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

# A quadratically convergent semismooth Newton method for nonlinear semidefinite programming without the subdifferential regularity

**Fuxiaoyue Feng · Chao Ding · Xudong Li**

This version: February 21, 2024

**Abstract** We introduce a quadratically convergent semismooth Newton method for nonlinear semidefinite programming that eliminates the need for the subdifferential regularity, a common yet stringent requirement in existing approaches. Our strategy revolves around identifying a single nonsingular element within the B(ouligand)-subdifferential, thus avoiding the standard requirement for nonsingularity across the entire subdifferential set, which is often too restrictive for practical applications. The theoretical framework is supported by the introduction of the weak second order condition (W-SOC) and the weak strict Robinson constraint qualification (W-SRCQ). These conditions not only guarantee the existence of a nonsingular element in the subdifferential but also forge a primal-dual connection in linearly constrained convex quadratic programming. The theoretical advancements further lay the foundation for the algorithmic designing of a novel semismooth Newton method, which integrates a correction step to address degenerate issues. Particularly, this correction step ensures the local convergence as well as a superlinear/quadratic convergence rate of the proposed method. Preliminary numerical experiments corroborate our theoretical findings and underscore the practical effectiveness of our method.

**Key words:** Semismooth Newton method, singualrity, subdifferential, nonlinear semidefinite programming, second order conditions, constraint qualifications.

**AMS subject classification: 49K10, 49J52, 90C30, 90C22, 90C33**

F.X.Y. Feng
Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, P.R. China, School of Mathematical Sciences, University of Chinese Academy of Science, Beijing, P.R. China.
E-mail: fengfuxiaoyue@amss.ac.cn

C. Ding
Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, P.R. China.
E-mail: dingchao@amss.ac.cn

X.D. Li
School of Data Science, Fudan University, Shanghai, P.R. China.
E-mail: lixudong@fudan.edu.cn

## 1 Introduction

Let $\mathbb{X}$ and $\mathbb{Y}$ be two Euclidean spaces each equipped with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$. Consider the following optimization problem (OP):

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & G(x) \in K, \end{aligned} \tag{1}$$

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

Let $L : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$ be the Lagrangian function of problem (1) defined by

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

For any $y \in \mathbb{Y}$, denote the derivative of $L(\cdot; y)$ at $x \in \mathbb{X}$ by $L'_x(x; y)$ and denote the adjoint of $L'_x(x; y)$ by $\nabla_x L(x; y)$. The Karush-Kuhn-Tucker (KKT) optimality conditions for problem (1) takes the following form:

$$\begin{cases} \nabla_x L(x, y) = 0, \\ y \in \mathcal{N}_K(G(x)), \end{cases} \tag{2}$$

where $\mathcal{N}_K(G(x))$ is the normal cone to $K$ at $G(x)$ in the context of convex analysis [24]. We say $(x, y) \in \mathbb{Z} := \mathbb{X} \times \mathbb{Y}$ is a KKT point of (1) if it solves (2). The set of Lagrange multipliers associated with a feasible point $x$ is defined by $M(x) = \{y \mid (x, y) \text{ is a KKT point}\}$. For a given feasible point $\overline{x}$, if $M(\overline{x}) \neq \emptyset$, then we call it a stationary point of (1). It is well-known that $(\overline{x}, \overline{y})$ is a KKT point of (1) if and only if it is a solution of the nonsmooth system

$$F(x, y) := \begin{bmatrix} \nabla_x L(x, y) \\ -G(x) + \Pi_K(G(x) + y) \end{bmatrix} = 0, \tag{3}$$

where $\Pi_K : \mathbb{Y} \to \mathbb{Y}$ is the metric projection operator onto $K$, i.e., $\Pi_K(z) = \arg\min_{y' \in K} \frac{1}{2}\|z - y'\|^2$ for $z \in \mathbb{Y}$.

When solving $F(x, y) = 0$, one may encounter challenges related to the nonsmoothness of $F$, primarily arising from the nonsmooth nature of the metric projection operator $\Pi_K(\cdot)$, especially when $K$ is not a subspace. Fortunately, for various problems, including nonlinear programming (NLP), nonlinear second-order cone programming, and nonlinear semidefinite programming (NLSDP), the corresponding nonsmooth map $F$ exhibits the so-called (strong) semismoothness (see Definition 1). This property is conducive to the use of nonsmooth Newton methods. Qi and Sun [21] introduced a semismooth Newton method, which takes the following semismooth Newton step in each iteration:

$$Z^{k+1} = Z^k - (U^k)^{-1} F(Z^k),$$

with $U^k$ being selected from the B(ouligand)-subdifferential $\partial_B F(Z^k)$ (see (4) for definition) or the Clarke-subdifferential $\partial_C F(Z^k)$ (see (5) for definition). The semismooth Newton method, renowned for its high efficiency, has found widespread applications across various domains. It has been effectively utilized for tackling variational inequalities and constrained optimization problems in function spaces [30], infinite-dimensional linear complementarity problems [7] and computing the nearest correlation matrix [18]. Notably, its efficacy has been particularly pronounced in solving augmented Lagrangian subproblems, whose versatility is underscored by its implementation in several influential software packages, including SDPNAL [33], SDPNAL+ [32], QSDPNAL [13], and SSNAL [12]. For an in-depth exploration, one may refer to [20, 7, 5].

It is well-known (see e.g., [21, 10, 19]) that the (quadratic) superlinear convergence of semismooth Newton methods hinge on the (strong) g-semismoothness (see Definition 1) of $F$ and the uniform boundedness of $\{(U^k)^{-1}\}$. To guarantee such uniform boundedness, typically, one has to assume the KKT nonsmooth mapping $F$ satisfies certain subdifferential regularity at the solution point $(\overline{x}, \overline{y})$, for example, all elements in some subdifferential sets of $F$ at $(\overline{x}, \overline{y})$ are nonsingular. Particularly, one may assume that every element in $\partial_B F(\overline{x}, \overline{y})$ or even $\partial_C F(\overline{x}, \overline{y})$ is nonsingular, known as the BD-regularity or CD-regularity, respectively. However, this condition is frequently demanding and may not always be fulfilled in practical applications. Thus, the following important question arises:

> *Can we design a locally superlinear or quadratic convergent semismooth Newton method for solving* (3) *without assuming the subdifferential regularity of $F$?*

In this paper, we provide an affirmative answer to this question. Specifically, for the NLSDP, we design a semismooth Newton method with a novel correction step to avoid the requirement of nonsingularity across the entire subdifferential of $F$, and prove its fast local convergence rates, thus advancing the traditional paradigms. Take [4, Example 3] (more details can be found in Example 4) as an illustrative example. It can be shown that for this special example, the corresponding B-subdifferential $\partial_B F(\overline{x}, \overline{y})$ at the KKT point $(\overline{x}, \overline{y})$ contains singular matrices. Figure 1 provides an illustration of the performance of our newly proposed semismooth Newton method with a correction step for solving this degenerate problem. Clearly, even in the absence of the BD-regularity, our method still exhibits a local quadratic convergence.

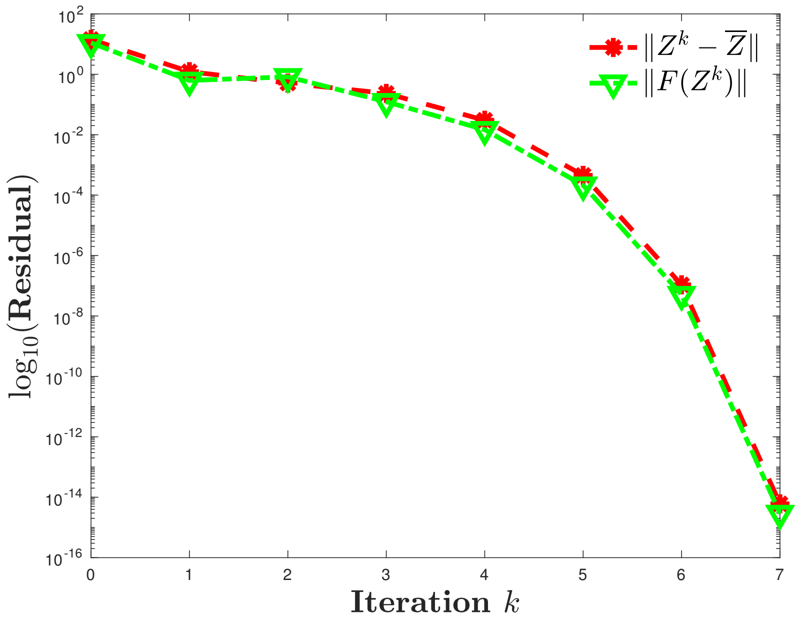


Fig. 1: Quadratic convergence of our semismooth Newton method for solving Example 4.

For the optimization problem (1), the characterization of the subdifferential regularity for the nonsmooth function $F$ is closely intertwined with perturbation theory, a fundamental topic in optimization. Indeed, the perturbation theory is crucial for both theoretical insights and practical algorithm design. Over the past four decades, significant advancements in this field has been well-documented in the literature [1, 9, 25, 5, 16], especially concerning (1) with $K$ being a polyhedral set (i.e., $K$ is the intersection of a finite number of half-spaces). However, less exploration has been done regarding non-polyhedral cases, such as $K$ being the set of positive semidefinite matrices.

In the realm of nonlinear semidefinite programming, Sun [26] established that under the Robinson constraint qualification (RCQ) (Definition 2), the nonsingularity of $\partial_C F(\overline{x}, \overline{y})$ (i.e., the CD-regularity of $F$) at a local optimal solution $\overline{x}$ and a associated multiplier $\overline{y}$ is equivalent to the strong second-order sufficient condition (S-SOSC) (Definition 6) and the constraint nondegeneracy (Definition 4), and is also equivalent with the strong regularity introduced by Robinson [23] of the solution to the KKT system (2). For linear SDP problems, Chan and Sun [2] provided deeper insights into the strong regularity, demonstrating that under RCQ, primal and dual constraint nondegeneracies, the nonsingularity of $\partial_B F(\overline{x}, \overline{y})$ (i.e., the BD-regularity of $F$), CD-regularity and the strong regularity at $(\overline{x}, \overline{y})$ are all equivalent. This comprehensive understanding underscores that the conditions for the nonsingularity of all elements in the subdifferentials of $F$ at $(\overline{x}, \overline{y})$ may be stringent and not consistently be satisfied in practice. For instance, the S-SOSC condition presupposes the isolatedness of a local optimal solution, a premise that is often violated in statistical optimization applications such as LASSO [29]. Consequently, it is crucial to develop a new framework of semismooth Newton methods that can bypass the subdifferential regularity and still achieve a local superlinear or quadratic convergence rate.

In this paper, we propose to relax the stringent subdifferential regularity conditions by identifying a single nonsingular element within the B-subdifferential. However, characterizing the nonsingularity of even a single element within the subdifferentials of $F$ at a KKT point $(\overline{x}, \overline{y})$, especially in non-polyhedral cases, remains an unresolved challenge. When $K$ is polyhedral, Izmailov and Solodov [8, Proposition 6] indicated that the weak strict Robinson constraint qualification (W-SRCQ) and the weak second-order condition (W-SOC) (see Definitions 7 and 8, respectively) guarantee the existence of a nonsingular element in the parametric-subdifferential. Here the parametric-subdifferential, defined as differentials of all possible parametric systems, contains $\partial_B F(\overline{x}, \overline{y})$. Indeed, a straightforward deduction from their findings indicates that the W-SRCQ and W-SOC also guarantee the presence of a nonsingular element in $\partial_B F(\overline{x}, \overline{y})$. However, these conditions are inadequate when applied to NLSDP. Specifically, the W-SRCQ and W-SOC are insufficient to ensure the existence of a nonsingular element in $\partial_B F(\overline{x}, \overline{y})$, as demonstrated by our counterexample (Example 2). Meanwhile, in NLSDP, the presence of a nonsingular subdifferential element, as in Example 4, does not necessarily imply calmness, an important property described in [25, Definition 9(30)]. The lack of calmness may lead to slow progress of algorithms for solving the KKT system (3), as the associated solution mapping exhibits heightened sensitivity to small perturbations.

In this paper, we show that the existence of a nonsingular element in $\partial_B F(\overline{x}, \overline{y})$ can be ensured under the W-SOC with constraint nondegeneracy or the S-SOSC with the W-SRCQ at a KKT point $(\overline{x}, \overline{y})$. Additionally, our results indicate that the W-SRCQ and W-SOC are sufficient for the existence of a nonsingular element in $\partial_C F(\overline{x}, \overline{y})$, and are also necessary for convex cases. More interestingly, we uncover a profound primal-dual connection between the W-SOC and W-SRCQ for linearly constrained convex quadratic semidefinite programming. These connections mirror those found between the strict Robinson constraint qualification (SRCQ) (Definition 3) and the second-order sufficient condition (SOSC) (Definition 5) for linearly constrained quadratic semidefinite programming [6], as well as those between the constraint nondegeneracy and the S-SOSC in the setting of linear semidefinite programming [2]. We summarize our results and connections in Figure 2. Based on the theoretical findings, we are able to design an inexact semismooth Newton method with a correction step. Each iteration commences with a semismooth Newton step, followed by a correction step designed to ensure the nonsingularity. Under the W-SOC and constraint nondegeneracy (or the S-SOSC and W-SRCQ), we show the local convergence and superliner convergence of the designed algorithm. Furthermore, if the function is twice Lipschitz continuously differentiable, the convergence rate can be further improved to quadratic.

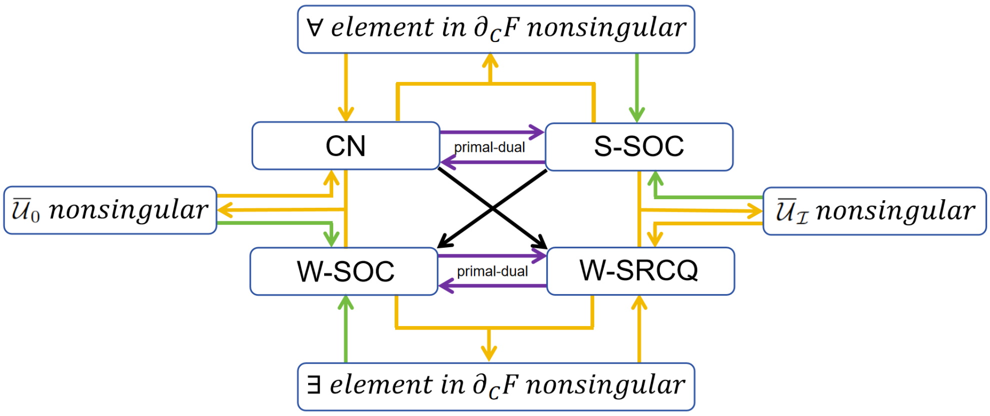


Fig. 2: Connections of second order conditions, constraint qualifications and nonsingularity of subdifferentials. Here the black arrow holds for (1); the orange arrow holds for NLSDP; the green arrow holds for CSDP; and the violet arrow holds for QSDP. Moreover, "primal-dual" means that the starting point of the arrow for the primal problem implies the arriving point of the arrow for the dual problem.

The remaining parts of this paper are organized as follows. Section 2 provides some basic definitions and preliminary results concerning the variational properties of the metric projection projector over the positive semidefinite cone. Section 3 delves into the sufficient and necessray conditions for the existence of one nonsingular element in the subdifferentials of $F$ for NLSDP. The primal-dual connections of the W-SOC and the W-SRCQ are explored in Section 4 for convex quadratic semidefinite programming. Then, we present a semismooth Newton method with a correction step in Section 5, and some preliminary numerical results in Section 6. Finally, we conclude this paper and highlight potential directions for future research in Section 7.

## 2 Preliminaries

We first give the following notation that will be used throughout the paper. Let $\mathbb{X}$ and $\mathbb{Y}$ be two real Euclidean spaces. Given a set $S \subseteq \mathbb{X}$ and a point $x \in \mathbb{X}$, the distance from $x$ to $S$ is denoted by

$$\mathrm{dist}(x, S) := \inf\{\|y - x\| \mid y \in S\}.$$

For a linear operator $\mathcal{A} : \mathbb{X} \to \mathbb{Y}$, $\mathcal{A}^*$ denotes the adjoint of the linear operator $\mathcal{A}$. We denote by $A^T$ the transpose of a given matrix $A$. For a mapping $\Phi : \mathbb{X} \to \mathbb{Y}$ and $x \in \mathbb{X}$, $\Phi'(x)$ stands for the classical derivative or the Jacobian of $\Phi$ at $x$ and $\nabla\Phi(x)$ the adjoint of the Jacobian. We use $\mathbb{S}^n$

to denote the linear space of all $n \times n$ real symmetric matrices and use $\mathbb{S}^n_+$ to denote the closed convex cone of all $n\times n$ positive semidefinite matrices in $\mathbb{S}^n$. For a function $g : \mathbb{X} \to \mathbb{R}$, we denote $g^+(x) := \max\{0, g(x)\}$ and if it is vector-valued then the maximum is taken componentwise. Some other notation are listed below:

- Let $\mathcal{O}^n$ be the set of all $n \times n$ orthogonal matrices.
- For any $Z \in \mathbb{R}^{m\times n}$, we denote by $Z_{ij}$ the $(i, j)$-th entry of $Z$.
- For any $Z \in \mathbb{R}^{m\times n}$ and a given index set $\mathcal{J} \subseteq \{1, \dots, n\}$, we use $Z_{\mathcal{J}}$ to denote the sub-matrix of $Z$ obtained by removing all the columns of $Z$ not in $\mathcal{J}$. In particular, we use $Z_j$ to represent the $j$-th column of $Z$, $j = 1, \dots, n$.
- Let $\mathcal{I} \subseteq \{1, \dots, m\}$ and $\mathcal{J} \subseteq \{1, \dots, n\}$ be two index sets. For any $Z \in \mathbb{R}^{m\times n}$, we use $Z_{\mathcal{I}\mathcal{J}}$ to denote the $|\mathcal{I}| \times |\mathcal{J}|$ sub-matrix of $Z$ obtained by removing all the rows of $Z$ not in $\mathcal{I}$ and all the columns of $Z$ not in $\mathcal{J}$.
- We use " $\circ$ " to denote the Hardamard product between matrices, i.e., for any two matrices $A$ and $B$ in $\mathbb{R}^{m\times n}$ the $(i, j)$-th entry of $Z := A \circ B \in \mathbb{R}^{m\times n}$ is $Z_{ij} = A_{ij}B_{ij}$.
- Let $\mathrm{Diag}(\cdot) : \mathbb{R}^n \to \mathbb{S}^n$ be the linear map such that for any $x \in \mathbb{R}^n$, $\mathrm{Diag}(x)$ denotes the diagonal matrix whose $i$-th diagonal entry is $x_i$, $i = 1, \dots, n$.
- Let $\mathrm{lin}(C)$ be the linearity space of a closed convex set $C \subseteq \mathbb{X}$, i.e., the largest subspace in $C$. Meanwhile, let $\mathrm{aff}(C)$ be the affine hull of $C$, i.e., the smallest subspace containing $C$.
- For a map $\Phi : \mathbb{X} \to \mathbb{Y}$, define $\mathrm{Im}(\Phi) := \{\Phi(x) \mid x \in \mathbb{X}\}$ and $\mathrm{Null}(\Phi) := \{x \in \mathbb{X} \mid \Phi(x) = 0\}$.
- The supprot function of a set $S \subseteq \mathbb{X}$ is defined by $\sigma(x, S) = \sup_{s\in S}\langle x, s\rangle$ for $x \in \mathbb{X}$.
- For $r > 0$ and $\overline{X} \in \mathbb{X}$, the closed ball centered at $\overline{X}$ with radius $r$ is denote by $\mathbb{B}(\overline{X}, r) = \{X \in \mathbb{X} \mid \|X - \overline{X}\| \le r\}$.
- For $f, g : \mathbb{X} \to \mathbb{R}$, $f(x) = \Theta(g(x))$ for $x \to \overline{x}$ means that $f(x) = O(g(x))$ and $g(x) = O(f(x))$.

