Pi'_{\mathcal{S}^p_+}(\overline{Z}; H) = \overline{P} \begin{bmatrix} \overline{P}^T_\alpha H \overline{P}_\alpha & \overline{P}^T_\alpha H \overline{P}_\beta & \Theta_{\alpha\gamma} \circ \overline{P}^T_\alpha H \overline{P}_\gamma \\ \overline{P}^T_\beta H \overline{P}_\alpha & \Pi_{\mathcal{S}^{|\beta|}_+}(\overline{P}^T_\beta H \overline{P}_\beta) & 0 \\ \overline{P}^T_\gamma H \overline{P}_\alpha \circ \Theta_{\gamma\alpha} & 0 & 0 \end{bmatrix} \overline{P}^T\,, \tag{8}$$

where " $\circ$ " denotes the Hadamard product [29,46]. When $\beta = \emptyset$, $\Pi_{\mathcal{S}^p_+}(\cdot)$ is Fréchet-differentiable at $\overline{Z}$ and (8) reduces to the classical result of Löwner [25]:

$$\mathcal{J}\Pi_{\mathcal{S}^p_+}(\overline{Z})H = \overline{P} \begin{bmatrix} \overline{P}^T_\alpha H \overline{P}_\alpha & \Theta_{\alpha\gamma} \circ \overline{P}^T_\alpha H \overline{P}_\gamma \\ \overline{P}^T_\gamma H \overline{P}_\alpha \circ \Theta_{\gamma\alpha} & 0 \end{bmatrix} \overline{P}^T \quad \forall\, H \in \mathcal{S}^p\,. \tag{9}$$

The tangent cone of $\mathcal{S}^p_+$ at $\overline{Z}_+$, denoted $\mathcal{T}_{\mathcal{S}^p_+}(\overline{Z}_+)$, can be completely characterized as follows

$$\mathcal{T}_{\mathcal{S}^p_+}(\overline{Z}_+) = \{B \in \mathcal{S}^p \mid B = \Pi'_{\mathcal{S}^p_+}(\overline{Z}_+; B)\} = \{B \in \mathcal{S}^p \mid [\overline{P}_\beta\; \overline{P}_\gamma]^T B [\overline{P}_\beta\; \overline{P}_\gamma] \succeq 0\}\,.$$

The characterization of $\mathcal{T}_{\mathcal{S}^p_+}(\overline{Z}_+)$ was first obtained by Arnold [1] by using a different approach from the above. The lineality space of $\mathcal{T}_{\mathcal{S}^p_+}(\overline{Z}_+)$, i.e., the largest linear space in $\mathcal{T}_{\mathcal{S}^p_+}(\overline{Z}_+)$, denoted by $\mathrm{lin}\left(\mathcal{T}_{\mathcal{S}^p_+}(\overline{Z}_+)\right)$, takes the following form:

$$\mathrm{lin}\left(\mathcal{T}_{\mathcal{S}^p_+}(\overline{Z}_+)\right) = \{B \in \mathcal{S}^p \mid [\overline{P}_\beta\; \overline{P}_\gamma]^T B [\overline{P}_\beta\; \overline{P}_\gamma] = 0\}\,.$$

The critical cone of $\mathcal{S}^p_+$ at $\overline{Z} \in \mathcal{S}^p$ associated with the problem of finding the metric projection of $\overline{Z}$ onto $\mathcal{S}^p_+$ (i.e., $\overline{Z}_+$) is defined as [8, Section 5.3]

$$\mathcal{C}(\overline{Z}; \mathcal{S}^p_+) \;:=\; \mathcal{T}_{\mathcal{S}^p_+}(\overline{Z}_+) \cap \{B \in \mathcal{S}^p \mid \langle B, \overline{Z}_+ - \overline{Z}\rangle = 0\}\,.$$

Thus, it holds that

$$\mathcal{C}(\overline{Z};\mathcal{S}^p_+) \;=\; \left\{ B \in \mathcal{S}^p \,\middle|\, \overline{P}_\beta^T B \overline{P}_\beta \succeq 0,\ \overline{P}_\beta^T B \overline{P}_\gamma = 0,\ \overline{P}_\gamma^T B \overline{P}_\gamma = 0 \right\}.$$

The affine hull of $\mathcal{C}(\overline{Z};\mathcal{S}^p_+)$, denoted by $\mathrm{aff}(\mathcal{C}(\overline{Z};\mathcal{S}^p_+))$, can then be written as

$$\mathrm{aff}\left(\mathcal{C}(\overline{Z};\mathcal{S}^p_+)\right) = \left\{ B \in \mathcal{S}^p \mid \overline{P}_\beta^T B \overline{P}_\gamma = 0,\ \overline{P}_\gamma^T B \overline{P}_\gamma = 0 \right\}. \qquad (10)$$

The following lemma on $\partial_B \Pi_{\mathcal{S}^p_+}(\overline{Z})$ is part of [45, Proposition 4], which is based on [29, Lemma 11].

Lemma 5 *Let $\Theta \in \mathcal{S}^p$ satisfy (7). Then $W \in \partial_B \Pi_{\mathcal{S}^p_+}(\overline{Z})$ if and only if there exists $W_0 \in \partial_B \Pi_{\mathcal{S}^{|\beta|}_+}(0)$ such that*

$$W(H) = \overline{P} \begin{bmatrix} \overline{P}_\alpha^T H \overline{P}_\alpha & \overline{P}_\alpha^T H \overline{P}_\beta & \Theta_{\alpha\gamma} \circ \overline{P}_\alpha^T H \overline{P}_\gamma \\ \overline{P}_\beta^T H \overline{P}_\alpha & W_0(\overline{P}_\beta^T H \overline{P}_\beta) & 0 \\ \overline{P}_\gamma^T H \overline{P}_\alpha \circ \Theta_{\gamma\alpha} & 0 & 0 \end{bmatrix} \overline{P}^T \quad \forall\, H \in \mathcal{S}^p.$$

Let $\mathcal{Q}$ be the set of all orthogonal matrices of order $|\beta| \times |\beta|$. Let

$$\mathcal{P} := \left\{ P \in \Re^{p\times p} \mid P = [P_\alpha\ P_\beta\ P_\gamma] = [\overline{P}_\alpha\ (\overline{P}_\beta Q)\ \overline{P}_\gamma],\ Q \in \mathcal{Q} \right\}. \qquad (11)$$

Note that all $P \in \mathcal{P}$ have the same P_α and P_γ. From the definition of $\partial_B \Pi_{\mathcal{S}^{|\beta|}_+}(0)$ and (9) we know that if $W_0 \in \partial_B \Pi_{\mathcal{S}^{|\beta|}_+}(0)$, then there exist matrices $Q \in \mathcal{Q}$ and $\Omega \in \mathcal{S}^{|\beta|}$ with entries $\Omega_{ij} \in [0,1]$ such that

$$W_0(D) = Q(\Omega \circ (Q^T D Q))Q^T \quad \forall\, D \in \mathcal{S}^{|\beta|}.$$

For an extension to the above result, see [9, Lemma 4.7]. By using Lemma 5 we obtain the following useful lemma, which does not need further explanation.

Lemma 6 *For any $W \in \partial_B \Pi_{\mathcal{S}^p_+}(\overline{Z})$, there exist two matrices $P \in \mathcal{P}$ and $\Theta \in \mathcal{S}^p$ satisfying (7) such that*

$$W(H) = P\left(\Theta \circ (P^T H P)\right)P^T \quad \forall\, H \in \mathcal{S}^p.$$

The following result, due to Debreu [13], is useful for the study of the Bouligand-subdifferential of $\nabla_x L_c(\cdot)$.

Lemma 7 *Let $\phi : X \mapsto \Re$ be continuous and positive homogeneous of degree two:*

$$\phi(td) = t^2\phi(d) \quad \forall\, t \ge 0 \text{ and } d \in X.$$

Suppose that there exists a positive number $\eta_0 > 0$ such that for any d satisfying $\mathcal{L}d = 0$, one has $\phi(d) \ge \eta_0\|d\|^2$, where $\mathcal{L} : X \mapsto Y$ is a given linear operator. Then there exist positive numbers $\underline{\eta} \in (0, \eta_0]$ and $c_0 > 0$ such that

$$\phi(d) + c_0\langle \mathcal{L}d, \mathcal{L}d\rangle \ge \underline{\eta}\langle d, d\rangle \quad \forall\, d \in X.$$

Next, we provide a technical result used in Section 4.

Lemma 8 *Let $a, b, c,$ and c_0 be four positive scalars with $c \geq c_0$. Let*

$$\psi(t; c, a, b, c_0) := a - \frac{1}{c}t + \frac{t^2}{b + (c - c_0)t}, \quad t \in [0, 1]. \tag{12}$$

Then, for any $c \geq \max\{c_0, (b - c_0)^2/c_0\}$, $\psi(\cdot; c, a, b, c_0)$ is a convex function on $[0, 1]$,

$$\min_{t\in[0,1]} \psi(t; c, a, b, c_0) = a - \frac{1}{c}\frac{b}{(\sqrt{c} + \sqrt{c_0})^2}, \tag{13}$$

and

$$\max_{t\in[0,1]} \psi(t; c, a, b, c_0) = \max\Big\{\psi(0; c, a, b, c_0),\ \psi(1; c, a, b, c_0)\Big\}. \tag{14}$$

Proof. By simple calculations for any $t \in [0, 1]$ we have

$$\nabla_t \psi(t; c, a, b, c_0) = \frac{c_0}{c(c - c_0)} - \frac{b^2}{(c - c_0)(b + (c - c_0)t)^2}$$

and

$$\nabla^2_{tt} \psi(t; c, a, b, c_0) = \frac{2b^2}{(b + (c - c_0)t)^3}.$$

Then, since for $c \geq c_0$, $\nabla^2_{tt}\psi(t, c, a, b, c_0) \geq 0$ for all $t \in [0, 1]$, $\psi(\cdot; c, a, b, c_0)$ is a convex function on $[0, 1]$. Consequently, (14) holds.

Let $\underline{t} := b/(c_0 + \sqrt{cc_0})$. Then $\nabla_t \psi(\underline{t}; c, a, b, c_0) = 0$. Since for any $c \geq \max\{c_0, (b - c_0)^2/c_0\}$, $\underline{t} \in (0, 1]$ and $\psi(\cdot; c, a, b, c_0)$ is convex on $[0, 1]$, we have

$$\min_{t\in[0,1]} \psi(t; c, a, b, c_0) = \psi(\underline{t}; c, a, b, c_0),$$

which, implies that (13) holds. □

3 General discussions on the rate of convergence

In this section, we always assume that the cone K presented in the optimization problem (OP) is a self-dual cone, i.e, $K = K^*$ and that $\Pi_K(\cdot)$ is semismooth everywhere. In particular, this is the case for any closed symmetric cone because a closed symmetric cone is always self-dual [16] and $\Pi_K(\cdot)$ is strongly semismooth everywhere [47]. The cones $\Re^p_+$ and $\mathcal{S}^p_+$ are special cases of symmetric cones. For more on symmetric cones, see Faraut and Korányi [16].

Let $c > 0$ and $\overline{x}$ be a stationary point of (OP). Then $\mathcal{M}(\overline{x})$, the set of Lagrange multipliers at $\overline{x}$, is nonempty. Since $f, h,$ and g are assumed to be twice continuously differentiable, we know from (2) and [51] that the augmented Lagrangian function $L_c(\cdot)$ is continuously differentiable and for any $(x, \zeta, \xi) \in X \times \Re^m \times Y$,

$$\nabla_x L_c(x, \zeta, \xi) = \nabla f(x) + \nabla h(x)(\zeta + ch(x)) - \nabla g(x)\Pi_K(\xi - cg(x)). \tag{15}$$

Therefore, from (1) and [15], we have $\nabla_x L_c(\overline{x},\zeta,\xi)=0$ for any $(\zeta,\xi)\in\mathcal{M}(\overline{x})$.

Define $F_c : X\times\Re^m\times Y\mapsto Y$ by

$$F_c(x,\zeta,\xi)=\xi-cg(x),\quad (x,\zeta,\xi)\in X\times\Re^m\times Y\,.$$

Since $\Pi_K(\cdot)$ is assumed to be semismooth everywhere, $\Pi_K(\cdot)$ is directionally differentiable at any point $y\in Y$. Hence, by using the fact that for any $(x,\zeta,\xi)\in X\times\Re^m\times Y$, $\mathcal{J}F_c(x,\zeta,\xi): X\times\Re^m\times Y\mapsto Y$ is onto, we know from Lemma 2 that

$$\partial_B(\Pi_K * F_c)(x,\zeta,\xi)=\partial_B\Pi_K(\xi-cg(x))\mathcal{J}F_c(x,\zeta,\xi). \tag{16}$$

For any $(x,\zeta,\xi)\in X\times\Re^m\times Y$, let

$$\Psi_c(x,\zeta,\xi):=\nabla g(x)(\Pi_K * F_c)(x,\zeta,\xi)=\nabla g(x)\Pi_K(\xi-cg(x))\,.$$

Let $(x,\zeta,\xi)\in X\times\Re^m\times Y$. Then from the semismoothness of $\Pi_K(\cdot)$ and Lemma 3 we obtain that for any $(\Delta x,\Delta\zeta,\Delta\xi)\in X\times\Re^m\times Y$,

$$\begin{aligned}&\partial_B\Psi_c(x,\zeta,\xi)(\Delta x,\Delta\zeta,\Delta\xi)=\nabla^2 g(x)(\Delta x)\Pi_K(\xi-cg(x))\\ &+\nabla g(x)\partial_B(\Pi_K * F_c)(x,\zeta,\xi)(\Delta x,\Delta\zeta,\Delta\xi)\,.\end{aligned} \tag{17}$$

From (15) and the definition of $\Psi_c(\cdot)$ we know that

$$\begin{aligned}\partial_B(\nabla_x L_c)(x,\zeta,\xi)&=(\nabla^2 f(x),0,0)-\partial_B\Psi_c(x,\zeta,\xi)\\ &\quad+\left(\sum_{i=1}^m(\zeta_i+ch_i(x))\nabla^2 h_i(x)+c\nabla h(x)\mathcal{J}h(x),\nabla h(x),0\right)\end{aligned}$$

which, together with (16) and (17), implies that for any $\Delta x\in X$,

$$\begin{aligned}&(\pi_x\partial_B(\nabla_x L_c)(x,\zeta,\xi))\,(\Delta x)\\ &=\nabla^2_{xx}L_0(x,\zeta+ch(x),\Pi_K(\xi-cg(x)))(\Delta x)+c\nabla h(x)\mathcal{J}h(x)(\Delta x)\\ &\quad+c\,\nabla g(x)\partial_B\Pi_K(\xi-cg(x))\mathcal{J}g(x)(\Delta x)\,,\end{aligned} \tag{18}$$

where

$$\begin{aligned}&\nabla^2_{xx}L_0(x,\zeta+ch(x),\Pi_K(\xi-cg(x)))(\Delta x)\\ &=\nabla^2 f(x)(\Delta x)+\nabla^2 h(x)(\Delta x)(\zeta+ch(x))-\nabla^2 g(x)(\Delta x)\Pi_K(\xi-cg(x))\,.\end{aligned}$$

Let $(\overline{\zeta},\overline{\xi})\in\mathcal{M}(\overline{x})$ be a Lagrange multiplier at $\overline{x}$. For any $W: Y\mapsto Y$, let

$$\mathcal{A}_c(\overline{\zeta},\overline{\xi},W):=\nabla^2_{xx}L_0(\overline{x},\overline{\zeta},\overline{\xi})+c\nabla h(\overline{x})\mathcal{J}h(\overline{x})+c\nabla g(\overline{x})W\mathcal{J}g(\overline{x})\,. \tag{19}$$

Then for any $\Delta x\in X$,

$$\begin{aligned}&\left(\pi_x\partial_B(\nabla_x L_c)(\overline{x},\overline{\zeta},\overline{\xi})\right)(\Delta x)\\ &=\left\{\mathcal{A}_c(\overline{\zeta},\overline{\xi},W)(\Delta x)\,|\,W\in\partial_B\Pi_K(\overline{\xi}-cg(\overline{x}))\right\}.\end{aligned} \tag{20}$$

Next, we make two basic assumptions for the constrained optimization problem (OP). The first one is about the positive definiteness of $\mathcal{A}_c(\overline{\zeta},\overline{\xi},\cdot)$.

Assumption B1. We assume that $(\overline{\zeta}, \overline{\xi})$ is the unique Lagrange multiplier at $\overline{x}$, i.e., $\mathcal{M}(\overline{x}) = \{(\overline{\zeta}, \overline{\xi})\}$ and that there exist two positive numbers c_0 and $\underline{\eta}$ such that for any $c \geq c_0$ and any $W \in \partial_B \Pi_K(\overline{\xi} - cg(\overline{x}))$,

$$\big\langle d, \mathcal{A}_c(\overline{\zeta}, \overline{\xi}, W)d \big\rangle \geq \underline{\eta}\, \langle d, d\rangle \quad \forall\, d \in X\,.$$

Assumption B1 is related to the sufficient conditions for the constrained optimization problem (OP). It will be shown in Proposition 4 that, under the constraint nondegeneracy condition and the strong second order sufficient condition, Assumption B1 is valid for (NLSDP).

