CONSTRAINT NONDEGENERACY, STRONG REGULARITY AND NONSINGULARITY IN SEMIDEFINITE PROGRAMMING*

ZI XIAN CHAN† AND DEFENG SUN‡

Abstract. It is known that the Karush-Kuhn-Tucker (KKT) conditions of semidefinite programming can be reformulated as a nonsmooth system via the metric projector over the cone of symmetric and positive semidefinite matrices. We show in this paper that the primal and dual constraint nondegeneracies, the strong regularity, the nonsingularity of the B-subdifferential of this nonsmooth system, and the nonsingularity of the corresponding Clarke's generalized Jacobian, at a KKT point are all equivalent. Moreover, we prove the equivalence between each of these conditions and the nonsingularity of Clarke's generalized Jacobian of the smoothed counterpart of this nonsmooth system used in several globally convergent smoothing Newton methods. In particular, we establish the quadratic convergence of these methods under the primal and dual constraint nondegeneracies, but without the strict complementarity.

Key words. Semidefinite programming, constraint nondegeneracy, strong regularity, nonsingularity, variational analysis, quadratic convergence

AMS subject classifications. 90C22, 90C25, 90C31, 65K05, 65K10

1. Introduction.

The standard semidefinite programming (SDP) problem takes the following form

$$\begin{array}{ll} \min & \langle C, X\rangle \\ \text{s.t.} & \mathcal{A}X = b\,, \\ & X \in \mathcal{S}^n_+\,, \end{array} \tag{1.1}$$

where $C \in \mathcal{S}^n$, the linear space of all $n \times n$ real symmetric matrices, $\langle \cdot, \cdot \rangle$ is the usual Frobenius inner product in $\mathcal{S}^n$, $\mathcal{A}$ is a linear operator from $\mathcal{S}^n$ to $\Re^m$, $b \in \Re^m$, and $\mathcal{S}^n_+$ is the cone of all $n \times n$ positive semidefinite matrices in $\mathcal{S}^n$. Let $\mathcal{A}^* : \Re^m \to \mathcal{S}^n$ be the adjoint of $\mathcal{A}$. The dual form of the SDP problem (1.1) is

$$\begin{array}{ll} \max & b^T y \\ \text{s.t.} & \mathcal{A}^* y + S = C\,, \\ & S \in \mathcal{S}^n_+\,. \end{array} \tag{1.2}$$

The Karush-Kuhn-Tucker (KKT) conditions, i.e., the first order optimality conditions, for the SDP problem (1.1) and its dual (1.2), are

$$\left\{ \begin{array}{l} \mathcal{A}^* y + S = C\,, \\ \mathcal{A}X = b\,, \\ \mathcal{S}^n_+ \ni X \perp S \in \mathcal{S}^n_+\,, \end{array} \right. \tag{1.3}$$

where "$X \perp S$" means that X and S are perpendicular to each other, i.e., $\langle X, S\rangle = 0$. Any point $(\overline{X}, \bar{y}, \overline{S}) \in \mathcal{S}^n \times \Re^m \times \mathcal{S}^n$ satisfying (1.3) is called a KKT point.

Due to its mathematical elegance and wide applications, the research on SDP has been extremely active after the discovery of polynomial time interior point algorithms

*This research was partially supported by the Academic Research Fund under Grant No. R-146-000-104-112 of the National University of Singapore.

†Department of Mathematics, National University of Singapore, Republic of Singapore (u0301479@alumni.nus.edu.sg).

‡Department of Mathematics and Risk Management Institute, National University of Singapore, Republic of Singapore (matsundf@nus.edu.sg).

[1, 26] for solving this problem. For an excellent survey on this, see [46]. Our research in this paper is motivated by [41] on various characterizations of strong regularity, one of the most important concepts in sensitivity and perturbation analysis, introduced by Robinson in his seminal paper [31], for a local optimal solution of the general nonlinear SDP problem. The basic question we want to ask here is:

What does the strong regularity mean for the SDP problem (1.1) *and its dual* (1.2)?

Certainly, all conditions equivalent to the strong regularity presented in [41] for the general nonlinear SDP problem apply to the SDP problem (1.1) too. However, due to the special structure of the SDP problem (1.1) and its dual, one may be able to obtain more insightful characterizations about the strong regularity. This is exactly the primary objective of this paper.

For the purpose of achieving this objective, we study the B-subdifferential and Clarke's generalized Jacobian of the nonsmooth system reformulated from (1.3). We show that the primal and dual constraint nondegeneracies, the strong regularity, the nonsingularity of the B-subdifferential of this nonsmooth system, and the nonsingularity of the corresponding Clarke's generalized Jacobian, at a KKT point $(\overline{X}, \bar{y}, \overline{S}) \in \mathcal{S}^n \times \Re^m \times \mathcal{S}^n$ are all equivalent. The equivalence of the nonsingularity of the B-subdifferential and the nonsingularity of Clarke's generalized Jacobian comes as a surprise, at least to the authors, as we know that the nonsingularity of the B-subdifferential is only a necessary condition for the strong regularity while the nonsingularity of Clarke's generalized Jacobian is a sufficient condition for the strong regularity (for more discussions, see [15, 27]). It is true, by [41, Theorem 4.1], that the nonsingularity of Clarke's generalized Jacobian is also necessary for the strong regularity in the context of SDP problems. However, it is never known if the nonsingularity of the B-subdifferential is sufficient too. Here, the unique structure exhibited in SDP problems (1.1) and (1.2) plays a key role for us to prove these conditions equivalent. Consequently, the quadratic convergence of some local nonsmooth Newton type methods studied in [18, 14] follows from any one of these equivalent conditions. In fact, by combining the two papers [18, 14], we know that the primal and dual constraint nondegeneracies are sufficient for the nonsingularity of the B-subdifferential. On the other hand, our equivalent results imply that they are also necessary for the nonsingularity of the B-subdifferential.

The second objective, largely motivated by the first one, of this paper is to study under what conditions the globally convergent smoothing Newton methods studied in [9, 10, 19, 45] for solving SDP problems (1.1) and (1.2) possess local quadratic convergence, without assuming the strict complementary condition. We achieve this objective by showing that the nonsingularity of the B-subdifferential of one smoothed system used in [9, 10, 19, 45] and the nonsingularity of Clarke's generalized Jacobian of this smoothed system are both equivalent to any of the above stated equivalent conditions, in particular, the primal and dual constraint nondegeneracies.

The organization of this paper is as follows. In Section 2, we study some useful properties of the B-subdifferential and Clarke's generalized Jacobian for Lipschitz functions, in particular for the metric projector over $\mathcal{S}^n_+$ and its smoothed counterpart. The promised equivalent conditions are given in Section 3. In Section 4, we prove the quadratic convergence of some smoothing Newton methods under the primal and dual constraint nondegenerate conditions, but without the strict complementarity condition. We give our conclusions in Section 5.