For a given map $\Phi : \mathbb{X} \to \mathbb{Y}$, the directional differential of $\Phi$ at point $x \in \mathbb{X}$ along a direction $d \in \mathbb{X}$ is defined by

$$\Phi'(x; d) = \lim_{t\downarrow 0} \frac{F(x + td) - F(x)}{t} \quad \text{if exists.}$$

$\Phi$ is directionally differentiable at $x \in \mathbb{X}$ if $\Phi'(x; d)$ exists for all $d \in \mathbb{X}$. Suppose that $\Phi : N \subseteq \mathbb{X} \to \mathbb{Y}$ is a locally Lipschitz continuous function on the open set $N$. Then, according to Rademacher's theorem [22], $\Phi$ is almost everywhere F-differentiable in $N$. Let $D_\Phi$ be the set of points in $N$ where $\Phi$ is differentiable. Let $\Phi'(x)$ be the derivative of $\Phi$ at $x \in D_\Phi$. Then the B-subdifferential of $\Phi$ at $x \in N$ is denoted by [19]:

$$\partial_B\Phi(x) := \left\{ \lim_{D_\Phi \ni x^k \to x} \Phi'(x^k) \right\}, \tag{4}$$

and the Clarke-subdifferential of $\Phi$ at $x \in N$ [3] takes the form:

$$\partial_C\Phi(x) = \mathrm{conv}\left\{\partial_B\Phi(x)\right\}, \tag{5}$$

where "conv" stands for the convex hull in the usual sense of convex analysis [24]. The following g-semismoothness of a locally Lipschitz continuous function is an extension of the semismoothness (originally defined by Mifflin [15] for real-valued functions and adopted by Qi and Sun [21] to the vector case.)

**Definition 1** Let $\Phi : N \subseteq \mathbb{X} \to \mathbb{Y}$ be a locally Lipschitz continuous function on the open set $N$. The function $\Phi$ is said to be g-semismooth at a point $x \in N$ if for any $y \to x$ and $V \in \partial_C\Phi(y)$,

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

The function $\Phi$ is said to be strongly g-semismooth at $x$ if for any $y \to x$ and $V \in \partial_C \Phi(y)$,

$$\Phi(y) - \Phi(x) - V(y - x) = O(||y - x||^2)\,.$$

Furthermore, the function $\Phi$ is said to be (strongly) semismooth at $x \in N$ if (i) $\Phi$ is directional differentiable at $x$; and (ii) $\Phi$ is (strongly) g-semismooth.

## 2.1 Background in variational analysis

In this subsection, we introduce some variational properties related to problem (1), which will be useful in our subsequent discussions.

For the closed convex set $K \subseteq \mathbb{Y}$, the tangent cone to $K$ at $y \in \mathbb{Y}$ is defined by

$$\mathcal{T}_K(y) = \{d \in \mathbb{Y} \mid \exists\, t^k \downarrow 0,\ \operatorname{dist}(y + t^k d, K) = o(t^k)\}. \tag{6}$$

Let $\overline{x}$ be a given feasible solution to (1). It is well-known [1, Theorem 3.9 and Proposition 3.17] that for the feasible solution $\overline{x}$, the following Robinson constraint qualification is equivalent to the nonemptyness and boundedness of the corresponding Lagrange multiplier set $M(\overline{x})$.

**Definition 2** The Robinson constraint qualification (RCQ) is said to hold at a feasible point $\overline{x}$ if

$$G'(\overline{x})\mathbb{X} + \mathcal{T}_K(G(\overline{x})) = \mathbb{Y}.$$

For a stationary solution $\overline{x}$ of (1), the strict Robinson constraint qualification condition is defined as follows.

**Definition 3** Let $\overline{x} \in \mathbb{X}$ be a stationary point of (1) with a Lagrange multiplier $\overline{y} \in M(\overline{x})$. The strict Robinson constraint qualification (SRCQ) is said to hold at $\overline{x}$ with respect to $\overline{y}$ if

$$G'(\overline{x})\mathbb{X} + \mathcal{T}_K(G(\overline{x})) \cap \overline{y}^{\perp} = \mathbb{Y},$$

where $y^{\perp} := \{s \in \mathbb{Y} \mid \langle s, y\rangle = 0\}$ for any vector $y \in \mathbb{Y}$.

It follows from [1, Proposition 4.50] that the set of Lagrange multipliers $M(\overline{x})$ is a singleton if the SRCQ hold.

For the feasible point $\overline{x}$, the critical cone $\mathcal{C}(\overline{x})$ of (1) at $\overline{x}$ is defined by

$$\mathcal{C}(\overline{x}) = \{d \in \mathbb{X} \mid G'(\overline{x})d \in \mathcal{T}_K(G(\overline{x})),\ f'(\overline{x})d = 0\}. \tag{7}$$

If $\overline{x}$ is a stationary point with $\overline{y} \in M(\overline{x})$, then the critical cone $\mathcal{C}(\overline{x})$ takes the following form:

$$\mathcal{C}(\overline{x}) = \{d \in \mathbb{X} \mid G'(\overline{x})d \in \mathcal{T}_K(G(\overline{x})) \cap \overline{y}^{\perp}\}. \tag{8}$$

The following definition on the constraint nondegeneracy of (1), introduced by Robinson [23], is stronger than the SRCQ since $\operatorname{lin}(\mathcal{T}_K(G(\overline{x}))) \subseteq \mathcal{T}_K(G(\overline{x})) \cap \overline{y}^{\perp}$ (cf. [1, Proposition 4.73]).

**Definition 4** The constraint nondegeneracy of (1) is said to hold at $\overline{x}$ if

$$G'(\overline{x})\mathbb{X} + \operatorname{lin}(\mathcal{T}_K(G(\overline{x}))) = \mathbb{Y}.$$

Recall that the inner and outer second order tangent sets ([1, (3.49) and (3.50)]) to the given closed set $K \in \mathbb{Y}$ in the direction $d \in \mathbb{Y}$ can be defined, respectively, by

$$\mathcal{T}_K^{i,2}(y,d) := \{w \in \mathbb{Y} \mid \operatorname{dist}(y + td + \frac{1}{2}t^2 w, C) = o(t^2),\ t \geq 0\}$$

and

$$\mathcal{T}_K^{2}(y,d) := \{w \in \mathbb{Y} \mid \exists\, t^k \downarrow 0,\ \operatorname{dist}(y + t^k d + \frac{1}{2}(t^k)^2 w, C) = o((t^k)^2)\}.$$

Note that in general, $\mathcal{T}_K^{i,2}(y,d) \neq \mathcal{T}_K^{2}(y,d)$ even for the convex set $K$ (cf. [1, Section 3.3]). However, it follows from [1, Proposition 3.136] that if $K$ is a $C^2$-cone reducible [1, Definition 3.135] convex set (e.g., the polyhedron, second-order cone, positive semidefinite matrix cone and their Cartesian products), then the equality always holds. In this case, $\mathcal{T}_K^{2}(y,d)$ will be simply called the second order tangent set to $K$ at $y \in \mathbb{Y}$ in the direction $d \in \mathbb{Y}$.

Suppose that $K$ in (1) is $C^2$-cone reducible. The following second order sufficient condition is adopted from [1, (3.276)].

**Definition 5** Let $\overline{x} \in \mathbb{X}$ be a stationary point of (1) with a Lagrange multiplier $\overline{y} \in M(\overline{x})$. The second order sufficient condition (SOSC) is said to hold at $(\overline{x}, \overline{y})$ if

$$\langle d, \nabla^2_{xx} L(\overline{x}, \overline{y}) d\rangle - \sigma(\overline{y}, \mathcal{T}_K^2(G(\overline{x}), G'(\overline{x})d)) > 0 \quad \forall\, d \in \mathcal{C}(\overline{x}) \backslash \{0\}. \tag{9}$$

The following definition of strong second order sufficient condition for (1) is an extension of [26, Definition 3.2] for NLSDP. It is worth to note when $M(\overline{x})$ is a singleton, both of them are the same.

**Definition 6** Let $\overline{x} \in \mathbb{X}$ be a stationary point of (1) with a Lagrange multiplier $\overline{y} \in M(\overline{x})$. The strong second order sufficient condition (S-SOSC) is said to hold at $(\overline{x}, \overline{y})$ if

$$\langle d, \nabla^2_{xx} L(\overline{x}, \overline{y}) d\rangle - \sigma(\overline{y}, \mathcal{T}_K^2(G(\overline{x}), G'(\overline{x})d)) > 0 \quad \forall\, d \in \operatorname{app}(\overline{x}, \overline{y}) \backslash \{0\},$$

where $\operatorname{app}(\overline{x}, \overline{y})$ is the out approximation of the affine hull of the critical cone $\mathcal{C}(\overline{x})$ of (1) with respect to $(\overline{x}, \overline{y})$, i.e.,

$$\operatorname{app}(\overline{x}, \overline{y}) := \{d \in \mathbb{X} \mid G'(\overline{x})d \in \operatorname{aff}(\mathcal{T}_K(G(\overline{x})) \cap \overline{y}^{\perp})\}. \tag{10}$$

## 2.2 Eigenvalue decomposition of symmetric matrices

In this subsection, we introduce some useful preliminary results on eigenvalue decompositions of real symmetric matrices and the differentiability of the metric projectors over the positive semidefinite matrix cone $\mathbb{S}^n_+$.

Let $A \in \mathbb{S}^n$ be given. We use $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_n(A)$ to denote the eigenvalues of $A$ (all real and counting multiplicity) arranging in nonincreasing order and use $\lambda(A)$ to denote the vector of the ordered eigenvalues of $A$. Let $\Lambda(A) := \operatorname{Diag}(\lambda(A))$. Consider the eigenvalue decomposition of $A$, i.e., $A = P\Lambda(A)P^T$, where $P \in \mathcal{O}^n$. By considering the index sets of positive, zero, and negative eigenvalues of $A$, we are able to write $A$ in the following form

$$A = \begin{bmatrix} P_\alpha & P_\beta & P_\gamma \end{bmatrix} \begin{bmatrix} \Lambda(A)_{\alpha\alpha} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Lambda(A)_{\gamma\gamma} \end{bmatrix} \begin{bmatrix} P_\alpha^T \\ P_\beta^T \\ P_\gamma^T \end{bmatrix}, \tag{11}$$

where $\alpha$, $\beta$ and $\gamma$ are index sets defined by

$$\alpha := \{i \mid \lambda_i > 0\}, \quad \beta := \{i \mid \lambda_i = 0\} \quad \text{and} \quad \gamma := \{i \mid \lambda_i < 0\}. \tag{12}$$

We use $\mathcal{O}^n(A)$ to denote the set of all orthogonal matrices $P \in \mathcal{O}^n$ satisfying (11).

The next proposition, which was stated in [28, Proposition 4.4], illustrates that the eigenvectors of symmetric matrices, though not continuous, are upper Lipschitz continuous.

**Proposition 1** *Given $A \in \mathbb{S}^n$, for any $H \in \mathbb{S}^n$, let $P$ be an orthogonal matrix such that $X+H = P\Lambda(A+H)P^T$. Then, as $H \to 0$, we have*

$$\mathrm{dist}(P, \mathcal{O}^n(A)) = O(\|H\|).$$

Let $\Pi_{\mathbb{S}^n_+} : \mathbb{S}^n \to \mathbb{S}^n$ be the metric projection operator over the positive semidefinite matrix cone $\mathbb{S}^n_+$, i.e.,

$$\Pi_{\mathbb{S}^n_+}(A) := \arg\min_{X \in \mathbb{S}^n_+} \left\{ \frac{1}{2}\|X - A\|^2 \right\}, \quad A \in \mathbb{S}^n. \tag{13}$$

It is clear that $\Pi_{\mathbb{S}^n_+}(\cdot)$ is globally Lipschtiz continuous. Furthermore, it follows from [27, Theorem 4.7] (see also [17, Proposition 9]) that $\Pi_{\mathbb{S}^n_+}(\cdot)$ is directionally differentiable at any $A \in \mathbb{S}^n$ and the directional derivative at $A$ along direction $H \in \mathbb{S}^n$ is given by

$$\Pi'_{\mathbb{S}^n_+}(A; H) = P \begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & \Xi_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} \\ \widetilde{H}^T_{\alpha\beta} & \Pi_{\mathcal{S}^{|\beta|}_+}(\widetilde{H}_{\beta\beta}) & 0 \\ \Xi^T_{\alpha\gamma} \circ \widetilde{H}^T_{\alpha\gamma} & 0 & 0 \end{bmatrix} P^T, \tag{14}$$

with $P \in \mathcal{O}^n(A)$, $\widetilde{H} := P^T H P$ and $\Xi \in \mathbb{S}^n$ is defined by

$$\Xi_{ij} := \frac{\max\{\lambda_i(A), 0\} - \max\{\lambda_j(A), 0\}}{\lambda_i(A) - \lambda_j(A)}, \quad i, j = 1, \ldots, n, \tag{15}$$

where 0/0 is defined to be 1. Meanwhile, it is well-known [27, Proposition 4.5 and Theorem 4.13] that the metric projection operator $\Pi_{\mathbb{S}^n_+}(\cdot)$ is strongly semismooth everywhere.

The following result on the characterizations of the B-subdifferential and Clarke-subdifferential of $\Pi_{\mathbb{S}^n_+}(\cdot)$ is taken from [17, Lemma 11].

**Proposition 2** *Given $A \in \mathbb{S}^n$. $\mathcal{V} \in \partial_B \Pi_{\mathbb{S}^n_+}(A)$ $(\partial_C \Pi_{\mathbb{S}^n_+}(A))$ if and only if there is $\mathcal{W} \in \partial_B \Pi_{\mathbb{S}^{|\beta|}_+}(0)$ $(\partial_C \Pi_{\mathbb{S}^{|\beta|}_+}(0))$ such that for any $H \in \mathbb{S}^n$,*

$$\mathcal{V}(H) = P \begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & \Xi_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} \\ \widetilde{H}_{\beta\alpha} & \mathcal{W}(\widetilde{H}_{\beta\beta}) & 0 \\ \Xi_{\gamma\alpha} \circ \widetilde{H}_{\gamma\alpha} & 0 & 0 \end{bmatrix} P^T,$$

*where $P \in \mathcal{O}^n(A)$ and $\widetilde{H} = P^T H P$. In particular, since both the zero mapping $\mathcal{W} \equiv 0$ and the identity mapping $\mathcal{W} \equiv \mathcal{I}$ from $\mathbb{S}^{|\beta|} \to \mathbb{S}^{|\beta|}$ are elements of $\partial_B \Pi_{\mathbb{S}^{|\beta|}_+}(0)$, the following two mappings $\mathcal{V}_0$ and $\mathcal{V}_{\mathcal{I}}$ defined by*

$$\mathcal{V}_0(H) = P \begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & \Xi_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} \\ \widetilde{H}_{\beta\alpha} & 0 & 0 \\ \Xi_{\gamma\alpha} \circ \widetilde{H}_{\gamma\alpha} & 0 & 0 \end{bmatrix} P^T, \quad H \in \mathbb{S}^n \tag{16}$$

*and*

$$\mathcal{V}_{\mathcal{I}}(H) = P \begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & \Xi_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} \\ \widetilde{H}_{\beta\alpha} & \widetilde{H}_{\beta\beta} & 0 \\ \Xi_{\gamma\alpha} \circ \widetilde{H}_{\gamma\alpha} & 0 & 0 \end{bmatrix} P^T, \quad H \in \mathbb{S}^n \tag{17}$$

*are elements of* $\partial_B \Pi_{\mathbb{S}^n_+}(A)$*. Moreover,* $\mathcal{V}_0$ *and* $\mathcal{V}_{\mathcal{I}}$ *are independent with the choice of* $P \in \mathcal{O}^n(A)$*.*

## 3 Existence of nonsingular elements of the subdifferentials

As mentioned in the introduction, our preliminary objective is to explore the existence of nonsingular elements within the subdifferentials of $F$ defined in (3). This investigation sets the stage for the subsequent development of a semismooth Newton method that achieves quadratic convergence without the BD-regularity. To commence, we introduce the concept of the weak strict Robinson constraint qualification (W-SRCQ), which serves as an extension to the SRCQ condtion (Definition 3).

**Definition 7** Let $\overline{x} \in \mathbb{X}$ be a stationary point of (1) with a Lagrange multiplier $\overline{y} \in M(\overline{x})$. The weak strict Robinson constraint qualification (W-SRCQ) is said to hold at $\overline{x}$ with respect to $\overline{y}$ if

$$G'(\overline{x})\mathbb{X} + \operatorname{aff}(\mathcal{T}_K(G(\overline{x})) \cap \overline{y}^{\perp}) = \mathbb{Y}.$$

*Remark 1* The W-SRCQ condition is referred to as the *weak linear independence constraint qualification* (W-LICQ) in the context of nonlinear variational inequalities, as explored in [8]. Nevertheless, we choose not to adopt the W-LICQ terminology in this context. This is because the W-SRCQ is not simply a weaker form of the classical LICQ; rather, it is a distinct concept that arises by incorporating the affine hull into the SRCQ.

As can be inferred directly from the definitions (Definitions 3 and 7), the W-SRCQ is weaker than the SRCQ. This implies that the W-SRCQ is also milder than the constraint non-degeneracy condition as outlined in Definition 4. Moreover, the example provided below demonstrates that W-SRCQ is, in fact, even weaker than the RCQ defined by Definition 2.