For the remaining part of this section, we suppose that Assumption B1 is satisfied. Let $\overline{y} := (\overline{\zeta}, \overline{\xi})$. Then $\nabla_x L_c(\overline{x}, \overline{y}) = 0$. Let c_0 and $\underline{\eta}$ be two positive numbers defined in Assumption B1 and $c \geq c_0$ be a positive number. Since by (20) and Assumption B1, every element in $\pi_x \partial_B (\nabla_x L_c)(\overline{x}, \overline{y})$ is positive definite, we know from Lemma 1 that there exist an open neighborhood $\mathcal{O}_{\overline{y}}$ of $\overline{y}$ and a locally Lipschitz continuous function $x_c(\cdot)$ defined on $\mathcal{O}_{\overline{y}}$ such that for any $y \in \mathcal{O}_{\overline{y}}$, $\nabla_x L_c(x_c(y), y) = 0$. Furthermore, since $\Pi_K(\cdot)$ is assumed to be semismooth everywhere, $x_c(\cdot)$ is semismooth (strongly semismooth if $\nabla^2 f, \nabla^2 g$, and $\nabla^2 h$ are locally Lipschitz continuous and $\Pi_K(\cdot)$ is strongly semismooth everywhere) at any point in $\mathcal{O}_{\overline{y}}$. Moreover, there exist two positive numbers $\varepsilon > 0$ and $\delta_0 > 0$ (both depending on c) such that for any $x \in \mathbf{B}_\varepsilon(\overline{x})$ and $y \in \mathbf{B}_{\delta_0}(\overline{y}) := \{y \in \Re^m \times Y \,|\, \|y - \overline{y}\| < \delta_0\} \subset \mathcal{O}_{\overline{y}}$, every element in $\pi_x \partial_B (\nabla_x L_c)(x, y)$ is positive definite. Thus, for any $y \in \mathbf{B}_{\delta_0}(\overline{y})$, $x_c(y)$ is the unique minimizer of $L_c(\cdot, y)$ over $\mathbf{B}_\varepsilon(\overline{x})$, i.e.,

$$\{x_c(y)\} = \operatorname{argmin} \Big\{ L_c(x, y) \,|\, x \in \mathbf{B}_\varepsilon(\overline{x}) \Big\}. \tag{21}$$

For ease of reference, we write these conclusions in the following proposition.

Proposition 1 *Suppose that Assumption B1 is satisfied. Let $c \geq c_0$. Then there exist two positive numbers $\varepsilon > 0$ and $\delta_0 > 0$ (both depending on c) and a locally Lipschitz continuous function $x_c(\cdot)$, given by (21), defined on the open ball $\mathbf{B}_{\delta_0}(\overline{y})$ such that the following conclusions hold:*

(i) The function $x_c(\cdot)$ is semismooth at any point in $\mathbf{B}_{\delta_0}(\overline{y})$.
(ii) If $\nabla^2 f, \nabla^2 g$, and $\nabla^2 h$ are locally Lipschitz continuous and $\Pi_K(\cdot)$ is strongly semismooth everywhere, then $x_c(\cdot)$ is strongly semismooth at any point in $\mathbf{B}_{\delta_0}(\overline{y})$.
(iii) For any $x \in \mathbf{B}_\varepsilon(\overline{x})$ and $y \in \mathbf{B}_{\delta_0}(\overline{y})$, every element in $\pi_x \partial_B (\nabla_x L_c)(x, y)$ is positive definite.
(iv) For any $y \in \mathbf{B}_{\delta_0}(\overline{y})$, $x_c(y)$ is the unique optimal solution to

$$\min \;\; L_c(x, y) \quad \text{s.t. } x \in \mathbf{B}_\varepsilon(\overline{x})\,.$$

Let $\vartheta_c : \Re^m \times Y \mapsto \Re$ be defined as

$$\vartheta_c(\zeta, \xi) := \min_{x \in \mathbf{B}_\varepsilon(\overline{x})} L_c(x, \zeta, \xi), \quad (\zeta, \xi) \in \Re^m \times Y\,. \tag{22}$$

Since for each fixed $x \in X$, $L_c(x, \cdot)$ is a concave function, $\vartheta_c(\cdot)$ is also a concave function as it is the minimum function of a family of concave functions.

By using the fact that for any $y \in \mathbf{B}_{\delta_0}(\overline{y})$, $x_c(y)$ is the unique minimizer of $L_c(\cdot, y)$ over $\mathbf{B}_\varepsilon(\overline{x})$, we have

$$\vartheta_c(y) = L_c(x_c(y), y)\,, \quad y \in \mathbf{B}_{\delta_0}(\overline{y})\,.$$

For any $y \in \mathbf{B}_{\delta_0}(\overline{y})$ with $y = (\zeta, \xi) \in \Re^m \times Y$, let

$$\begin{pmatrix} \zeta_c(y) \\ \xi_c(y) \end{pmatrix} := \begin{pmatrix} \zeta + ch(x_c(y)) \\ \Pi_K(\xi - cg(x_c(y))) \end{pmatrix}. \tag{23}$$

Then we have

$$\nabla_x L_0(x_c(y), \zeta_c(y), \xi_c(y)) = \nabla_x L_c(x_c(y), y) = 0\,, \quad y \in \mathbf{B}_{\delta_0}(\overline{y})\,. \tag{24}$$

Proposition 2 *Suppose that Assumption B1 is satisfied. Let $c \geq c_0$. Then the concave function $\vartheta_c(\cdot)$ defined by (22) is continuously differentiable on $\mathbf{B}_{\delta_0}(\overline{y})$ with*

$$\nabla\vartheta_c(y) = \begin{pmatrix} h(x_c(y)) \\ -c^{-1}\xi + c^{-1}\Pi_K(\xi - cg(x_c(y))) \end{pmatrix}, \quad y = (\zeta, \xi) \in \mathbf{B}_{\delta_0}(\overline{y})\,. \tag{25}$$

Moreover, $\nabla\vartheta_c(\cdot)$ is semismooth at any point in $\mathbf{B}_{\delta_0}(\overline{y})$. It is strongly semismooth at any point in $\mathbf{B}_{\delta_0}(\overline{y})$ if $\nabla^2 f, \nabla^2 g$, and $\nabla^2 h$ are locally Lipschitz continuous and $\Pi_K(\cdot)$ is strongly semismooth everywhere.

Proof. Let $y = (\zeta, \xi) \in \mathbf{B}_{\delta_0}(\overline{y})$. Then from (24) and [10, Theorem 2.6.6] we have for any $(\Delta\zeta, \Delta\xi) \in \Re^m \times Y$ that

$$\begin{aligned}\partial\vartheta_c(y)(\Delta\zeta, \Delta\xi) &= \mathcal{J}_x L_c(x_c(y), y)(\partial x_c(y)(\Delta\zeta, \Delta\xi)) \\ &\quad + \mathcal{J}_\zeta L_c(x_c(y), y)(\Delta\zeta) + \mathcal{J}_\xi L_c(x_c(y), y)(\Delta\xi) \\ &= \langle h(x_c(y)), \Delta\zeta\rangle - c^{-1}\langle \xi, \triangle\xi\rangle + \langle c^{-1}\Pi_K(\xi - cg(x_c(y))), \Delta\xi\rangle\,.\end{aligned}$$

Thus, $\partial\vartheta_c(y)(\Delta\zeta, \Delta\xi)$ is a singleton for each $(\Delta\zeta, \Delta\xi) \in \Re^m \times Y$. This implies that $\partial\vartheta_c(y)$ is a singleton. Therefore, $\vartheta_c(\cdot)$ is Fréchet-differentiable at y and $\nabla\vartheta_c(y)$ is given by (25). The continuity of $\nabla\vartheta_c(\cdot)$ follows from the continuity of $x_c(\cdot)$.

The properties on the (strong) semismoothness of $\nabla\vartheta_c(\cdot)$ at y follows directly from (25) and Proposition 1. $\square$

For any $c \geq c_0$ and $\Delta y := (\Delta\zeta, \Delta\xi) \in \Re^m \times Y$, define

$$\begin{aligned}&\overline{\mathcal{V}}_c(\Delta y) \\ &:= \left\{ \begin{pmatrix} \mathcal{J}h(\overline{x}) \\ -W\mathcal{J}g(\overline{x}) \end{pmatrix} \mathcal{A}_c(\overline{y}, W)^{-1}\left(-\nabla h(\overline{x})(\Delta\zeta) + \nabla g(\overline{x})W(\Delta\xi)\right) \right. \\ &\qquad \left. + \begin{pmatrix} 0 \\ -c^{-1}\Delta\xi + c^{-1}W(\Delta\xi) \end{pmatrix} \,\middle|\, W \in \partial_B \Pi_K(\overline{\xi} - cg(\overline{x})) \right\}.\end{aligned} \tag{26}$$

Since by Assumption B1, $\mathcal{A}_c(\overline{y}, W)$ is positive definite for any $W \in \partial_B\Pi_K(\overline{\xi} - cg(\overline{x}))$, $\overline{\mathcal{V}}_c(\cdot)$ is well defined. The next proposition establishes an important relationship between $\overline{\mathcal{V}}_c(\cdot)$ and $\partial_B(\nabla\vartheta_c)(\overline{y})(\cdot)$. Note that the function $\nabla\vartheta_c(\cdot)$ given by (25) involves two nonsmooth functions $\Pi_K(\cdot)$ and $x_c(\cdot)$, which are related to each other.

Proposition 3 *Suppose that Assumption B1 is satisfied. Let $c \geq c_0$. Then for any $\Delta y := (\Delta\zeta, \Delta\xi) \in \Re^m \times Y$,*

$$\partial_B(\nabla\vartheta_c)(\overline{y})(\Delta y) \subseteq \overline{\mathcal{V}}_c(\Delta y)\,. \tag{27}$$

Proof. Let $\Delta y = (\Delta\zeta, \Delta\xi) \in \Re^m \times Y$ be an arbitrary but fixed point. From Proposition 2, we know that $\nabla\vartheta_c(\cdot)$ is semismooth at any point $y \in \mathbf{B}_{\delta_0}(\overline{y})$. Let $\mathcal{D}_{\nabla\vartheta_c}$ denote the set of all Fréchet-differentiable points of $\nabla\vartheta_c(\cdot)$ in $\mathbf{B}_{\delta_0}(\overline{y})$. Then for any $y = (\zeta, \xi) \in \mathcal{D}_{\nabla\vartheta_c}$, we have

$$\begin{aligned}&\nabla^2\vartheta_c(y)(\Delta y)\\ &= \begin{pmatrix} \mathcal{J}h(x_c(y))(x_c)'(y;\Delta y) \\ -c^{-1}\Delta\xi + c^{-1}\Pi_K'\Big(\xi - cg(x_c(y)); \Delta\xi - c\mathcal{J}g(x_c(y))(x_c)'(y;\Delta y)\Big)\end{pmatrix}.\end{aligned} \tag{28}$$

Let $y \in \mathbf{B}_{\delta_0}(\overline{y})$. Now, we derive the formula for $(x_c)'(y;\Delta y)$. From (24) and (23) we have

$$\begin{aligned}0 = {}&\nabla_{xx}L_0(x_c(y), \zeta_c(y), \xi_c(y))(x_c)'(y;\Delta y)\\ &+c\nabla h(x_c(y))\mathcal{J}h(x_c(y))(x_c)'(y;\Delta y) + \nabla h(x_c(y))(\Delta\zeta)\\ &-\nabla g(x_c(y))\Pi_K'\Big(\xi - cg(x_c(y)); \Delta\xi - c\mathcal{J}g(x_c(y))(x_c)'(y;\Delta y)\Big).\end{aligned} \tag{29}$$

Since $\Pi_K(\cdot)$ is semismooth everywhere, there exists an element $\widehat{W} \in \partial_B\Pi_K(\xi - cg(x_c(y)))$ such that

$$\begin{aligned}&\Pi_K'\Big(\xi - cg(x_c(y)); \Delta\xi - c\mathcal{J}g(x_c(y))(x_c)'(y;\Delta y)\Big)\\ &= \widehat{W}(\Delta\xi - c\mathcal{J}g(x_c(y))(x_c)'(y;\Delta y)).\end{aligned} \tag{30}$$

For any $W \in \partial_B\Pi_K(\xi - cg(x_c(y)))$, let

$$\begin{aligned}\mathcal{A}_c(y, W) := {}&\nabla^2_{xx}L_0(x_c(y), \zeta_c(y), \xi_c(y)) + c\nabla h(x_c(y))\mathcal{J}h(x_c(y))\\ &+c\nabla g(x_c(y))W\mathcal{J}g(x_c(y))\,.\end{aligned}$$

From (18) and the definition of δ_0, $\mathcal{A}_c(y, W)$ is positive definite for any $W \in \partial_B\Pi_K(\xi - cg(x_c(y)))$. Then from (29) and (30) we obtain that

$$(x_c)'(y;\Delta y) = \mathcal{A}_c(y, \widehat{W})^{-1}\left(-\nabla h(x_c(y))(\Delta\zeta) + \nabla g(x_c(y))\widehat{W}(\Delta\xi)\right). \tag{31}$$

Therefore, we have from (31) and (28) that for any $y = (\zeta, \xi) \in \mathcal{D}_{\nabla\vartheta_c}$,

$$\begin{aligned}\nabla^2\vartheta_c(y)(\Delta y) \in \Bigg\{&\begin{pmatrix}\mathcal{J}h(x_c(y)) \\ -W\mathcal{J}g(x_c(y))\end{pmatrix}\mathcal{A}_c(y, W)^{-1}\left(-\nabla h(x_c(y))(\Delta\zeta)\right.\\ &\left.+\nabla g(x_c(y))W(\Delta\xi)\right) + \begin{pmatrix}0 \\ -c^{-1}(\Delta\xi) + c^{-1}W(\Delta\xi)\end{pmatrix}\Bigg|\; W \in \partial_B\Pi_K(\xi - cg(x_c(y)))\Bigg\},\end{aligned}$$

which, together with the continuity of $x_c(\cdot)$ and the upper semicontinuity of $\partial_B\Pi_K(\cdot)$, implies that for any $V \in \partial_B(\nabla\vartheta_c)(\overline{y})$, one has $V(\Delta y) \in \overline{\mathcal{V}}_c(\Delta y)$. Consequently, (27) holds. $\square$

The second basic assumption needed in this section is stated below.
Assumption B2. There exist positive numbers $\overline{c} \geq c_0$, $\mu_0 > 0$, $\varrho_0 > 0$, and $\tau > 1$ such that for any $c \geq \overline{c}$ and $\Delta y \in \Re^m \times Y$,

$$\|(x_c)'(\overline{y}; \Delta y))\| \leq \varrho_0 \|\Delta y\|/c \tag{32}$$

and

$$\big\langle V(\Delta y) + c^{-1}\Delta y,\, \Delta y \big\rangle \in \mu_0\,[-1, 1]\,\|\Delta y\|^2/c^\tau \quad \forall\, V(\Delta y) \in \overline{\mathcal{V}}_c(\Delta y)\,. \tag{33}$$

Relation (32) in Assumption B2 is about an estimate of the directional derivative of $x_c(\cdot)$ at $\overline{y}$ while (33) pertains to the generalized Jacobian of $\nabla\vartheta_c(\cdot)$ at $\overline{y}$. It will be shown in the next section that Assumption B2 is valid for (NLSDP) when the constraint nondegeneracy condition and the strong second order sufficient condition are satisfied.

Under Assumptions B1 and B2, we are ready to give the main result on the rate of convergence of the augmented Lagrangian method for the constrained optimization problem (OP).

Theorem 1 *Suppose that K is a self-dual cone and that $\Pi_K(\cdot)$ is semismooth everywhere. Let Assumptions B1 and B2 be satisfied. Let c_0, η, $\overline{c}$, μ_0, ϱ_0, and τ be the positive numbers defined in these assumptions. Define*

$$\varrho_1 := 2\varrho_0 \quad \text{and} \quad \varrho_2 := 4\mu_0.$$

Then for any $c \geq \overline{c}$, there exist two positive numbers ε and δ (both depending on c) such that for any $(\zeta, \xi) \in \mathbf{B}_\delta(\overline{\zeta}, \overline{\xi})$, the problem

$$\min\ L_c(x, \zeta, \xi) \quad \text{s.t. } x \in \mathbf{B}_\varepsilon(\overline{x}) \tag{34}$$

has a unique solution denoted $x_c(\zeta, \xi)$. The function $x_c(\cdot, \cdot)$ is locally Lipschitz continuous on $\mathbf{B}_\delta(\overline{\zeta}, \overline{\xi})$ and is semismooth at any point in $\mathbf{B}_\delta(\overline{\zeta}, \overline{\xi})$, and for any $(\zeta, \xi) \in \mathbf{B}_\delta(\overline{\zeta}, \overline{\xi})$, we have

$$\|x_c(\zeta, \xi) - \overline{x}\| \leq \varrho_1 \|(\zeta, \xi) - (\overline{\zeta}, \overline{\xi})\|/c \tag{35}$$

and

$$\|(\zeta_c(\zeta, \xi), \xi_c(\zeta, \xi)) - (\overline{\zeta}, \overline{\xi})\| \leq \varrho_2 \|(\zeta, \xi) - (\overline{\zeta}, \overline{\xi})\|/c^{\tau-1}\,, \tag{36}$$

where $\zeta_c(\zeta, \xi)$ and $\xi_c(\zeta, \xi))$ are defined by (23), i.e.,

$$\zeta_c(\zeta, \xi) := \zeta + ch(x_c(\zeta, \xi)) \quad \text{and} \quad \xi_c(\zeta, \xi) := \Pi_K(\xi - cg(x_c(\zeta, \xi)))\,.$$

Proof. Let $c \geq \overline{c}$. From Proposition 1 we have already known that there exist two positive numbers $\varepsilon > 0$ and $\delta_0 > 0$ (both depending on c) and a locally Lipschitz continuous function $x_c(\cdot, \cdot)$ defined on $\mathbf{B}_{\delta_0}(\overline{\zeta}, \overline{\xi})$ such that the function $x_c(\cdot, \cdot)$ is semismooth at any point in $\mathbf{B}_{\delta_0}(\overline{\zeta}, \overline{\xi})$ and for any $(\zeta, \xi) \in \mathbf{B}_{\delta_0}(\overline{\zeta}, \overline{\xi})$, $x_c(\zeta, \xi)$ is the unique solution to (34).