2. Generalized Jacobians. Assume that $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ are three finite dimensional real vector spaces each equipped with a scalar product $\langle\cdot,\cdot\rangle$ and its induced norm $\|\cdot\|$, $\mathcal{O}$ is an open set in $\mathcal{Y}$, and $\Xi : \mathcal{O} \subseteq \mathcal{Y} \to \mathcal{Z}$ is a locally Lipschitz continuous function on the open set $\mathcal{O}$. By the well known Rademacher's theorem [36, Section 9.J], we know that Ξ is almost everywhere F(réchet)-differentiable in $\mathcal{O}$. Denote by $\mathcal{D}_\Xi$ the set of all points in $\mathcal{O}$ where Ξ is F-differentiable. Then Clarke's generalized Jacobian of Ξ at $y \in \mathcal{O}$ is defined as follows [12]:

$$\partial\Xi(y) := \text{conv}\{\, \partial_B\Xi(y)\, \},$$

where "conv" denotes the convex hull and the B-subdifferential $\partial_B\Xi(y)$, a name coined by Qi in [28], of Ξ at y takes the form

$$\partial_B\Xi(y) := \{V \,:\, V = \lim_{k\to\infty} \Xi'(y^k)\,,\; y^k \to y\,,\; y^k \in \mathcal{D}_\Xi\}.$$

The next lemma, which is originally proven in [41, Lemma 2.1] under the additional assumption of directional diffrentiability, is a useful property about characterizing the B-subdifferential of composite functions. Here we drop the condition of directional differentiability and provide a self-contained proof as it may have applications in other places where the directional differentiability is not readily available.

LEMMA 2.1. *Let $\Psi : \mathcal{X} \to \mathcal{Y}$ be a continuously differentiable function on an open neighborhood $\widehat{N}$ of $\bar{x}$ and $\Xi : \mathcal{O} \subseteq \mathcal{Y} \to \mathcal{Z}$ be a locally Lipschitz continuous function on an open set $\mathcal{O}$ containing $\bar{y} := \Psi(\bar{x})$. Define $\Phi : \widehat{N} \to \mathcal{Z}$ by defined by $\Phi(x) := \Xi(\Psi(x))$, $x \in \widehat{N}$. Suppose that $\Psi'(\bar{x}) : \mathcal{X} \to \mathcal{Y}$ is onto. Then there exists an open neighborhood of $\bar{x}$ such that Φ is F-differentiable at x in this neighborhood if and only if Ξ is F-differentiable at $\Psi(x)$ and*

$$\partial_B\Phi(\bar{x}) = \partial_B\Xi(\bar{y})\Psi'(\bar{x})\,. \tag{2.1}$$

Proof. Shrink $\widehat{N}$, if necessary, assume that $\Psi(\widehat{N}) \subseteq \mathcal{O}$ and for each $x \in \widehat{N}$, $\Psi'(x)$ is onto. Then Φ is Lipschitz continuous on $\widehat{N}$.

We shall first show that Φ is F-differentiable at $x \in \widehat{N}$ if and only if Ξ is F-differentiable at $\Psi(x)$, which, by the definition of the B-subdifferential, implies

$$\partial_B\Phi(\bar{x}) \subseteq \partial_B\Xi(\bar{y})\Psi'(\bar{x}).$$

By the definition of Φ, we know that if Ξ is F-differentiable at $\Psi(x)$, then Φ is F-differentiable at $x \in \widehat{N}$. Now, assume that Φ is F-differentiable at $x \in \widehat{N}$. Since $A := \Psi'(x)$ is onto, AA^* is invertible, where $A^* : \mathcal{Y} \to \mathcal{X}$ is the adjoint of A. For any $\Delta y \in \mathcal{Y}$, let

$$\Delta x := A^*(AA^*)^{-1}\Delta y\,.$$

Then, for any $\mathcal{Y} \ni \Delta y \to 0$, we have

$$\begin{aligned}
&\|\Xi(\Psi(x) + \Delta y) - \Xi(\Psi(x)) - \Phi'(x)(A^*(AA^*)^{-1}\Delta y)\| \\
&= \|\Xi(\Psi(x) + A\Delta x) - \Phi(x) - \Phi'(x)(\Delta x)\| \\
&\le \|\Phi(x + \Delta x) - \Phi(x) - \Phi'(x)(\Delta x)\| + O(\|\Xi(\Psi(x + \Delta x)) - \Xi(\Psi(x) + A\Delta x)\|) \\
&= o(\|\Delta x\|) + O(\|\Psi(x + \Delta x) - (\Psi(x) + A\Delta x)\|) \\
&= o(\|\Delta x\|) + O(\|\Psi(x + \Delta x) - \Psi(x) - \Psi'(x)(\Delta x)\|) \\
&= o(\|\Delta x\|) = o(\|\Delta y\|)\,,
\end{aligned}$$

which, implies that Ξ is F-differentiable at $\Psi(x)$. This proves the first part of our conclusion.

Next, we show that the following inclusion holds:

$$\partial_B \Phi(\bar{x}) \supseteq \partial_B \Xi(\bar{y}) \Psi'(\bar{x}).$$

This part's proof follows exactly the proof of the second part of Lemma 2.1 in [41]. Let $W \in \partial_B \Xi(\bar{y})$ be an arbitrary element. Then there exists a sequence $\{y^k\}$ in $\mathcal{O}$ converging to $\bar{y}$ such that Ξ is F-differentiable at y^k and $W = \lim_{k\to\infty} \Xi'(y^k)$. Let $\overline{A} := \Psi'(\bar{x})$. By applying the classical Inverse Function Theorem to

$$\Psi\big(\bar{x} + \overline{A}^*(y - \bar{y})\big) - \Psi(\bar{x}) = 0,$$

we obtain that there exists a sequence $\{\tilde{y}^k\}$ in $\mathcal{O}$ converging to $\bar{y}$ such that

$$\Psi\big(\bar{x} + \overline{A}^*(\tilde{y}^k - \bar{y})\big) - \Psi(\bar{x}) = y^k - \Psi(\bar{x})$$

for all k sufficiently large. Let $\tilde{x}^k := \bar{x} + \overline{A}^*(\tilde{y}^k - \bar{y})$. Then $y^k = \Psi(\tilde{x}^k)$ and Φ is F-differentiable at $\tilde{x}^k$ with

$$\Phi'(\tilde{x}^k) = \Xi'(y^k)\Psi'(\tilde{x}^k).$$

By using the fact that $\tilde{y}^k \to \bar{y}$ implies $\tilde{x}^k \to \bar{x}$, we know that there exists a $V \in \partial_B \Phi(\bar{x})$ such that

$$W\Psi'(\bar{x}) = \lim_{k\to\infty} \Xi'(y^k) \lim_{k\to\infty} \Psi'(\tilde{x}^k) = \lim_{k\to\infty} \Phi'(\tilde{x}^k) = V \in \partial_B \Phi(\bar{x}).$$

The proof is completed. □

For any nonempty closed convex set $K \subseteq \mathcal{Z}$, let $\Pi_K : \mathcal{Z} \to \mathcal{Z}$ denote the metric projector over K. That is, for any $y \in \mathcal{Z}$, $\Pi_K(y)$ is the unique optimal solution to the convex programming problem:

$$\begin{array}{ll} \min & \frac{1}{2}\langle z - y, z - y\rangle \\ \text{s.t.} & z \in K. \end{array} \tag{2.2}$$