*Example 1* Consider the following quadratic SDP problem:

$$\begin{aligned} \min \quad & \frac{1}{2}\|X_{11}\|^2 - \frac{1}{2}\|X_{22}\|^2 \\ \text{s.t.} \quad & X_{12} = 0, \quad X_{22} = 0, \\ & X \in \mathbb{S}^n_+, \end{aligned} \tag{18}$$

where $X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} \in \mathbb{S}^n$ with $X_{11} \in \mathbb{R}^{l_1 \times l_1}$ and $X_{22} \in \mathbb{R}^{l_2 \times l_2}$. It is easy to verify that $\overline{X} = 0$ is a local optimal solution with the following Lagrange multiplier set

$$M(\overline{X}) = \left\{ (\xi_{12}, \xi_{22}, \Gamma) \in \mathbb{R}^{l_1 \times l_2} \times \mathbb{R}^{l_2 \times l_2} \times \mathbb{S}^n \mid \begin{bmatrix} 0 & \xi_{12} \\ \xi_{12}^T & \xi_{22} \end{bmatrix} + \Gamma = 0 \right\}.$$

Thus, $(\overline{\xi}_{12}, \overline{\xi}_{22}, \overline{\Gamma}) = (0, 0, 0) \in M(\overline{X})$. Consider the KKT point $(\overline{X}, \overline{\xi}_{12}, \overline{\xi}_{22}, \overline{\Gamma}) = (0, 0, 0, 0)$. We know that the index sets defined by (12) are $\alpha = \gamma = \emptyset$ and $\beta = \{1, \ldots, n\}$. It then can be checked directly by Definition 7 that the W-SRCQ holds at $\overline{X}$ with respect to $(\overline{\xi}_{12}, \overline{\xi}_{22}, \overline{\Gamma})$. Meanwhile, since the multiplier set $M(\overline{X})$ is unbounded, we know that the RCQ does not hold at $\overline{X}$.

Next, we shall introduce the concept of the weak second order condition for a KKT solution $(\overline{x}, \overline{y})$ of (1), which is a generalization of the SOSC (Definition 5).

**Definition 8** Let $\overline{x} \in \mathbb{X}$ be a stationary point of (1) with a Lagrange multiplier $\overline{y} \in M(\overline{x})$. The weak second order condition (W-SOC) is said to hold at $(\overline{x}, \overline{y})$ if

$$\langle d, \nabla^2_{xx} L(\overline{x}, \overline{y}) d\rangle - \sigma(\overline{y}, \mathcal{T}^2_K(G(\overline{x}), G'(\overline{x})d)) > 0 \quad \forall\, d \in \mathrm{appl}(\overline{x}, \overline{y}) \backslash \{0\}, \tag{19}$$

where $\mathrm{appl}(\overline{x}, \overline{y})$ is defined by

$$\mathrm{appl}(\overline{x}, \overline{y}) := \left\{ d \in \mathbb{X} \mid G'(\overline{x})d \in \mathrm{lin}(\mathcal{T}_K(G(\overline{x})) \cap \overline{y}^{\perp}) \right\}. \tag{20}$$

*Remark 2* The W-SOC is referred to as the *second order condition* for nonlinear variational inequalities in [8]. However, to emphasize Definition 8 is a weaker variant of the SOSC, we prefer the term of W-SOC here. Moreover, it can be seen from the definitions that at a KKT point, the following implications hold:

$$\text{S-SOSC} \quad \Longrightarrow \quad \text{SOSC} \quad \Longrightarrow \quad \text{W-SOC}.$$

Furthermore, it is important to note that the W-SOC does not suffice to ensure local optimality for general optimization problems. Additionally, within the domain of convex problems, the W-SOC alone does not necessarily lead to the isolation of the optimal solution.

Henceforth, our primary focus will be on the following nonlinear semidefinite programming (NLSDP) problem:

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & h(x) = 0, \\ & g(x) \in \mathbb{S}^n_+. \end{array} \tag{21}$$

It should be noted that the results derived herein can be similarly extended through straightforward generalizations to cases where $K$ in (1) represents the Cartesian product of a finite number of positive semidefinite cones and zero vectors. Furthermore, considering that $\mathbb{R}^n$ can be represented as the Cartesian product of $n$ one-dimensional positive semidefinite cones $\mathbb{S}^1$, our analysis encompasses the classical nonlinear programming (NLP) as well. In addiction, we call the NLSDP (21) the convex semidefinite programming (CSDP) if $f$ is convex, $h$ is affine and $g$ is $\mathbb{S}^n_+$-convex, i.e., for any $x, y \in \mathbb{X}$ and $t \in (0,1)$,

$$g(tx + (1-t)y) - tg(x) - (1-t)g(y) \in \mathbb{S}^n_+.$$

For the NLSDP (21), the nonsmooth system (3) which is equivalent to the KKT optimality condition (2), takes the following form:

$$F(x, \xi, \Gamma) = \begin{bmatrix} \nabla_x L(x, \xi, \Gamma) \\ h(x) \\ -g(x) + \Pi_{\mathbb{S}^n_+}(g(x) + \Gamma) \end{bmatrix} = 0, \tag{22}$$

where the associated Lagrange function is given by $L(x, \xi, \Gamma) = f(x) + \langle \xi, h(x)\rangle + \langle \Gamma, g(x)\rangle$ and $\Pi_{\mathbb{S}^n_+}(\cdot)$ is the metric projection operator over $\mathbb{S}^n_+$ defined by (13).

Let $\overline{x}$ be a stationery solution of (22) with the Lagrange multiplers $(\overline{\xi}, \overline{\Gamma}) \in M(\overline{x})$. Denote $\overline{A} = g(\overline{x}) + \overline{\Gamma}$. Consider the corresponding eigenvalue decomposition (11) of $\overline{A}$ with the index sets $\alpha$, $\beta$ and $\gamma$ defined by (12), i.e.,

$$A = g(\overline{x}) + \overline{\Gamma} = \overline{P} \begin{bmatrix} \overline{\Lambda}_{\alpha\alpha} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \overline{\Lambda}_{\gamma\gamma} \end{bmatrix} \overline{P}^T, \tag{23}$$

where $\overline{\Lambda} = \Lambda(A) = \mathrm{Diag}(\overline{\lambda}_i)$ with $\overline{\lambda} := \lambda(A)$ and $\overline{P} \in \mathcal{O}^n(A)$. Thus, the tangent cone $\mathcal{T}_{\mathbb{S}^n_+}(g(\overline{x}))$ of $\mathbb{S}^n_+$ defined by (6) has the following explicit expression (cf. e.g., [26, (17)]):

$$\mathcal{T}_{\mathbb{S}^n_+}(g(\overline{x})) = \left\{ \overline{P} B \overline{P}^T \in \mathbb{S}^n \mid \begin{bmatrix} B_{\beta\beta} & B_{\beta\gamma} \\ B_{\gamma\beta} & B_{\gamma\gamma} \end{bmatrix} \succeq 0 \right\}. \tag{24}$$

Moreover, by simply calculations, we have

$$\mathcal{T}_{\mathbb{S}^n_+}(g(\overline{x})) \cap \overline{\Gamma}^{\perp} = \left\{ \overline{P} B \overline{P}^T \in \mathbb{S}^n \mid B_{\beta\beta} \succeq 0,\ B_{\beta\gamma} = 0,\ B_{\gamma\gamma} = 0 \right\}, \tag{25}$$

$$\mathrm{lin}(\mathcal{T}_{\mathbb{S}^n_+}(g(\overline{x}))) = \left\{ \overline{P} B \overline{P}^T \in \mathbb{S}^n \mid B_{\beta\beta} = 0,\ B_{\beta\gamma} = 0,\ B_{\gamma\gamma} = 0 \right\}, \tag{26}$$

$$\mathrm{aff}(\mathcal{T}_{\mathbb{S}^n_+}(g(\overline{x})) \cap \overline{\Gamma}^{\perp}) = \left\{ \overline{P} B \overline{P}^T \in \mathbb{S}^n \mid B_{\beta\gamma} = 0,\ B_{\gamma\gamma} = 0 \right\}. \tag{27}$$

For the notationally simplicity, we further define the linear operator $\overline{\mathcal{B}} : \mathbb{X} \to \mathbb{S}^n$ by

$$\overline{\mathcal{B}}(d) = \overline{P}^T g'(\overline{x}) d \overline{P}, \quad d \in \mathbb{X}. \tag{28}$$

Thus, together with (8), we know from (25) that the critical cone $\mathcal{C}(\overline{x})$ of the NLSDP (21) at the stationary point $\overline{x}$ takes the following form:

$$\mathcal{C}(\overline{x}) = \left\{ d \in \mathbb{X} \mid h'(\overline{x}) d = 0,\ \overline{\mathcal{B}}(d)_{\beta\beta} \succeq 0,\ \overline{\mathcal{B}}(d)_{\beta\gamma} = 0,\ \overline{\mathcal{B}}(d)_{\gamma\gamma} = 0 \right\}. \tag{29}$$

It follows from (27) and (26) that for the stationary point $\overline{x} \in \mathbb{X}$ of (21) with a Lagrange multiplier $(\overline{\xi}, \overline{\Gamma}) \in M(\overline{x})$, we have

$$\mathrm{app}(\overline{x}, \overline{\xi}, \overline{\Gamma}) = \left\{ d \in \mathbb{X} \mid h'(\overline{x}) d = 0,\ \overline{\mathcal{B}}(d)_{\beta\gamma} = 0,\ \overline{\mathcal{B}}(d)_{\gamma\gamma} = 0 \right\} \tag{30}$$

and

$$\mathrm{appl}(\overline{x}, \overline{\xi}, \overline{\Gamma}) = \left\{ d \in \mathbb{X} \mid h'(\overline{x}) d = 0,\ \overline{\mathcal{B}}(d)_{\beta\beta} = 0,\ \overline{\mathcal{B}}(d)_{\beta\gamma} = 0,\ \overline{\mathcal{B}}(d)_{\gamma\gamma} = 0 \right\}, \tag{31}$$

where the sets $\mathrm{app}(\overline{x}, \overline{\xi}, \overline{\Gamma})$ and $\mathrm{appl}(\overline{x}, \overline{\xi}, \overline{\Gamma})$ are defined by (10) and (20), respectively. Moreover, by employing the explicit formulas (27) and (26), we obtain the following useful characterizations of the W-SRCQ (Definition 7) and constarint nondegeneracy (Definition 4) for the NLSDP (21). For simplicity, we omit the detail proof here.

**Lemma 1** *Let $\overline{x} \in \mathbb{X}$ be a stationary point of* (21) *with a Lagrange multiplier $(\overline{\xi}, \overline{\Gamma}) \in M(\overline{x})$.*

*(i) The W-SRCQ holds at $\overline{x}$ with respect to $(\overline{\xi}, \overline{\Gamma})$ if and only if*

$$\begin{cases} h'(\overline{x})^* \Delta\xi + \overline{\mathcal{B}}^* \Delta\Gamma = 0, \\ \Delta\Gamma_{\alpha\alpha} = 0,\ \Delta\Gamma_{\alpha\beta} = 0,\ \Delta\Gamma_{\alpha\gamma} = 0,\ \Delta\Gamma_{\beta\beta} = 0 \end{cases} \Longrightarrow \quad (\Delta\xi, \Delta\Gamma) = 0. \tag{32}$$

*(ii) The constraint nondegeneracy holds at $\overline{x}$ if and only if*

$$\begin{cases} h'(\overline{x})^* \Delta\xi + \overline{\mathcal{B}}^* \Delta\Gamma = 0, \\ \Delta\Gamma_{\alpha\alpha} = 0,\ \Delta\Gamma_{\alpha\beta} = 0,\ \Delta\Gamma_{\alpha\gamma} = 0 \end{cases} \Longrightarrow \quad (\Delta\xi, \Delta\Gamma) = 0. \tag{33}$$

For the stationary point $\overline{x} \in \mathbb{X}$ of the NLSDP (21) with a Lagrange multiplier $(\overline{\xi}, \overline{\Gamma}) \in M(\overline{x})$, the support function of the second order tangent set in the second order condtions (Definitions 5, 6 and 8) has the following explicit characterization, which takes from [26, Lemma 3.1]:

$$\sigma(\overline{y}, \mathcal{T}_K^2(G(\overline{x}), G'(\overline{x})d)) = \sigma(\overline{\Gamma}, \mathcal{T}_{\mathbb{S}_+^n}^2(g(\overline{x}), g'(\overline{x})d)) = 2\sum_{i\in\alpha}\sum_{j\in\gamma}\frac{\overline{\lambda}_j}{\overline{\lambda}_i}\overline{\mathcal{B}}(d)_{ij}^2, \quad d \in \mathbb{X}, \tag{34}$$

where $K = \{0\} \times \mathbb{S}_+^n$, $G(x) = (h(x), g(x))$ and $\overline{y} = (\overline{\xi}, \overline{\Gamma})$.

Define $\mathbb{Z} := \mathbb{X} \times \mathbb{R}^m \times \mathbb{S}^n$. Let $(\overline{x}, \overline{\xi}, \overline{\Gamma})$ be a KKT solution of the NLSDP (21). Consider the B-subdifferential $\partial_B F(\overline{x}, \overline{\xi}, \overline{\Gamma})$ and Clarke-subdifferential $\partial_C F(\overline{x}, \overline{\xi}, \overline{\Gamma})$ of the KKT nonsmooth mapping $F$ given by (22). It follows from [26, Lemma 2.1] (see also [2, Lemma 1]) that $\overline{\mathcal{U}} \in \partial_B F(\overline{x}, \overline{\xi}, \overline{\Gamma})$ $(\partial_C F(\overline{x}, \overline{\xi}, \overline{\Gamma}))$ if and only if there exists $\overline{\mathcal{V}} \in \partial_B \Pi_{\mathbb{S}_+^m}(g(\overline{x}) + \overline{\Gamma})$ $(\partial_C \Pi_{\mathbb{S}_+^m}(g(\overline{x}) + \overline{\Gamma}))$ such that for any $(\Delta x, \Delta\xi, \Delta\Gamma) \in \mathbb{Z}$,

$$\overline{\mathcal{U}}(\Delta x, \Delta\xi, \Delta\Gamma) = \begin{bmatrix} \nabla_{xx}^2 L(\overline{x}, \overline{\xi}, \overline{\Gamma})\Delta x + h'(\overline{x})^*\Delta\xi + g'(\overline{x})^*\Delta\Gamma \\ h'(\overline{x})\Delta x \\ -g'(\overline{x})\Delta x + \overline{\mathcal{V}}(g'(\overline{x})\Delta x + \Delta\Gamma) \end{bmatrix}. \tag{35}$$

In particular, let $\overline{\mathcal{V}}_0$ and $\overline{\mathcal{V}}_{\mathcal{I}}$ be two mappings defined by (16) and (17) for $A = g(\overline{x}) + \overline{\Gamma}$, respectively. Define the following two elements $\overline{\mathcal{U}}_0$ and $\overline{\mathcal{U}}_{\mathcal{I}}$ in $\partial_B F(\overline{x}, \overline{\xi}, \overline{\Gamma})$ with $\overline{\mathcal{V}}_0$ and $\overline{\mathcal{V}}_{\mathcal{I}}$ by (35), respectively, i.e., for any $(\Delta x, \Delta\xi, \Delta\Gamma) \in \mathbb{Z}$,

$$\overline{\mathcal{U}}_0(\Delta x, \Delta\xi, \Delta\Gamma) = \begin{bmatrix} \nabla_{xx}^2 L(\overline{x}, \overline{\xi}, \overline{\Gamma})\Delta x + h'(\overline{x})^*\Delta\xi + g'(\overline{x})^*\Delta\Gamma \\ h'(\overline{x})\Delta x \\ -g'(\overline{x})\Delta x + \overline{\mathcal{V}}_0(g'(\overline{x})\Delta x + \Delta\Gamma) \end{bmatrix} \tag{36}$$

and

$$\overline{\mathcal{U}}_{\mathcal{I}}(\Delta x, \Delta\xi, \Delta\Gamma) = \begin{bmatrix} \nabla_{xx}^2 L(\overline{x}, \overline{\xi}, \overline{\Gamma})\Delta x + h'(\overline{x})^*\Delta\xi + g'(\overline{x})^*\Delta\Gamma \\ h'(\overline{x})\Delta x \\ -g'(\overline{x})\Delta x + \overline{\mathcal{V}}_{\mathcal{I}}(g'(\overline{x})\Delta x + \Delta\Gamma) \end{bmatrix}. \tag{37}$$

In the following two subsections, we shall study the existence of nonsingularity elements in subdifferentials at a KKT point $(\overline{x}, \overline{\xi}, \overline{\Gamma})$.

## 3.1 Sufficient conditions for the existence of nonsingularity elements in subdifferentials

In this subsection, we study the sufficient conditions to ensure the existence of nonsingularity elements in subdifferentials of the KKT nonsmooth mapping $F$ (22). The following results illustrate that under the W-SOC and constraint nondegeneracy (or the S-SOSC and W-SRCQ), the mapping $\overline{\mathcal{U}}_0$ (or $\overline{\mathcal{U}}_{\mathcal{I}}$) defined in (36) (or (37)) is indeed a nonsingular element of the B-subdifferential of $F$.

**Theorem 1** *Let $\overline{x} \in \mathbb{X}$ be a stationary point of (21) with a Lagrange multiplier $(\overline{\xi}, \overline{\Gamma}) \in M(\overline{x})$.*

*(i) If the W-SOC and constraint nondegeneracy hold at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$, then $\overline{\mathcal{U}}_0 \in \partial_B F(\overline{x}, \overline{\xi}, \overline{\Gamma})$ in (36) is nonsingular.*

*(ii) If the S-SOSC and W-SRCQ hold at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$, then $\overline{\mathcal{U}}_{\mathcal{I}} \in \partial_B F(\overline{x}, \overline{\xi}, \overline{\Gamma})$ in (37) is nonsingular.*

*Proof* Let $A = g(\overline{x})+\overline{\Gamma}$ satisfy the eigenvalue decomposition (23) and $\overline{P} \in \mathcal{O}^n(A)$. We only show (i) holds, as (ii) can be established through a similar argument.