Denote $y := (\zeta, \xi) \in \Re^m \times Y$. Since $x_c(\cdot)$ is locally Lipschitz continuous on $\mathbf{B}_{\delta_0}(\overline{y})$ and is directionally differentiable at $\overline{y}$, by [41] we know that $x_c(\cdot)$

is Bouligand-differentiable at $\overline{y}$, i.e., $x_c(\cdot)$ is directionally differentiable at $\overline{y}$ and

$$\lim_{y\to\overline{y}} \frac{\|x_c(y) - x_c(\overline{y}) - (x_c)'(y; y - \overline{y})\|}{\|y - \overline{y}\|} = 0\,.$$

By Proposition 2, $\nabla\vartheta_c(\cdot)$ is semismooth at $\overline{y}$, and thus is also Bouligand-differentiable at $\overline{y}$. Then there exists $\delta \in (0, \delta_0]$ such that for any $y \in \mathbf{B}_\delta(\overline{y})$,

$$\|x_c(y) - x_c(\overline{y}) - (x_c)'(\overline{y}; y - \overline{y})\| \le \varrho_0\|y - \overline{y}\|/c \tag{37}$$

and

$$\|\nabla\vartheta_c(y) - \nabla\vartheta_c(\overline{y}) - (\nabla\vartheta_c)'(\overline{y}; y - \overline{y})\| \le \mu_0\|y - \overline{y}\|/c^\tau\,. \tag{38}$$

Let $y := (\zeta, \xi) \in \mathbf{B}_\delta(\overline{y})$ be an arbitrary point. From (32), (37), and the fact that $x_c(\overline{y}) = \overline{x}$, we have

$$\|x_c(y) - \overline{x}\| \le \|(x_c)'(\overline{y}; y - \overline{y})\| + \varrho_0\|y - \overline{y}\|/c = \varrho_1\|y - \overline{y}\|/c\,,$$

which, shows that (35) holds.

Since $\nabla\vartheta_c(\cdot)$ is semismooth at $\overline{y}$, there exists an element $V \in \partial_B(\nabla\vartheta_c)(\overline{y})$ such that $(\nabla\vartheta_c)'(\overline{y}; y-\overline{y}) = V(y-\overline{y})$. By using the fact that V is self-adjoint (see Lemma 4), we know from (33) in Assumption B2 and Proposition 3 that

$$\|V(y - \overline{y}) + c^{-1}(y - \overline{y})\| \le 3\mu_0\|y - \overline{y}\|/c^\tau\,. \tag{39}$$

Therefore, we have from (38) and (39)

$$\begin{aligned}
&\|y + c\nabla\vartheta_c(y) - \overline{y}\|\\
&= c\|\nabla\vartheta_c(y) - \nabla\vartheta_c(\overline{y}) - (\nabla\vartheta_c)'(\overline{y}; y - \overline{y}) + (\nabla\vartheta_c)'(\overline{y}; y - \overline{y}) + c^{-1}(y - \overline{y})\|\\
&\le c\|\nabla\vartheta_c(y) - \nabla\vartheta_c(\overline{y}) - (\nabla\vartheta_c)'(\overline{y}; y - \overline{y})\| + c\|V(y - \overline{y}) + c^{-1}(y - \overline{y})\|\\
&\le \mu_0\|y - \overline{y}\|/c^{\tau-1} + 3\mu_0\|y - \overline{y}\|/c^{\tau-1} = \varrho_2\|y - \overline{y}\|/c^{\tau-1}\,,
\end{aligned}$$

which, together with (25) and the definitions of $\zeta_c(\zeta, \xi)$ and $\xi_c(\zeta, \xi)$), proves (36). The proof is completed. □

Under Assumptions B1 and B2, Theorem 1 shows that if for all k sufficiently large with $c_k \equiv c$ larger than a threshold and if (x^k, ζ^k, ξ^k) is sufficiently close to $(\overline{x}, \overline{\zeta}, \overline{\xi})$, then the augmented Lagrangian method can locally be regarded as the gradient ascent method applied to the dual problem

$$\max\ \vartheta_c(\zeta, \xi) \quad \text{s.t. } (\zeta, \xi) \in \Re^m \times Y$$

with a constant step-length c, i.e., for all k sufficiently large

$$\begin{pmatrix} \zeta^{k+1} \\ \xi^{k+1} \end{pmatrix} = \begin{pmatrix} \zeta^k \\ \xi^k \end{pmatrix} + c\nabla\vartheta_c(\zeta^k, \xi^k)\,.$$

By (32) in Assumption B2, we see that locally the augmented Lagrangian method can also be treated as an approximate generalized Newton method applied to the following nonsmooth equation

$$\nabla\vartheta_c(\zeta, \xi) = 0$$

with $-c^{-1}\mathcal{I}$ as a good estimate to elements in $\partial\nabla\vartheta_c(\zeta^k,\xi^k)$ for all (ζ^k,ξ^k) sufficiently close to $(\overline{\zeta},\overline{\xi})$ as every element in $\partial\nabla\vartheta_c(\overline{\zeta},\overline{\xi})$ is in the form of $-c^{-1}I+O(c^{-\tau})$, where $\mathcal{I}$ is the identity operator in $\Re^m\times Y$. Since $\nabla\vartheta_c(\cdot,\cdot)$ is semismooth at $(\overline{\zeta},\overline{\xi})$ (cf. Proposition 2), the fast local convergence of the augmented Lagrangian method comes no surprise for those who are familiar with the theory developed by Kummer [22] and Qi and Sun [32] on the superlinear convergence of the generalized Newton method for semismooth equations.

The local rate of convergence for $\{(\zeta^k,\xi^k)\}$ established in Theorem 1 is proportional to $1/c^{\tau-1}$, which tends to zero as $c\to\infty$. However, to increase the value of c may force the convergence sphere to shrink. In the next section, we shall check whether Assumptions B1 and B2 imposed in this section can be satisfied by nonlinear semidefinite programming.

4 The case for nonlinear semidefinite programming

This section is devoted to studying the following nonlinear semidefinite programming

$$\text{(NLSDP)} \qquad \min\ f(x) \quad \text{s.t. } h(x)=0\,,\ g(x)\in\mathcal{S}^p_+\,,$$

where $f:\Re^n\mapsto\Re$, $h:\Re^n\mapsto\Re^m$, and $g:\Re^n\mapsto\mathcal{S}^p$ are twice continuously differentiable. Nonlinear semidefinite programming (NLSDP) is a special case of (OP) with $X:=\Re^n$, $Y:=\mathcal{S}^p$ and $K:=\mathcal{S}^p_+$. The Lagrangian function for (NLSDP) is

$$L_0(x,\zeta,\Xi)=f(x)+\langle\zeta,h(x)\rangle-\langle\Xi,g(x)\rangle\,,\quad (x,\zeta,\Xi)\in\Re^n\times\Re^m\times\mathcal{S}^p\,,$$

where we use Ξ instead of ξ to represent the Lagrange multiplier corresponding to the constraint $g(x)\in\mathcal{S}^p_+$. Then for any $(x,\zeta,\Xi)\in\Re^n\times\Re^m\times\mathcal{S}^p$,

$$\nabla_x L_0(x,\zeta,\Xi)=\nabla f(x)+\nabla h(x)\zeta-\nabla g(x)\Xi\,.$$

Let $(\overline{x},\overline{\zeta},\overline{\Xi})\in\Re^n\times\Re^m\times\mathcal{S}^p$ be a given KKT point. Then, $(\overline{x},\overline{\zeta},\overline{\Xi})$ satisfies

$$\nabla_x L_0(\overline{x},\overline{\zeta},\overline{\Xi})=0\,,\quad h(\overline{x})=0\,,\quad \overline{\Xi}\succeq 0\,,\quad g(\overline{x})\succeq 0\,,\quad \text{and}\quad \langle\overline{\Xi},g(\overline{x})\rangle=0\,. \tag{40}$$

Let $\overline{Z}:=\overline{\Xi}-g(\overline{x})$. Suppose that $\overline{Z}$ has the spectral decomposition as in (6), i.e,

$$\overline{Z}=\overline{P}\Lambda\overline{P}^T\,,$$

where Λ is the diagonal matrix of eigenvalues of $\overline{Z}$ and $\overline{P}$ is a corresponding orthogonal matrix of orthonormal eigenvectors. Define three index sets of positive, zero, and negative eigenvalues of $\overline{Z}$, respectively, as

$$\alpha:=\{i\,|\,\lambda_i>0\},\ \beta:=\{i\,|\,\lambda_i=0\},\ \gamma:=\{i\,|\,\lambda_i<0\}.$$

Write

$$\Lambda=\begin{bmatrix}\Lambda_\alpha & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & \Lambda_\gamma\end{bmatrix}\ \text{and}\ \overline{P}=[\,\overline{P}_\alpha\ \overline{P}_\beta\ \overline{P}_\gamma\,]$$

with $\overline{P}_\alpha \in \Re^{p\times|\alpha|}$, $\overline{P}_\beta \in \Re^{p\times|\beta|}$, and $\overline{P}_\gamma \in \Re^{p\times|\gamma|}$. From (40), we know that $\overline{\Xi}g(\overline{x}) = g(\overline{x})\overline{\Xi} = 0$. Thus, we have

$$\overline{\Xi} = \overline{P}\begin{bmatrix} \Lambda_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\overline{P}^T, \quad g(\overline{x}) = \overline{P}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\Lambda_\gamma \end{bmatrix}\overline{P}^T$$

$$\overline{\Xi} - tg(\overline{x}) = \overline{P}\begin{bmatrix} \Lambda_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & t\Lambda_\gamma \end{bmatrix}\overline{P}^T. \tag{41}$$

Let

$$\underline{\nu}_0 := \min_{i\in\alpha, j\in\gamma} \lambda_i/|\lambda_j|, \quad \overline{\nu}_0 := \max_{i\in\alpha, j\in\gamma} \lambda_i/|\lambda_j|. \tag{42}$$

Let $\mathcal{Q}$ be the set of all orthogonal matrices of order $|\beta| \times |\beta|$. Define $\mathcal{P}$ by (11), i.e,

$$\mathcal{P} = \left\{ P \in \Re^{p\times p} \mid P = [\overline{P}_\alpha \ (\overline{P}_\beta Q) \ \overline{P}_\gamma], \ Q \in \mathcal{Q} \right\}.$$

We now introduce the conditions needed in this section.

Assumption (nlsdp-A1). The constraint nondegeneracy condition holds at $\overline{x}$:

$$\begin{pmatrix} \mathcal{J}h(\overline{x}) \\ \mathcal{J}g(\overline{x}) \end{pmatrix} \Re^n + \begin{pmatrix} \{0\} \\ \operatorname{lin}\left(\mathcal{T}_{\mathcal{S}^p_+}(g(\overline{x}))\right) \end{pmatrix} = \begin{pmatrix} \Re^m \\ \mathcal{S}^p \end{pmatrix}. \tag{43}$$

Assumption (nlsdp-A1) is the analogue to the linear independence constraint qualification for nonlinear programming [33,42]. It also implies that $\mathcal{M}(\overline{x})$ is a singleton [8, Proposition 4.50].

Assumption (nlsdp-A2) The strong second order sufficient condition holds at $\overline{x}$ [45]:

$$\left\langle d, \nabla^2_{xx} L_0(\overline{x}, \overline{\zeta}, \overline{\Xi}) d \right\rangle + \Upsilon_{g(\overline{x})}(\overline{\Xi}, \mathcal{J}g(\overline{x})d) > 0 \quad \forall\, d \in \operatorname{app}(\overline{\zeta}, \overline{\Xi}) \setminus \{0\},$$

where

$$\operatorname{app}(\overline{\zeta}, \overline{\Xi}) := \left\{ d \mid \mathcal{J}h(\overline{x})d = 0, \ \mathcal{J}g(\overline{x})d \in \operatorname{aff}(\mathcal{C}(g(\overline{x}) - \overline{\Xi}; \mathcal{S}^p_+)) \right\} \tag{44}$$

and for any given $B \in \mathcal{S}^p$, the linear-quadratic function $\Upsilon_B(\cdot,\cdot)$ is defined as

$$\Upsilon_B(\Gamma, C) := 2\left\langle \Gamma, CB^\dagger C \right\rangle, \quad (\Gamma, C) \in \mathcal{S}^p \times \mathcal{S}^p$$

with $B^\dagger$ being the Moore-Penrose pseudo-inverse of B.

Note that if the strict complementarity condition (i.e., $\beta = \emptyset$) holds, then the strong second order sufficient condition made in Assumption (nlsdp-A2) reduces to the so called "no gap" second order sufficient optimality condition [8, Section 5.3.5] as in this case the two sets $\operatorname{aff}(\mathcal{C}(g(\overline{x})-\overline{\Xi}; \mathcal{S}^p_+))$ and $\mathcal{C}(g(\overline{x})-\overline{\Xi}; \mathcal{S}^p_+)$ coincide. In the general case, as its name suggests, the strong second order sufficient condition is a stronger condition than the second order sufficient optimality condition. See [43,45] for many conditions equivalent or related to Assumptions (nlsdp-A1) and (nlsdp-A2).

Let $P \in \mathcal{P}$. Then there exists $Q \in \mathcal{Q}$ such that $P = [\overline{P}_\alpha \ (\overline{P}_\beta Q) \ \overline{P}_\gamma]$. For index sets $\chi, \chi' \in \{\alpha, \beta, \gamma\}$, let

$$C_{(\chi,\chi')}(P) := \Big(\mathrm{vec}(P_\chi^T \mathcal{J}_{x_1} g(\overline{x}) P_{\chi'}) \ \cdots \ \mathrm{vec}(P_\chi^T \mathcal{J}_{x_n} g(\overline{x}) P_{\chi'})\Big)$$

and

$$\widehat{C}_{(\chi,\chi)}(P) := \Big(\mathrm{svec}(P_\chi^T \mathcal{J}_{x_1} g(\overline{x}) P_{\chi}) \ \cdots \ \mathrm{svec}(P_\chi^T \mathcal{J}_{x_n} g(\overline{x}) P_{\chi})\Big),$$

where $\mathrm{vec}(B)$ denotes the vector obtained by stacking up all the columns of a given matrix B and $\mathrm{svec}(B)$ denotes the vector obtained by stacking up all the columns of the upper triangular part of a given symmetric matrix B. Since $P_\alpha = \overline{P}_\alpha$ and $P_\gamma = \overline{P}_\gamma$, we write $C_{(\chi,\chi')}$ and $\widehat{C}_{(\chi,\chi)}$ instead of $C_{(\chi,\chi')}(P)$ and $\widehat{C}_{(\chi,\chi)}(P)$, respectively if $\chi, \chi' \in \{\alpha, \gamma\}$. Define

$$n_1 := m + |\alpha|(|\alpha|+1)/2\,, \ n_2 := n_1 + |\beta|(|\beta|+1)/2 + |\alpha||\beta|\,, \ n_3 := n - n_2\,,$$

and

$$A(P) := \begin{pmatrix} \mathcal{J}h(\overline{x}) \\ -\widehat{C}_{(\alpha,\alpha)} \\ -\widehat{C}_{(\beta,\beta)}(P) \\ -C_{(\alpha,\beta)}(P) \end{pmatrix}.$$

Suppose that Assumption (nlsdp-A1) holds. Then by (43) in Assumption (nlsdp-A1) we know that $A(P)$ is of full row rank[1]. Let $A(P)$ have the following singular value decomposition:

$$A(P) = U[\Sigma(P) \quad 0]R^T\,, \tag{45}$$

where $U \in \Re^{n_2 \times n_2}$ and $R \in \Re^{n \times n}$ are orthogonal matrices, $\Sigma(P) = \mathrm{Diag}\Big(\sigma_1(A(P)), \cdots, \sigma_{n_2}(A(P))\Big)$, and $\sigma_1(A(P)) \geq \sigma_2(A(P)) \geq \cdots \geq \sigma_{n_2}(A(P)) > 0$ are the singular values of $A(P)$. It should be pointed out here that U and R also depend on P. But for the sake of notational simplification, we drop the argument P from U and R in our analysis below.

Let

$$\underline{\sigma} := \min\left\{1, \min_{P\in\mathcal{P}} \min_{1\le i\le n_2} \sigma_i^{-2}(A(P))\right\} \text{ and } \overline{\sigma} := \max\left\{1, \max_{P\in\mathcal{P}} \max_{1\le i\le n_2} \sigma_i^{-2}(A(P))\right\}.$$

Then, since $\mathcal{P}$ is a compact set and $\Sigma(P)$ changes continuously with respect to P, both $\underline{\sigma}$ and $\overline{\sigma}$ are finite positive numbers. Thus there exist two positive numbers $\underline{\nu}$ and $\overline{\nu}$ such that for any $P \in \mathcal{P}$ and $s \in \Re^{|\alpha||\gamma|}$,

$$\underline{\nu}\|s\|^2 \le \max\left\{\left\langle s, \widetilde{C}_{(\alpha,\gamma)}(P)(\widetilde{C}^T_{(\alpha,\gamma)}(P))s\right\rangle, \left\langle s, C_{(\alpha,\gamma)} C^T_{(\alpha,\gamma)} s\right\rangle\right\} \le \overline{\nu}\|s\|^2, \tag{46}$$

[1] One may consult [8, Proposition 5.71] for a proof, where Bonnans and Shapiro only considered the case that $g(x) \in \mathcal{S}^p_+$. However, it is easy to modify their arguments to include the equality constraint $h(x) = 0$.

where

$$\widetilde{C}_{(\alpha,\gamma)}(P) := C_{(\alpha,\gamma)}\widetilde{R} \quad \text{and} \quad \widetilde{R} := R\begin{bmatrix} \Sigma(P)^{-1}U^T & 0 \\ 0 & I_{n_3} \end{bmatrix}.$$

When no ambiguity arises, we often drop P from $A(P)$, $C_{(\chi,\chi')}(P)$, and $\widetilde{C}_{(\alpha,\gamma)}(P)$.