Since the metric projector $\Pi_K(\cdot)$ is globally Lipschitz continuous with modulus 1 [48], $\Pi_K(\cdot)$ is F-differentiable almost everywhere in $\mathcal{Z}$. Thus, for any $y \in Z$, $\partial \Pi_K(y)$ is well defined. In particular, it is shown in [24, Proposition 1] that for any $y \in \mathcal{Z}$, $V \in \partial \Pi_K(y)$ is self-adjoint and satisfies

$$V \succeq V^2, \quad \text{i.e.,} \ \langle d, Vd\rangle \geq \big\langle d, V^2 d\big\rangle \quad \forall\, d \in \mathcal{Z}\,. \tag{2.3}$$

In our subsequent analysis, we need a finer characterization about the B-subdifferential and Clarke's generalized Jacobian of $\Pi_{\mathcal{S}^n_+}(\cdot)$ and its smoothed counterpart. We write $A \succeq 0$ and $A \succ 0$ to mean that A is a symmetric positive semidefinite matrix and a symmetric positive definite matrix, respectively. For any $A \in \mathcal{S}^n$, let $A_+ := \Pi_{\mathcal{S}^n_+}(A)$ be the metric projection of A onto $\mathcal{S}^n_+$ under the usual Frobenius inner product in $\mathcal{S}^n$. Assume that A has the following spectral decomposition

$$A = P\Lambda P^T, \tag{2.4}$$

where Λ is the diagonal matrix of eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ of A and P is a corresponding orthogonal matrix of orthonormal eigenvectors. Then

$$A_+ = P\Lambda_+ P^T,$$

where Λ_+ is the diagonal matrix whose diagonal entries are the nonnegative parts of the respective diagonal entries of Λ. The formula for A_+ has been used by statisticians for several decades, e.g., [37, Theorem 1]. Higham [16] and Tseng [47] brought it to the attention of the optimization community. Define three index sets of positive, zero, and negative eigenvalues of A, respectively, as

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

Write

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

with $P_\alpha \in \Re^{n\times|\alpha|}$, $P_\beta \in \Re^{n\times|\beta|}$, and $P_\gamma \in \Re^{n\times|\gamma|}$. For this eigenvalue vector $\lambda \in \Re^n$, define the corresponding symmetric matrix $U \in \mathcal{S}^n$ with entries

$$U_{ij} := \frac{\max\{\lambda_i, 0\} + \max\{\lambda_j, 0\}}{|\,\lambda_i\,| + |\,\lambda_j\,|}, \quad i, j = 1, \ldots, n, \tag{2.5}$$

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

We know from Bonnans, Cominetti, and Shapiro [5, 6] that $\Pi_{\mathcal{S}^n_+}$ is directionally differentiable everywhere in $\mathcal{S}^n$, and from Sun and Sun [42], $\Pi_{\mathcal{S}^n_+}$ is strongly semismooth everywhere in $\mathcal{S}^n$ and the directional derivative $\Pi'_{\mathcal{S}^n_+}(A; H)$ of $\Pi_{\mathcal{S}^n_+}$ at A with direction $H \in \mathcal{S}^n$ is given by

$$\Pi'_{\mathcal{S}^n_+}(A; H) = P \begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & U_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} \\ \widetilde{H}^T_{\alpha\beta} & \Pi_{\mathcal{S}^{|\beta|}_+}(\widetilde{H}_{\beta\beta}) & 0 \\ \widetilde{H}^T_{\alpha\gamma} \circ U^T_{\alpha\gamma} & 0 & 0 \end{bmatrix} P^T, \tag{2.6}$$

where $\widetilde{H} := P^T H P$ and "$\circ$" denotes the Hadamard product. For a general discussion on (strongly) semismooth functions, see [25, 28, 30]. The tangent cone of $\mathcal{S}^n_+$ at A_+, in the sense of convex analysis [35], can be characterized as

$$\mathcal{T}_{\mathcal{S}^n_+}(A_+) = \{B \in \mathcal{S}^n \,:\, B = \Pi'_{\mathcal{S}^n_+}(A_+; B)\} = \{B \in \mathcal{S}^n \,:\, [\,P_\beta \; P_\gamma\,]^T B\, [\,P_\beta \; P_\gamma\,] \succeq 0\}\,.$$

Note, however, that the characterization of $\mathcal{T}_{\mathcal{S}^n_+}(A_+)$ was first obtained by Arnold [3] without using the directional derivative $\Pi'_{\mathcal{S}^n_+}(A_+; H)$. The lineality space of $\mathcal{T}_{\mathcal{S}^n_+}(A_+)$, i.e., the largest linear space in $\mathcal{T}_{\mathcal{S}^n_+}(A_+)$, denoted by $\mathrm{lin}\left(\mathcal{T}_{\mathcal{S}^n_+}(A_+)\right)$, then takes the following form:

$$\mathrm{lin}\left(\mathcal{T}_{\mathcal{S}^n_+}(A_+)\right) = \{B \in \mathcal{S}^n \,:\, P^T_\beta B P_\beta = 0, \;\; P^T_\beta B P_\gamma = 0, \;\; P^T_\gamma B P_\gamma = 0\}\,. \tag{2.7}$$

The critical cone of $\mathcal{S}^n_+$ at $A \in \mathcal{S}^n$, associated with the convex optimization problem (2.2) with $K = \mathcal{S}^n_+$, is defined as

$$\begin{aligned} \mathcal{C}(A; \mathcal{S}^n_+) : &= \mathcal{T}_{\mathcal{S}^n_+}(A_+) \cap (A_+ - A)^\perp \\ &= \{B \in \mathcal{S}^n \,:\, P^T_\beta B P_\beta \succeq 0, \; P^T_\beta B P_\gamma = 0, \; P^T_\gamma B P_\gamma = 0\}\,, \end{aligned} \tag{2.8}$$

where $(A_+ - A)^\perp := \{B \in \mathcal{S}^n : \langle B, A_+ - A\rangle = 0\}$. Therefore, the affine hull of $\mathcal{C}(A;\mathcal{S}^n_+)$, which we denote $\mathrm{aff}(\mathcal{C}(A;\mathcal{S}^n_+))$, can be written as

$$\mathrm{aff}\left(\mathcal{C}(A;\mathcal{S}^n_+)\right) = \left\{B \in \mathcal{S}^n \,:\, P_\beta^T B P_\gamma = 0,\; P_\gamma^T B P_\gamma = 0\right\}. \tag{2.9}$$

In the case that $\beta = \emptyset$ holds, i.e., the case that A is nonsingular, $\Pi_{\mathcal{S}^n_+}(\cdot)$ is F-differentiable at A and (2.6) reduces to the famous result of Löwner [21]:

$$\Pi'_{\mathcal{S}^n_+}(A)H = P\begin{bmatrix} \widetilde{H}_{\alpha\alpha} & U_{\alpha\gamma}\circ \widetilde{H}_{\alpha\gamma} \\ \widetilde{H}_{\alpha\gamma}^T \circ U_{\alpha\gamma}^T & 0 \end{bmatrix} P^T \quad \forall\, H \in \mathcal{S}^n. \tag{2.10}$$