Suppose that $\overline{\mathcal{U}}_0 \in \partial_B F(\overline{x}, \overline{\xi}, \overline{\Gamma})$ is singular, i.e., there is $0 \neq (\Delta x, \Delta\xi, \Delta\Gamma) \in \mathbb{Z}$ such that $\overline{\mathcal{U}}_0(\Delta x, \Delta\xi, \Delta\Gamma) = 0$. It then follows from (36) and Proposition 2 that

$$\nabla^2_{xx} L(\overline{x}, \overline{\xi}, \overline{\Gamma})\Delta x + h'(\overline{x})^*\Delta\xi + g'(\overline{x})^*\Delta\Gamma = 0, \tag{38}$$
$$h'(\overline{x})\Delta x = 0, \tag{39}$$
$$\widetilde{\Delta\Gamma}_{\alpha\alpha} = 0, \quad \widetilde{\Delta\Gamma}_{\alpha\beta} = 0, \tag{40}$$
$$\overline{\mathcal{B}}(\Delta x)_{\beta\beta} = 0, \quad \overline{\mathcal{B}}(\Delta x)_{\beta\gamma} = 0, \quad \overline{\mathcal{B}}(\Delta x)_{\gamma\gamma} = 0, \tag{41}$$
$$(E_{\alpha\gamma} - \overline{\Xi}_{\alpha\gamma}) \circ \overline{\mathcal{B}}(\Delta x)_{\alpha\gamma} - \overline{\Xi}_{\alpha\gamma} \circ \widetilde{\Delta\Gamma}_{\alpha\gamma} = 0, \tag{42}$$

where $\widetilde{\Delta\Gamma} := \overline{P}^T \Delta\Gamma \overline{P}$, the linear operator $\overline{\mathcal{B}} : \mathbb{X} \to \mathbb{S}^n$ is defined by (28), the matrix $\overline{\Xi} \in \mathbb{S}^n$ is given by (15) for $A$ and $E \in \mathbb{S}^n$ denotes the matrix whose elements are all ones. Then, by (38)-(42) and (34), we obtain that

$$\begin{aligned} 0 &= \left\langle \Delta x, \nabla^2_{xx} L(\overline{x}, \overline{\xi}, \overline{\Gamma})\Delta x + h'(\overline{x})^*\Delta\xi + g'(\overline{x})^*\Delta\Gamma \right\rangle \\ &= \left\langle \Delta x, \nabla^2_{xx} L(\overline{x}, \overline{\xi}, \overline{\Gamma})\Delta x \right\rangle + \langle h'(\overline{x})\Delta x, \Delta\xi\rangle + \left\langle \overline{\mathcal{B}}(\Delta x), \widetilde{\Delta\Gamma} \right\rangle \\ &= \langle \Delta x, \nabla^2_{xx} L(\overline{x}, \overline{\xi}, \overline{\Gamma})\Delta x\rangle + 2\sum_{i\in\alpha}\sum_{j\in\gamma} \frac{-\overline{\lambda}_j}{\overline{\lambda}_i} (\overline{\mathcal{B}}(\Delta x)_{ij})^2 \\ &= \langle \Delta x, \nabla^2_{xx} L(\overline{x}, \overline{\xi}, \overline{\Gamma})\Delta x\rangle - \sigma(\overline{\Gamma}, \mathcal{T}^2_{\mathbb{S}^n_+}(g(\overline{x}), g'(\overline{x})\Delta x)). \end{aligned} \tag{43}$$

Meanwhile, by (41) and (39), we know from (31) that $\Delta x \in \mathrm{appl}(\overline{x}, \overline{\xi}, \overline{\Gamma})$. Since the W-SOC holds at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$, we know from (43) that $\Delta x = 0$. It then follows from (42) that $\widetilde{\Delta\Gamma}_{\alpha\gamma} = 0$. This, together with (40) and (38), yields that

$$\begin{cases} h'(\overline{x})^*\Delta\xi + \overline{\mathcal{B}}^*\widetilde{\Delta\Gamma} = 0, \\ \widetilde{\Delta\Gamma}_{\alpha\alpha} = 0, \quad \widetilde{\Delta\Gamma}_{\alpha\beta} = 0, \quad \widetilde{\Delta\Gamma}_{\alpha\gamma} = 0. \end{cases}$$

Since the constraint nondegeneracy hold at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$, we know from (33) in Lemma 1 (ii) that $(\Delta\xi, \Delta\Gamma) = 0$, which contradicts with $(\Delta x, \Delta\xi, \Delta\Gamma) \neq 0$. The proof is then completed. □

*Remark 3* It should be noted that Theorem 1 can be easiliy extended to cases where $K$ in (1) is a Cartesian product of a finite number of positive semidefinite cones and zero vectors. By regarding a polyhedron as a Cartesian product of a finite number of one dimensional positive semidefinite cones and zero vectors, the results derived in Theorem 1 are consistent with the findings presented in [8, P644].

It is natural to question whether the W-SRCQ and W-SOC are sufficient for the existence of nonsingular elements in the B-subdifferential $\partial_B F(\overline{x}, \overline{\xi}, \overline{\Gamma})$. We address this question with the following proposition, which confirms the sufficiency under the condition that $|\beta| \leq 1$, i.e., the number of zero eigenvalues of $g(\overline{x}) + \overline{\Gamma}$ is less than or equal to one.

**Proposition 3** *Let $\overline{x} \in \mathbb{X}$ be a stationary point of (21) with a Lagrange multiplier $(\overline{\xi}, \overline{\Gamma}) \in M(\overline{x})$. Suppose that $|\beta| \leq 1$. If the W-SOC and W-SRCQ hold at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$, then at least one of $\{\overline{\mathcal{U}}_0, \overline{\mathcal{U}}_{\mathcal{I}}\}$ is nonsingular.*

*Proof* Consider the following two cases.
**Case 1**: $|\beta| = 0$. In this case, from Lemma 1, we know that the W-SRCQ is equivalent to the constraint nondegeneracy. Moreover, (30) and (31) imply that $\mathrm{app}(\overline{x}, \overline{\xi}, \overline{\Gamma}) = \mathrm{appl}(\overline{x}, \overline{\xi}, \overline{\Gamma})$, and thus further imply that the W-SOC is equivalent to the S-SOSC. We also have from Proposition 2 that $\partial_B F(\overline{x}, \overline{\xi}, \overline{\Gamma}) = \{F'(\overline{x}, \overline{\xi}, \overline{\Gamma})\} = \{\overline{\mathcal{U}}_0\} = \{\overline{\mathcal{U}}_{\mathcal{I}}\}$. The desired result follows directly from Theorem 1.
**Case 2**: $|\beta| = 1$. Since the W-SOC holds at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$, we know from (i) of Theorem 1 that if the constraint nondegeneracy hold at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$, then $\overline{\mathcal{U}}_0$ is nonsingular. Suppose that the constraint nondegeneracy does not hold at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$. Since the W-SRCQ holds at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$, we know from Lemma 1 that there exists $(\Delta\xi, \Delta\Gamma) \neq 0$ such that

$$\begin{cases} h'(\overline{x})^*\Delta\xi + \overline{\mathcal{B}}^*\Delta\Gamma = 0, \\ \Delta\Gamma_{\alpha\alpha} = 0, \ \Delta\Gamma_{\alpha\beta} = 0, \ \Delta\Gamma_{\alpha\gamma} = 0, \\ \Delta\Gamma_{\beta\beta} \neq 0. \end{cases}$$

Next, we show that $\mathrm{app}(\overline{x}, \overline{\xi}, \overline{\Gamma}) = \mathrm{appl}(\overline{x}, \overline{\xi}, \overline{\Gamma})$ with their explicit expressions given in (30) and (31), respectively. Clearly, $\mathrm{appl}(\overline{x}, \overline{\xi}, \overline{\Gamma}) \subseteq \mathrm{app}(\overline{x}, \overline{\xi}, \overline{\Gamma})$. For any $d \in \mathrm{app}(\overline{x}, \overline{\xi}, \overline{\Gamma})$, we have

$$0 = \langle d, h'(\overline{x})^*\Delta\xi + \overline{\mathcal{B}}^*\Delta\Gamma\rangle = \langle h'(\overline{x})d, \Delta\xi\rangle + \langle \overline{\mathcal{B}}(d), \Delta\Gamma\rangle = \langle \overline{\mathcal{B}}(d)_{\beta\beta}, \Delta\Gamma_{\beta\beta}\rangle, \tag{44}$$

where the last equation follows from (30). Since $|\beta| = 1$, by noting that $\Delta\Gamma_{\beta\beta} \neq 0$, we know that $\overline{\mathcal{B}}(d)_{\beta\beta} = 0$. Therefore, we obtain from (31) that $\mathrm{appl}(\overline{x}, \overline{\xi}, \overline{\Gamma}) \subseteq \mathrm{app}(\overline{x}, \overline{\xi}, \overline{\Gamma})$. Then, by Definitions 8 and 6, the W-SOC and S-SOSC are equivalent. Therefore, we know from (ii) of Theorem 1 that $\overline{\mathcal{U}}_{\mathcal{I}}$ is nonsingular. The proof is then completed. □

*Remark 4* It is worth noting that the condition $|\beta| \leq 1$ is always satisfied for classic nonlinear programming (NLP). Actually, as we mentioned before, the positive orthant $\mathbb{R}^n_+$ can be represented as the Cartesian product of $n$ one-dimensional positive semidefinite cones. Consequently, based on Proposition 3, the existence of a nonsingular element in the Bouligand subdifferential $\partial_B F(\overline{x}, \overline{\xi}, \overline{\Gamma})$ is ensured under the W-SOC and the W-SRCQ for the conventional NLP, which recovers the corresponding result obtained by Izmailov and Solodov [8, Propostion 6].

However, the following simple example illustrates that for the NLSDP (21), if the number of zero eigenvalues of $g(\overline{x}) + \overline{\Gamma}$ is greater than one, i.e., $|\beta| > 1$, the W-SRCQ and W-SOC are not sufficient to guarantee the existence of nonsingular elements in $\partial_B F(\overline{x}, \overline{\xi}, \overline{\Gamma})$. This showcases the fundamental differences between the NLSDP and the NLP.

*Example 2* Consider the following SDP problem:

$$\begin{aligned} \min \quad & 0 \\ \text{s.t.} \quad & x_{12} = 0, \\ & X \in \mathbb{S}^2_+, \end{aligned}$$

where $X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \in \mathbb{S}^2$. It is clear that $(\overline{X}, \overline{\xi}, \overline{\Gamma}) = (0, 0, 0)$ is a KKT point. We have $\beta = \{1, 2\}$ and $\alpha = \gamma = \emptyset$. It is easy to verify that the W-SRCQ and W-SOC hold at $(\overline{X}, \overline{\xi}, \overline{\Gamma})$. Moreover, it follows from [17, Lemma 11] that $\mathcal{U} \in \partial_B F(\overline{X}, \overline{\xi}, \overline{\Gamma})$ if and only if there exists $\Omega \in R$ such that for any $(\Delta X, \Delta\xi, \Delta\Gamma) \in \mathbb{S}^2 \times \mathbb{R} \times \mathbb{S}^2$,

$$\mathcal{U}(\Delta X, \Delta\xi, \Delta\Gamma) = \begin{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & \Delta\xi \\ \Delta\xi & 0 \end{bmatrix} + \Delta\Gamma \\ \Delta X_{12} \\ -\Delta X + \Omega \circ (\Delta X + \Delta\Gamma) \end{bmatrix}, \tag{45}$$

where $R \in \mathbb{S}^n$ is the set of matrices defined by

$$R := \left\{ \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \right\} \bigcup \left\{ \begin{bmatrix} 0 & t \\ t & 1 \end{bmatrix} \mid t \in [0,1] \right\} \bigcup \left\{ \begin{bmatrix} 1 & t \\ t & 0 \end{bmatrix} \mid t \in [0,1] \right\}.$$

It is easy to see that all elements in $\partial_B F(\overline{X}, \overline{\xi}, \overline{\Gamma})$ are singular. Indeed, let $\mathcal{U} \in \partial_B F(\overline{X}, \overline{\xi}, \overline{\Gamma})$ be arbitrarily given. Then,

if $\Omega = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, then $\Delta X = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, $\Delta\xi = -1$ and $\Delta\Gamma = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is a nonzero solution of $\mathcal{U}(\Delta X, \Delta\xi, \Delta\Gamma) = 0$;

if $\Omega = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, then $\Delta X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\Delta\xi = 0$ and $\Delta\Gamma = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ is a nonzero solution of $\mathcal{U}(\Delta X, \Delta\xi, \Delta\Gamma) = 0$;

if $\Omega = \begin{bmatrix} 0 & t \\ t & 1 \end{bmatrix}$ with $t \in [0,1]$, then $\Delta X = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $\Delta\xi = 0$ and $\Delta\Gamma = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ is a nonzero solution of $\mathcal{U}(\Delta X, \Delta\xi, \Delta\Gamma) = 0$;

if $\Omega = \begin{bmatrix} 1 & t \\ t & 0 \end{bmatrix}$ with $t \in [0,1]$, then $\Delta X = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, $\Delta\xi = 0$ and $\Delta\Gamma = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ is a nonzero solution of $\mathcal{U}(\Delta X, \Delta\xi, \Delta\Gamma) = 0$.

Next, let us consider the Clarke-subdifferential defined by (5) of the KKT nonsmooth mapping (22). Interestingly, one may observe that for Example 2, the convex combination $t\mathcal{U}_0 + (1-t)\mathcal{U}_{\mathcal{I}} \in \partial_C F(\overline{X}, \overline{\xi}, \overline{\Gamma})$ for any $t \in (0,1)$ is actually nonsingular, where $\mathcal{U}_0$ and $\mathcal{U}_{\mathcal{I}}$ are elements in $\partial_B F(\overline{X}, \overline{\xi}, \overline{\Gamma})$ defined by (45) with $\Omega = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, respectively. In fact, we have the following general results, which demonstrates that the W-SOC and W-SRCQ guarantee the existence of a nonsingular element in the Clarke-subdifferential.

**Proposition 4** *Let $\overline{x} \in \mathbb{X}$ be a stationary point of* (21) *with a Lagrange multiplier $(\overline{\xi}, \overline{\Gamma}) \in M(\overline{x})$. Suppose that the W-SOC and W-SRCQ hold at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$. Let $\overline{\mathcal{U}}_0$ and $\overline{\mathcal{U}}_{\mathcal{I}}$ be two elements defined in* (36) *and* (37)*, respectively.*

*(i) There exists $t \in (0,1)$ such that*

$$t\overline{\mathcal{U}}_0 + (1-t)\overline{\mathcal{U}}_{\mathcal{I}} \in \partial_C F(\overline{x}, \overline{\xi}, \overline{\Gamma}) \text{ is nonsingular.}$$

*(ii) If, additionally, the NLSDP* (21) *is convex, that is, it becomes a CSDP, then for any $t \in (0,1)$,*

$$t\overline{\mathcal{U}}_0 + (1-t)\overline{\mathcal{U}}_{\mathcal{I}} \in \partial_C F(\overline{x}, \overline{\xi}, \overline{\Gamma}) \text{ is nonsingular.}$$

*Proof* Let $A = g(\overline{x}) + \overline{\Gamma}$ satisfy the eigenvalue decomposition (23) and $\overline{P} \in \mathcal{O}^n(A)$ with the index sets $\alpha$, $\beta$ and $\gamma$ defined by (12). Recall the linear operator $\overline{\mathcal{B}} : \mathbb{X} \to \mathbb{S}^n$ defined in (28).
**(i)** Assume that for any $t \in (0,1)$, $t\overline{\mathcal{U}}_0 + (1-t)\overline{\mathcal{U}}_{\mathcal{I}} \in \partial_C F(\overline{x}, \overline{\xi}, \overline{\Gamma})$ is singular. Define a sequence of elements $\{\overline{\mathcal{U}}^k\}$ as $\overline{\mathcal{U}}^k = \frac{k}{k+1}\overline{\mathcal{U}}_0 + \frac{1}{k+1}\overline{\mathcal{U}}_{\mathcal{I}} \in \partial_C F(\overline{x}, \overline{\xi}, \overline{\Gamma})$ for each $k$. Since $\overline{\mathcal{U}}^k$ is singular, we know that for each $k$, there are $0 \neq (\Delta x^k, \Delta\xi^k, \Delta\Gamma^k) \in \mathbb{Z}$ such that $\overline{\mathcal{U}}^k(\Delta x^k, \Delta\xi^k, \Delta\Gamma^k) = 0$.

Thus, for each $k$, we know from (35) and Proposition 2 that

$$\nabla^2_{xx}L(\overline{x},\overline{\xi},\overline{\Gamma})\Delta x^k + h'(\overline{x})^*\Delta\xi^k + g'(\overline{x})^*\Delta\Gamma^k = 0, \tag{46}$$

$$h'(\overline{x})\Delta x^k = 0, \tag{47}$$

$$\widetilde{\Delta\Gamma}^k_{\alpha\alpha} = 0, \quad \widetilde{\Delta\Gamma}^k_{\alpha\beta} = 0, \tag{48}$$

$$\overline{\mathcal{B}}(\Delta x^k)_{\beta\gamma} = 0, \quad \overline{\mathcal{B}}(\Delta x^k)_{\gamma\gamma} = 0, \tag{49}$$

$$(E_{\alpha\gamma} - \overline{\Xi}_{\alpha\gamma}) \circ \overline{\mathcal{B}}(\Delta x^k)_{\alpha\gamma} - \overline{\Xi}_{\alpha\gamma} \circ \widetilde{\Delta\Gamma}^k_{\alpha\gamma} = 0, \tag{50}$$

$$k\overline{\mathcal{B}}(\Delta x^k)_{\beta\beta} - \widetilde{\Delta\Gamma}^k_{\beta\beta} = 0, \tag{51}$$

where $\widetilde{\Delta\Gamma}^k = \overline{P}^T\Delta\Gamma^k\overline{P}$ with $\overline{P} \in \mathcal{O}^n(A)$ and $A = g(\overline{x}) + \overline{\Gamma}$, the matrix $\overline{\Xi} \in \mathbb{S}^n$ is given by (15) and $E \in \mathbb{S}^n$ denotes the matrix whose elements are all ones. We then conclude that for each $k$, $\Delta x^k \neq 0$. Indeed, if there exists some $\bar{k}$ such that $\Delta x^{\bar{k}} = 0$, then by (48), (50) and (51), we obtain that

$$\widetilde{\Delta\Gamma}^{\bar{k}}_{\alpha\alpha} = 0, \quad \widetilde{\Delta\Gamma}^{\bar{k}}_{\alpha\beta} = 0, \quad \widetilde{\Delta\Gamma}^{\bar{k}}_{\beta\beta} = 0 \quad \text{and} \quad \widetilde{\Delta\Gamma}^{\bar{k}}_{\alpha\gamma} = 0. \tag{52}$$

Thus, combining with (46), we know that

$$\begin{cases} h'(\overline{x})^*\Delta\xi^{\bar{k}} + \overline{\mathcal{B}}^*\Delta\widetilde{\Gamma}^{\bar{k}} = 0, \\ \widetilde{\Delta\Gamma}^{\bar{k}}_{\alpha\alpha} = 0, \quad \widetilde{\Delta\Gamma}^{\bar{k}}_{\alpha\beta} = 0, \quad \widetilde{\Delta\Gamma}^{\bar{k}}_{\beta\beta} = 0, \quad \widetilde{\Delta\Gamma}^{\bar{k}}_{\alpha\gamma} = 0. \end{cases} \tag{53}$$

It then follows from (32) in Lemma 1 (i) that $(\Delta\xi^{\bar{k}}, \Delta\Gamma^{\bar{k}}) = 0$, which, together with $\Delta x^{\bar{k}} = 0$, contradicts with the assumption. Now, since $\Delta x^k \neq 0$ for all $k$, without loss of generality, we may assume that $\|\Delta x^k\| = 1$ and $\Delta x^k \to \Delta x^\infty \neq 0$ as $k \to \infty$. By taking the inner product with $\Delta x^k$ in (46), we obtain from (47)-(51) that for each $k$,

$$\langle \Delta x^k, \nabla^2_{xx}L(\overline{x},\overline{\xi},\overline{\Gamma})\Delta x^k\rangle + 2\sum_{i\in\alpha}\sum_{j\in\gamma}\frac{-\overline{\lambda}_j}{\overline{\lambda}_i}\overline{\mathcal{B}}(\Delta x^k)^2_{ij} + \sum_{i\in\beta}\sum_{j\in\beta}k\overline{\mathcal{B}}(\Delta x^k)^2_{ij} = 0. \tag{54}$$

Let $\mathcal{Q} : \mathbb{X} \to \mathbb{X}$ be the linear map such that

$$\langle d, \mathcal{Q}(d)\rangle = \langle d, \nabla^2_{xx}L(\overline{x},\overline{\xi},\overline{\Gamma})d\rangle + 2\sum_{i\in\alpha}\sum_{j\in\gamma}\frac{-\overline{\lambda}_j}{\overline{\lambda}_i}\overline{\mathcal{B}}(d)^2_{ij}, \quad \forall\, d \in \mathbb{X}.$$

Then there exists $q > 0$ that $|\mathcal{Q}(d)| \leq q\|d\|^2$ for any $d \in \mathbb{X}$. Therefore, we have from (54) that for each $k$,

$$|\overline{\mathcal{B}}(\Delta x^k)_{ij}|^2 \leq \frac{1}{k}q, \quad (i,j) \in \beta\times\beta.$$

As a result, $\overline{\mathcal{B}}(\Delta x^\infty)_{\beta\beta} = 0$, which, together with (49), implies $\Delta x^\infty \in \operatorname{appl}(\overline{x},\overline{\xi},\overline{\Gamma})$. Further note that (54) implies that

$$\langle \Delta x^k, \mathcal{Q}\Delta x^k\rangle \leq 0.$$

Taking the limit as $k \to +\infty$, we obtain $\langle \Delta x^\infty, \mathcal{Q}\Delta x^\infty\rangle \leq 0$, which contradicts the W-SOC given that $0 \neq \Delta x^\infty \in \operatorname{appl}(\overline{x},\overline{\xi},\overline{\Gamma})$. Therefore, there exists a nonsingular $\overline{\mathcal{U}}^k$.