Let $c > 0$ and $W \in \partial_B \Pi_{\mathcal{S}^p_+}(\overline{\Xi} - cg(\overline{x}))$. Define $\lambda_c \in \Re^p$ as

$$(\lambda_c)_i := \begin{cases} \lambda_i & \text{if } i \in \alpha \cup \beta\,, \\ c\lambda_i & \text{if } i \in \gamma\,. \end{cases}$$

Then it follows from Lemma 6 that there exist two matrices $Q \in \mathcal{Q}$ with $P = [\overline{P}_\alpha \ (\overline{P}_\beta Q) \ \overline{P}_\gamma]$ and $\Theta_c \in \mathcal{S}^p$ such that

$$W(H) = P\left(\Theta_c \circ (P^T H P)\right)P^T \quad \forall\, H \in \mathcal{S}^p \tag{47}$$

with the entries of Θ_c being given by

$$\begin{cases} (\Theta_c)_{ij} = \dfrac{\max\{(\lambda_c)_i, 0\} + \max\{(\lambda_c)_j, 0\}}{|\,(\lambda_c)_i\,| + |\,(\lambda_c)_j\,|} & \text{if } (i,j) \notin \beta\times\beta\,, \\ (\Theta_c)_{ij} \in [0,1] & \text{if } (i,j) \in \beta\times\beta\,. \end{cases} \tag{48}$$

For index sets $\chi, \chi' \in \{\alpha, \beta, \gamma\}$, we introduce the following notation:

$$(\Theta_c)_{(\chi,\chi')} = \mathrm{Diag}\left(\mathrm{vec}((\Theta_c)_{\chi\chi'})\right)\,, \quad (\widehat{\Theta}_c)_{(\chi,\chi)} = \mathrm{Diag}\left(\mathrm{svec}((\Theta_c)_{\chi\chi} \circ E_{\chi\chi})\right)\,,$$

where "$\circ$" is the Hadamard product and E is a matrix in $\mathcal{S}^p$ with entries being given by

$$E_{ij} := \begin{cases} 1 \text{ if } i = j\,, \\ 2 \text{ if } i \neq j\,. \end{cases}$$

Let

$$D_c := \begin{bmatrix} I_m & 0 & 0 & 0 \\ 0 & (\widehat{\Theta}_c)_{(\alpha,\alpha)} & 0 & 0 \\ 0 & 0 & (\widehat{\Theta}_c)_{(\beta,\beta)} & 0 \\ 0 & 0 & 0 & 2I_{|\alpha||\beta|} \end{bmatrix}.$$

Let $\mathcal{A}_c(\overline{\zeta}, \overline{\xi}, W)$ be defined as (19), i.e,

$$\mathcal{A}_c(\overline{\zeta}, \overline{\Xi}, W) = \nabla^2_{xx} L_0(\overline{x}, \overline{\zeta}, \overline{\Xi}) + c\nabla h(\overline{x})\mathcal{J}h(\overline{x}) + c\nabla g(\overline{x}) W \mathcal{J}g(\overline{x})\,.$$

A compact formula for $\mathcal{A}_c(\overline{\zeta}, \overline{\xi}, W)$ is given in the next lemma.

Lemma 9 *The matrix $\mathcal{A}_c(\overline{\zeta}, \overline{\xi}, W)$ can be expressed equivalently as*

$$\begin{aligned} \mathcal{A}_c(\overline{\zeta}, \overline{\Xi}, W) = \nabla^2_{xx} L_0(\overline{x}, \overline{\zeta}, \overline{\Xi}) &+ c\,(\nabla h(\overline{x})\mathcal{J}h(\overline{x}) \\ &+ \widehat{C}^T_{(\alpha,\alpha)}(\widehat{\Theta}_c)_{(\alpha,\alpha)}\widehat{C}_{(\alpha,\alpha)} + 2C^T_{(\alpha,\beta)}C_{(\alpha,\beta)} \\ &+ 2C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}C_{(\alpha,\gamma)} + \widehat{C}^T_{(\beta,\beta)}(\widehat{\Theta}_c)_{(\beta,\beta)}\widehat{C}_{(\beta,\beta)}\Big)\,. \end{aligned} \tag{49}$$

Proof. Let d be an arbitrary point in $\Re^n$. By the definition of $\nabla g(\overline{x})$, we have for $H := \mathcal{J}g(\overline{x})d$ that

$$\nabla g(\overline{x})W(H) = \begin{bmatrix} \langle \mathcal{J}_{x_1} g(\overline{x}), W(H)\rangle \\ \langle \mathcal{J}_{x_2} g(\overline{x}), W(H)\rangle \\ \vdots \\ \langle \mathcal{J}_{x_n} g(\overline{x}), W(H)\rangle \end{bmatrix} . \tag{50}$$

Noting that from (47) and (48), for any $1 \le l \le p$,

$$\begin{aligned}
&\langle \mathcal{J}_{x_l} g(\overline{x}), W(H)\rangle = \Big\langle P^T \mathcal{J}_{x_l} g(\overline{x})P, \sum_{i=1}^{p} P^T W(\mathcal{J}_{x_i} g(\overline{x})d_i)P \Big\rangle \\
&= \Big\langle P^T \mathcal{J}_{x_l} g(\overline{x})P, (\Theta_c) \circ \Big(\sum_{i=1}^{p} P^T \mathcal{J}_{x_i} g(\overline{x})P d_i \Big) \Big\rangle \\
&= \langle (C_{(\alpha,\alpha)})_l, C_{(\alpha,\alpha)} d\rangle + \langle (C_{(\beta,\alpha)})_l, C_{(\beta,\alpha)} d\rangle + \langle (C_{(\alpha,\beta)})_l, C_{(\alpha,\beta)} d\rangle \\
&\quad + \langle (C_{(\gamma,\alpha)})_l, (\Theta_c)_{(\gamma,\alpha)} C_{(\gamma,\alpha)} d\rangle + \langle (C_{(\alpha,\gamma)})_l, (\Theta_c)_{(\alpha,\gamma)} C_{(\alpha,\gamma)} d\rangle \\
&\quad + \langle (C_{(\beta,\beta)})_l, (\Theta_c)_{(\beta,\beta)} C_{(\beta,\beta)} d\rangle \\
&= \langle (\widehat{C}_{(\alpha,\alpha)})_l, (\widehat{\Theta}_c)_{(\alpha,\alpha)} \widehat{C}_{(\alpha,\alpha)} d\rangle + \langle (C_{(\beta,\alpha)})_l, C_{(\beta,\alpha)} d\rangle + \langle (C_{(\alpha,\beta)})_l, C_{(\alpha,\beta)} d\rangle \\
&\quad + \langle (C_{(\gamma,\alpha)})_l, (\Theta_c)_{(\gamma,\alpha)} C_{(\gamma,\alpha)} d\rangle + \langle (C_{(\alpha,\gamma)})_l, (\Theta_c)_{(\alpha,\gamma)} C_{(\alpha,\gamma)} d\rangle \\
&\quad + \langle (\widehat{C}_{(\beta,\beta)})_l, (\widehat{\Theta}_c)_{(\beta,\beta)} \widehat{C}_{(\beta,\beta)} d\rangle ,
\end{aligned}$$

we have from (50) that

$$\begin{aligned}
\nabla g(\overline{x})W(\mathcal{J}_x g(\overline{x})d) = \Big(& \widehat{C}^T_{(\alpha,\alpha)} (\widehat{\Theta}_c)_{(\alpha,\alpha)} \widehat{C}_{(\alpha,\alpha)} + 2C^T_{(\alpha,\beta)} C_{(\alpha,\beta)} \\
& + 2C^T_{(\alpha,\gamma)} (\Theta_c)_{(\alpha,\gamma)} C_{(\alpha,\gamma)} + \widehat{C}^T_{(\beta,\beta)} (\widehat{\Theta}_c)_{(\beta,\beta)} \widehat{C}_{(\beta,\beta)} \Big) d .
\end{aligned}$$

Since d is arbitrarily chosen, (49) holds. $\square$

Lemma 9 shows that $\mathcal{A}_c(\overline{\zeta}, \overline{\Xi}, W)$ can be written as

$$\mathcal{A}_c(\overline{\zeta}, \overline{\Xi}, W) = \nabla^2_{xx} L_0(\overline{x}, \overline{\zeta}, \overline{\Xi}) + cA^T D_c A + 2cC^T_{(\alpha,\gamma)} (\Theta_c)_{(\alpha,\gamma)} C_{(\alpha,\gamma)} . \tag{51}$$

For any $c', c > 0$, let

$$\mathcal{B}_{c',c}(\overline{\zeta}, \overline{\Xi}, W) := \nabla^2_{xx} L_0(\overline{x}, \overline{\zeta}, \overline{\Xi}) + c' A^T D_c A + 2cC^T_{(\alpha,\gamma)} (\Theta_c)_{(\alpha,\gamma)} C_{(\alpha,\gamma)} . \tag{52}$$

The following proposition shows that, under Assumptions (nlsdp-A1) and (nlsdp-A2), the basic Assumption B1 made in Section 3 is satisfied by nonlinear semidefinite programming.

Proposition 4 *Suppose that Assumptions (nlsdp-A1) and (nlsdp-A2) are satisfied. Then there exist two positive numbers c_0 and $\underline{\eta}$ such that for any $c \ge c_0$ and $W \in \partial_B \Pi_{\mathcal{S}^p_+}(\overline{\Xi} - cg(\overline{x}))$,*

$$\big\langle d, \mathcal{A}_c(\overline{\zeta}, \overline{\Xi}, W)d \big\rangle \ge \big\langle d, \mathcal{B}_{c_0,c}(\overline{\zeta}, \overline{\Xi}, W)d \big\rangle \ge \underline{\eta} \langle d, d\rangle \quad \forall\, d \in \Re^n .$$

Proof. It follows from Assumption (nlsdp-A2) that there exists $\eta_0 > 0$ such that

$$\langle d, \nabla_{xx}^2 L_0(\overline{x}, \overline{\zeta}, \overline{\Xi})d\rangle + \Upsilon_{g(\overline{x})}(\overline{\Xi}, \mathcal{J}g(\overline{x})d) \geq \eta_0\|d\|^2 \quad \forall\, d \in \mathrm{app}(\overline{\zeta}, \overline{\Xi})\,. \tag{53}$$

By (10), (41), (44), and the fact that $g(\overline{x}) - \overline{\Xi} = -\overline{Z}$, we have

$$\mathrm{app}(\overline{\zeta}, \overline{\Xi}) = \{d \,|\, \mathcal{J}h(\overline{x})d = 0\,,\ \overline{P}_\alpha^T(\mathcal{J}g(\overline{x})d)\overline{P}_\alpha = 0\,,\ \overline{P}_\alpha^T(\mathcal{J}g(\overline{x})d)\overline{P}_\beta = 0\}$$

or equivalently

$$\mathrm{app}\ (\overline{\zeta}, \overline{\Xi}) = \{d \,|\, \mathcal{J}h(\overline{x})d = 0\,,\ \widehat{C}_{(\alpha,\alpha)}d = 0\,,\ C_{(\alpha,\beta)}(\overline{P})d = 0\}\,. \tag{54}$$

Since (53) and (54) hold, by using Lemma 7 with ϕ and $\mathcal{L}$ being defined by $\phi(d) := \langle d, \nabla_{xx}^2 L_0(\overline{x}, \overline{\zeta}, \overline{\Xi})d\rangle + \Upsilon_{g(\overline{x})}(\Xi, \mathcal{J}g(\overline{x})d)$ and $\mathcal{L}(d) := (\mathcal{J}h(\overline{x})d; \widehat{C}_{(\alpha,\alpha)}d; C_{(\alpha,\beta)}(\overline{P})d)$ for any $d \in \Re^n$, respectively, we know that there exist two positive numbers c_1 and $\underline{\eta} \in (0, \eta_0/2]$ such that for any $c \geq c_1$,

$$\begin{aligned}&\langle d, \nabla_{xx}^2 L_0(\overline{x}, \overline{\zeta}, \overline{\Xi})d\rangle + \Upsilon_{g(\overline{x})}(\Xi, \mathcal{J}g(\overline{x})d)\\ &\quad + c\|\mathcal{J}h(\overline{x})d\|^2 + c\|\widehat{C}_{(\alpha,\alpha)}d\|^2 + c\|C_{(\alpha,\beta)}(\overline{P})d\|^2 \geq 2\underline{\eta}\|d\|^2 \quad \forall\, d \in \Re^n\,.\end{aligned} \tag{55}$$

Let $c_0 \geq c_1$ be such that for any $c \geq c_0$,

$$\max_{1\leq l\leq n} \|\mathcal{J}_{x_l} g(\overline{x})\|^2 \sum_{i\in\gamma, j\in\alpha} \frac{\lambda_j^2}{|\lambda_i|(\lambda_j + c|\lambda_i|)} \leq \underline{\eta}/2\,. \tag{56}$$

Let $c \geq c_0$ and $W \in \partial_B \Pi_{\mathcal{S}_+^p}(\overline{\Xi} - cg(\overline{x}))$. Then there exist two matrices $Q \in \mathcal{Q}$ with $P = [\overline{P}_\alpha\ (\overline{P}_\beta Q)\ \overline{P}_\gamma]$ and $\Theta_c \in \mathcal{S}^p$ satisfying (48) such that (47) holds, i.e.,

$$W(H) = P\left(\Theta_c \circ (P^T H P)\right) P^T \quad \forall\, H \in \mathcal{S}^p\,.$$

It is easy to see from (56) that for any $c \geq c_0$ and $d \in \Re^n$ we have

$$\begin{aligned}
&\Upsilon_{g(\overline{x})}(\overline{\Xi}, \mathcal{J}g(\overline{x})d) - 2c\left\langle d, C_{(\alpha,\gamma)}^T(\Theta_c)_{(\alpha,\gamma)}C_{(\alpha,\gamma)}d\right\rangle\\
&= 2\left\langle \overline{\Xi}, \mathcal{J}g(\overline{x})dg(\overline{x})^\dagger \mathcal{J}g(\overline{x})d\right\rangle - 2c\left\langle d, C_{(\alpha,\gamma)}^T(\Theta_c)_{(\alpha,\gamma)}C_{(\alpha,\gamma)}d\right\rangle\\
&= 2\sum_{i\in\gamma, j\in\alpha} \frac{\lambda_j}{|\lambda_i|}\left(\sum_{l=1}^n P_i^T \mathcal{J}_{x_l} g(\overline{x})P_j d_l\right)^2\\
&\quad -2c\sum_{i\in\gamma, j\in\alpha} \frac{\lambda_j}{\lambda_j + c|\lambda_i|}\left(\sum_{l=1}^n P_i^T \mathcal{J}_{x_l} g(\overline{x})P_j d_l\right)^2\\
&\leq 2\sum_{i\in\gamma, j\in\alpha}\left[\frac{\lambda_j^2}{|\lambda_i|(\lambda_j + c|\lambda_i|)}\sum_{l=1}^n \|\mathcal{J}_{x_l} g(\overline{x})\|^2\|P_i\|^2\|P_j\|^2 d_l^2\right]\\
&\leq 2\max_{1\leq l\leq n}\|\mathcal{J}_{x_l} g(\overline{x})\|^2 \sum_{i\in\gamma, j\in\alpha}\frac{\lambda_j^2}{|\lambda_i|(\lambda_j + c|\lambda_i|)}\|d\|^2\\
&\leq \underline{\eta}\|d\|^2\,,
\end{aligned}$$

which, together with (55), implies that for any $c \geq c_0$ we have

$$\begin{aligned}\langle d, \nabla^2_{xx} L_0(\overline{x}, \overline{\zeta}, \overline{\Xi}) d\rangle + 2c \left\langle d, C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)} C_{(\alpha,\gamma)} d \right\rangle \\ + c_0 \|\mathcal{J}h(\overline{x})d\|^2 + c_0 \|\widehat{C}_{(\alpha,\alpha)} d\|^2 + c_0 \|C_{(\alpha,\beta)}(\overline{P})d\|^2 \ \geq \underline{\eta}\|d\|^2 \quad \forall\, d \in \Re^n .\end{aligned} \tag{57}$$

Let "$\otimes$" denote the Kronecker product. Since for any $d \in \Re^n$,

$$\begin{aligned}
& \|C_{(\alpha,\beta)}(P)d\|^2 = \left\langle C_{(\alpha,\beta)}(P)d, C_{(\alpha,\beta)}(P)d \right\rangle \\
= \ & \left\langle \sum_{l=1}^n \mathrm{vec}(P^T_\alpha \mathcal{J}_{x_l} g(\overline{x}) P_\beta) d_l, \sum_{l=1}^n \mathrm{vec}(P^T_\alpha \mathcal{J}_{x_l} g(\overline{x}) P_\beta) d_l \right\rangle \\
= \ & \left\langle \sum_{l=1}^n \mathrm{vec}(\overline{P}^T_\alpha \mathcal{J}_{x_l} g(\overline{x}) \overline{P}_\beta Q) d_l, \sum_{l=1}^n \mathrm{vec}(\overline{P}^T_\alpha \mathcal{J}_{x_l} g(\overline{x}) \overline{P}_\beta Q) d_l \right\rangle \\
= \ & \left\langle (Q^T \otimes I_{|\alpha|}) \sum_{l=1}^n \mathrm{vec}(\overline{P}^T_\alpha \mathcal{J}_{x_l} g(\overline{x}) \overline{P}_\beta) d_l, (Q^T \otimes I_{|\alpha|}) \sum_{l=1}^n \mathrm{vec}(\overline{P}^T_\alpha \mathcal{J}_{x_l} g(\overline{x}) \overline{P}_\beta) d_l \right\rangle \\
= \ & \left\langle \sum_{l=1}^n \mathrm{vec}(\overline{P}^T_\alpha \mathcal{J}_{x_l} g(\overline{x}) \overline{P}_\beta) d_l, \sum_{l=1}^n \mathrm{vec}(\overline{P}^T_\alpha \mathcal{J}_{x_l} g(\overline{x}) \overline{P}_\beta) d_l \right\rangle \\
= \ & \left\langle C_{(\alpha,\beta)}(\overline{P})d, C_{(\alpha,\beta)}(\overline{P})d \right\rangle = \|C_{(\alpha,\beta)}(\overline{P})d\|^2 ,
\end{aligned}$$

from (52), (57), and the fact that $\widehat{C}^T_{(\beta,\beta)}(\widehat{\Theta}_c)_{(\beta,\beta)}\widehat{C}_{(\beta,\beta)} \succeq 0$, we can see that for any $c \geq c_0$,

$$\left\langle d, \mathcal{B}_{c_0,c}(\overline{\zeta}, \overline{\Xi}, W)d \right\rangle \geq \underline{\eta}\|d\|^2 \quad \forall\, d \in \Re^n .$$