From (2.10), one may compute the B-subdifferential and Clarke's generalized Jacobian of $\Pi_{\mathcal{S}^n_+}(\cdot)$ by their definitions.[1] This has been done by a number of authors [9, 19, 22, 23, 27]. One difficulty in obtaining good formulas for $\partial_B \Pi_{\mathcal{S}^n_+}(A)$ and $\partial \Pi_{\mathcal{S}^n_+}(A)$ is that they both depend on the orthogonal matrices P in the spectral decomposition of A. This difficulty can be overcome by employing the following link developed by Pang et al. [27] on $\partial_B \Pi_{\mathcal{S}^n_+}(A)$ and the B-subdifferential of $\Theta(\cdot) := \Pi'_{\mathcal{S}^n_+}(A;\cdot)$ at the origin

$$\partial_B \Pi_{\mathcal{S}^n_+}(A) = \partial_B \Theta(0). \tag{2.11}$$

This link leads to the following useful result on $\partial_B \Pi_{\mathcal{S}^n_+}(A)$ and $\partial \Pi_{\mathcal{S}^n_+}(A)$. See Sun [41, Proposition 2.2] for a short proof.

PROPOSITION 2.2. *Suppose that $A \in \mathcal{S}^n$ has the spectral decomposition as in* (2.4). *Then a $V \in \partial_B \Pi_{\mathcal{S}^n_+}(A)$ (respectively, $\partial \Pi_{\mathcal{S}^n_+}(A)$) if and only if there exists a $V_{|\beta|} \in \partial_B \Pi_{\mathcal{S}^{|\beta|}_+}(0)$ (respectively, $\partial \Pi_{\mathcal{S}^{|\beta|}_+}(0)$) such that*

$$V(H) = P\begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & U_{\alpha\gamma}\circ \widetilde{H}_{\alpha\gamma} \\ \widetilde{H}_{\alpha\beta}^T & V_{|\beta|}(\widetilde{H}_{\beta\beta}) & 0 \\ \widetilde{H}_{\alpha\gamma}^T \circ U_{\alpha\gamma}^T & 0 & 0 \end{bmatrix} P^T \quad \forall\, H \in \mathcal{S}^n, \tag{2.12}$$

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

Proposition 2.2 simply says that in order to compute $\partial_B \Pi_{\mathcal{S}^n_+}(A)$ and $\partial \Pi_{\mathcal{S}^n_+}(A)$, one only needs to fix an arbitrary orthogonal matrix P satisfying (2.4) and compute the corresponding "caged" part $\partial_B \Pi_{\mathcal{S}^{|\beta|}_+}(0)$ (hence $\partial \Pi_{\mathcal{S}^{|\beta|}_+}(0)$), which is much easier to handle. To see this, let $\mathcal{Q}_{|\beta|}$ be the set of all orthogonal matrices of order $|\beta| \times |\beta|$ and

$$\Re^{|\beta|}_> := \left\{z \in \Re^{|\beta|} \,:\, z_1 \geq \cdots \geq z_{|\beta|} \ \text{ and } z_i \neq 0 \ \forall\, i\right\}.$$

Let $p : \Re \to \Re$ be the "plus" function defined by $p(t) \equiv \max(0,t)$, $t \in \Re$. For any $z \in \Re^{|\beta|}_>$, let $p^{[1]}(z)$ represent the first divided difference matrix used in matrix analysis

[1] Note that in numerical computations it is generally impossible to compute exactly the spectral decomposition of A as in (2.4). Instead, the right hand side of (2.4) is the true spectral decomposition of a nearby matrix of A [8]. Consequently, the numerically computed subdifferentials are actually for this nearby matrix. In this paper, we will not address this numerical issue further.

for $p(\cdot)$ at z [4]

$$(2.13)\qquad \big[\, p^{[1]}(z)\,\big]_{ij} = \begin{cases} \dfrac{p(z_i)-p(z_j)}{z_i - z_j} \in [0,1] & \text{if } z_i \neq z_j \\ p'(z_i) \in \{0,1\} & \text{if } z_i = z_j \end{cases}, \quad i,j = 1,\ldots,n\,.$$

Then, by (2.6) and (2.10), one can readily draw the conclusion that $V_{|\beta|} \in \partial_B \Pi_{\mathcal{S}_+^{|\beta|}}(0)$ if and only if there exist $Q \in \mathcal{Q}_{|\beta|}$ and $\Omega \in \mathcal{U}_{|\beta|}$ such that

$$(2.14)\qquad V_{|\beta|}(Z) = Q\,[\,\Omega \circ (Q^T Z Q)\,]\,Q^T \quad \forall\, Z \in \mathcal{S}^{|\beta|}\,,$$

where

$$\mathcal{U}_{|\beta|} := \big\{ \Omega \,:\, \Omega = \lim_{k\to\infty} p^{[1]}(z^k), \quad z^k \to 0,\ z^k \in \Re^{|\beta|}_{>} \,\big\}\,.$$

In [22], Malick and Sendov gave a detailed account on the structure of $\mathcal{U}_{|\beta|}$. In this paper, we do not need the exact structure of $\mathcal{U}_{|\beta|}$ except the following fact that for any $\Omega \in \mathcal{U}_{|\beta|}$,

$$\Omega_{ij} \in [0,1], \quad i,j = 1,\ldots,|\beta|\,.$$

Note that both the zero mapping $V^0_{|\beta|} \equiv 0$ and the identity mapping $V^{\mathcal{I}}_{|\beta|} = \mathcal{I}$ from $\mathcal{S}^{|\beta|} \to \mathcal{S}^{|\beta|}$ are elements in $\partial_B \Pi_{\mathcal{S}_+^{|\beta|}}(0)$. Let V^0 and $V^{\mathcal{I}}$ be defined by (2.12) with $V_{|\beta|}$ being replaced by $V^0_{|\beta|}$ and $V^{\mathcal{I}}_{|\beta|}$, respectively. Define

$$(2.15)\qquad \mathrm{ex}\,(\partial_B \Pi_{\mathcal{S}^n_+}(A)) := \big\{\, V^0,\ V^{\mathcal{I}} \,\big\}.$$

Using the fact that both V^0 and $V^{\mathcal{I}}$ are elements in $\partial_B \Pi_{\mathcal{S}^n_+}(A)$, we have

$$\mathrm{ex}\,(\partial_B \Pi_{\mathcal{S}^n_+}(A)) \subseteq \partial_B \Pi_{\mathcal{S}^n_+}(A)\,.$$

Since $\Pi_{\mathcal{S}^n_+}(\cdot)$ is not differentiable everywhere, several papers [9, 10, 19, 45] on smoothing Newton methods, for solving the SDP problem and beyond, consider the following smoothed counterpart of $\Pi_{\mathcal{S}^n_+}(\cdot)$:

$$(2.16)\qquad \Phi(\varepsilon, A) := \big[\, A + \sqrt{\varepsilon^2 I + A^2}\,\big]/2, \quad (\varepsilon, A) \in \Re \times \mathcal{S}^n\,,$$

where we use I to represent the identity matrix of appropriate dimension. Note that the function $\Phi(\cdot,\cdot)$ is continuously differentiable around any $(\varepsilon, A) \in \Re\times\mathcal{S}^n$ if $\varepsilon^2 I + A^2$ is nonsingular and when $\varepsilon = 0$, $\Phi(0, A) = \Pi_{\mathcal{S}^n_+}(A)$. Further more, $\Phi(\cdot,\cdot)$ is globally Lipschitz continuous and strongly semismooth at any $(0, A) \in \Re \times \mathcal{S}^n$ [45]. For some extensions on these properties, see [43].