**(ii)** For the CSDP, if there is $t \in (0,1)$ that $\overline{\mathcal{U}}_t = t\overline{\mathcal{U}}_0 + (1-t)\overline{\mathcal{U}}_{\mathcal{I}}$ is singular. Then, there is $(\Delta x, \Delta\xi, \Delta\Gamma) \neq 0$ such that $\overline{\mathcal{U}}_t(\Delta x, \Delta\xi, \Delta\Gamma) = 0$. Similarly as the proof of (i), we obtain from (52)-(54) that $\Delta x \neq 0$ and

$$\langle \Delta x, \nabla^2_{xx} L(\overline{x}, \overline{\xi}, \overline{\Gamma})\Delta x\rangle + 2\sum_{i\in\alpha}\sum_{j\in\gamma} \frac{-\overline{\lambda}_j}{\overline{\lambda}_i}\overline{\mathcal{B}}(\Delta x)^2_{ij} + \sum_{i\in\beta}\sum_{j\in\beta}\frac{t}{1-t}\overline{\mathcal{B}}(\Delta x)^2_{ij} = 0. \tag{55}$$

By convexity, we know that $\nabla^2_{xx} L(\overline{x}, \overline{\xi}, \overline{\Gamma})$ is positive semidefinite. Thus, $\overline{\mathcal{B}}(\Delta x)_{\beta\beta} = 0$. This, together with (47) and (49), yields $0 \neq \Delta x \in \mathrm{appl}(\overline{x}, \overline{\xi}, \overline{\Gamma})$, and futher implies the contradiction between (55) the W-SOC. Therefore, we conclude that $\overline{\mathcal{U}}_t = t\overline{\mathcal{U}}_0 + (1-t)\overline{\mathcal{U}}_{\mathcal{I}}$ is nonsingular for any $t \in (0,1)$. □

*Remark 5* According to the proof of part (i) of Proposition 4, there exists only a finite number of singular elements on the line segment connecting $\overline{\mathcal{U}}_0$ and $\overline{\mathcal{U}}_{\mathcal{I}}$. In fact, by considering the matrix representations of $\overline{\mathcal{U}}_0$ and $\overline{\mathcal{U}}_{\mathcal{I}}$ and invoking the fundamental theorem of algebra, we are able to deduce that the non-Zero polynomial $\det(t\overline{\mathcal{U}}_0 + (1-t)\overline{\mathcal{U}}_{\mathcal{I}})$ has a number of roots that does not exceed its degree.

## 3.2 Necessary conditions for the existence of nonsingularity elements in subdifferentials

We first show that when the NLSDP (21) is convex, i.e., for the CSDP, the conditions of Theorem 1 are actually necessary. In fact, we have the following characterizations of the nonsingularity of $\overline{\mathcal{U}}_0$ and $\overline{\mathcal{U}}_{\mathcal{I}}$ defined by (36) and (37), respectively.

**Proposition 5** *Let $\overline{x} \in \mathbb{X}$ be a stationary point of the CSDP with a Lagrange multiplier $(\overline{\xi}, \overline{\Gamma}) \in M(\overline{x})$.*

*(i) The W-SOC and constraint nondegeneracy hold at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$ if and only if $\overline{\mathcal{U}}_0$ defined in (36) is nonsingular;*

*(ii) the S-SOSC and W-SRCQ hold at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$ if and only if $\overline{\mathcal{U}}_{\mathcal{I}}$ defined in (37) is nonsingular.*

*Proof* Let $A = g(\overline{x}) + \overline{\Gamma}$ satisfy the eigenvalue decomposition (23) and $\overline{P} \in \mathcal{O}^n(A)$ with the index sets $\alpha$, $\beta$ and $\gamma$ defined by (12). Recall the linear operator $\overline{\mathcal{B}} : \mathbb{X} \to \mathbb{S}^n$ defined by (28). We only show (i), as (ii) can be established through a similar argument.

By Theorem 1, we only need to show that the nonsingularity of $\overline{\mathcal{U}}_0$ implies the W-SOC and constraint nondegeneracy. Suppose $\overline{\mathcal{U}}_0(\Delta x, \Delta\xi, \Delta\Gamma) = 0$. We know from (36) and Proposition 2 that

$$\nabla^2_{xx} L(\overline{x}, \overline{\xi}, \overline{\Gamma})\Delta x + h'(\overline{x})^*\Delta\xi + g'(\overline{x})^*\Delta\Gamma = 0, \tag{56}$$

$$h'(\overline{x})\Delta x = 0, \tag{57}$$

$$\widetilde{\Delta\Gamma}_{\alpha\alpha} = 0, \quad \widetilde{\Delta\Gamma}_{\alpha\beta} = 0, \tag{58}$$

$$\overline{\mathcal{B}}(\Delta x)_{\beta\beta} = 0, \quad \overline{\mathcal{B}}(\Delta x)_{\beta\gamma} = 0, \quad \overline{\mathcal{B}}(\Delta x)_{\gamma\gamma} = 0, \tag{59}$$

$$(E_{\alpha\gamma} - \overline{\Xi}_{\alpha\gamma}) \circ \overline{\mathcal{B}}(\Delta x)_{\alpha\gamma} - \overline{\Xi}_{\alpha\gamma} \circ \widetilde{\Delta\Gamma}_{\alpha\gamma} = 0, \tag{60}$$

where $\widetilde{\Delta\Gamma} = \overline{P}^T \Delta\Gamma \overline{P}$, $\overline{\Xi}$ is defined in (15) for $A$ and $E$ represents the matrix consisting of all ones in $\mathbb{S}^n$. Suppose that the constraint nondegeneracy does not hold. Then, by (ii) of Lemma 1, we know that there exist $0 \neq (\Delta\xi, \Delta\Gamma)$ such that

$$\begin{cases} h'(\overline{x})^*\Delta\xi + \overline{\mathcal{B}}^*\widetilde{\Delta\Gamma} = 0, \\ \widetilde{\Delta\Gamma}_{\alpha\alpha} = 0, \quad \widetilde{\Delta\Gamma}_{\alpha\beta} = 0, \quad \widetilde{\Delta\Gamma}_{\alpha\gamma} = 0. \end{cases}$$

Thus, it follows from (56)-(60) that $(\Delta x, \Delta\xi, \Delta\Gamma) = (0, \Delta\xi, \Delta\Gamma) \neq 0$ satisfy $\overline{\mathcal{U}}_0(\Delta x, \Delta\xi, \Delta\Gamma) = 0$, which contradicts with the nonsingularity of $\overline{\mathcal{U}}_0$. Thus, the constraint nondegeneracy holds at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$.

On the other hand, suppose that the W-SOC does not hold at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$. Then, there exists $d \neq 0$ that

$$\begin{cases} \langle d, \nabla^2_{xx} L(\overline{x}, \overline{\xi}, \overline{\Gamma})d\rangle + 2\sum_{i\in\alpha}\sum_{j\in\gamma} \frac{-\overline{\lambda}_j}{\overline{\lambda}_i}\overline{\mathcal{B}}(d)^2_{ij} = 0, \\ h'(\overline{x})d = 0, \\ \overline{\mathcal{B}}(d)_{\beta\beta} = 0, \quad \overline{\mathcal{B}}(d)_{\beta\gamma} = 0, \quad \overline{\mathcal{B}}(d)_{\gamma\gamma} = 0. \end{cases}$$

By convexity, we know that $\nabla^2_{xx} L(\overline{x}, \overline{\xi}, \overline{\Gamma})$ is positive semidefinite. Together with $\overline{\lambda}_j < 0$ for $j \in \gamma$ and $\overline{\lambda}_i > 0$ for $i \in \alpha$, we conclude that

$$\begin{cases} \nabla^2_{xx} L(\overline{x}, \overline{\xi}, \overline{\Gamma})d = 0, \\ h'(\overline{x})d = 0, \\ \overline{\mathcal{B}}(d)_{\alpha\gamma} = 0, \quad \overline{\mathcal{B}}(d)_{\beta\beta} = 0, \quad \overline{\mathcal{B}}(d)_{\beta\gamma} = 0, \quad \overline{\mathcal{B}}(d)_{\gamma\gamma} = 0, \end{cases} \tag{61}$$

This, yields that $\overline{\mathcal{U}}_0(d, 0, 0) = 0$, which again contradicts with the nonsingularity of $\overline{\mathcal{U}}_0$. The proof of part (i) has been established. □

*Remark 6* In the context of linear semidefinite programming, a particular instance of the CSDP, Chan and Sun [2, Proposition 17] demonstrated that the nonsingularity of $\overline{\mathcal{U}}_0$ ($\overline{\mathcal{U}}_{\mathcal{I}}$) implies primal (dual) constraint nondegeneracy. In contrast, Proposition 5 in the current work provides a complete characterization of the nonsingularity of $\overline{\mathcal{U}}_0$ and $\overline{\mathcal{U}}_{\mathcal{I}}$ for more general CSDP.

*Remark 7* For the general NLSDP, one can still use the first part of the proof to show that the the nonsingularity of $\overline{\mathcal{U}}_0$ (or $\overline{\mathcal{U}}_{\mathcal{I}}$) implies the constraint nondegeneracy (or the W-SRCQ).

Finally, we shall establish in the forthcoming proposition that, for general NLSDP, the W-SRCQ is necessary to ensure the existence of nonsingular elements in the Clarke subdifferential. Additionally, for the CSDP, both the W-SRCQ and the W-SOC are requisite.

**Proposition 6** *Let $\overline{x} \in \mathbb{X}$ be a stationary point of the NLSDP* (21) *with a Lagrange multiplier $(\overline{\xi}, \overline{\Gamma}) \in M(\overline{x})$.*

*(i) If there is a nonsingular element $\overline{\mathcal{U}} \in \partial_C F(\overline{x}, \overline{\xi}, \overline{\Gamma})$, then the W-SRCQ holds at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$.*
*(ii) For the CSDP, if there is a nonsingular element $\overline{\mathcal{U}} \in \partial_C F(\overline{x}, \overline{\xi}, \overline{\Gamma})$, then the W-SRCQ and W-SOC hold at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$.*

*Proof* Let $A = g(\overline{x}) + \overline{\Gamma}$ satisfy the eigenvalue decomposition (23) and $\overline{P} \in \mathcal{O}^n(A)$ with the index sets $\alpha$, $\beta$ and $\gamma$ defined by (12). Recall the linear operator $\overline{\mathcal{B}} : \mathbb{X} \to \mathbb{S}^n$ defined by (28).
**(i)** Let $(\Delta\xi, \Delta\Gamma) \in \mathbb{R}^m \times \mathbb{S}^n$ be a solution to the following system:

$$\begin{cases} h'(\overline{x})^*\Delta\xi + \overline{\mathcal{B}}^*\widetilde{\Delta\Gamma} = 0, \\ \widetilde{\Delta\Gamma}_{\beta\beta} = 0, \quad \widetilde{\Delta\Gamma}_{\alpha\alpha} = 0, \quad \widetilde{\Delta\Gamma}_{\alpha\beta} = 0, \quad \widetilde{\Delta\Gamma}_{\alpha\gamma} = 0, \end{cases}$$

where $\widetilde{\Delta\Gamma} = \overline{P}^T\Delta\Gamma\overline{P}$. Then, we know from (35) and Proposition 2 that

$$\overline{\mathcal{U}}(0, \Delta\xi, \Delta\Gamma) = \begin{bmatrix} h'(\overline{x})^*\Delta\xi + \overline{\mathcal{B}}^*\widetilde{\Delta\Gamma} \\ 0 \\ \overline{P}\begin{bmatrix} \widetilde{\Delta\Gamma}_{\alpha\alpha} & \widetilde{\Delta\Gamma}_{\alpha\beta} & \Xi_{\alpha\gamma} \circ \widetilde{\Delta\Gamma}_{\alpha\gamma} \\ \widetilde{\Delta\Gamma}_{\beta\alpha} & \mathcal{W}(\widetilde{\Delta\Gamma}_{\beta\beta}) & 0 \\ \Xi_{\gamma\alpha} \circ \widetilde{\Delta\Gamma}_{\gamma\alpha} & 0 & 0 \end{bmatrix}\overline{P}^T \end{bmatrix} = 0,$$

where $\mathcal{W} \in \partial_C \Pi_{\mathbb{S}_+^{|\beta|}}(0)$. Since $\overline{\mathcal{U}}$ is nonsingular, we obtain that $(\Delta\xi, \Delta\Gamma) = 0$. Therefore, we know from (32) in Lemma 1 (i) that the W-SRCQ holds at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$.

**(ii)** We only need to show that the W-SOC (Definition (8)) holds at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$. Suppose on the contrary that the W-SOC does not hold. Then, we know from (34) and (31) that there exists $0 \neq \Delta x$ such that

$$\begin{cases} \langle \Delta x, \nabla_{xx}^2 L(\overline{x}, \overline{\xi}, \overline{\Gamma})\Delta x\rangle + 2\sum_{i\in\alpha}\sum_{j\in\beta} \dfrac{-\overline{\lambda}_j}{\overline{\lambda}_i}\overline{\mathcal{B}}(\Delta x)_{ij}^2 = 0, \\ h'(\overline{x})\Delta x = 0, \\ \overline{\mathcal{B}}(\Delta x)_{\beta\beta} = 0, \quad \overline{\mathcal{B}}(\Delta x)_{\beta\gamma} = 0, \quad \overline{\mathcal{B}}(\Delta x)_{\gamma\gamma} = 0. \end{cases}$$

Since $\nabla_{xx}^2 L(\overline{x}, \overline{\xi}, \overline{\Gamma})$ is positive semidefinite and $\dfrac{-\overline{\lambda}_j}{\overline{\lambda}_i} > 0$ for any $i \in \alpha$ and $j \in \gamma$, we have

$$\begin{cases} \nabla_{xx}^2 L(\overline{x}, \overline{\xi}, \overline{\Gamma})\Delta x = 0 \\ h'(\overline{x})\Delta x = 0, \\ \overline{\mathcal{B}}(\Delta x)_{\beta\beta} = 0, \quad \overline{\mathcal{B}}(\Delta x)_{\beta\gamma} = 0, \quad \overline{\mathcal{B}}(\Delta x)_{\gamma\gamma} = 0, \quad \overline{\mathcal{B}}(\Delta x)_{\alpha\gamma} = 0. \end{cases}$$

This, together with (35) and Proposition 2, implies that

$$\overline{\mathcal{U}}(\Delta x, 0, 0) = \begin{bmatrix} \nabla_{xx}^2 L(\overline{x}, \overline{\xi}, \overline{\Gamma})\Delta x \\ h'(\overline{x})\Delta x \\ -\overline{\mathcal{B}}\Delta x + \begin{bmatrix} (\overline{\mathcal{B}}\Delta x)_{\alpha\alpha} & (\overline{\mathcal{B}}\Delta x)_{\alpha\beta} & \overline{\Xi}_{\alpha\gamma} \circ (\overline{\mathcal{B}}\Delta x)_{\alpha\gamma} \\ (\overline{\mathcal{B}}\Delta x)_{\beta\alpha} & W((\overline{\mathcal{B}}\Delta x)_{\beta\beta}) & 0 \\ \overline{\Xi}_{\gamma\alpha} \circ (\overline{\mathcal{B}}\Delta x)_{\gamma\alpha} & 0 & 0 \end{bmatrix} \end{bmatrix} = 0,$$

which contricts with the nonsingularity of $\overline{\mathcal{U}}$. The proof is then completed. □

## 4 The primal-dual characterizations for convex quadratic semidefinite programming

In this section, we shall study the primal-dual characterizations of the W-SRCQ (Definition 7) and W-SOC (Definition 8) for the following convex quadratic semidefinite programming (QSDP):

$$\begin{array}{ll} \min & \frac{1}{2}\langle x, \mathcal{Q}x\rangle + \langle c, x\rangle \\ \text{s.t.} & \mathcal{H}x - p = 0, \\ & \mathcal{G}x - q \in \mathbb{S}_+^n, \end{array} \tag{62}$$

where $c \in \mathbb{X}$, $p \in \mathbb{R}^m$, $q \in \mathbb{S}^n$, the self-adjoint linear operator $\mathcal{Q} : \mathbb{X} \to \mathbb{X}$ is positive semi-definite, and $\mathcal{H} : \mathbb{X} \to \mathbb{R}^m$ and $\mathcal{G} : \mathbb{X} \to \mathbb{S}^n$ are two given linear operators. The dual problem (in the restricted Wolfe sense [31,13]) of the QSDP (62) takes the form:

$$\begin{array}{ll} \min\limits_{\omega\in\mathbb{W}, \xi\in\mathbb{R}^m, \Gamma\in\mathbb{S}^n} & \frac{1}{2}\langle \omega, \mathcal{Q}\omega\rangle + \langle p, \xi\rangle + \langle q, \Gamma\rangle \\ \text{s.t.} & -\mathcal{Q}\omega - c - \mathcal{H}^*\xi - \mathcal{G}^*\Gamma = 0, \\ & \Gamma \in \mathbb{S}_-^n, \end{array} \tag{63}$$

where $\mathbb{W} \subseteq \mathbb{X}$ is a linear subspace containing $\mathrm{Im}(\mathcal{Q})$. Thus, the KKT system of (63) is given by

$$\begin{cases} \begin{bmatrix} \mathcal{Q}\omega \\ p \\ q \end{bmatrix} + \begin{bmatrix} -\mathcal{Q} \\ -\mathcal{H} \\ -\mathcal{G} \end{bmatrix} \tau + \begin{bmatrix} 0 \\ 0 \\ \mathcal{I} \end{bmatrix} \Upsilon = 0, \\ \mathcal{Q}\omega + c + \mathcal{H}^*\xi + \mathcal{G}^*\Gamma = 0, \\ \Upsilon \in \mathcal{N}_{\mathbb{S}^n_-}(\Gamma), \\ \omega \in \mathbb{W}, \quad \xi \in \mathbb{R}^m, \quad \Gamma \in \mathbb{S}^n, \quad \tau \in \mathbb{X}, \quad \Upsilon \in \mathbb{S}^n. \end{cases} \tag{64}$$

**Proposition 7** *Consider the convex quadratic semidefinite programming* (62) *and its dual problem* (63)*.*

*(i) Suppose that $\mathbb{W} = \mathbb{X}$. Let $\overline{x} \in \mathbb{X}$ be a stationary point of the convex quadratic semidefinite programming* (62) *with a Lagrange multiplier $(\overline{\xi}, \overline{\Gamma}) \in M(\overline{x})$. The primal W-SOC holds at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$ if and only if the dual W-SRCQ holds at $(\overline{x}, \overline{\xi}, \overline{\Gamma}, \overline{x}, \mathcal{G}\overline{x} - q)$.*

*(ii) Suppose that $\mathbb{W} = \mathrm{Im}(\mathcal{Q})$. Let $(\overline{\omega}, \overline{\xi}, \overline{\Gamma}) \in \mathbb{W} \times \mathbb{R}^m \times \mathbb{S}^n$ be a stationary point of the dual problem* (63) *with a Lagrange multiplier $(\overline{x}, \overline{\Upsilon}) \in M(\overline{\omega}, \overline{\xi}, \overline{\Gamma})$. The dual W-SOC holds at $(\overline{\omega}, \overline{\xi}, \overline{\Gamma}, \overline{x}, \overline{\Upsilon})$ if and only if the primal W-SRCQ holds at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$.*

*Proof* We demonstrate (i) here, as (ii) follows from it with only minor modifications.