By noting the fact that

$$\mathcal{A}_c(\overline{\zeta}, \overline{\Xi}, W) = \mathcal{B}_{c_0,c}(\overline{\zeta}, \overline{\Xi}, W) + (c - c_0)A^T D_c A ,$$

we complete the proof. □

Let Assumptions (nlsdp-A1) and (nlsdp-A2) be satisfied. Let the two positive numbers c_0 and $\underline{\eta}$ be defined as in Proposition 4. Let $c \geq c_0$. Then, by Propositions 1 and 4 and the fact that $\Pi_{\mathcal{S}^p_+}(\cdot)$ is strongly semismooth everywhere, there exist two positive numbers $\varepsilon > 0$ and $\delta_0 > 0$ (both depending on c) and a locally Lipschitz continuous function $x_c(\cdot,\cdot)$ defined on $\mathbf{B}_{\delta_0}(\overline{\zeta}, \overline{\Xi})$ such that for any $(\zeta, \Xi) \in \mathbf{B}_{\delta_0}(\overline{\zeta}, \overline{\Xi})$, $x_c(\zeta, \Xi)$ is the unique minimizer of $L_c(\cdot, \zeta, \Xi)$ over $\mathbf{B}_\varepsilon(\overline{x})$ and $x_c(\cdot,\cdot)$ is semismooth at (ζ, Ξ). Let $\vartheta_c : \Re^m \times \mathcal{S}^p \mapsto \Re$ be defined as (22), i.e.,

$$\vartheta_c(\zeta, \Xi) := \min_{x \in \mathbf{B}_\varepsilon(\overline{x})} L_c(x, \zeta, \Xi), \quad (\zeta, \Xi) \in \Re^m \times \Re^p .$$

Then it holds that

$$\vartheta_c(\zeta, \Xi) = L_c(x_c(\zeta, \Xi), \zeta, \Xi) , \quad (\zeta, \Xi) \in \mathbf{B}_{\delta_0}(\overline{\zeta}, \overline{\Xi}) .$$

Furthermore, it follows from Propositions 2 and 4 that the concave function $\vartheta_c(\cdot,\cdot)$ is continuously differentiable on $\mathbf{B}_{\delta_0}(\overline{\zeta},\overline{\Xi})$ with

$$\nabla\vartheta_c(\zeta,\Xi) = \begin{pmatrix} h(x_c(\zeta,\Xi)) \\ c^{-1}\Big(-\Xi + \Pi_{\mathcal{S}^p_+}(\xi - cg(x_c(\zeta,\Xi)))\Big) \end{pmatrix}, \quad (\zeta,\Xi)\in\mathbf{B}_{\delta_0}(\overline{\zeta},\overline{\Xi}).$$

For any $(\Delta\zeta,\Delta\Xi)\in\Re^m\times\mathcal{S}^p$, let $\overline{\mathcal{V}}_c(\Delta\zeta,\Delta\Xi)$ be defined as in (26). By Propositions 3 and 4, we have for any $(\Delta\zeta,\Delta\Xi)\in\Re^m\times\mathcal{S}^p$ that

$$\partial_B(\nabla\vartheta_c)(\overline{\zeta},\overline{\Xi})(\Delta\zeta,\Delta\Xi)\subseteq\overline{\mathcal{V}}_c(\Delta\zeta,\Delta\Xi).$$

Since

$$\lim_{c\to\infty} c(\Theta_c)_{ij} = \lim_{c\to\infty} c\frac{\lambda_i}{\lambda_i + c|\lambda_j|} = \frac{\lambda_i}{|\lambda_j|} \quad \forall\,(i,j)\in\alpha\times\gamma,$$

we know that there exists a positive number $\overline{\eta}$ such that

$$\big\langle d,\mathcal{B}_{c_0,c}(\overline{\zeta},\overline{\Xi},W)d\big\rangle \le \overline{\eta}\langle d,d\rangle \quad \forall\,d\in\Re^n,\ c\ge c_0,\ \text{and}\ W\in\partial_B\Pi_{\mathcal{S}^p_+}(\overline{\Xi}-cg(\overline{x})). \tag{58}$$

Let $c\ge c_0$ and $W\in\partial_B\Pi_{\mathcal{S}^p_+}(\overline{\Xi}-cg(\overline{x}))$. Then there exist two matrices $Q\in\mathcal{Q}$ with $P=[\overline{P}_\alpha\ (\overline{P}_\beta Q)\ \overline{P}_\gamma]$ and $\Theta_c\in\mathcal{S}^p$ satisfying (48) such that (47) holds. Let $A(P)$ have the singular value decomposition as in (45), i.e.,

$$A(P) = U[\Sigma(P)\quad 0]R^T. \tag{59}$$

Let $\overline{y}:=(\overline{\zeta},\overline{\Xi})$. Then we have the following result for $\mathcal{A}_c(\overline{y},W)$.

Lemma 10 *Let $c>c_0$ and $W\in\partial_B\Pi_{\mathcal{S}^p_+}(\overline{\Xi}-cg(\overline{x}))$. Suppose that Assumptions (nlsdp-A1) and (nlsdp-A2) are satisfied. Then we have*

$$\mathcal{A}_c(\overline{y},W)^{-1}\preceq R\begin{bmatrix}\Sigma^{-1}U^T\Big(\underline{\sigma}\underline{\eta}I_{n_2}+(c-c_0)D_c\Big)^{-1}U\Sigma^{-1} & 0\\ 0 & \underline{\sigma}^{-1}\underline{\eta}^{-1}I_{n_3}\end{bmatrix}R^T, \tag{60}$$

$$\mathcal{A}_c(\overline{y},W)^{-1}\succeq R\begin{bmatrix}\Sigma^{-1}U^T\Big(\overline{\sigma}\overline{\eta}I_{n_2}+(c-c_0)D_c\Big)^{-1}U\Sigma^{-1} & 0\\ 0 & \overline{\sigma}^{-1}\overline{\eta}^{-1}I_{n_3}\end{bmatrix}R^T, \tag{61}$$

and

$$\|\mathcal{A}_c(\overline{y},W)^{-1}A^TD_cu\|\le\sqrt{2}\left(\overline{\sigma}+(\underline{\sigma}\underline{\eta})^{-2}(\overline{\sigma}\overline{\eta})^2\right)\|u\|/(c-c_0)\ \ \forall\,u\in\Re^{n_2}, \tag{62}$$

where $\Sigma:=\Sigma(P)$.

Proof. Let $\hat{c} := c - c_0$. By (51), (52), and the singular value decomposition (59) of $A := A(P)$, we have

$$
\begin{aligned}
&\mathcal{A}_c(\overline{y}, W)^{-1} = \Big(\mathcal{B}_{c_0,c}(\overline{y}, W) + \hat{c} A^T D_c A\Big)^{-1} \\
&= \big(\mathcal{B}_{c_0,c}(\overline{y}, W) + \hat{c} R[\Sigma \quad 0]^T U^T D_c U [\Sigma \quad 0] R^T\big)^{-1} \\
&= R \left(R^T \mathcal{B}_{c_0,c}(\overline{y}, W) R + \hat{c} \begin{bmatrix} \Sigma & 0 \\ 0 & I_{n_3} \end{bmatrix} \begin{bmatrix} U^T D_c U & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & I_{n_3} \end{bmatrix} \right)^{-1} R^T \\
&= R \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & I_{n_3} \end{bmatrix} \left(\mathcal{G}_{c_0}(\overline{y}, W) + \hat{c} \begin{bmatrix} U^T D_c U & 0 \\ 0 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & I_{n_3} \end{bmatrix} R^T, \qquad (63)
\end{aligned}
$$

where

$$
\mathcal{G}_{c_0}(\overline{y}, W) := \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & I_{n_3} \end{bmatrix} R^T \mathcal{B}_{c_0,c}(\overline{y}, W) R \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & I_{n_3} \end{bmatrix}.
$$

It follows from Proposition 4, the definitions of $\underline{\sigma}$ and $\overline{\sigma}$, and (58) that

$$
\mathcal{G}_{c_0}(\overline{y}, W) \succeq \underline{\eta} \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & I_{n_3} \end{bmatrix}^2 \succeq \underline{\sigma}\underline{\eta} I_n \qquad (64)
$$

and

$$
\mathcal{G}_{c_0}(\overline{y}, W) \preceq \overline{\eta} \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & I_{n_3} \end{bmatrix}^2 \preceq \overline{\sigma}\overline{\eta} I_n \,. \qquad (65)
$$

Therefore, (60) and (61) follow from (63).

Now we turn to the proof of (62). Let

$$
\overline{\mathcal{G}}_{c_0}(\overline{y}, W) := \begin{bmatrix} U & 0 \\ 0 & I_{n_3} \end{bmatrix} \mathcal{G}_{c_0}(\overline{y}, W) \begin{bmatrix} U^T & 0 \\ 0 & I_{n_3} \end{bmatrix} \text{ and } \overline{\mathcal{H}}_{c_0}(\overline{y}, W) := \overline{\mathcal{G}}_{c_0}(\overline{y}, W)^{-1}.
$$

Partition $\overline{\mathcal{H}}_{c_0}(\overline{y}, W)$ as

$$
\overline{\mathcal{H}}_{c_0}(\overline{y}, W) = \begin{bmatrix} H_1(W) & H_2(W)^T \\ H_2(W) & H_3(W) \end{bmatrix}
$$

with $H_1(W) \in \mathcal{S}^{n_2}$, $H_2(W) \in \Re^{n_3 \times n_2}$, and $H_3(W) \in \mathcal{S}^{n_3}$. Then, it follows from (64) and (65) that

$$
\|H_1(W)\|_2 \le (\underline{\sigma}\underline{\eta})^{-1}, \ \ \|H_1(W)^{-1}\|_2 \le \overline{\sigma}\overline{\eta}, \ \text{ and } \ \|H_2(W) H_1(W)^{-1}\|_2 \le (\underline{\sigma}\underline{\eta})^{-1} \overline{\sigma}\overline{\eta}\,. \qquad (66)
$$

For any $\varepsilon > 0$, let

$$
D_{c,\varepsilon} := D_c + \varepsilon I_{n_2}, \quad \mathcal{A}_{c,\varepsilon}(\overline{y}, W) := \mathcal{B}_{c_0,c}(\overline{y}, W) + \hat{c} A^T D_{c,\varepsilon} A.
$$

Let $\varepsilon > 0$. By referring to (63), we obtain

$$
\mathcal{A}_{c,\varepsilon}(\overline{y}, W)^{-1} = R \begin{bmatrix} \Sigma^{-1} U^T & 0 \\ 0 & I_{n_3} \end{bmatrix} \left(\overline{\mathcal{G}}_{c_0}(\overline{y}, W) + \hat{c} \begin{bmatrix} D_{c,\varepsilon} & 0 \\ 0 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} U \Sigma^{-1} & 0 \\ 0 & I_{n_3} \end{bmatrix} R^T,
$$

which, together with (59) and the Sherman-Morrison-Woodbury formula (cf. [18, Section 2.1]), implies

$$
\begin{aligned}
&\mathcal{A}_{c,\varepsilon}(\overline{y}, W)^{-1} A^T D_{c,\varepsilon} \\
&= R \begin{bmatrix} \Sigma^{-1} U^T & 0 \\ 0 & I_{n_3} \end{bmatrix} \begin{bmatrix} \left(H_1(W)^{-1} + \hat{c} D_{c,\varepsilon}\right)^{-1} D_{c,\varepsilon} \\ H_2(W) H_1(W)^{-1} \left(H_1(W)^{-1} + \hat{c} D_{c,\varepsilon}\right)^{-1} D_{c,\varepsilon} \end{bmatrix}.
\end{aligned}
$$

Since, it follows from the Sherman-Morrison-Woodbury formula that

$$
\begin{aligned}
&\left(H_1(W)^{-1} + \hat{c} D_{c,\varepsilon}\right)^{-1} D_{c,\varepsilon} \\
&= \left(\hat{c} I_{n_2} + D_{c,\varepsilon}^{-1} H_1(W)^{-1}\right)^{-1} \\
&= \hat{c}^{-1} I_{n_2} - \hat{c}^{-2} D_{c,\varepsilon}^{-1} \left(I_{n_2} + \hat{c}^{-1} H_1(W)^{-1} D_{c,\varepsilon}^{-1}\right)^{-1} H_1(W)^{-1} \\
&= \hat{c}^{-1} I_{n_2} - \hat{c}^{-1} \left(\hat{c} D_{c,\varepsilon} + H_1(W)^{-1}\right)^{-1} H_1(W)^{-1},
\end{aligned}
$$

we have

$$
\begin{aligned}
&\mathcal{A}_c(\overline{y}, W)^{-1} A^T D_c = \lim_{\varepsilon \downarrow 0} \mathcal{A}_{c,\epsilon}(\overline{y}, W)^{-1} A^T D_{c,\epsilon} \\
&= R \begin{bmatrix} \Sigma^{-1} U^T \\ H_2(W) H_1(W)^{-1} \end{bmatrix} \left(\hat{c}^{-1} I_{n_2} - \hat{c}^{-1} \left(\hat{c} D_c + H_1(W)^{-1}\right)^{-1} H_1(W)^{-1}\right).
\end{aligned}
$$

Therefore, from the definition of $\overline{\sigma}$ and (66) we have for any $u \in \Re^{n_2}$ that

$$
\begin{aligned}
&\|\mathcal{A}_c(\overline{y}, W)^{-1} A^T D_c u\|^2 \\
&\leq \left(\overline{\sigma} + (\underline{\sigma\eta})^{-2} (\overline{\sigma\eta})^2\right) \left\| \left(\hat{c}^{-1} I_{n_2} - \hat{c}^{-1} \left(\hat{c} D_c + H_1(W)^{-1}\right)^{-1} H_1(W)^{-1}\right) u \right\|^2 \\
&\leq \left(\overline{\sigma} + (\underline{\sigma\eta})^{-2} (\overline{\sigma\eta})^2\right) \left(\hat{c}^{-1} \|u\| + \hat{c}^{-1} \left\| \left(\hat{c} D_c + H_1(W)^{-1}\right)^{-1} \right\|_2 \|H_1(W)^{-1}\|_2 \|u\| \right)^2 \\
&\leq \left(\overline{\sigma} + (\underline{\sigma\eta})^{-2} (\overline{\sigma\eta})^2\right) \hat{c}^{-2} \left(1 + \|H_1(W)\|_2 \|H_1(W)^{-1}\|_2\right)^2 \|u\|^2 \\
&\leq \left(\overline{\sigma} + (\underline{\sigma\eta})^{-2} (\overline{\sigma\eta})^2\right) \hat{c}^{-2} \left(1 + (\underline{\sigma\eta})^{-1} (\overline{\sigma\eta})\right)^2 \|u\|^2,
\end{aligned}
$$

which, together with the fact that $\overline{\sigma} \geq 1$, proves (62). □

Let

$$
\overline{c} := \max\left\{(2 + \sqrt{2}) c_0,\ (\overline{\sigma\eta} - c_0)^2 / c_0,\ (\underline{\sigma\eta}/2 - c_0)^2 / c_0\right\} \tag{67}
$$

and

$$
\varrho_0 := \left(\max\left\{4 \overline{\nu\sigma} \underline{\sigma}^{-2} \underline{\eta}^{-2} \overline{\nu}_0^2,\ 4\kappa_0^2\right\}\right)^{1/2}. \tag{68}
$$

where

$$
\kappa_0 := \sqrt{2} \left(\overline{\sigma} + (\underline{\sigma\eta})^{-2} (\overline{\sigma\eta})^2\right).
$$

Proposition 5 *Suppose that Assumptions (nlsdp-A1) and (nlsdp-A2) are satisfied. Then there exists a positive number μ_0 such that for any $c \geq \overline{c}$ and $\Delta y \in \Re^m \times \mathcal{S}^p$,*

$$
\|(x_c)'(\overline{y}; \Delta y)\| \leq \varrho_0 \|\Delta y\| / c \tag{69}
$$

and

$$
\left\langle V(\Delta y) + c^{-1} \Delta y, \Delta y \right\rangle \in \mu_0 [-1, 1] \|\Delta y\|^2 / c^2 \quad \forall V(\Delta y) \in \overline{\mathcal{V}}_c(\Delta y). \tag{70}
$$

Proof. Let $c \geq \bar{c}$. Let $\Delta y := (\Delta\zeta, \Delta\Xi) \in \Re^m \times \mathcal{S}^p$. From the proof of Proposition 3 we know that there exists an element $W \in \partial_B \Pi_{\mathcal{S}^p_+}(\overline{\Xi} - cg(\overline{x}))$ such that

$$(x_c)'(\overline{y}; \Delta y) = \mathcal{A}_c(\overline{y}, W)^{-1} \left(-\nabla h(\overline{x})(\Delta\zeta) + \nabla g(\overline{x}) W(\Delta\Xi)\right) . \tag{71}$$

For this $W \in \partial_B \Pi_{\mathcal{S}^p_+}(\overline{\Xi} - cg(\overline{x}))$, there exist two matrices $P \in \mathcal{P}$ and $\Theta_c \in \mathcal{S}^p$ satisfying (48) such that

$$W(H) = P\left(\Theta_c \circ (P^T H P)\right) P^T \quad \forall\, H \in \mathcal{S}^p .$$

Let $A := A(P)$ have the singular value decomposition as in (45), i.e.,

$$A = U[\Sigma \quad 0] R^T , \tag{72}$$

where $\Sigma := \Sigma(P)$. For any two index sets $\chi, \chi' \in \{\alpha, \beta, \gamma\}$, let

$$\omega_{(\chi,\chi')} := \mathrm{vec}(P_\chi^T \Delta\Xi P_{\chi'}) , \quad \widehat{\omega}_{(\chi,\chi)} := \mathrm{svec}(P_\chi^T \Delta\Xi P_\chi) .$$

Define

$$\Delta d_0 := \begin{pmatrix} \Delta\zeta \\ \widehat{\omega}_{(\alpha,\alpha)} \\ \widehat{\omega}_{(\beta,\beta)} \\ \omega_{(\alpha,\beta)} \end{pmatrix} , \quad \Delta d := \begin{pmatrix} \Delta d_0 \\ \omega_{(\alpha,\gamma)} \end{pmatrix} .$$

Then, from (71), we have

$$\begin{aligned}
&\langle (x_c)'(\overline{y}; \Delta y), (x_c)'(\overline{y}; \Delta y) \rangle \\
&= \left\langle [A^T D_c \quad 2C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}]\Delta d, \mathcal{A}_c(\overline{y}, W)^{-2} [A^T D_c \quad 2C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}]\Delta d \right\rangle \\
&\leq 2\left\langle A^T D_c \Delta d_0, \mathcal{A}_c(\overline{y}, W)^{-2} A^T D_c \Delta d_0 \right\rangle \\
&\quad + 8\left\langle C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)} \omega_{(\alpha,\gamma)}, \mathcal{A}_c(\overline{y}, W)^{-2} C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)} \omega_{(\alpha,\gamma)} \right\rangle .
\end{aligned} \tag{73}$$