Let $\phi : \Re^2 \to \Re$ be defined by

$$\phi(\varepsilon, t) = \big[\, t + \sqrt{\varepsilon^2 + t^2}\,\big]/2, \quad (\varepsilon, t) \in \Re \times \Re\,.$$

Let A have the spectral decomposition as in (2.4). Then, by matrix analysis [4, 17], we have

$$\Phi(\varepsilon, A) = P \begin{bmatrix} \phi(\varepsilon,\lambda_1) & & \\ & \ddots & \\ & & \phi(\varepsilon,\lambda_n) \end{bmatrix} P^T\,.$$

For any $(\varepsilon, x) \in \Re \times \Re^n$ such that $\varepsilon^2 + x_i^2 > 0$ for all i, we use $\widehat{U}(\varepsilon, x) \in \mathcal{S}^n$ to represent the first divided difference matrix for $\phi(\varepsilon, \cdot)$ at x given by

$$(2.17)\quad \big[\,\widehat{U}(\varepsilon, x)\,\big]_{ij} = \begin{cases} \dfrac{\phi(\varepsilon, x_i) - \phi(\varepsilon, x_j)}{x_i - x_j} \in [0,1] & \text{if } x_i \neq x_j \\ \phi'_{x_i}(\varepsilon, x_i) \in [0,1] & \text{if } x_i = x_j \end{cases}, \; i, j = 1, \ldots, n\,.$$

Then, according to Lemma 2.3 in [45], we know that for any $\varepsilon \in \Re$ such that $\varepsilon^2 + \lambda_i^2 > 0$ for all i (i.e., $\varepsilon^2 I + A^2$ is nonsingular), and any $(\tau, H) \in \Re \times \mathcal{S}^n$, we have

$$(2.18)\qquad \Phi'(\varepsilon, A)(\tau, H) = P\,[\,\widehat{U}(\varepsilon, \lambda) \circ (P^T H P) + \tau D(\varepsilon, \lambda)\,]\,P^T$$

and

$$(2.19)\quad \Phi'((0, A); (\tau, H)) = P \begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & U_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} \\ \widetilde{H}_{\alpha\beta}^T & \Phi_{|\beta|}(\tau, \widetilde{H}_{\beta\beta}) & 0 \\ \widetilde{H}_{\alpha\gamma}^T \circ U_{\alpha\gamma}^T & 0 & 0 \end{bmatrix} P^T,$$

where $\widetilde{H} = P^T H P$, $D(\varepsilon, \lambda) \in \mathcal{S}^n$ is the diagonal matrix given by

$$(2.20)\qquad D(\varepsilon, \lambda) = \begin{bmatrix} \phi'_\varepsilon(\varepsilon, \lambda_1) & & \\ & \ddots & \\ & & \phi'_\varepsilon(\varepsilon, \lambda_n) \end{bmatrix},$$

$U \in \mathcal{S}^n$ is defined by (2.5), and for any $(t, Z) \in \Re \times \mathcal{S}^{|\beta|}$,

$$(2.21)\qquad \Phi_{|\beta|}(t, Z) := \big[\, Z + \sqrt{t^2 I + Z^2}\,\big]/2\,.$$

Define $\Psi : \Re \times \mathcal{S}^n \to \Re \times \mathcal{S}^n$ by

$$\Psi(\tau, H) := (\tau, P^T H P), \quad (\tau, H) \in \Re \times \mathcal{S}^n$$

and $\Xi : \Re \times \mathcal{S}^n \to \mathcal{S}^n$ by

$$(2.22)\qquad \Xi(t, M) := P \begin{bmatrix} M_{\alpha\alpha} & M_{\alpha\beta} & U_{\alpha\gamma} \circ M_{\alpha\gamma} \\ M_{\alpha\beta}^T & \Phi_{|\beta|}(t, M_{\beta\beta}) & 0 \\ M_{\alpha\gamma}^T \circ U_{\alpha\gamma}^T & 0 & 0 \end{bmatrix} P^T,$$

where $(t, M) \in \Re \times \mathcal{S}^n$. Write $\Gamma(\cdot, \cdot) \equiv \Phi'((0, A); (\cdot, \cdot))$. Then, we have

$$(2.23)\qquad \Gamma(\tau, H) = \Xi(\Psi(\tau, H)), \quad (\tau, H) \in \Re \times \mathcal{S}^n\,.$$

Since for any $(\tau, H) \in \Re \times \mathcal{S}^n$, $\Psi'(\tau, H) : \Re \times \mathcal{S}^n \to \Re \times \mathcal{S}^n$ is onto, we know from the first part of Lemma 2.1 that Γ is F-differentiable at $(\tau, H) \in \Re \times \mathcal{S}^n$ if and only if Ξ is F-differentiable at $\Psi(\tau, H)$, which is equivalent to the nonsingularity of $\tau^2 I + (\widetilde{H}_{\beta\beta})^2$, where $\widetilde{H} = P^T H P$. Thus, we have

Lemma 2.3. *For any $(\tau, H) \in \Re \times \mathcal{S}^n$, let $\widetilde{H} = P^T H P$. Then $\Gamma(\cdot, \cdot) \equiv \Phi'((0, A); (\cdot, \cdot))$ is F-differentiable at $(\tau, H) \in \Re \times \mathcal{S}^n$ if and only if $\tau^2 I + (\widetilde{H}_{\beta\beta})^2$ is nonsingular.*

The following lemma establishes the equivalence between $\partial_B\Phi(0,A)$ and $\partial_B\Gamma(0,0)$, which is analogous to (2.11) for operators $\Pi_{\mathcal{S}^n_+}$ and Θ. Its proof largely follows that given in [27, Lemma 11], but with new difficulties to overcome.