Let $A = g(\overline{x}) + \overline{\Gamma} = \mathcal{G}\overline{x} - q + \overline{\Gamma}$ satisfy the eigenvalue decomposition (23) with $\overline{P} \in \mathcal{O}^n(A)$, the index sets $\alpha$, $\beta$ and $\gamma$ defined by (12), and $\overline{\mathcal{B}} : \mathbb{X} \to \mathbb{S}^n$ defined by (28). By the convexity, we know that the primal W-SOC holds at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$ if and only if

$$\begin{cases} \mathcal{Q}d = 0, \quad \mathcal{H}d = 0, \\ \overline{\mathcal{B}}(d)_{\alpha\gamma} = 0, \quad \overline{\mathcal{B}}(d)_{\beta\beta} = 0, \quad \overline{\mathcal{B}}(d)_{\beta\gamma} = 0, \quad \overline{\mathcal{B}}(d)_{\gamma\gamma} = 0 \end{cases} \Longrightarrow \quad d = 0. \tag{65}$$

On the other side, it is obvious that $(\overline{x}, \overline{\xi}, \overline{\Gamma}, \overline{x}, \mathcal{G}\overline{x} - q)$ is a KKT point of (63) by (64). By Definition 7, we know that the W-SRCQ of the dual problem (63) at $(\overline{x}, \overline{\xi}, \overline{\Gamma}, \overline{x}, \mathcal{G}\overline{x} - q)$ takes the form:

$$\begin{bmatrix} \mathcal{Q} & \mathcal{H}^* & \mathcal{G}^* \\ 0 & 0 & \mathcal{I} \end{bmatrix} \begin{bmatrix} \mathbb{X} \\ \mathbb{R}^m \\ \mathbb{S}^n \end{bmatrix} + \begin{bmatrix} \{0\} \\ \mathrm{aff}(\mathcal{T}_{\mathbb{S}^n_-}(\overline{\Gamma}) \cap [\mathcal{G}\overline{x} - q]^\perp) \end{bmatrix} = \begin{bmatrix} \mathbb{X} \\ \mathbb{S}^n \end{bmatrix}. \tag{66}$$

Since $\mathrm{aff}(\mathcal{T}_{\mathbb{S}^n_-}(\overline{\Gamma}) \cap [\mathcal{G}\overline{x} - q]^\perp) = \{\overline{P} B \overline{P}^T \in \mathbb{S}^n \mid B_{\alpha\beta} = 0, \ B_{\alpha\alpha} = 0\}$, we know that (66) is equivalent to

$$\mathrm{Null}\Big( \begin{bmatrix} \mathcal{Q} & 0 \\ \mathcal{H} & 0 \\ \overline{\mathcal{B}} & \mathcal{I} \end{bmatrix} \Big) \bigcap \begin{bmatrix} \mathbb{X} \\ \{B \in \mathbb{S}^n \mid B_{\alpha\beta} = 0, \ B_{\alpha\alpha} = 0\}^\perp \end{bmatrix} = \{0\}. \tag{67}$$

Meanwhile, it is not difficult to observe that (67) and (65) are equivalent. Consequently, the primal W-SOC is satisfied at $(\overline{x}, \overline{\xi}, \overline{\Gamma})$ if and only if the dual W-SRCQ is satisfied at $(\overline{x}, \overline{\xi}, \overline{\Gamma}, \overline{x}, \mathcal{G}\overline{x} - q)$. This completes the proof. □

*Remark 8* By employing the similar argument, we are able to obtain the primal-dual characterizations of the S-SOSC (Definition 6) and the constraint nondegeneracy (Definition 4) for the QSDP (62). This extends the corresponding results obtained in Chan and Sun [2] for the linear SDP. Also, for the QSDP (62), a primal-dual characterizations of the SRCQ (Definition 3) and SOSC (Definition 5) is established by [6].

## 5 A semismooth Newton method with correction

In this section, based on the theoretical results obtained in previous sections, we introduce a new semismooth Newton method with a correction step, which enjoys the locally quadratically convergence rate even without the subdifferential regularity.

Denote $\mathbb{Z} := \mathbb{X} \times \mathbb{R}^m \times \mathbb{S}^n$. Let $\delta > 0$ be some given constant. Define the correction mapping $\mathcal{P}_\delta : \mathbb{Z} \to \mathbb{Z}$ by

$$\mathcal{P}_\delta(Z) := \big(x, \xi, \Gamma - \sum_{i\in\theta} \lambda_i(A) P_i P_i^T\big), \quad Z = (x, \xi, \Gamma) \in \mathbb{Z}, \tag{68}$$

where $A := g(x) + \Gamma$ has the eigenvalue decomposition (23) with $P \in \mathcal{O}^n(A)$ and $\theta := \{i \mid |\lambda_i(A)| \le \delta\}$.

Given $Z = (x, \xi, \Gamma) \in \mathbb{Z}$, define two elements $\mathcal{U}_0$ and $\mathcal{U}_\mathcal{I} \in \partial_B F(Z)$ at $Z$ similarly as those in (36) and (37), i.e., for any $(\Delta x, \Delta\xi, \Delta\Gamma) \in \mathbb{Z}$,

$$\mathcal{U}_0(\Delta x, \Delta\xi, \Delta\Gamma) = \begin{bmatrix} \nabla^2_{xx} L(x,\xi,\Gamma)\Delta x + h'(x)^*\Delta\xi + g'(x)^*\Delta\Gamma \\ h'(x)\Delta x \\ -g'(x)\Delta x + \mathcal{V}_0(g'(x)\Delta x + \Delta\Gamma) \end{bmatrix} \tag{69}$$

and

$$\mathcal{U}_\mathcal{I}(\Delta x, \Delta\xi, \Delta\Gamma) = \begin{bmatrix} \nabla^2_{xx} L(x,\xi,\Gamma)\Delta x + h'(x)^*\Delta\xi + g'(x)^*\Delta\Gamma \\ h'(x)\Delta x \\ -g'(x)\Delta x + \mathcal{V}_\mathcal{I}(g'(x)\Delta x + \Delta\Gamma) \end{bmatrix}, \tag{70}$$

where $\mathcal{V}_0$ and $\mathcal{V}_\mathcal{I} \in \partial_B \Pi_{\mathbb{S}^m_+}(A)$ is given by (16) and (17) in Proposition 2, respectively. The following proposition addresses the legitimacy of the correction mapping. It essentially shows that under weaker regularity conditions, when $\widehat{Z}$ is sufficiently close to $\overline{Z}$, the corresponding $\mathcal{U}_0$ (or $\mathcal{U}_\mathcal{I}$) in the B-subdifferential of $F$ at the corrected point $Z = \mathcal{P}_\delta(\widehat{Z})$ with properly chosen $\delta$ is nonsingular and its inverse is uniformly bounded. As one can see later, this will be crucial for the local convergence analysis of the proposed semismooth Newton algorithm.

**Proposition 8** *Let $\overline{x} \in \mathbb{X}$ be a stationary point of the NLSDP* (21) *with a Lagrange multiplier $(\overline{\xi}, \overline{\Gamma}) \in M(\overline{x})$. Let $\delta$ be any fixed constant in* $(0, \min(\overline{\lambda}_\alpha, |\overline{\lambda}_\gamma|))$.

*(i) If the W-SOC (Definition 8) and constraint nondegeneracy (Definition 4) hold at $\overline{Z} = (\overline{x}, \overline{\xi}, \overline{\Gamma}) \in \mathbb{Z}$, then there exists $\kappa_0 > 0$ such that for any $\widehat{Z} = (\widehat{x}, \widehat{\xi}, \widehat{\Gamma})$ sufficient close to $\overline{Z}$ and $Z = \mathcal{P}_\delta(\widehat{Z})$, the corresponding $\mathcal{U}_0 \in \partial_B F(Z)$ defined by* (69) *with respect to $Z$ is nonsingular and*

$$\kappa_0^{-1}\|\Delta Z\| \le \|\mathcal{U}_0(\Delta Z)\| \quad \forall\, \Delta Z \in \mathbb{Z}. \tag{71}$$

*(ii) If the S-SOSC (Definition 6) and W-SRCQ (Definition 7) hold at $\overline{Z} = (\overline{x}, \overline{\xi}, \overline{\Gamma}) \in \mathbb{Z}$, then there exists $\kappa_1 > 0$ such that for any $\widehat{Z} = (\widehat{x}, \widehat{\xi}, \widehat{\Gamma})$ sufficient close to $\overline{Z}$ and $Z = \mathcal{P}_\delta(\widehat{Z})$, the corresponding $\mathcal{U}_\mathcal{I} \in \partial_B F(Z)$ defined by* (70) *with respect to $Z$ is nonsingular and*

$$\kappa_1^{-1}\|\Delta Z\| \le \|\mathcal{U}_\mathcal{I}(\Delta Z)\| \quad \forall\, \Delta Z \in \mathbb{Z}. \tag{72}$$

*Proof* We focus on the result (i), since (ii) can be obtained similarly.

It suffices to show that the inequality (71) holds, as the nonsingularity of $\mathcal{U}_0$ is a direct implication of (71). Assume, on the contrary, there exist a sequence $\{\widehat{Z}^\nu\}$ converging to $\overline{Z}$ and a nonzero sequence $\{\Delta Z^\nu = (\Delta x^\nu, \Delta\xi^\nu, \Delta\Gamma^\nu)\} \subseteq \mathbb{Z}$ such that $\mathcal{U}_0^\nu \in \partial_B F(Z^\nu)$ defined by (69) satisfy

$$\|\mathcal{U}_0^\nu(\Delta Z^\nu)\| < 1/\nu \|\Delta Z^\nu\|.$$

Without loss of generality, we may assume that $\|\Delta Z^\nu\| = 1$ for each $\nu$ and $\Delta Z^\nu \to \Delta Z^\infty \neq 0$ as $\nu \to \infty$. For each $\nu$, let $\widehat{A}^\nu = g(\widehat{x}^\nu) + \widehat{\Gamma}^\nu$ enjoy the eigenvalue decomposition (23) with $\widehat{P}^\nu \in \mathcal{O}^n(\widehat{A}^\nu)$. For each $\nu$, denote

$$\zeta^\nu := \{i \mid \widehat{\lambda}_i^\nu \geq \delta\}, \quad \theta^\nu := \{i \mid |\widehat{\lambda}_i^\nu| \leq \delta\} \quad \text{and} \quad \kappa^\nu := \{i \mid \widehat{\lambda}_i^\nu \leq -\delta\},$$

where $\widehat{\lambda}^\nu := \lambda(\widehat{A}^\nu)$ is the eigenvalues of $\widehat{A}^\nu$. Since $\widehat{A}^\nu \to \overline{A} = g(\overline{x}) + \overline{\Gamma}$ as $\nu \to \infty$, we know from the continuity of eigenvalues [11] that $\widehat{\lambda}^\nu \to \lambda(A)$ as $\nu \to \infty$. Notice that $\delta \in (0, \min(\overline{\lambda}_\alpha, |\overline{\lambda}_\gamma|))$, for $\nu$ large enough,

$$\zeta^\nu \equiv \alpha, \quad \theta^\nu \equiv \beta \quad \text{and} \quad \kappa^\nu \equiv \gamma, \tag{73}$$

where $\alpha$, $\beta$ and $\gamma$ are the index sets defined by (11) for $\overline{A}$. Since $Z^\nu = \mathcal{P}_\delta(\widehat{Z}^\nu)$, it holds that

$$\|Z^\nu - \widehat{Z}^\nu\| = \|\mathcal{P}_\delta(\widehat{Z}^\nu) - \widehat{Z}^\nu\| = \|\sum_{i\in\beta} \lambda_i(\widehat{A}^\nu)\widehat{P}_i^\nu(\widehat{P}_i^\nu)^T\| \leq \|\widehat{\lambda}_\beta^\nu\| \to 0, \text{ as } \nu \to \infty.$$

Thus, we know that $Z^\nu \to \overline{Z}$ as $\nu \to \infty$. It also holds that for each $\nu$,

$$A^\nu = g(x^\nu) + \Gamma^\nu = \widehat{P}^\nu \begin{bmatrix} \widehat{\Lambda}_{\alpha\alpha}^\nu & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \widehat{\Lambda}_{\gamma\gamma}^\nu \end{bmatrix} (\widehat{P}^\nu)^T \tag{74}$$

with $\widehat{\Lambda}^\nu = \mathrm{Diag}(\widehat{\lambda}^\nu)$, and $A^\nu \to \overline{A}$ as $\nu \to \infty$.

Let $\overline{\mathcal{U}}_0$ be given by (36) with $\overline{\mathcal{V}}_0$ defined by (16). By noting that $\|\mathcal{U}_0^\nu(\Delta Z^\nu)\| < 1/\nu$, $\|\Delta Z^\nu\| = 1$, $f$, $h$ and $g$ are twice continous differentiable and $Z^\nu \to \overline{Z}$, we have

$$\begin{aligned}
&\|\overline{\mathcal{U}}_0(\Delta Z^\infty)\| \leq \|\overline{\mathcal{U}}_0(\Delta Z^\infty) - \mathcal{U}_0^\nu(\Delta Z^\nu)\| + \|\mathcal{U}_0^\nu(\Delta Z^\nu)\| \\
&\leq \|\overline{\mathcal{U}}_0(\Delta Z^\infty) - \overline{\mathcal{U}}_0(\Delta Z^\nu)\| + \|\overline{\mathcal{U}}_0(\Delta Z^\nu) - \mathcal{U}_0^\nu(\Delta Z^\nu)\| + 1/\nu \\
&\leq O(\|\Delta Z^\infty - \Delta Z^\nu\|) + \|\overline{\mathcal{V}}_0(g'(\overline{x})\Delta x^\nu + \Delta\Gamma^\nu) - \mathcal{V}_0^\nu(g'(x^\nu)\Delta x^\nu + \Delta\Gamma^\nu)\| + 1/\nu,
\end{aligned} \tag{75}$$

where for each $\nu$, $\mathcal{V}_0^\nu$ is defined by (16) with respect to $A^\nu$. By Proposition 1 and taking subsequence if necessary, we may further assume that $\widehat{P}^\nu \to P^\infty \in \mathcal{O}^n(\overline{A})$. Then, (74) and Proposition 2 imply that $\lim_{\nu\to\infty} \mathcal{V}_0^\nu = \overline{\mathcal{V}}_0$. Since $\mathcal{V}_0^\nu$ is uniformly bounded [14, Proposition 1], we thus have as $\nu \to \infty$,

$$\begin{aligned}
&\|\mathcal{V}_0^\nu(g'(x^\nu)\Delta x^\nu + \Delta\Gamma^\nu) - \overline{\mathcal{V}}_0(g'(\overline{x})\Delta x^\nu + \Delta\Gamma^\nu)\| \\
&\leq \|\mathcal{V}_0^\nu(g'(x^\nu)\Delta x^\nu + \Delta\Gamma^\nu) - \mathcal{V}_0^\nu(g'(\overline{x})\Delta x^\nu + \Delta\Gamma^\nu)\| \\
&\quad + \|\mathcal{V}_0^\nu(g'(\overline{x})\Delta x^\nu + \Delta\Gamma^\nu) - \overline{\mathcal{V}}_0(g'(\overline{x})\Delta x^\nu + \Delta\Gamma^\nu)\| \\
&\leq O(\|(g'(x^\nu) - g'(\overline{x}))\Delta x^\nu\|) + o(\|g'(\overline{x})\Delta x^\nu + \Delta\Gamma^\nu\|) = o(1).
\end{aligned} \tag{76}$$

Combining (75) and (76), we obtain $\|\overline{\mathcal{U}}_0(\Delta Z^\infty)\| = 0$, which implies that $\overline{\mathcal{U}}_0$ is singular. This contradicts to Theorem 1 (i), since the W-SOC and constraint nondegeneracy are assumed. The proof is then completed. □

Using Proposition 8, we can further establish a local error bound for the nonsmooth map $F$ at the corrected points around $\overline{Z}$ in the following proposition.

**Proposition 9** *Let $\overline{x} \in \mathbb{X}$ be a stationary point of the NLSDP (21) with a Lagrange multiplier $(\overline{\xi}, \overline{\Gamma}) \in M(\overline{x})$. Recall the definitions of constants $\delta$, $\kappa_0$ and $\kappa_1$ in Proposition 8. Then,* *(i) if the W-SOC and the constraint nondegeneracy hold at* $\overline{Z} = (\overline{x}, \overline{\xi}, \overline{\Gamma}) \in \mathbb{Z}$*, then there exists a neighborhood* $\mathcal{N}$ *of* $\overline{Z}$ *such that for any* $\widehat{Z} \in \mathcal{N}$ *and* $Z = \mathcal{P}_\delta(\widehat{Z})$*,*

$$\|F(Z)\| \geq \frac{1}{2\kappa_0}\|Z - \overline{Z}\|;$$

*(ii) if the S-SOSC and the W-SRCQ hold at* $\overline{Z} = (\overline{x}, \overline{\xi}, \overline{\Gamma}) \in \mathbb{Z}$*, then there exists a neighborhood* $\mathcal{N}$ *of* $\overline{Z}$ *such that for any* $\widehat{Z} \in \mathcal{N}$ *and* $Z = \mathcal{P}_\delta(\widehat{Z})$*,*

$$\|F(Z)\| \geq \frac{1}{2\kappa_1}\|Z - \overline{Z}\|.$$

*Proof* We only show (i), as (ii) can be proved via similar arguments.