From (62), we have for $c \geq \bar{c} \quad (\geq (2+\sqrt{2})c_0)$ that

$$\begin{aligned}
&\left\langle A^T D_c \Delta d_0, \mathcal{A}_c(\overline{y}, W)^{-2} A^T D_c \Delta d_0 \right\rangle \\
&= \|\mathcal{A}_c(\overline{y}, W)^{-1} A^T D_c \Delta d_0\|^2 \\
&\leq \kappa_0^2 \hat{c}^{-2} \left(\|\Delta d_0\|\right)^2 \\
&\leq \kappa_0^2 \hat{c}^{-2} \left(\|(\Delta\zeta, \omega_{(\alpha,\alpha)}, \omega_{(\beta,\beta)})\|^2 + 2\|\omega_{(\alpha,\beta)}\|^2\right) \\
&\leq \frac{1}{2} \varrho_0^2 c^{-2} \left(\|(\Delta\zeta, \omega_{(\alpha,\alpha)}, \omega_{(\beta,\beta)})\|^2 + 2\|\omega_{(\alpha,\beta)}\|^2\right) .
\end{aligned} \tag{74}$$

Let

$$\underline{\mathcal{E}}_c := (\overline{\sigma}\overline{\eta} I_{n_2} + (c - c_0) D_c)^{-1} , \quad \overline{\mathcal{E}}_c := \left(\underline{\sigma}\underline{\eta} I_{n_2} + (c - c_0) D_c\right)^{-1}$$

and

$$\underline{\mathcal{H}}_c := \begin{bmatrix} \underline{\mathcal{E}}_c & 0 \\ 0 & \overline{\sigma}^{-1}\overline{\eta}^{-1} I_{n_3} \end{bmatrix} , \quad \overline{\mathcal{H}}_c := \begin{bmatrix} \overline{\mathcal{E}}_c & 0 \\ 0 & \underline{\sigma}^{-1}\underline{\eta}^{-1} I_{n_3} \end{bmatrix} . \tag{75}$$

By recalling that

$$\widetilde{C}_{(\alpha,\gamma)} = C_{(\alpha,\gamma)}\widetilde{R} \quad \text{and} \quad \widetilde{R} = R\begin{bmatrix} \Sigma^{-1}U^T & 0 \\ 0 & I_{n_3} \end{bmatrix},$$

we know from Lemma 10, (75), (46), and (42) (by denoting $\bar{M} = \mathcal{A}_c(\overline{y}, W)$) that

$$\begin{aligned}
&\left\langle C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}, \mathcal{A}_c(\overline{y}, W)^{-2}C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\right\rangle \\
&= \left\langle \bar{M}^{-1/2}C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}, \bar{M}^{-1}\bar{M}^{-1/2}C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\right\rangle \\
&\le \left\langle \bar{M}^{-1/2}C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}, \widetilde{R}\overline{\mathcal{H}}_c\widetilde{R}^T\bar{M}^{-1/2}C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\right\rangle \\
&\le \overline{\sigma}\underline{\sigma}^{-1}\underline{\eta}^{-1}\left\langle C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}, \mathcal{A}_c(\overline{y}, W)^{-1}C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\right\rangle \\
&\le \overline{\sigma}\underline{\sigma}^{-2}\underline{\eta}^{-2}\left\langle \widetilde{C}^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}, \widetilde{C}^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\right\rangle \\
&\le \overline{\nu}\overline{\sigma}\underline{\sigma}^{-2}\underline{\eta}^{-2}\|(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\|^2 \\
&\le \overline{\nu}\overline{\sigma}\underline{\sigma}^{-2}\underline{\eta}^{-2}\left(\max_{i\in\alpha, j\in\gamma}\lambda_i/(\lambda_i + c|\lambda_j|)\right)^2\|\omega_{(\alpha,\gamma)}\|^2 \\
&\le \overline{\nu}\overline{\sigma}\underline{\sigma}^{-2}\underline{\eta}^{-2}\overline{\nu}_0^2(\overline{\nu}_0 + c)^{-2}\|\omega_{(\alpha,\gamma)}\|^2 \\
&\le \overline{\nu}\overline{\sigma}\underline{\sigma}^{-2}\underline{\eta}^{-2}\overline{\nu}_0^2c^{-2}\|\omega_{(\alpha,\gamma)}\|^2\,, \\
&\le \frac{1}{8}\varrho_0^2c^{-2}(2\|\omega_{(\alpha,\gamma)}\|^2)\,,
\end{aligned} \tag{76}$$

which, together with (73) and (74), implies

$$\langle (x_c)'(\overline{y}; \Delta y), (x_c)'(\overline{y}; \Delta y)\rangle \le \varrho_0^2\|\Delta y\|^2/c^2.$$

Thus (69) holds.

Let $V(\Delta y) \in \overline{\mathcal{V}}_c(\Delta y)$. Then from the definition of $\overline{\mathcal{V}}_c(\Delta y)$, there exists $W \in \partial_B\Pi_{\mathcal{S}^p_+}(\overline{\Xi} - cg(\overline{x}))$ such that

$$\begin{aligned}
V(\Delta y) = &\begin{pmatrix} \mathcal{J}h(\overline{x}) \\ -W\mathcal{J}g(\overline{x}) \end{pmatrix}\mathcal{A}_c(\overline{y}, W)^{-1}\Big(-\nabla h(\overline{x})\Delta\zeta + \nabla g(\overline{x})W(\Delta\Xi)\Big) \\
&+ \begin{pmatrix} 0 \\ -c^{-1}\Delta\Xi + c^{-1}W(\Delta\Xi) \end{pmatrix}.
\end{aligned}$$

For notational convenience, we assume that this $W \in \partial_B\Pi_{\mathcal{S}^p_+}(\overline{\Xi} - cg(\overline{x}))$ is the same as in (71). After direct calculations (cf. Lemma 9), we obtain

$$\begin{aligned}
&-\langle V(\Delta y), \Delta y\rangle \\
&= \left\langle [A^TD_c \;\; 2C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}]\Delta d, \mathcal{A}_c(\overline{y}, W)^{-1}[A^TD_c \;\; 2C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}]\Delta d\right\rangle \\
&\quad + c^{-1}\|\Delta\Xi\|^2 - c^{-1}\langle\Delta\Xi, W(\Delta\Xi)\rangle \\
&= \left\langle A^TD_c\Delta d_0, \bar{M}^{-1}A^TD_c\Delta d_0\right\rangle + 4\left\langle A^TD_c\Delta d_0, \bar{M}^{-1}C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\right\rangle \\
&\quad + 4\left\langle C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}, \bar{M}^{-1}C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\right\rangle \\
&\quad + c^{-1}\|\Delta\Xi\|^2 - c^{-1}\langle\Delta\Xi, W(\Delta\Xi)\rangle\,.
\end{aligned} \tag{77}$$

Next, we estimate the lower and upper bounds of the right hand side of (77). By using (72) and Lemma 10 we obtain

$$\underline{\mathcal{E}}_c \preceq A\mathcal{A}_c(\overline{y},W)^{-1}A^T \preceq \overline{\mathcal{E}}_c \,.$$

Thus, we have

$$\begin{aligned}
&\left\langle A^T D_c \Delta d_0, \mathcal{A}_c(\overline{y},W)^{-1}A^T D_c\Delta d_0\right\rangle \geq \left\langle D_c\Delta d_0, \underline{\mathcal{E}}_c D_c \Delta d_0\right\rangle \\
&\geq (\overline{\sigma\eta}+(c-c_0))^{-1}\,\|(\Delta\zeta,\omega_{(\alpha,\alpha)})\|^2 + 4\,(\overline{\sigma\eta}+2(c-c_0))^{-1}\,\|\omega_{(\alpha,\beta)}\|^2 \\
&\quad + \left\langle (\widehat{\Theta}_c)_{(\beta,\beta)}\widehat{\omega}_{(\beta,\beta)}, \left(\overline{\sigma\eta} I_{|\beta|(|\beta|+1)/2} + (c-c_0)(\widehat{\Theta}_c)_{(\beta,\beta)}\right)^{-1} (\widehat{\Theta}_c)_{(\beta,\beta)}\widehat{\omega}_{(\beta,\beta)}\right\rangle \\
&\geq (\overline{\sigma\eta}+(c-c_0))^{-1}\,\|(\Delta\zeta,\omega_{(\alpha,\alpha)})\|^2 + 4\,(\overline{\sigma\eta}+2(c-c_0))^{-1}\,\|\omega_{(\alpha,\beta)}\|^2 \\
&\quad + \left\langle (\Theta_c)_{(\beta,\beta)}\omega_{(\beta,\beta)}, \left(\overline{\sigma\eta} I_{|\beta|} + (c-c_0)(\Theta_c)_{(\beta,\beta)}\right)^{-1} (\Theta_c)_{(\beta,\beta)}\omega_{(\beta,\beta)}\right\rangle
\end{aligned} \tag{78}$$

and

$$\begin{aligned}
&\left\langle A^T D_c \Delta d_0, \mathcal{A}_c(\overline{y},W)^{-1}A^T D_c\Delta d_0\right\rangle \leq \left\langle D_c\Delta d_0, \overline{\mathcal{E}}_c D_c \Delta d_0\right\rangle \\
&\leq \left(\underline{\sigma\eta}/2+(c-c_0)\right)^{-1}\,\|(\Delta\zeta,\omega_{(\alpha,\alpha)})\|^2 + 4\left(\underline{\sigma\eta}+2(c-c_0)\right)^{-1}\,\|\omega_{(\alpha,\beta)}\|^2 \\
&\quad + \left\langle (\widehat{\Theta}_c)_{(\beta,\beta)}\widehat{\omega}_{(\beta,\beta)}, \left(\underline{\sigma\eta} I_{|\beta|(|\beta|+1)/2} + (c-c_0)(\widehat{\Theta}_c)_{(\beta,\beta)}\right)^{-1} (\widehat{\Theta}_c)_{(\beta,\beta)}\widehat{\omega}_{(\beta,\beta)}\right\rangle \\
&\leq \left(\underline{\sigma\eta}/2+(c-c_0)\right)^{-1}\,\|(\Delta\zeta,\omega_{(\alpha,\alpha)})\|^2 + 4\left(\underline{\sigma\eta}+2(c-c_0)\right)^{-1}\,\|\omega_{(\alpha,\beta)}\|^2 \\
&\quad + \left\langle (\Theta_c)_{(\beta,\beta)}\omega_{(\beta,\beta)}, \left((\underline{\sigma\eta}/2) I_{|\beta|} + (c-c_0)(\Theta_c)_{(\beta,\beta)}\right)^{-1} (\Theta_c)_{(\beta,\beta)}\omega_{(\beta,\beta)}\right\rangle .
\end{aligned} \tag{79}$$

From Lemma 10, (75), (46), and (42) we know that

$$\begin{aligned}
&\left\langle C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}, \mathcal{A}_c(\overline{y},W)^{-1}C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\right\rangle \\
&\geq \left\langle \widetilde{C}^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}, \underline{\mathcal{H}}_c\widetilde{C}^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\right\rangle \\
&\geq (\overline{\sigma\eta}+2(c-c_0))^{-1}\left\langle \widetilde{C}^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}, \widetilde{C}^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\right\rangle \\
&\geq \underline{\nu}\,(\overline{\sigma\eta}+2(c-c_0))^{-1}\,\|(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\|^2 \\
&\geq \underline{\nu}\,(\overline{\sigma\eta}+2(c-c_0))^{-1}\left(\min_{i\in\alpha,j\in\gamma}\lambda_i/(\lambda_i+c|\lambda_j|)\right)^2\|\omega_{(\alpha,\gamma)}\|^2 \\
&\geq \underline{\nu}\,(\overline{\sigma\eta}+2(c-c_0))^{-1}\,\underline{\nu}_0^2(\underline{\nu}_0+c)^{-2}\|\omega_{(\alpha,\gamma)}\|^2 ,
\end{aligned} \tag{80}$$

$$\begin{aligned}
&\left\langle C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}, \mathcal{A}_c(\overline{y},W)^{-1}C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\right\rangle \\
&\leq \left\langle \widetilde{C}^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}, \overline{\mathcal{H}}_c\widetilde{C}^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\right\rangle \\
&\leq \underline{\sigma}^{-1}\underline{\eta}^{-1}\left\langle \widetilde{C}^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}, \widetilde{C}^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\right\rangle \\
&\leq \overline{\nu}\,\underline{\sigma}^{-1}\underline{\eta}^{-1}\|(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\|^2 \\
&\leq \overline{\nu}\,\underline{\sigma}^{-1}\underline{\eta}^{-1}\left(\max_{i\in\alpha,j\in\gamma}\lambda_i/(\lambda_i+c|\lambda_j|)\right)^2\|\omega_{(\alpha,\gamma)}\|^2 \\
&\leq \overline{\nu}\,\underline{\sigma}^{-1}\underline{\eta}^{-1}\overline{\nu}_0^2(\overline{\nu}_0+c)^{-2}\|\omega_{(\alpha,\gamma)}\|^2 ,
\end{aligned} \tag{81}$$

and

$$\begin{aligned}
&\left\langle C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}, C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\right\rangle \\
&\le \overline{\nu}\|(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\|^2 \\
&\le \overline{\nu}\left(\max_{i\in\alpha,j\in\gamma}\lambda_i/(\lambda_i+c|\lambda_j|)\right)^2\|\omega_{(\alpha,\gamma)}\|^2 \\
&\le \overline{\nu}\overline{\nu}_0^2 c^{-2}\|\omega_{(\alpha,\gamma)}\|^2 .
\end{aligned} \tag{82}$$

By using (74) and (82) we have

$$\begin{aligned}
&\left|\left\langle A^T D_c \Delta d_0, \mathcal{A}_c(\overline{y},W)^{-1}C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\right\rangle\right| \\
&\le \|\mathcal{A}_c(\overline{y},W)^{-1}A^T D_c\Delta d_0\|\,\|C^T_{(\alpha,\gamma)}(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\| \\
&\le \frac{\varrho_0}{\sqrt{2}}c^{-1}\left(\|(\Delta\zeta,\omega_{(\alpha,\alpha)},\omega_{(\beta,\beta)})\|^2+2\|\omega_{(\alpha,\beta)}\|^2\right)^{1/2}\left(\overline{\nu}_0\sqrt{\overline{\nu}}c^{-1}\|\omega_{(\alpha,\gamma)}\|\right) \\
&\le \frac{\varrho_0\overline{\nu}_0\sqrt{\overline{\nu}}}{4}c^{-2}\left(\|(\Delta\zeta,\omega_{(\alpha,\alpha)},\omega_{(\beta,\beta)})\|^2+2\|\omega_{(\alpha,\beta)}\|^2+2\|\omega_{(\alpha,\gamma)}\|^2\right) .
\end{aligned} \tag{83}$$

By direct calculations we have

$$\begin{aligned}
&\|\Delta\Xi\|^2-\langle\Delta\Xi,W(\Delta\Xi)\rangle \\
&=\left(\|\omega_{(\gamma,\gamma)}\|^2+2\|\omega_{(\beta,\gamma)}\|^2\right)+2\left(\|\omega_{(\alpha,\gamma)}\|^2-\langle\omega_{(\alpha,\gamma)},(\Theta_c)_{(\alpha,\gamma)}\omega_{(\alpha,\gamma)}\rangle\right) \\
&\quad+\left(\|\omega_{(\beta,\beta)}\|^2-\langle\omega_{(\beta,\beta)},(\Theta_c)_{(\beta,\beta)}\omega_{(\beta,\beta)}\rangle\right) .
\end{aligned} \tag{84}$$

Now we are ready to estimate the lower and upper bounds of $-\langle V(\Delta y),\Delta y\rangle$. In light of (77), (78), (80), (83), and (84), we have

$$\begin{aligned}
-\langle V(\Delta y),\Delta y\rangle \ge\; & c^{-1}\left(\|\omega_{(\gamma,\gamma)}\|^2+2\|\omega_{(\beta,\gamma)}\|^2\right) \\
&+\kappa_1(c)\|(\Delta\zeta,\omega_{(\alpha,\alpha)})\|^2+2\kappa_2(c)\|\omega_{(\alpha,\beta)}\|^2 \\
&+2\kappa_3(c)\|\omega_{(\alpha,\gamma)}\|^2+\kappa_4(c)\|\omega_{(\beta,\beta)}\|^2,
\end{aligned} \tag{85}$$

where

$$\begin{aligned}
\kappa_1(c) &:= \left(\overline{\sigma\eta}+(c-c_0)\right)^{-1}-\varrho_0\overline{\nu}_0\sqrt{\overline{\nu}}c^{-2} , \\
\kappa_2(c) &:= \left(\overline{\sigma\eta}/2+(c-c_0)\right)^{-1}-\varrho_0\overline{\nu}_0\sqrt{\overline{\nu}}c^{-2} \\
\kappa_3(c) &:= c^{-1}[1-\overline{\nu}_0(\overline{\nu}_0+c)^{-1}]+2\underline{\nu}\left(\overline{\sigma\eta}+2(c-c_0)\right)^{-1}\underline{\nu}_0^2(\underline{\nu}_0+c)^{-2}-\varrho_0\overline{\nu}_0\sqrt{\overline{\nu}}c^{-2} ,
\end{aligned}$$

and

$$\kappa_4(c) := \min_{t\in[0,1]}\psi(t;c,a_c,b_c,c_0)$$

with $\psi(\cdot;\cdot)$ being defined as (12) in Lemma 8 and

$$a_c := c^{-1}-\varrho_0\overline{\nu}_0\sqrt{\overline{\nu}}c^{-2} , \quad b_c := \overline{\sigma\eta} .$$

It follows from (13) in Lemma 8 that for $c\ge\overline{c}$,

$$\kappa_4(c)=c^{-1}-\rho_0\overline{\nu}_0\sqrt{\overline{\nu}}c^{-2}-\frac{\overline{\sigma\eta}}{c(\sqrt{c}+\sqrt{c_0})^2} .$$