LEMMA 2.4. *Suppose that $A \in \mathcal{S}^n$ has the spectral decomposition as in* (2.4). *For $\Gamma(\cdot,\cdot) \equiv \Phi'((0,A);(\cdot,\cdot))$, it holds that*

$$\partial_B\Phi(0,A) = \partial_B\Gamma(0,0)\,. \tag{2.24}$$

Proof. Let $V \in \partial_B\Phi(0,A)$. Then, by (2.18), (2.19), and the definition of $\partial_B\Phi(0,A)$, there exists a sequence $\{(\varepsilon_k, A^k)\}$ in $\Re \times \mathcal{S}^n$ converging to $(0,A)$ with $\varepsilon_k^2 I + (A^k)^2$ being nonsingular such that $V = \lim_{k\to\infty} \Phi'(\varepsilon_k, A^k)$. Let $A^k \equiv P^k\Lambda^k(P^k)^T$ be the orthogonal decomposition of A^k, where Λ^k is the diagonal matrix whose diagonal entries are the eigenvalues $\lambda_1^k \geq \cdots \geq \lambda_n^k$ of A^k and P^k is a corresponding matrix of orthonormal eigenvectors. Writing each Λ^k in the same form as Λ:

$$\Lambda^k = \begin{bmatrix} \Lambda_\alpha^k & 0 & 0 \\ 0 & \Lambda_\beta^k & 0 \\ 0 & 0 & \Lambda_\gamma^k \end{bmatrix},$$

we have $\Lambda = \lim_{k\to\infty} \Lambda^k$, which implies that Λ_α^k and Λ_γ^k are nonsingular matrices for all k sufficiently large and $\lim_{k\to\infty} \Lambda_\beta^k = 0$. For each k, let $U^k \equiv \widehat{U}(\varepsilon_k, \lambda^k)$ be defined by (2.17) and $D^k \equiv D(\varepsilon_k, \lambda^k)$ be defined by (2.20), respectively. Then, for an arbitrarily chosen $(\tau, H) \in \Re \times \mathcal{S}^n$ with $\widetilde{H}^k = (P^k)^T H P^k$, we obtain from (2.18) that

$$\Phi'(\varepsilon_k, A^k)(\tau, H) = P^k \left[U^k \circ (P^k)^T H P^k + \tau D^k \right] (P^k)^T\,. \tag{2.25}$$

By taking a subsequence if necessary, we may assume that $\{P^k\}$ is a convergent sequence with limit $P^\infty \equiv \lim_{k\to\infty} P^k$. This matrix P^∞ will play the role of the matrix P in the spectral decomposition (2.4). Without causing any confusion, we will simply use P, rather than P^∞, in our subsequent analysis. Since both $\{U^k\}$ and $\{D^k\}$ are uniformly bounded, by further taking subsequences if necessary, we may assume that both sequences $\{U^k\}$ and $\{D^k\}$ converge. Taking limits on both sides of (2.25), we obtain

$$P^T V(\tau, H) P = \begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & U_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} \\ \widetilde{H}_{\alpha\beta}^T & \lim_{k\to\infty} U_{\beta\beta}^k \circ \widetilde{H}_{\beta\beta} & 0 \\ \widetilde{H}_{\alpha\gamma}^T \circ U_{\alpha\gamma}^T & 0 & 0 \end{bmatrix} + \tau \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lim_{k\to\infty} D_\beta^k & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

where

$$D_\beta^k = \begin{bmatrix} \phi'_\varepsilon(\varepsilon_k, \lambda_{|\alpha|+1}^k) & & \\ & \ddots & \\ & & \phi'_\varepsilon(\varepsilon_k, \lambda_{|\alpha|+|\beta|}^k) \end{bmatrix}.$$

For each k, define

$$M^k := P \begin{bmatrix} 0 & 0 & 0 \\ 0 & \Lambda_\beta^k & 0 \\ 0 & 0 & 0 \end{bmatrix} P^T\,.$$

Let $\widetilde{M}^k := P^T M^k P$. Because $\varepsilon_k^2 I + (\widetilde{M}^k_{\beta\beta})^2 = \varepsilon_k^2 I + (\Lambda^k_\beta)^2$ is nonsingular, Γ is F-differentiable at (ε_k, M^k) with

$$\Gamma'(\varepsilon_k, M^k)(\tau, H) = \lim_{t\downarrow 0}\left\{\frac{\Gamma(\varepsilon_k + t\tau, M^k + tH) - \Gamma(\varepsilon_k, M^k)}{t}\right\}$$

$$= P\begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & U_{\alpha\gamma}\circ\widetilde{H}_{\alpha\gamma} \\ \widetilde{H}^T_{\alpha\beta} & \displaystyle\lim_{t\downarrow 0}\frac{\Phi_{|\beta|}(\varepsilon_k + t\tau, \Lambda^k_\beta + t\widetilde{H}_{\beta\beta}) - \Phi_{|\beta|}(\varepsilon_k, \Lambda^k_\beta)}{t} & 0 \\ \widetilde{H}^T_{\alpha\gamma}\circ U^T_{\alpha\gamma} & 0 & 0 \end{bmatrix} P^T$$

$$= P\begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & U_{\alpha\gamma}\circ\widetilde{H}_{\alpha\gamma} \\ \widetilde{H}^T_{\alpha\beta} & U^k_{\beta\beta}\circ\widetilde{H}_{\beta\beta} + \tau D^k_\beta & 0 \\ \widetilde{H}^T_{\alpha\gamma}\circ U^T_{\alpha\gamma} & 0 & 0 \end{bmatrix} P^T\,,$$

where we have applied (2.18) to $\Phi_{|\beta|}$ defined by (2.21) at $(\varepsilon_k, \Lambda^k_\beta)$. Thus,

$$V(\tau, H) = \lim_{k\to\infty}\Gamma'(\varepsilon_k, M^k)(\tau, H)\,.$$

Since $(\tau, H) \in \Re\times\mathcal{S}^n$ is arbitrary, it follows that $V \in \partial_B\Gamma(0,0)$.

Conversely, let $V \in \partial_B\Gamma(0,0)$. Since, from Lemma 2.3, Γ is F-differentiable at $(\varepsilon, M) \in \Re\times\mathcal{S}^n$ if and only if $\varepsilon^2 I + (\widetilde{M}_{\beta\beta})^2$ is nonsingular with $\widetilde{M} = P^T M P$, there exists a sequence $\{(\varepsilon_k, M^k)\} \in \Re\times\mathcal{S}^n$ converging to $(0,0)$ such that $\varepsilon_k^2 I + (\widetilde{M}^k_{\beta\beta})^2$ is nonsingular for each k and $V = \lim_{k\to\infty}\Gamma'(\varepsilon_k, M^k)$, where $\widetilde{M}^k = P^T M^k P$. Let $\widetilde{M}^k_{\beta\beta}$ have the following spectral decomposition

$$\widetilde{M}^k_{\beta\beta} = Q^k\widetilde{\Lambda}^k_\beta(Q^k)^T\,,$$

where $Q^k \in \mathcal{Q}_{|\beta|}$ is an orthogonal matrix in $\mathcal{S}^{|\beta|}$ and $\widetilde{\Lambda}^k_\beta$ is the diagonal matrix whose diagonal entries are the eigenvalues $\tilde{z}^k_1 \ge \cdots \ge \tilde{z}^k_{|\beta|}$ of $\widetilde{M}^k_{\beta\beta}$. Let $\tilde{\lambda}^k \in \Re^n$ be such that if $i \in \alpha\cup\gamma$, then $\tilde{\lambda}^k_i = \lambda_i$ and if $i \in \beta$, $\tilde{\lambda}^k_i$ is the $(i-|\alpha|)$-th eigenvalue of $\widetilde{M}^k_{\beta\beta}$, i.e., $\tilde{z}^k_{(i-|\alpha|)}$. Then, by (2.19), for any $(\tau, H) \in \Re\times\mathcal{S}^n$ we have