By the semismoothness of $F$ and Proposition 8 (i), we know that there exists a neighborhood $\mathcal{N}$ of $\overline{Z}$ such that for any $\widehat{Z} \in \mathcal{N}$ and $Z = \mathcal{P}_\delta(\widehat{Z})$,

$$\|F(Z) - \mathcal{U}_0(Z - \overline{Z})\| \leq \frac{1}{2\kappa_0}\|Z - \overline{Z}\| \quad \text{and} \quad \|\mathcal{U}_0\| \geq \frac{1}{\kappa_0},$$

where $\mathcal{U}_0 \in \partial_B F(Z)$ is given in (69). Thus,

$$\|F(Z)\| \geq \|\mathcal{U}_0(Z - \overline{Z})\| - \|F(Z) - \mathcal{U}_0(Z - \overline{Z})\| \geq \frac{1}{\kappa_0}\|Z - \overline{Z}\| - \frac{1}{2\kappa_0}\|Z - \overline{Z}\| = \frac{1}{2\kappa_0}\|Z - \overline{Z}\|.$$

The proof is completed. □

Now we are ready to present our inexact semismooth Newton method with correction in Algorithm 1 for solving the nonsmooth system $F(Z) = 0$ with $F$ given in (22). As one can observe, the algorithm is quite similar to the classic inexact semismooth Newton method [5]. The main difference is that in each iteration, an extra correction step, i.e., step 6 in Algorithm 1, involving the correction mapping $\mathcal{P}_\delta$ is taken. This correction step, based on Proposition 8, ensures locally the nonsingularity of $\mathcal{U}_0^k$ (or $\mathcal{U}_\mathcal{I}^k$), and hence the well-definedness of the proposed algorithm. Meanwhile, we shall emphasize that the additional computational costs brought by the correction step are negligible, as the eigendecompostion of $\widehat{A}^{k+1} = g(\widehat{x}^{k+1}) + \widehat{\Gamma}^{k+1}$ in step 6 can be re-used to obtain $\mathcal{U}^{k+1} \in \partial_B F(Z^{k+1})$ in step 3. Moreover, the inexact computation in step 4 indicates that our proposed algorithm has great potential for solving large-scale problems.

**Algorithm 1** The semismooth Newton method with correction using $\mathcal{U}_0$ (or $\mathcal{U}_\mathcal{I}$)

1: Initialize $\widehat{Z}^0 = (\widehat{x}^0, \widehat{\xi}^0, \widehat{\Gamma}^0)$, $\delta > 0$, $\eta \geq 0$, $\tau \in (0, 1]$ and $Z^0 = \mathcal{P}_\delta(\widehat{Z}^0)$.
2: **for** $k = 0, 1, \ldots$ **do**
3: Get $\mathcal{U}^k = \mathcal{U}_0^k \in \partial_B F(Z^k)$ by (69) ($\mathcal{U}^k = \mathcal{U}_\mathcal{I}^k \in \partial_B F(Z^k)$ by (70)).
4: Find an approximate solution $d^k$ to the linear system $\mathcal{U}^k d + F(Z^k) = 0$ such that

$$\|\mathcal{U}^k d^k + F(Z^k)\| \leq \min(\eta, \|F(Z^k)\|^\tau)\|F(Z^k)\|.$$

5: $\widehat{Z}^{k+1} = Z^k + d^k$.
6: $Z^{k+1} = \mathcal{P}_\delta(\widehat{Z}^{k+1})$.
7: **end for**

As promised, in the following theorem, we establish the well-definedness and local convergence results of Algorithm 1 without assuming the subdifferential regularity.

**Theorem 2** *Let $\overline{x} \in \mathbb{X}$ be a stationary point of the NLSDP (21) with a Lagrange multiplier $(\overline{\xi}, \overline{\Gamma}) \in M(\overline{x})$. Let $\delta \in (0, \min(\overline{\lambda}_\alpha, |\overline{\lambda}_\gamma|))$. Then,*

*(i) if the W-SOC and the constraint nondegeneracy hold at $\overline{Z} = (\overline{x}, \overline{\xi}, \overline{\Gamma}) \in \mathbb{Z}$, then there exists $\overline{\eta} > 0$ such that for any $\widehat{Z}^0$ sufficiently close to $\overline{Z}$ and $\eta \le \overline{\eta}$, Algorithm 1 using $\mathcal{U}_0$ is well defined and the sequence $\{Z^k\}$ so generated converges to $\overline{Z}$ superlinearly;*

*(ii) if the S-SOSC and W-SRCQ hold at $\overline{Z} = (\overline{x}, \overline{\xi}, \overline{\Gamma}) \in \mathbb{Z}$, then there exists $\overline{\eta} > 0$ such that for any $\widehat{Z}^0$ sufficiently close to $\overline{Z}$ and $\eta \le \overline{\eta}$, Algorithm 1 using $\mathcal{U}_{\mathcal{I}}$ is well-defined and the sequence $\{Z^k\}$ so generated converges to $\overline{Z}$ superlinearly.*

*Moreover, if $f$, $g$ and $h$ are twice Lipschitz continuously differentiable, i.e., they have Lipschitz continuous second order differentials, and $\tau = 1$, then the corresponding convergence rates obtained in (i) and (ii) are quadratic.*

*Proof* Here, we focus on the case (i) where the W-SOC and the constraint nondegeneracy hold at $\overline{Z}$ and $\mathcal{U}_0$ is employed. The proof for the case (ii) where the S-SOSC and W-SRCQ hold at $\overline{Z}$ can be obtained similarly.

By the semismoothness of $\Pi_{\mathbb{S}^n_+}(\cdot)$ on $\mathbb{S}^n$ and the twice continously differentiability of $f$ $g$ and $h$, we know $F$ is semismooth at $\overline{Z}$. This, together with Proposition 8, implies that we can find a positive constant $\varsigma$, a neighborhood $\mathbb{B}(\overline{Z}, \varsigma)$ of $\overline{Z}$ and a positve constant $\kappa_0$ such that for any given $\widehat{Z} \in \mathbb{B}(\overline{Z}, \varsigma)$ and $Z = \mathcal{P}_\delta(\widehat{Z})$,

$$\|F(Z) - F(\overline{Z}) - \mathcal{U}_0(Z - \overline{Z})\| \le \frac{1}{4(L_g + 2)\kappa_0}\|Z - \overline{Z}\| \quad \text{and} \quad \|(\mathcal{U}_0)^{-1}\| \le \kappa_0, \tag{77}$$

where $\mathcal{U}_0 \in \partial_B F(Z)$ is given by (69) and $L_g > 0$ is the Lipschitz constant of $g$ over $\mathbb{B}(\overline{Z}, \varsigma)$. Recall the eigenvalue decomposition of $\overline{A} = g(\overline{x}) + \overline{\Gamma}$ in (23) with index sets $\alpha, \beta$ and $\gamma$ defined in (12).

Meanwhile, for any $\widehat{Z} \in \mathbb{B}(\overline{Z}, \varsigma)$ and $Z = \mathcal{P}_\delta(\widehat{Z})$, let $\widehat{A} = g(\widehat{x}) + \widehat{\Gamma}$ satisfy the eigen value decompostion (11) with $\widehat{P} \in \mathcal{O}^n(\widehat{A})$ and $\widehat{\Lambda} = \Lambda(\widehat{A})$. Then, similar to (73) and (74), by shrinking $\mathbb{B}(\overline{Z}, \varsigma)$ if necessary, we have

$$A = g(x) + \Gamma = \widehat{P}\begin{bmatrix} \widehat{\Lambda}_{\alpha\alpha} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \widehat{\Lambda}_{\gamma\gamma} \end{bmatrix}(\widehat{P})^T.$$

Thus,

$$\begin{aligned} \|Z - \overline{Z}\| \le& \|Z - \widehat{Z}\| + \|\widehat{Z} - \overline{Z}\| \le \|\widehat{\Lambda}_{\beta\beta}\| + \|\widehat{Z} - \overline{Z}\| \le \|\lambda(\widehat{A}) - \lambda(\overline{A})\| + \|\widehat{Z} - \overline{Z}\| \\ \le& \|g(\widehat{x}) + \widehat{\Gamma} - g(\overline{x}) - \overline{\Gamma}\| + \|\widehat{Z} - \overline{Z}\| \le (L_g + 2)\|\widehat{Z} - \overline{Z}\|, \end{aligned} \tag{78}$$

where the last inequality is due to the Lipschitz continuity of $g$ over $\mathbb{B}(\overline{Z}, \varsigma)$.

Now, let $\widehat{Z}^0$ be any point in $\mathbb{B}(\overline{Z}, \varsigma)$. Recall that $Z^0 = \mathcal{P}_\delta(\widehat{Z}^0)$. Then, (77) implies that $\mathcal{U}_0^0 \in \partial_B F(Z^0)$ is nonsingular, and therefore $d^0$ and $\widehat{Z}^1$ are well-defined. Since $F(\overline{Z}) = 0$, we further have

$$\begin{aligned} \|\widehat{Z}^1 - \overline{Z}\| &= \|Z^0 - \overline{Z} - (\mathcal{U}_0^0)^{-1}F(Z^0) + d^0 + (\mathcal{U}_0^0)^{-1}F(Z^0)\| \\ &\le \| - (\mathcal{U}_0^0)^{-1}\left[F(Z^0) - F(\overline{Z}) - \mathcal{U}_0^0(Z^0 - \overline{Z}) - \mathcal{U}_0^0 d^0 - F(Z^0)\right]\| \\ &\le \kappa_0\|F(Z^0) - F(\overline{Z}) - \mathcal{U}_0^0(Z^0 - \overline{Z})\| + \eta\kappa_0\|F(Z^0) - F(\overline{Z})\| \\ &\le \Big(\frac{1}{4(L_g + 2)} + \eta\kappa_0 L_F\Big)\|Z^0 - \overline{Z}\|, \end{aligned} \tag{79}$$

where the last inequality follows from (77) and the Lipschitz continuity of $F$ over $\mathbb{B}(\overline{Z}, \varsigma)$ with $L_F$ being the corresponding Lipschitz constant. Meanwhile, we know from (78) that $\|Z^0 - \overline{Z}\| \leq (L_g + 2)\|\widehat{Z}^0 - \overline{Z}\| \leq (L_g + 2)\varsigma$. Then, for any $0 \leq \eta \leq \overline{\eta} = 1/(4\kappa_0 L_F(L_g + 2))$, we see from (79) that $\|\widehat{Z}^1 - \overline{Z}\| \leq \|Z^0 - \overline{Z}\|/(2(L_g + 2)) \leq \varsigma$ and

$$\|Z^1 - \overline{Z}\| \leq (L_g + 2)\|\widehat{Z}^1 - \overline{Z}\| \leq \frac{1}{2}\|Z^0 - \overline{Z}\|. \tag{80}$$

In summary, the derivation above shows that if $\widehat{Z}^0$ belongs to $\mathbb{B}(\overline{Z}, \varsigma)$, then every $\widehat{Z}^k$ and $Z^k$ produced by Algorithm 1 satisfy

$$\widehat{Z}^k \in \mathbb{B}(\overline{Z}, \varsigma) \quad \text{and} \quad Z^k \in \mathbb{B}(\overline{Z}, (L_g + 2)\varsigma),$$

$\{Z^k\}$ converges to $\overline{Z}$ due to the contraction inequality (80). Moreover, the semismoothness of $F$, (77) and (79) further imply that

$$\begin{aligned} \|\widehat{Z}^{k+1} - \overline{Z}\| &= \| - (\mathcal{U}_0^k)^{-1}[F(Z^k) - F(\overline{Z}) - \mathcal{U}_0^k(Z^k - \overline{Z}) - \mathcal{U}_0^k d^k - F(Z^k)]\| \\ &= o(\|Z^k - \overline{Z}\|) + O(\|Z^k - \overline{Z}\|^{1+\tau}). \end{aligned} \tag{81}$$

This, together with the fact that $\|Z^{k+1} - \overline{Z}\| \leq (L_g + 2)\|\widehat{Z}^{k+1} - \overline{Z}\|$, implies the superlinear convergence of $\{Z^k\}$.

If $f, g, h$ are twice Lipschitz continuously differentiable, by the strong semismoothness of $\Pi_{\mathbb{S}^n_+}(\cdot)$ on $\mathbb{S}^n$ and the Lipshitz continuously differentiability of $\nabla_x L$, $g$ and $h$, we know $F$ is strongly semismooth at $\overline{Z}$. The strong semismoothness of $F$ and $\tau = 1$ allow us to replace (81) by the following:

$$\begin{aligned} \|\widehat{Z}^{k+1} - \overline{Z}\| &= \| - (\mathcal{U}_0^k)^{-1}[F(Z^k) - F(\overline{Z}) - \mathcal{U}_0^k(Z^k - \overline{Z}) - \mathcal{U}_0^k d^k - F(Z^k)]\| \\ &= O(\|Z^k - \overline{Z}\|^2). \end{aligned}$$

Then, the quadratic convergence of $\{Z^k\}$ follows directly. □

*Remark 9* Let $\{Z^k\}$ be the sequence generated by Algorithm 1. For a given tolerance $\varepsilon > 0$, Proposition 9 implies that $F(Z^k) < \varepsilon$ is a reasonable stopping criterion for Algorithm 1. In fact, the local Lipschitz continuity of $F$ around $\overline{Z}$ and Proposition 9 imply that $F(Z^k) = \Theta(\|Z^k - \overline{Z}\|)$.

*Remark 10 (Choice of $\widehat{Z}^0$ and $\delta$)* It is not difficult to observe from Theorem 2 and its proof that the choice of $\widehat{Z}^0$ and $\delta$ is crucial to the performance of Algorithm 1. A larger $\delta$ may lead to a more restrictive choice of $\widehat{Z}^0$ in order to ensure the nonsingularity of $\mathcal{U}^0$. Meanwhile, accurately determining the value of $\min(\overline{\lambda}_\alpha, |\lambda_\gamma|)$ for general NLSDP is often challenging in practical settings, though obtaining a rough estimate may be feasible. In this study, our primary emphasis lies in the local convergence analysis of Algorithm 1. We defer the considerations regarding the efficient and robust implementation of the algorithm, including the selection of $\widehat{Z}^0$ and $\delta$, to be explored in future researches.

## 6 Numerical results

In this section, we present a series of examples to empirically validate our theoretical findings and demonstrate the fast local convergence of our proposed algorithm. In our numerical experiments, we consider a range of problems, including both convex and nonconvex nonlinear SDPs. It's worth noting that the subdifferential regularity condition does not hold for all the test instances, thereby rendering classic convergence results for semismooth Newton methods inapplicable.

As preliminary numerical results, we solve the Newton linear system involved in each iteration of Algorithm 1 exactly, i.e., the parameter $\eta$ is set to be zero. Following Remark 9, we assess the accuracy of an approximate solution $Z$ to the NLSDP (21) by measuring the residual $\|F(Z)\|$. For a given stopping tolerance $\epsilon > 0$, our algorithm terminates when $\|F(Z)\| < \epsilon$. Throughout our experiments, we fix $\epsilon = 10^{-10}$. All computational results are generated using MATLAB (version 9.10) running on a Mac mini (Apple M1, 16 G RAM).

While our focus has been on investigating the local convergence properties of our newly proposed semismooth Newton method, our numerical tests indicate that the algorithm is quite robust to the selection of the starting point. To facilitate reproducibility, we adhere to the following procedures for generating the starting point:

(i) Fix the random seed and select the perturbation constant $C_{\text{perturb}}$. In our tests, we fix the random seed via `rng(2,''twister'')` and set $C_{\text{perturb}} = 10$ for examples 3, 4 and 5, and set $C_{\text{perturb}} = 1$ for example 6.

(ii) Generate random noises $(\Delta x, \Delta\xi, \Delta\Gamma)$ with each entry following a standard normal distribution $N(0,1)$, e.g.,

$$\mathtt{\Delta\Gamma = randn(n)}, \quad \mathtt{\Delta\Gamma = 0.5 * (\Delta\Gamma + \Delta\Gamma')}.$$

(iii) Set the starting point $\widehat{Z}^0$ as

$$\widehat{Z}^0 = \overline{Z} + C_{\text{perturb}} \times \left( \frac{\Delta x}{\|\Delta x\|}, \frac{\Delta\xi}{\|\Delta\xi\|}, \frac{\Delta\Gamma}{\|\Delta\Gamma\|} \right),$$

where $\overline{Z}$ is the known optimal solution to (21).

It is important to note that the perturbation constant $C_{\text{perturb}}$ employed is not insignificantly small. Remarkably, varying the random seeds and the perturbation constant $C_{\text{perturb}}$ results in comparable quadratic convergence behavior.