Thus, there exists a positive number μ_1 such that for $c \geq \overline{c}$ we have

$$\min\{\kappa_1(c), \kappa_2(c), \kappa_3(c), \kappa_4(c)\} \geq c^{-1} - \mu_1 c^{-2}\,.$$

Therefore, from (85) we have

$$-\langle V(\varDelta y), \varDelta y\rangle \geq \min\{c^{-1}, \kappa_1(c), \kappa_2(c), \kappa_3(c), \kappa_4(c)\}\|\varDelta y\|^2 \geq (c^{-1}-\mu_1 c^{-2})\|\varDelta y\|^2\,. \tag{86}$$

On the other hand, in light of (77), (79), (81), (83), and (84), we have

$$\begin{aligned} -\langle V(\varDelta y), \varDelta y\rangle \leq c^{-1}\left(\|\omega_{(\gamma,\gamma)}\|^2 + 2\|\omega_{(\beta,\gamma)}\|^2\right) & \\ +\mu_1(c)\|(\varDelta\zeta, \omega_{(\alpha,\alpha)})\|^2 + 2\mu_2(c)\|\omega_{(\alpha,\beta)}\|^2 & \\ +2\mu_3(c)\|\omega_{(\alpha,\gamma)}\|^2 + \mu_4(c)\|\omega_{(\beta,\beta)}\|^2, & \end{aligned} \tag{87}$$

where

$$\begin{aligned} \mu_1(c) :=\ & \left(\underline{\sigma}\underline{\eta}/2 + (c-c_0)\right)^{-1} + \varrho_0\overline{\nu}_0\sqrt{\overline{\nu}}c^{-2}\,, \\ \mu_2(c) :=\ & \mu_1(c)\,, \\ \mu_3(c) :=\ & c^{-1}[1 - \underline{\nu}_0(\underline{\nu}_0 + c)^{-1}] + 2\overline{\nu}\underline{\sigma}^{-1}\underline{\eta}^{-1}\overline{\nu}_0^2\left(\overline{\nu}_0 + (c-c_0)\right)^{-2} + \varrho_0\overline{\nu}_0\sqrt{\overline{\nu}}c^{-2}\,, \end{aligned}$$

and

$$\mu_4(c) := \max_{t\in[0,1]} \psi(t; c, a_c', b_c', c_0)$$

with

$$a_c' := c^{-1} + \varrho_0\overline{\nu}_0\sqrt{\overline{\nu}}c^{-2}\,, \quad b_c' := \underline{\sigma}\underline{\eta}/2\,.$$

It follows from (14) in Lemma 8 that for $c \geq \overline{c}$,

$$\begin{aligned} \mu_4(c) &= \max\{\psi(0; c, a_c', b_c', c_0), \psi(1; c, a_c', b_c', c_0)\} \\ &= \varrho_0\overline{\nu}_0\sqrt{\overline{\nu}}c^{-2} + \max\{c^{-1}, \left(\underline{\sigma}\underline{\eta}/2 + (c-c_0)\right)^{-1}\}\,. \end{aligned} \tag{88}$$

Then there exists a positive number $\mu_0 \geq \mu_1$ such that for $c \geq \overline{c}$ we have

$$\max\{\mu_1(c), \mu_2(c), \mu_3(c), \mu_4(c)\} \leq c^{-1} + \mu_0 c^{-2}\,.$$

Therefore, from (87) we have

$$-\langle V(\varDelta y), \varDelta y\rangle \leq \max\{c^{-1}, \mu_1(c), \mu_2(c), \mu_3(c), \mu_4(c)\}\|\varDelta y\|^2 \leq (c^{-1}+\mu_0 c^{-2})\|\varDelta y\|^2\,. \tag{89}$$

By (86) and (89), noting that $\mu_0 \geq \mu_1$, we obtain that

$$\mu_0 c^{-2}\|\varDelta y\|^2 \geq -\langle V(\varDelta y) + c^{-1}\varDelta y, \varDelta y\rangle \geq -\mu_0 c^{-2}\|\varDelta y\|^2\,.$$

This shows that (70) holds. The proof is completed. $\square$

Now we are ready to state our main result on the rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming.

Theorem 2 *Suppose that Assumptions (nlsdp-A1) and (nlsdp-A2) are satisfied. Let c_0 and $\underline{\eta}$ be two positive numbers obtained by* Proposition 4. *Let $\overline{\eta}$, $\overline{c}$, and ϱ_0 be defined as in* (58), (67), *and* (68), *respectively. Let μ_0 be obtained by* Proposition 5. *Define*

$$\varrho_1 := 2\varrho_0 \quad \text{and} \quad \varrho_2 := 4\mu_0.$$

Then for any $c \geq \overline{c}$, there exist two positive numbers ε and δ (both depending on c) such that for any $(\zeta, \varXi) \in \mathbf{B}_\delta(\overline{\zeta}, \overline{\varXi})$, the problem

$$\min \; L_c(x, \zeta, \varXi) \quad \text{s.t. } x \in \mathbf{B}_\varepsilon(\overline{x})$$

has a unique solution denoted $x_c(\zeta, \varXi)$. The function $x_c(\cdot,\cdot)$ is locally Lipschitz continuous on $\mathbf{B}_\delta(\overline{\zeta}, \overline{\varXi})$ and is semismooth at any point in $\mathbf{B}_\delta(\overline{\zeta}, \overline{\varXi})$, and for any $(\zeta, \varXi) \in \mathbf{B}_\delta(\overline{\zeta}, \overline{\varXi})$, we have

$$\|x_c(\zeta, \varXi) - \overline{x}\| \leq \varrho_1 \|(\zeta, \varXi) - (\overline{\zeta}, \overline{\varXi})\|/c$$

and

$$\|(\zeta_c(\zeta, \varXi), \varXi_c(\zeta, \varXi)) - (\overline{\zeta}, \overline{\varXi})\| \leq \varrho_2 \|(\zeta, \varXi) - (\overline{\zeta}, \overline{\varXi})\|/c\,,$$

where $\zeta_c(\zeta, \varXi)$ and $\xi_c(\zeta, \varXi))$ are defined as

$$\zeta_c(\zeta, \varXi) := \zeta + ch(x_c(\zeta, \varXi)) \quad \text{and} \quad \varXi_c(\zeta, \varXi) := \varPi_{\mathcal{S}^p_+}(\xi - cg(x_c(\zeta, \varXi)))\,.$$

Proof. If Assumptions (nlsdp-A1) and (nlsdp-A2) are satisfied, then from Propositions 4 and 5 we know that both Assumption B1 and Assumption B2 (with $\tau = 2$) made in Section 3 are satisfied. Then the conclusions in this theorem follow from Theorem 1. □

Before closing this section, we make a final comment. Note that if the strict complementarity condition is satisfied, then the result on the rate of convergence can be deduced in a much more straightforward way. For a comparison, we present the corresponding analysis, i.e. Theorem 3 below, as an appendix. Another purpose of the appendix is to point out that one can adopt the proof used in Theorem 3 to deal with (NLP). By doing so, one can actually give a corrected proof of the approach sketched in Bertsekas [4] for (NLP) (compared with the proof given in the appendix, the missing parts in Bertsekas' approach [4, Section 3.1] can be readily seen). Just as in the case for (NLP) [4], the second order sufficient condition in Theorem 3 automatically implies the strict complementarity condition. For (NLP), when the strict complementarity condition holds, an approach to derive results similar to (94) and (95) was also suggested by Golshtein and Tretyakov [17, Chapter 7].

The conditions imposed in Theorem 3 are equivalent to Assumptions (nlsdp-A1) and (nlsdp-A2) plus the strict complementarity condition, i.e., $\beta = \emptyset$. Compared Theorem 2 with Theorem 3, we can see that there is no loss on the rate of convergence of the augmented Lagrangian method for (NLSDP) even the strict complementarity condition fails to hold. However, different from Theorem 3, the convergence region in Theorem 2 may depend on c. It would be interesting to know if this dependence can be removed under Assumptions (nlsdp-A1) and (nlsdp-A2) only.

5 Conclusions

This paper provides an analysis on the rate of convergence of the augmented Lagrangian method for solving nonlinear semidefinite programming. By assuming that K is a self-dual cone and that $\Pi_K(\cdot)$ is semismooth everywhere, we first establish a general result on the rate of convergence of the augmented Lagrangian method for a class of general optimization problems. Then we apply this general result to nonlinear semidefinite programming under the constraint nondegeneracy condition and the strong second order sufficient condition. This procedure suggests that our result may be used to deal with other optimization problems. For example, it seems possible to apply our general result to nonlinear second order cone programming by using the strong second order sufficient condition recently proposed in [7] for second order cone programming.

Acknowledgments. The authors are grateful to the Associate Editor and the referees for their helpful comments and suggestions on improving the quality of this paper. In particular, the authors are in debt to a referee for pointing out a technical error in the proof of Proposition 5 in an earlier version of this paper. Thanks also go to Zheng-Hai Huang, Weizhe Gu, and Nan Lu at Tianjin University and Xinyuan Zhao at National University of Singapore for their comments on this paper.

References

1. Arnold, V.I.: On matrices depending on parameters. Russian Mathematical Surveys **26**, 29–43 (1971)
2. Arrow, K.J., Solow, R.M.: Gradient methods for constrained maxima with weakened assumptions. In: K.J. Arrow, L. Hurwicz, H. Uzawa (eds.), Studies in Linear and Nonlinear Programming, Stanford University Press, Stanford, pp. 165–176 (1958)
3. Bertsekas, D.P.: On penalty and multiplier methods for constrained minimization. SIAM Journal Control and Optimization **14**, 216-235 (1976)
4. Bertsekas, D.P.: Constrained Optimization and Lagrange Multiplier Methods. Academic Press, New York, 1982
5. Bonnans, J.F., Cominetti, R., Shapiro, A.: Sensitivity analysis of optimization problems under second order regularity constraints. Mathematics of Operations Research **23**, 803–832 (1998)
6. Bonnans, J.F., Cominetti, R., Shapiro, A.: Second order optimality conditions based on parabolic second order tangent sets. SIAM Journal on Optimization **9**, 466-493 (1999)
7. Bonnans, J.F., Ramírez C., H.: Perturbation analysis of second order cone programming problems. Mathematical Programming, Series B **104**, 205–227 (2005)
8. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer-Verlag, New York, 2000
9. Chen, X., Qi, H.-D., Tseng, P.: Analysis of nonsmooth symmetric-matrix functions with applications to semidefinite complementarity problems. SIAM Journal on Optimization **13**, 960-985 (2003)
10. Clarke, F. H.: Optimization and Nonsmooth Analysis. John Wiley and Sons, New York, 1983

11. Conn, A.R., Gould, N.I.M., Toint, Ph. L.: A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. SIAM Journal on Numerical Analysis **28**, 545–572 (1991)
12. Contesse-Becker, L.: Extended convergence results for the method of multipliers for non-strictly binding inequality constraints. Journal of Optimization Theory and Applications **79**, 273–310 (1993)
13. Debreu, G.: Definite and semidefinite quadratic forms. Econometrica **20**, 295-300 (1952)
14. Diehl, M., Jarre, F., Vogelbusch C.: Loss of superlinear convergence for an SQP-type method with conic constraints. SIAM Journal on Optimization **16**, 1201–1210 (2006)
15. Eaves, B.C.: On the basic theorem of complementarity. Mathematical Programming **1**, 68–75 (1971)
16. Faraut, J. and Korányi, A.: Analysis on Symmetric Cones. Clarendon Press, London, 1994
17. Golshtein, E.G. and Tretyakov, N.V.: Modified Lagrangians and Monotone Maps in Optimization. John Wiley and Sons, New York, 1989
18. Golub, G. and Van Loan, C.F.: Matrix Computations. The Johns Hopkins University Press, Baltimore, 1996
19. Hestenes, M.R.: Multiplier and gradient methods. Journal of Optimization Theory and Applications **4**, 303-320 (1969)
20. Higham, N.J.: Computing a nearest symmetric positive semidefinite matrix. Linear Algebra and Applications **103**, 103–118 (1998)
21. Ito, K., Kunisch, K.: The augmented Lagrangian method for equality and inequality constraints in Hilbert spaces. Mathematical Programming **46**, 341–360 (1990)
22. Kummer, B., Newton's method for non-differentiable functions. In: J. Guddat et al.(eds.), Advances in Mathematical Optimization, Akademie-Verlag, Berlin, 1988, pp. 114–125
23. Kummer, B.: Lipschitzian inverse functions, directional derivatives, and applications in $C^{1,1}$-optimization. Journal of Optimization Theory and Applications **70**, 559-580 (1991)
24. Leibfritz, F.: COMP$l_e$ib 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
25. Löwner, K.: Über monotone matrixfunktionen. Mathematische Zeitschrift **38**, 177–216 (1934)
26. Kocvara, M., Stingl, M.: Solving nonconvex SDP problems of structural optimization with stability control. Optimization Methods and Software **19**, 595-609 (2004)
27. Meng, F., Sun, D., Zhao, G.: Semismoothness of solutions to generalized equations and the Moreau-Yosida regularization. Mathematical Programming, Series B **104**, 561–581 (2005)
28. Mifflin, R.: Semismooth and semiconvex functions in constrained optimization. SIAM Journal on Control and Optimization **15**, 957-972 (1977)
29. Pang, J.-S., Sun, D., Sun, J.: Semismooth homeomorphisms and strong stability of semidefinite and Lorentz complementarity problems. Mathematics of Operations Research **28**, 39–63 (2003)
30. Pennanen, T.: Local convergence of the proximal point algorithm and multiplier methods without monotonicity. Mathematics of Operations Research **27**, 170–191 (2002)
31. Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In R. Fletcher, Editor, Optimization, Academic Press, New York, 1972, pp. 283–298.
32. Qi, L., Sun, J.: A nonsmooth version of Newton's method. Mathematical Programming **58**, 353–367 (1993)
33. Robinson, S.M.: Local structure of feasible sets in nonlinear programming, Part II: Nondegeneracy. Mathematical Programming Study **22**, 217–230 (1984)

34. Rockafellar, R.T.: A dual approach to solving nonlinear programming problems by unconstrained optimization. Mathematical Programming **5**, 354-373 (1973)
35. Rockafellar, R.T.: The multiplier method of Hestenes and Powell applied to convex programming. Journal of Optimization Theory and Applications **12**, 555-562 (1973)
36. Rockafellar, R.T.: Augmented Lagrange multiplier functions and duality in nonconvex programming. SIAM Journal on Control **12**, 268-285 (1974)
37. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM Journal on Control and Optimization **14**, 877-898 (1976)
38. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Mathematics of Operations Research **1**, 97-116 (1976)
39. Rockafellar, R.T.: Lagrange multipliers and optimality. SIAM Review **35**, 183-238 (1993)
40. Rockafellar, R.T., Wets, R. J.-B.: Variational Analysis. Springer-Verlag, New York, 1998
41. Shapiro, A.: On concepts of directional differentiability. Journal of Optimization Theory and Applications **66**, 477–487 (1990)
42. Shapiro, A.: Sensitivity analysis of generalized equations. Journal of Mathematical Sciences **115**, 2554–2565 (2003)
43. Shapiro, A., Sun, J.: Some properties of the augmented Lagrangian in cone constrained optimization. Mathematics of Operations Research **29**, 479-491 (2004)
44. Sun, D.: A further result on an implicit function theorem for locally Lipschitz functions. Operations Research Letters **28**, 193-198 (2001)
45. Sun, D.: The strong second order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. Mathematics of Operations Research **31**, 761-776 (2006)
46. Sun, D., Sun, J.: Semismooth matrix valued functions. Mathematics of Operations Research **27**, 150-169 (2002)
47. Sun, D., Sun, J.: Löwner's operator and spectral functions in Euclidean Jordan algebras. Department of Mathematics, National University of Singapore, December 2004
48. Tretyakov, N.V.: A method of penalty estimates for convex programming problems. Èkonomika i Matematicheskie Metody **9**, 525–540 (1973)
49. Tseng, P.: Merit functions for semi-definite complementarity problems. Mathematical Programming **83**, 159–185 (1998)
50. Tsing, N.-K., Fan, M.K.H., Verriest, E.I.: On analyticity of functions involving eigenvalues. Linear Algebra and Its Applications **207**, 159–180 (1994)
51. Zarantonello, E.H.: Projections on convex sets in Hilbert space and spectral theory I and II. In E.H. Zarantonello, editor, Contributions to Nonlinear Functional Analysis, Academic Press, 1971, pp. 237–424.