$$\Gamma'(\varepsilon_k, M^k)(\tau, H) = \lim_{t\downarrow 0}\left\{\frac{\Gamma(\varepsilon_k + t\tau, M^k + tH) - \Gamma(\varepsilon_k, M^k)}{t}\right\}$$

$$= P\begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & U_{\alpha\gamma}\circ\widetilde{H}_{\alpha\gamma} \\ \widetilde{H}^T_{\alpha\beta} & \displaystyle\lim_{t\downarrow 0}\frac{\Phi_{|\beta|}(\varepsilon_k + t\tau, \widetilde{M}^k_{\beta\beta} + t\widetilde{H}_{\beta\beta}) - \Phi_{|\beta|}(\varepsilon_k, \widetilde{M}^k_{\beta\beta})}{t} & 0 \\ \widetilde{H}^T_{\alpha\gamma}\circ U^T_{\alpha\gamma} & 0 & 0 \end{bmatrix} P^T \tag{2.26}$$

with $\widetilde{H} = P^T H P$ and

$$\lim_{t\downarrow 0}\frac{\Phi_{|\beta|}(\varepsilon_k + t\tau, \widetilde{M}^k_{\beta\beta} + t\widetilde{H}_{\beta\beta}) - \Phi_{|\beta|}(\varepsilon_k, \widetilde{M}^k_{\beta\beta})}{t}$$

$$= Q^k\Big(\lim_{t\downarrow 0}\frac{\Phi_{|\beta|}(\varepsilon_k + t\tau, \widetilde{\Lambda}^k_\beta + t(Q^k)^T\widetilde{H}_{\beta\beta}Q^k) - \Phi_{|\beta|}(\varepsilon_k, \widetilde{\Lambda}^k_\beta)}{t}\Big)(Q^k)^T$$

$$= Q^k\big[\,\widetilde{\Omega}^k\circ(\,(Q^k)^T\widetilde{H}_{\beta\beta}Q^k\,) + \tau\widetilde{S}^k\,\big]\,(Q^k)^T\,, \tag{2.27}$$

where we have used (2.18) for $\Phi_{|\beta|}$ and the fact that $\Phi_{|\beta|}$ is F-differentiable at $(\varepsilon_k, \widetilde{\Lambda}^k_\beta)$ because $\varepsilon_k^2 I + (\widetilde{\Lambda}^k_\beta)^2$ is nonsingular,

$$(\widetilde{\Omega}^k)_{ij} = \begin{cases} \dfrac{\phi(\varepsilon_k, \tilde{z}_i^k) - \phi(\varepsilon_k, \tilde{z}_j^k)}{\tilde{z}_i^k - \tilde{z}_j^k} & \text{if } \tilde{z}_i^k \neq \tilde{z}_j^k \\ \phi'_{\tilde{z}_i^k}(\varepsilon, \tilde{z}_i^k) & \text{if } \tilde{z}_i^k = \tilde{z}_j^k \end{cases}, \quad i, j = 1, \ldots, |\beta|,$$

and

$$\widetilde{S}^k = \begin{bmatrix} \phi'_\varepsilon(\varepsilon_k, \tilde{z}_1^k) & & \\ & \ddots & \\ & & \phi'_\varepsilon(\varepsilon_k, \tilde{z}_{|\beta|}^k) \end{bmatrix}.$$

Define

$$A^k = A + P \begin{bmatrix} 0 & 0 & 0 \\ 0 & \widetilde{M}^k_{\beta\beta} & 0 \\ 0 & 0 & 0 \end{bmatrix} P^T \quad \text{and} \quad \widetilde{A}^k = P^T A^k P = \begin{bmatrix} \Lambda_\alpha & 0 & 0 \\ 0 & \widetilde{M}^k_{\beta\beta} & 0 \\ 0 & 0 & \Lambda_\gamma \end{bmatrix}.$$

Since, for each k, $\varepsilon_k^2 I + (\widetilde{M}^k_{\beta\beta})^2$ is nonsingular, the matrix $\varepsilon_k^2 I + (A^k)^2 = P[\varepsilon_k^2 I + (\widetilde{A}^k)^2]P^T$ is also nonsingular. Thus, Φ is F-differentiable at (ε_k, A^k). Let

$$P^k \equiv [\; P^k_\alpha \quad P^k_\beta \quad P^k_\gamma \;] = [\; P_\alpha \quad P_\beta Q^k \quad P_\gamma \;]$$

and $\widetilde{\Lambda}^k$ be the diagonal matrix whose diagonal entries are components of $\tilde{\lambda}^k$. Then

$$A^k = P^k \widetilde{\Lambda}^k (P^k)^T,$$

which, together with (2.18), implies that for any $(\tau, H) \in \Re \times \mathcal{S}^n$, we have

$$\Phi'(\varepsilon_k, A^k)(\tau, H) = P^k\, [\, \widetilde{U}^k \circ ((P^k)^T H P^k) + \tau \widetilde{D}^k \,]\, (P^k)^T, \tag{2.28}$$

where $\widetilde{U}^k \equiv \widehat{U}(\varepsilon_k, \tilde{\lambda}^k)$ and $\widetilde{D}^k \equiv D(\varepsilon_k, \tilde{\lambda}^k)$. Since $\{Q^k\}$, $\{\widetilde{U}^k\}$, and $\{\widetilde{D}^k\}$ are all uniformly bounded, by taking subsequences if necessary, we may assume that all these three sequences converge. By simple computations, we obtain

$$\lim_{k\to\infty} \widetilde{U}^k_{ij} = \begin{cases} 1 & \text{if } i \in \alpha,\ j \in \alpha \cup \beta \\ U_{ij} & \text{if } i \in \alpha,\ j \in \gamma \\ \lim_{k\to\infty} (\widetilde{\Omega}^k)_{(i-|\alpha|)(j-|\alpha|)} & \text{if } i \in \beta,\ j \in \beta \\ 0 & \text{if } i \in \beta \cup \gamma,\ j \in \gamma \end{cases}, \quad i, j = 1, \ldots, n$$

and

$$\lim_{k\to\infty} \widetilde{D}^k = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lim_{k\to\infty} \widetilde{S}^k & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

which, together with (2.28), (2.26), and (2.27), imply that for any $(\tau, H) \in \Re \times \mathcal{S}^n$,

$$\lim_{k\to\infty} (P^k)^T \big[\Gamma'(\varepsilon_k, M^k)(\tau, H) - \Phi'(\varepsilon_k, A^k)(\tau, H)\big] P^k = 0\,.$$

Consequently, we can conclude $V(\tau, H) = \lim_{k\to\infty} \Phi'(\varepsilon_k, A^k)(\tau, H)$ for all $(\tau, H) \in \Re \times \mathcal{S}^n$, which implies $V \in \partial_B \Phi(0, A)$. Hence, (2.24) holds. □

Lemma 2.4 allows us to completely characterize $\partial_B \Phi(0, A)$ (hence, $\partial \Phi(0, A)$).