*Example 3* Consider the following convex quadratic SDP problem:

$$\begin{aligned} \min \quad & \frac{1}{2}\|x_{11} - 1\|^2 + \frac{1}{2}\|x_{22} - x_{12} - x_{21}\|^2 \\ \text{s.t.} \quad & \langle E, X \rangle \le 1, \\ & X \in \mathbb{S}^2_+, \end{aligned} \tag{82}$$

where $X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \in \mathbb{S}^2$. It is not difficult to observe that $\overline{X} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ is an optimal solution and $(\overline{\xi}, \overline{\Gamma}) = (0, 0)$ is the unique multiplier. For such $\overline{Z} = (\overline{X}, \overline{\xi}, \overline{\Gamma})$, one can check that the W-SOC and the constraint nondegeneracy holds at $\overline{Z}$, but the S-SOSC fails. Then, Theorem 1 implies the nonsingularity of $\overline{\mathcal{U}}_0$, which can also be verified directly by noting that the linear system

$$\overline{\mathcal{U}}_0(\Delta Z) = \begin{bmatrix} \begin{bmatrix} \Delta x_{11} & -\Delta x_{22} + \Delta x_{12} + \Delta x_{21} \\ -\Delta x_{22} + \Delta x_{12} + \Delta x_{21} & \Delta x_{22} - \Delta x_{12} - \Delta x_{21} \end{bmatrix} + \Delta\xi E + \Delta\Gamma \\ \langle E, \Delta X \rangle \\ \begin{bmatrix} \Delta\Gamma_{11} & \Delta\Gamma_{12} \\ \Delta\Gamma_{21} & -\Delta x_{22} \end{bmatrix} \end{bmatrix} = 0$$

with $\Delta Z = (\Delta X, \Delta\xi, \Delta\Gamma) = \left(\begin{bmatrix}\Delta x_{11} & \Delta x_{12}\\ \Delta x_{21} & \Delta x_{22}\end{bmatrix}, \Delta\xi, \begin{bmatrix}\Delta\Gamma_{11} & \Delta\Gamma_{12}\\ \Delta\Gamma_{21} & \Delta\Gamma_{22}\end{bmatrix}\right) \in \mathbb{S}^2 \times \mathbb{R} \times \mathbb{S}^2$ has only trivial zero solution $\Delta Z = 0$. Thus, Algorithm 1 using $\mathcal{U}_0$ can be applied to solve (82). In the test, we set the parameter $\delta = 0.5$, which lies between 0 and the smallest nonzero magnitude eigenvalue of $\overline{A}$. We report the detailed performance of our algorithm in Figure 3 and Table 1. In the table,

| Iteration $k$ | $\\|F(Z^k)\\|$ | $\\|Z^k - \overline{Z}\\|$ | $\sigma_{\min}(\mathcal{U}^k)$ |
|---|---|---|---|
| 0 | 2.51e+01 | 1.43e+01 | 4.03e-01 |
| 1 | 7.50e-01 | 9.36e-01 | 3.07e-01 |
| 2 | 4.19e-02 | 2.10e-02 | 3.04e-01 |
| 3 | 3.33e-05 | 2.88e-05 | 3.04e-01 |
| 4 | 2.06e-14 | 1.80e-14 | 3.04e-01 |

Table 1: Convergence details of Algorithm 1 for solving (82).

we report residuals $\|F(Z^k)\|$ and $\|Z^k - \overline{Z}\|$, and the minimum singular value of $\mathcal{U}^k$, denoted by $\sigma_{\min}(\mathcal{U}^k)$. The detailed results in Table 1 demonstrate the quadratic convergence of Algorithm 1, the locally uniform boundedness of $(\mathcal{U}^k)^{-1}$, and the relation $F(Z^k) = \Theta(\|Z^k - \overline{Z}\|)$ asserted in Remark 9.

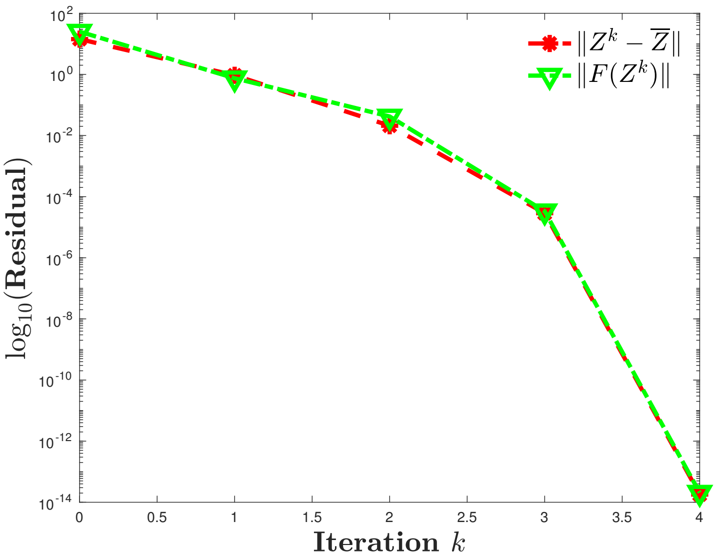


Fig. 3: Quadratic convergence of Algorithm 1 for solving (82).

The following example illustrates the efficiency of our algorithm in solving nonlinear SDPs even in the absence of calmness, as defined in [25, Definition 9(30)].

*Example 4 (Example 3 in [4]; see also [34])* Let $B = \begin{bmatrix} 3/2 & -2 \\ -2 & 3 \end{bmatrix}$ and $b = B^{-1/2}\begin{bmatrix} 5/2 \\ -1 \end{bmatrix}$. Consider the following convex quadratic SDP problem:

$$\begin{array}{ll} \min & \frac{1}{2}\|x+b\|^2 + t \\ \text{s.t.} & \mathcal{A}^* x + tE + I + \Delta \in \mathbb{S}^2_+, \\ & t \geq 0, \end{array} \tag{83}$$

and its dual

$$\begin{array}{ll} \max & -\frac{1}{2}\|\mathcal{A}Y - b\|^2 - \langle I + \Delta, Y\rangle \\ \text{s.t.} & \langle E, Y\rangle \leq 1, \\ & Y \in \mathbb{S}^2_+, \end{array} \tag{84}$$

where $\mathcal{A}Y = B^{1/2}\mathrm{Diag}(Y)$ for all $Y \in \mathbb{S}^2$, $\mathcal{A}^*x = \mathrm{Diag}(B^{1/2}x)$ for all $x \in \mathbb{R}^2$ and $\Delta = \mathrm{Diag}(-\varepsilon, \varepsilon)$ with $\varepsilon \geq 0$. As shown in [4], the solution map of the KKT system associated with problems (83) and (84), denoted by $S_{\mathrm{KKT}}$ in [4], is not calm with respect to the perturbation $\varepsilon$. When $\varepsilon = 0$, it is not difficult to see that $\overline{Y} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ is the unique optimal solution to (84) with the unique multipler $(0,0) \in \mathbb{R} \times \mathbb{S}^2$. Then, one can check that the W-SOC and constraint nondegeneracy hold at $\overline{Z} = (\overline{Y}, 0, 0)$, but the S-SOSC fails. From Theorem 1 and Proposition 5, we know that $\overline{\mathcal{U}}_0$ is nonsingular but $\overline{\mathcal{U}}_{\mathcal{I}}$ is singular. Thus, we can apply Algorithm 1 with $\mathcal{U}_0$ to solve (84). In the test, we set parameter $\delta = 0.5$. The convergence details of our algorithm are reported in Table 2 and Figure 4. Again, the quadratic convergence of our algorithm is confirmed.

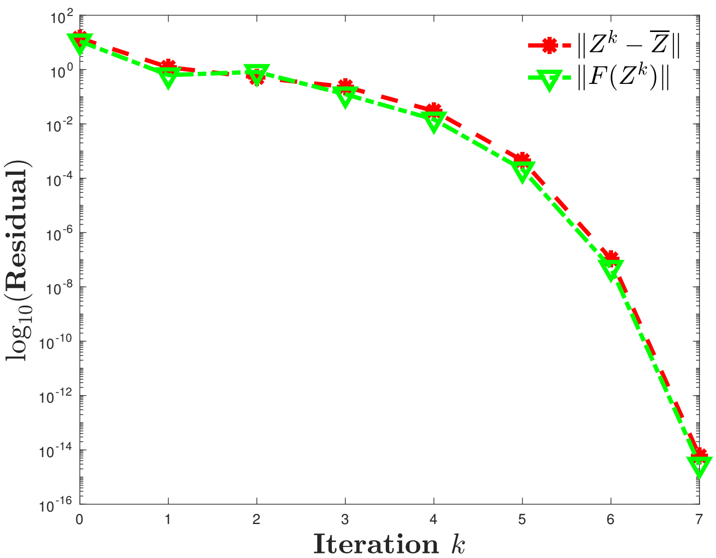


Fig. 4: Quadratic convergence of Algorithm 1 for solving (84).

| Iteration $k$ | $\|F(Z^k)\|$ | $\|Z^k - \overline{Z}\|$ | $\sigma_{\min}(\mathcal{U}^k)$ |
|---|---|---|---|
| 0 | 1.10e+01 | 1.43e+01 | 4.16e-02 |
| 1 | 6.26e-01 | 1.22e+00 | 1.48e-01 |
| 2 | 8.23e-01 | 5.12e-01 | 1.38e-01 |
| 3 | 1.23e-01 | 2.33e-01 | 3.70e-01 |
| 4 | 1.46e-02 | 3.02e-02 | 2.85e-01 |
| 5 | 2.13e-04 | 4.51e-04 | 2.76e-01 |
| 6 | 5.08e-08 | 1.08e-07 | 2.76e-01 |
| 7 | 2.89e-15 | 6.05e-15 | 2.76e-01 |

Table 2: Convergence details of Algorithm 1 for solving (84).

In the next two examples, we evaluate the performance of our algorithm for solving medium size nonlinear SDPs.

*Example 5* Consider the following convex quadratic SDP problem:

$$\begin{array}{ll} \min & \frac{1}{2}\|X_{11} - I\|^2 + \frac{1}{2}\|X_{12}\|^2 + \frac{1}{2}\|X_{21}\|^2 \\ \text{s.t.} & X \in \mathbb{S}^n_+, \end{array} \tag{85}$$

where $X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}$ is the block representation with $X_{11} \in \mathbb{R}^{l_1 \times l_1}$ and $X_{22} \in \mathbb{R}^{l_2 \times l_2}$. We know that $\overline{X} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$ is an optimal solution and $\overline{\Gamma} = 0$ is the unique multiplier. One can check that the W-SOC and the constraint nondegeneracy holds at $\overline{Z} = (\overline{X}, \overline{\Gamma})$, but the SOSC fails since $\overline{X}$ is not isolated. Then, we have from Theorem 1 that $\overline{\mathcal{U}}_0$ is nonsingualr and Algorithm 1 with $\mathcal{U}_0$ can be applied to solve (85). Here, the test instance is constructed by setting $l_1 = 60$, $l_2 = 40$, and the parameter $\delta$ is set to 0.5. Table Figure 5 shows our algorithm converges rapidly.

| Iteration $k$ | $\|F(Z^k)\|$ | $\|Z^k - \overline{Z}\|$ | $\sigma_{\min}(\mathcal{U}^k)$ |
|---|---|---|---|
| 0 | 8.39e-01 | 7.06e-01 | 6.02e-01 |
| 1 | 3.54e-02 | 3.54e-02 | 6.18e-01 |
| 2 | 1.85e-14 | 1.75e-14 | 6.18e-01 |

Table 3: Convergence details of Algorithm 1 for solving (85).

*Example 6* Recall problem (18) in Example 1. Note that problem (18) is nonconvex, and $\overline{X} = 0$ is a local optimal solution with unbounded Lagrange multipler set $M(\overline{X})$. It can be verified that the S-SOSC holds at $\overline{Z} = (\overline{X}, \overline{\xi}_{12}, \overline{\xi}_{22}, \overline{\Gamma}) = (0, 0, 0, 0) \in \mathbb{S}^n \times \mathbb{R}^{l_1 \times l_2} \times \mathbb{R}^{l_2 \times l_2} \times \mathbb{S}^n$ and the W-SRCQ holds at $\overline{X}$ with respect to $(\overline{\xi}_{12}, \overline{\xi}_{22}, \overline{\Gamma})$. Then, it is known from Theorem 1 that $\overline{\mathcal{U}}_{\mathcal{I}}$ is nonsingular. Thus, Algorithm 1 with $\mathcal{U}_{\mathcal{I}}$ can be applied to solve (18). In our test, we set the parameter $\delta = 0.5$, and $l_1 = 60$ and $l_2 = 40$. The detailed performance of our algorithm is reported in Table 4. As one can observe, our algorithm converges in one iteration from a randomly generated starting point with initial distance $\|Z^0 - \overline{Z}\| = 1.73$.

## 7 Conclusions

In this paper, we explore the weak second-order condition (W-SOC) and the weak strict Robinson constraint qualification (W-SRCQ) within the framework of general optimization problems

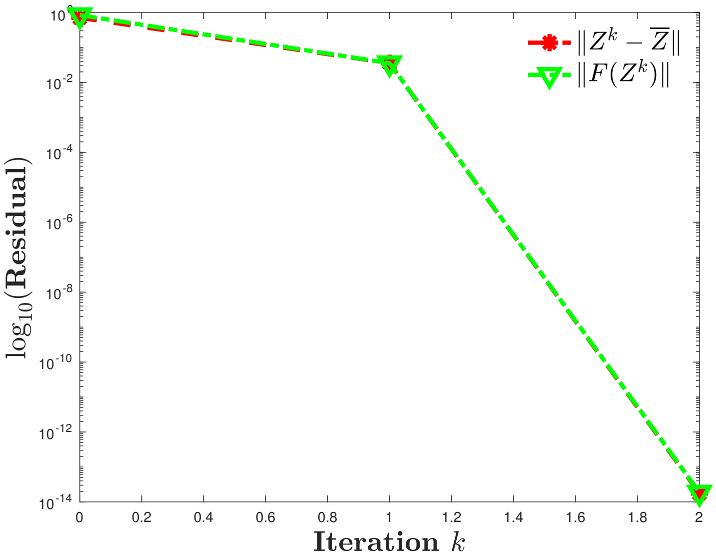


Fig. 5: Quadratic convergence of Algorithm 1 for solving (85).

| Iteration $k$ | $\|F(Z^k)\|$ | $\|Z^k - \overline{Z}\|$ | $\sigma_{\min}(\mathcal{U}^k)$ |
|---|---|---|---|
| 0 | 1.85e+00 | 1.73e+00 | 3.27e-01 |
| 1 | 2.98e-15 | 4.18e-15 | 3.27e-01 |

Table 4: Convergence details of Algorithm 1 for solving (18).

(1). We establish the sufficient and necessary conditions of the existence of a nonsingular element in the subdifferentials of the KKT residual map $F$ at a given KKT point. Moreover, we uncover a profound primal-dual connection between the W-SOC and the W-SRCQ for the convex quadratic SDP. Drawing upon these theoretical advancements, we propose a novel semismooth Newton method with a correction step, ensuring the local convergence and yielding a (quadratic) superlinear convergence rate without relying on subdifferential regularity. There are still many issues that deserve to be explored further. For example, addressing the globalization of the proposed algorithm and developing highly efficient implementations present significant challenges. It is also desirable to extend our theoretical results to more general sets.

## References


1. Bonnans, J.F., Shapiro, A.: Perturbation analysis of optimization problems. Springer Verlag (2000)
2. Chan, Z.X., Sun, D.: Constraint nondegeneracy, strong regularity, and nonsingularity in semidefinite programming. SIAM Journal on optimization **19**(1), 370–396 (2008)
3. Clarke, F.H.: Optimization and Nonsmooth Analysis. John Wiley & Sons, Inc., New York (1983)
4. Ding, C., Sun, D., Zhang, L.: Characterization of the robust isolated calmness for a class of conic programming problems. SIAM Journal on Optimization **27**(1), 67–90 (2017)
5. Facchinei, F., Pang, J.S.: Finite-dimensional variational inequalities and complementarity problems. Springer (2003)
6. Han, D., Sun, D., Zhang, L.: Linear rate convergence of the alternating direction method of multipliers for convex composite programming. Mathematics of Operations Research **43**(2), 622–637 (2018)

7. Hintermüller, M.: Semismooth newton methods and applications. Department of Mathematics, Humboldt-University of Berlin (2010)
8. Izmailov, A.F., Solodov, M.V.: Karush-kuhn-tucker systems: regularity conditions, error bounds and a class of newton-type methods. Mathematical Programming **95**, 631–650 (2003)
9. Klatte, D., Kummer, B.: Nonsmooth Equations in Optimization, *Springer Science & Business Media*, vol. 60. Kluwer Academic Publishers (2002). DOI 10.1007/b130810
10. Kummer, B.: Advances in Optimization, Proceedings of the 6th French-German Colloquium on Optimization Held at Lambrecht, FRG, June 2-8, 1991 pp. 171–194 (1992). DOI 10.1007/978-3-642-51682-5_12
11. Lancaster, P.: On eigenvalues of matrices dependent on a parameter. Numerische Mathematik **6**(1), 377–387 (1964)
12. Li, X., Sun, D., Toh, K.C.: A highly efficient semismooth newton augmented lagrangian method for solving lasso problems. SIAM Journal on Optimization **28**(1), 433–458 (2018)
13. Li, X., Sun, D., Toh, K.C.: Qsdpnal: A two-phase augmented lagrangian method for convex quadratic semidefinite programming. Mathematical Programming Computation **10**, 703–743 (2018)
14. Meng, F., Sun, D., Zhao, G.: Semismoothness of solutions to generalized equations and the Moreau-Yosida regularization. Mathematical Programming **104**(2-3), 561 – 581 (2005)
15. Mifflin, R.: Semismooth and semiconvex functions in constrained optimization. SIAM Journal on Control and Optimization **15**(6), 959–972 (1977)
16. Mordukhovich, B.S.: Variational analysis and generalized differentiation II: Applications, vol. 331. Springer (2006)
17. Pang, J.S., Sun, D., Sun, J.: Semismooth homeomorphisms and strong stability of semidefinite and lorentz complementarity problems. Mathematics of Operations Research **28**(1), 39–63 (2003)
18. Qi, H., Sun, D.: A quadratically convergent newton method for computing the nearest correlation matrix. SIAM journal on matrix analysis and applications **28**(2), 360–385 (2006)
19. Qi, L.: Convergence analysis of some algorithms for solving nonsmooth equations. Mathematics of operations research **18**(1), 227–244 (1993)
20. Qi, L., Sun, D.: A Survey of Some Nonsmooth Equations and Smoothing Newton Methods, pp. 121–146. Springer US, Boston, MA (1999). DOI 10.1007/978-1-4613-3285-5_7. URL `https://doi.org/10.1007/978-1-4613-3285-5_7`
21. Qi, L., Sun, J.: A nonsmooth version of newton's method. Mathematical programming **58**(1-3), 353–367 (1993)
22. Rademacher, H.: über partielle und totale differenzierbarkeit von funktionen mehrerer variabeln und über die transformation der doppelintegrale. Mathematische Annalen **79**(2), 340–359 (1919)
23. Robinson, S.M.: Strongly regular generalized equations. Mathematics of Operations Research **5**(1), 43–62 (1980)
24. Rockafellar, R.T.: Convex analysis, vol. 11. Princeton university press (1970)
25. Rockafellar, R.T., Wets, R.J.B.: Variational analysis, vol. 317. Springer Science & Business Media (1997)
26. Sun, D.: The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. Mathematics of Operations Research **31**(4), 761–776 (2006)
27. Sun, D., Sun, J.: Semismooth matrix-valued functions. Mathematics of Operations Research **27**(1), 150–169 (2002)
28. Sun, D., Sun, J.: Strong semismoothness of eigenvalues of symmetric matrices and its application to inverse eigenvalue problems. SIAM Journal on Numerical Analysis **40**(6), 2352–2367 (2002)
29. Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology **58**(1), 267–288 (1996)
30. Ulbrich, M.: Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces. SIAM (2011)
31. Wolfe, P.: A duality theorem for non-linear programming. Quarterly of applied mathematics **19**(3), 239–244 (1961)
32. Yang, L., Sun, D., Toh, K.C.: Sdpnal+: a majorized semismooth newton-cg augmented lagrangian method for semidefinite programming with nonnegative constraints. Mathematical Programming Computation **7**(3), 331–366 (2015)
33. Zhao, X.Y., Sun, D., Toh, K.C.: A newton-cg augmented lagrangian method for semidefinite programming. SIAM Journal on Optimization **20**(4), 1737–1765 (2010)
34. Zhou, Z., So, A.M.C.: A unified approach to error bounds for structured convex optimization problems. Mathematical Programming **165**, 689–728 (2017)