6 Appendix: An analysis under strict complementarity

In this appendix, we shall provide a direct analysis, with strict complementarity, on the rate of convergence of the augmented Lagrangian method for solving nonlinear semidefinite programming

$$\text{(NLSDP)} \qquad \min \ f(x) \quad \text{s.t. } h(x) = 0\,, \ g(x) \in \mathcal{S}_+^p\,,$$

where $f : \Re^n \mapsto \Re$, $h : \Re^n \mapsto \Re^m$, and $g : \Re^n \mapsto \mathcal{S}^p$ are twice continuously differentiable. For any two matrices C and D in $\Re^{m\times n}$, we write

$$\langle C, D\rangle := \mathrm{Tr}\left(C^T D\right)$$

for the Frobenius inner product between C and D, where "Tr" denotes the trace of a square matrix. Since in this case $\mathcal{S}^p_+ = \{z^2 \mid z \in \mathcal{S}^p\}$, (NLSDP) can be transformed into the following equality constrained optimization problem:

$$\min \ f(x) \quad \text{s.t.}\ h(x) = 0\,,\ z^2 - g(x) = 0\,,\ (x, z) \in X \times \mathcal{S}^p\,. \tag{90}$$

Let $c > 0$. Let $\widehat{L}_c(x, z, \zeta, \xi)$ be the augmented Lagrangian function for (90). Define $\widetilde{L}_c : X \times \mathcal{S}^p \times \Re^m \times \mathcal{S}^p \mapsto \Re$ by

$$\widetilde{L}_c(x, v, \zeta, \xi) := f(x) + \langle \zeta, h(x)\rangle + \frac{c}{2}\|h(x)\|^2 + \langle \xi, v - g(x)\rangle + \frac{c}{2}\|v - g(x)\|^2\,.$$

Then for any $(x, z, \zeta, \xi) \in X \times \mathcal{S}^p \times \Re^m \times \mathcal{S}^p$, we have

$$\begin{aligned}\widehat{L}_c(x, z, \zeta, \xi) &= \widetilde{L}_c(x, z^2, \zeta, \xi)\\ &= f(x) + \langle \zeta, h(x)\rangle + \frac{c}{2}\|h(x)\|^2 + \langle \xi, z^2 - g(x)\rangle + \frac{c}{2}\|z^2 - g(x)\|^2\,.\end{aligned}$$

For any $(x, \zeta, \xi) \in X \times \Re^m \times \mathcal{S}^p$, let

$$\widetilde{v}(x, \xi, c) := \Pi_{\mathcal{S}^p_+}(g(x) - c^{-1}\xi) = \big(g(x) - c^{-1}\xi\big) + c^{-1}\Pi_{\mathcal{S}^p_+}(\xi - cg(x))\,.$$

Since

$$\begin{aligned}&\inf_{v \in \mathcal{S}^p_+} \widetilde{L}_c(x, v, \zeta, \xi)\\ &= f(x) + \langle \zeta, h(x)\rangle + \frac{c}{2}\|h(x)\|^2 + \inf_{v \in \mathcal{S}^p_+} \langle \xi, v - g(x)\rangle + \frac{c}{2}\|v - g(x)\|^2\\ &= f(x) + \langle \zeta, h(x)\rangle + \frac{c}{2}\|h(x)\|^2 + \langle \xi, \widetilde{v}(x, \xi, c) - g(x)\rangle + \frac{c}{2}\|\widetilde{v}(x, \zeta, \xi) - g(x)\|^2\\ &= \widetilde{L}_c(x, \widetilde{v}(x, \xi, c), \zeta, \xi)\\ &= L_c(x, \zeta, \xi),\end{aligned}$$

we have for any $(x, \zeta, \xi) \in X \times \Re^m \times \mathcal{S}^p$

$$L_c(x, \zeta, \xi) = \widehat{L}_c(x, z_c(x, \zeta, \xi), \zeta, \xi) = \widetilde{L}(x, z_c^2(x, \zeta, \xi), \zeta, \xi)\,, \tag{91}$$

where $z_c(x, \zeta, \xi)$ is the square root of $\Pi_{\mathcal{S}^p_+}\big(g(x) - c^{-1}\xi\big)$, i.e.,

$$z_c(x, \zeta, \xi) := \sqrt{\Pi_{\mathcal{S}^p_+}(g(x) - c^{-1}\xi)}\,. \tag{92}$$

Define $\widehat{h} : X \times \mathcal{S}^p \mapsto \Re^m \times \mathcal{S}^p$ by

$$\widehat{h}(x, z) := \begin{pmatrix} h(x) \\ z^2 - g(x) \end{pmatrix}, \quad (x, z) \in X \times \mathcal{S}^p\,.$$

Then we have the following conclusion on the rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming.

Theorem 3 *Consider (NLSDP) and its equivalent problem* (90)*. Let* $(\overline{x}, \overline{z}, \overline{\zeta}, \overline{\xi}) \in X \times \mathcal{S}^p \times \Re^m \times \mathcal{S}^p_+$ *be a KKT point of* (90) *with* $\overline{z} := \sqrt{g(\overline{x})}$*. Suppose that* $\mathcal{J}\widehat{h}(\overline{x}, \overline{z}) : X \times \mathcal{S}^p \mapsto \Re^m \times \mathcal{S}^p$ *is onto and that the second order sufficient condition is satisfied at* $(\overline{x}, \overline{z}, \overline{\zeta}, \overline{\xi})$*. Let* $\hat{c}$ *be a positive scalar such that*

$$\left\langle d, \nabla^2_{(x,z)(x,z)} \widehat{L}_{\hat{c}}(\overline{x}, \overline{z}, \overline{\zeta}, \overline{\xi}) d \right\rangle > 0 \quad \forall\, 0 \neq d \in X \times \mathcal{S}^p .$$

Then there exist positive scalars $\overline{c} \geq \hat{c}$*,* δ*,* ε*, and* ϱ_0 *such that*

(i) For all (ζ, ξ, c) *in the set* $D \subset \Re^m \times \mathcal{S}^p \times \Re$ *defined as*

$$D := \{(\zeta, \xi, c) \in \Re^m \times \mathcal{S}^p \times \Re \,|\, \|(\zeta, \xi) - (\overline{\zeta}, \overline{\xi})\| < \delta c\,, \ \ \overline{c} \leq c\}\,,$$

the problem

$$\min \ L_c(x, \zeta, \xi) \quad \text{s.t. } x \in \mathbf{B}_\varepsilon(\overline{x}) \tag{93}$$

has a unique solution denoted $x(\zeta, \xi, c)$*. The function* $x(\cdot, \cdot, \cdot)$ *is continuously differentiable in the interior of* D*, and, for all* $(\zeta, \xi, c) \in D$*, we have*

$$\|x(\zeta, \xi, c) - \overline{x}\| \leq \varrho_0 \|(\zeta, \xi) - (\overline{\zeta}, \overline{\xi})\|/c\,. \tag{94}$$

(ii) For all $(\zeta, \xi, c) \in D$*, we have*

$$\| \left(\widetilde{\zeta}(\zeta, \xi, c), \widetilde{\xi}(\zeta, \xi, c)\right) - (\overline{\zeta}, \overline{\xi})\| \leq \varrho_0 \|(\zeta, \xi) - (\overline{\zeta}, \overline{\xi})\|/c\,, \tag{95}$$

where

$$\widetilde{\zeta}(\zeta, \xi, c) := \zeta + ch\left(x(\zeta, \xi, c)\right)\,, \quad \widetilde{\xi}(\zeta, \xi, c) := \Pi_{\mathcal{S}^p_+}\left(\xi - cg(x(\zeta, \xi, c))\right)\,.$$

Proof. It follows from [3] (also see [4, Section 2.2]) that there exist positive numbers $\hat{\delta}, \hat{\varepsilon}$, and ϱ_0 such that for all (ζ, ξ, c) in the set $\widehat{D} \subset \Re^m \times \mathcal{S}^p \times \Re$ defined as

$$\widehat{D} := \{(\zeta, \xi, c) \big| \|(\zeta, \xi) - (\overline{\zeta}, \overline{\xi})\| < \hat{\delta} c, \ \ \hat{c} \leq c\},$$

the problem

$$\min \ \widehat{L}_c(x, z, \zeta, \xi) \quad \text{s.t. } (x, z) \in \mathbf{B}_{\hat{\varepsilon}}(\overline{x}) \times \mathbf{B}_{\hat{\varepsilon}}(\overline{z}) \tag{96}$$

has a unique solution denoted $(\hat{x}(\zeta, \xi, c), \hat{z}(\zeta, \xi, c))$ satisfying

$$\begin{aligned} &\|(\hat{x}(\zeta, \xi, c), \hat{z}(\zeta, \xi, c)) - (\overline{x}, \overline{z})\| \leq \varrho_0 \|(\zeta, \xi) - (\overline{\zeta}, \overline{\xi})\|/c\,, \\ &\|(\hat{\zeta}(\zeta, \xi, c), \hat{\xi}(\zeta, \xi, c)) - (\overline{\zeta}, \overline{\xi})\| \leq \varrho_0 \|(\zeta, \xi) - (\overline{\zeta}, \overline{\xi})\|/c\,, \end{aligned} \tag{97}$$

where

$$\hat{\zeta}(\zeta, \xi, c) := \zeta + ch\left(\hat{x}(\zeta, \xi, c)\right)\,, \quad \hat{\xi}(\zeta, \xi, c) := \xi + c\left(\hat{z}^2(\zeta, \xi, c) - g(\hat{x}(\zeta, \xi, c))\right)\,. \tag{98}$$

Assume that rank $(g(\overline{x})) = r_0$ and $g(\overline{x})$ has the following spectral decomposition

$$g(\overline{x}) = P \begin{bmatrix} 0 & 0 \\ 0 & \Lambda_0 \end{bmatrix} P^T\,,$$

where $\Lambda_0 \in \mathcal{S}^{r_0}$ is a diagonal matrix whose diagonal elements are the r_0 positive eigenvalues of $g(\overline{x})$ and P is an orthogonal matrix. Then the mapping $G : \mathcal{S}^{r_0} \mapsto \mathcal{S}^{r_0}$ defined as

$$G(\Lambda) := \sqrt{\sqrt{\Lambda^2}}, \quad \Lambda \in \mathcal{S}^{r_0}$$

is analytic at Λ_0 [50, Theorem 3.1]. Therefore there exists a positive number $\hat{\varepsilon}_1 \in (0, \hat{\varepsilon})$ such that for any $\Lambda \in \mathcal{S}^{r_0}$ with $\|\Lambda - \Lambda_0\| \leq 2\hat{\varepsilon}_1$, Λ is positive definite and

$$\|\sqrt{\Lambda} - \sqrt{\Lambda_0}\| = \|G(\Lambda) - G(\Lambda_0)\| \leq 2\|\mathcal{J}G(\Lambda_0)\|\|\Lambda - \Lambda_0\| \,. \tag{99}$$

Let $\hat{\varepsilon}_2 \in (0, \hat{\varepsilon}_1]$ be such that

$$3\sqrt{p}\hat{\varepsilon}_2 + 16\|\mathcal{J}G(\Lambda_0)\|^2\hat{\varepsilon}_2^2 \leq \hat{\varepsilon}_1^2 \,. \tag{100}$$

Since g is continuously differentiable, there exists $l_g > 0$ such that

$$\|g(x) - g(\overline{x})\| \leq l_g\|x - \overline{x}\| \quad \forall\ x \in \mathbf{B}_{\hat{\varepsilon}}(\overline{x}) \,.$$

Let $\varepsilon \in (0, \hat{\varepsilon}]$, $\delta \in (0, \hat{\delta}]$, and $\overline{c} \geq \hat{c}$ be such that

$$l_g\varepsilon + \delta + \|\overline{\xi}\|/\overline{c} \leq \hat{\varepsilon}_2 \,. \tag{101}$$

Define

$$D := \{(\zeta, \xi, c) \in \Re^m \times \mathcal{S}^p \times \Re \,\big|\, \|(\zeta, \xi) - (\overline{\zeta}, \overline{\xi})\| < \delta c \,,\ \overline{c} \leq c\} \,.$$

Then, for any $(\zeta, \xi, c) \in D$ and $x \in X$ such that $\|x - \overline{x}\| \leq \varepsilon$, it follows from (92) and (101) that

$$\begin{aligned}
\|z_c^2(x, \zeta, \xi) - g(\overline{x})\| &= \|\Pi_{\mathcal{S}_+^p}(g(x) - \xi/c) - \Pi_{\mathcal{S}_+^p}(g(\overline{x}))\| \\
&\leq \|g(x) - g(\overline{x}) - (\xi - \overline{\xi})/c - \overline{\xi}/c\| \\
&\leq \|g(x) - g(\overline{x})\| + \|(\xi - \overline{\xi})/c\| + \|\overline{\xi}/c\| \\
&\leq l_g\|x - \overline{x}\| + \delta + \|\overline{\xi}\|/c \leq \hat{\varepsilon}_2 \,.
\end{aligned} \tag{102}$$

Let $v \in \mathcal{S}_+^p$ be such that $v \in \mathbf{B}_{\hat{\varepsilon}_2}(g(\overline{x}))$. Define $Z_{11} \in \mathcal{S}^{p-r_0}$, $Z_{12} \in \Re^{(p-r_0)\times r_0}$, and $Z_{22} \in \mathcal{S}^{r_0}$ by

$$\begin{bmatrix} Z_{11} & Z_{12} \\ Z_{12}^T & Z_{22} \end{bmatrix} := P^T\sqrt{v}P \,.$$

Then

$$\left\| \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{12}^T & Z_{22} \end{bmatrix}^2 - \begin{bmatrix} 0 & 0 \\ 0 & \Lambda_0 \end{bmatrix} \right\| = \|P^TvP - P^Tg(\overline{x})P\| = \|v - g(\overline{x})\| \leq \hat{\varepsilon}_2 \,.$$

Therefore,

$\langle Z_{11}^2, Z_{11}^2\rangle \leq \hat{\varepsilon}_2^2$, $\langle Z_{12}Z_{12}^T, Z_{12}Z_{12}^T\rangle \leq \hat{\varepsilon}_2^2$, $\langle Z_{22}^2 + Z_{12}^TZ_{12} - \Lambda_0, Z_{22}^2 + Z_{12}^TZ_{12} - \Lambda_0\rangle \leq \hat{\varepsilon}_2^2$, which, implies

$$\mathrm{Tr}(Z_{11}^2) \le \sqrt{p}\hat{\varepsilon}_2, \ \ \mathrm{Tr}(Z_{12}Z_{21}^T) = \mathrm{Tr}\ (Z_{12}^T Z_{12}) \le \sqrt{p}\hat{\varepsilon}_2\,. \tag{103}$$

Let $\Gamma_0 := Z_{22}^2 - \Lambda_0$. Then

$$\|\Gamma_0\| = \|\left(Z_{22}^2 + Z_{12}^T Z_{12} - \Lambda_0\right) - Z_{12}^T Z_{12}\| \le \hat{\varepsilon}_2 + \|Z_{12}Z_{12}^T\| \le 2\hat{\varepsilon}_2\,.$$

Hence, by (99),

$$\|Z_{22} - \sqrt{\Lambda_0}\| = \|\sqrt{\Lambda_0 + \Gamma_0} - \sqrt{\Lambda_0}\| \le 2\|\mathcal{J}G(\Lambda_0)\|\|\Gamma_0\| \le 4\hat{\varepsilon}_2\|\mathcal{J}G(\Lambda_0)\|\,. \tag{104}$$

Therefore, by (103), (104), and (100), we have

$$\begin{aligned}
&\left\|\sqrt{v} - \sqrt{g(\overline{x})}\right\|^2 = \left\|\begin{bmatrix} Z_{11} & Z_{12} \\ Z_{12}^T & Z_{22}\end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{\Lambda_0}\end{bmatrix}\right\|^2 \\
&= \left\|\begin{bmatrix} Z_{11} & Z_{12} \\ Z_{12}^T & Z_{22} - \sqrt{\Lambda_0}\end{bmatrix}\right\|^2 \\
&= \mathrm{Tr}(Z_{11}^2 + Z_{12}Z_{12}^T) + \mathrm{Tr}(Z_{12}^T Z_{12}) + \mathrm{Tr}((Z_{22} - \sqrt{\Lambda_0})^2) \\
&\le 3\sqrt{p}\hat{\varepsilon}_2 + 16\hat{\varepsilon}_2^2\|\mathcal{J}G(\Lambda_0)\|^2 \le \hat{\varepsilon}_1^2\,,
\end{aligned}$$

which implies that for any $v \in \mathcal{S}_+^p$ satisfying $v \in \mathbf{B}_{\hat{\varepsilon}_2}(g(\overline{x}))$ we have

$$\|\sqrt{v} - \overline{z}\| \le \hat{\varepsilon}_1\,.$$

In particular, by (102), for any $(\zeta, \xi, c) \in D$ and $x \in X$ such that $\|x - \overline{x}\| \le \varepsilon$, we have

$$z_c(x, \zeta, \xi) \in \mathbf{B}_{\hat{\varepsilon}_1}(\overline{z})\,.$$

Thus, from (91) we know that for any $(\zeta, \xi, c) \in D$,

$$\begin{aligned}
\min_{x \in \mathbf{B}_\varepsilon(\overline{x})} L_c(x, \zeta, \xi) &= \min_{x \in \mathbf{B}_\varepsilon(\overline{x})} \widehat{L}_c(x, z_c(x, \zeta, \xi), \zeta, \xi) \\
&\ge \min_{(x,z) \in \mathbf{B}_\varepsilon(\overline{x}) \times \mathbf{B}_{\hat{\varepsilon}_1}(\overline{z})} \widehat{L}_c(x, z, \zeta, \xi) \\
&\ge \min_{x \in \mathbf{B}_\varepsilon(\overline{x})} \left(\min_{z \in \mathcal{S}^p} \widehat{L}_c(x, z, \zeta, \xi)\right) \\
&= \min_{x \in \mathbf{B}_\varepsilon(\overline{x})} \left(\min_{v \in \mathcal{S}_+^p} \widetilde{L}_c(x, v, \zeta, \xi)\right) \\
&= \min_{x \in \mathbf{B}_\varepsilon(\overline{x})} L_c(x, \zeta, \xi)\,.
\end{aligned} \tag{105}$$

Let

$$x(\zeta, \xi, c) := \hat{x}(\zeta, \xi, c)\,, \quad (\zeta, \xi, c) \in D\,.$$

Then, by (105), for any $(\zeta, \xi, c) \in D$, $x(\zeta, \xi, c)$ is a solution to problem (93). The uniqueness of $x(\zeta, \xi, c)$ follows from the uniqueness of $(\hat{x}(\zeta, \xi, c), \hat{z}(\zeta, \xi, c))$.

For any $(\zeta, \xi, c) \in D$, by using (98), (105), and the fact that $(\hat{x}(\zeta, \xi, c), \hat{z}(\zeta, \xi, c))$ is the unique solution to (96), we know that

$$\hat{\zeta}(\zeta, \xi, c) = \zeta + ch(\hat{x}(\zeta, \xi, c)) = \zeta + ch(x(\zeta, \xi, c)) = \widetilde{\zeta}(\zeta, \xi, c)$$

and
$$\begin{aligned}\hat{\xi}(\zeta,\xi,c) &= \xi + c\left(\hat{z}^2(\zeta,\xi,c) - g(\hat{x}(\zeta,\xi,c))\right)\\ &= \xi + cz_c^2(x(\zeta,\xi,c),\zeta,\xi) - cg(x(\zeta,\xi,c))\\ &= \xi - cg(x(\zeta,\xi,c)) + c\Pi_{\mathcal{S}_+^p}\left(g(x(\zeta,\xi,c)) - c^{-1}\xi\right)\\ &= \Pi_{\mathcal{S}_+^p}\left(\xi - cg(x(\zeta,\xi,c))\right) = \widetilde{\xi}(\zeta,\xi,c)\,.\end{aligned}$$

Finally, the estimates (94) and (95) follow from (97). $\square$