PROPOSITION 2.5. *Suppose that $A \in \mathcal{S}^n$ has the spectral decomposition as in* (2.4). *Then a $V \in \partial_B \Phi(0, A)$* (*respectively,* $\partial \Phi(0, A)$) *if and only if there exists a $V_{|\beta|} \in \partial_B \Phi_{|\beta|}(0, 0)$* (*respectively,* $\partial \Phi_{|\beta|}(0, 0)$) *such that*

$$V(\tau, H) = P \begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & U_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} \\ \widetilde{H}_{\alpha\beta}^T & V_{|\beta|}(\tau, \widetilde{H}_{\beta\beta}) & 0 \\ \widetilde{H}_{\alpha\gamma}^T \circ U_{\alpha\gamma}^T & 0 & 0 \end{bmatrix} P^T \tag{2.29}$$

for all $(\tau, H) \in \Re \times \mathcal{S}^n$, where $\widetilde{H} = P^T H P$.

Proof. We only need to prove that (2.29) holds for $V \in \partial_B \Phi(0, A)$ and $V_{|\beta|} \in \partial_B \Phi_{|\beta|}(0, 0)$ as the case for Clarke's generalized Jacobian can be done similarly.

Let $\Psi(\tau, H) := (\tau, P^T H P)$ for any $(\tau, H) \in \Re \times \mathcal{S}^n$ and $\Xi : \Re \times \mathcal{S}^n \to \mathcal{S}^n$ be defined by (2.22). Then, since $\Psi'(\tau, H) : \Re \times \mathcal{S}^n \to \Re \times \mathcal{S}^n$ is onto, we know from Lemma 2.1 that

$$\partial_B \Gamma(0, 0) = \partial_B \Xi(0, 0) \Psi'(0, 0)\,,$$

which, together with (2.24) in Lemma 2.4, completes the proof. □

Just as in the case for the metric projector $\Pi_{\mathcal{S}^n_+}$, Proposition 2.5 says that in order to compute $\partial_B \Phi(0, A)$ and $\partial \Phi(0, A)$, one only needs to fix P and compute the corresponding easy part $\partial_B \Phi_{|\beta|}(0, 0)$ (hence, $\partial \Phi_{|\beta|}(0, 0)$). For any $(\varepsilon, z) \in \Re \times \Re^{|\beta|}$ with $\varepsilon^2 + z_i^2 > 0\; \forall\, i$, let $\widehat{\Omega}(\varepsilon, z)$ be defined by (2.17) with n and x being replaced by $|\beta|$ and z, respectively. Then, by (2.19) and (2.18), one can readily draw the conclusion that $V_{|\beta|} \in \partial_B \Phi_{|\beta|}(0, 0)$ if and only if there exist $Q \in \mathcal{Q}_{|\beta|}$ and $\Omega \in \widehat{\mathcal{U}}_{|\beta|}$ such

$$V_{|\beta|}(0, Z) = Q\, [\, \Omega \circ (Q^T Z Q)\,]\, Q^T \quad \forall\, Z \in \mathcal{S}^{|\beta|}\,, \tag{2.30}$$

where

$$\widehat{\mathcal{U}}_{|\beta|} := \big\{\Omega \,:\, \Omega = \lim_{k\to\infty} \widehat{\Omega}(\varepsilon_k, z^k),\; (\varepsilon_k, z^k) \to (0, 0),\; (z^k)_1 \geq \cdots \geq (z^k)_{|\beta|},\; \varepsilon_k^2 + (z_i^k)^2 > 0\, \forall\, i\big\}.$$

Note that for any $\Omega \in \widehat{\mathcal{U}}^\beta$, it holds that $\Omega_{ij} \in [0, 1], \;\; i, j = 1, \ldots, |\beta|$.

The next proposition establishes a link between $\partial_B \Pi_{\mathcal{S}^n_+}(A)$ and $\partial_B \Phi(0, A)$, and so a link between $\partial \Pi_{\mathcal{S}^n_+}(A)$ and $\partial \Phi(0, A)$.

PROPOSITION 2.6. *For any $V_0 \in \partial_B \Pi_{\mathcal{S}^n_+}(A)$, there exists $V \in \partial_B \Phi(0, A)$ such that*

$$V_0(H) = V(0, H) \quad \forall\, H \in \mathcal{S}^n\,. \tag{2.31}$$

Proof. By comparing Proposition 2.2, together with (2.14), with Proposition 2.5, together with (2.30), we can derive the conclusion directly. □

We conclude this section by presenting a useful inequality for elements in $\partial\Phi(0,A)$, which is analogous to (2.3) for the metric projector Π_K with $K = \mathcal{S}^n_+$.

PROPOSITION 2.7. *For any $V \in \partial\Phi(0,A)$, it holds that*

$$\langle H - V(0,H), V(0,H)\rangle \geq 0 \quad \forall\, H \in \mathcal{S}^n\,. \tag{2.32}$$

Proof. Let $V \in \partial\Phi(0,A)$. Then, by Carathéodory's theorem, there exist a positive integer κ and $V^i \in \partial_B\Phi(0,A)$, $i = 1,\ldots,\kappa$ such that V is the convex combination of $V^1,\ldots,V^\kappa$. Let $t_1,\ldots,t_\kappa$ be such that $V = \sum_{i=1}^{\kappa} t_i V^i$, where $t_i \geq 0$, $i = 1,\ldots,\kappa$ and $\sum_{i=1}^{\kappa} t_i = 1$.

From Sun et al [45, Proposition 3.1], we know that for each $i \in \{1,\ldots,\kappa\}$,

$$\langle H - V^i(0,H), V^i(0,H)\rangle \geq 0 \quad \forall\, H \in \mathcal{S}^n\,. \tag{2.33}$$

In order to prove that (2.32) holds for V, let $\theta(X) := \langle X, X\rangle$, $X \in \mathcal{S}^n$. By the convexity of θ, we have for any $H \in \mathcal{S}^n$ that

$$\theta(V(0,H)) = \theta\Big(\sum_{i=1}^{\kappa} t_i V^i(0,H)\Big) \leq \sum_{i=1}^{\kappa} t_i\, \theta(V^i(0,H)) = \sum_{i=1}^{\kappa} t_i \langle V^i(0,H), V^i(0,H)\rangle\,,$$

which, together with (2.33) and the definition of θ, implies

$$\langle V(0,H), V(0,H)\rangle \leq \sum_{i=1}^{\kappa} t_i \langle H, V^i(0,H)\rangle = \Big\langle H, \sum_{i=1}^{\kappa} t_i V^i(0,H)\Big\rangle = \langle H, V(0,H)\rangle\,.$$

Thus, (2.32) holds. □

3. Equivalent Conditions. Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite dimensional real vector spaces each equipped with a scalar product $\langle\cdot,\cdot\rangle$ and its induced norm $\|\cdot\|$. Let $g : \mathcal{X} \to \mathcal{Y}$ be a continuously differentiable function and K be a nonempty and closed convex set in $\mathcal{Y}$. Consider the following feasible problem

$$g(x) \in K\,, \quad x \in \mathcal{X}\,. \tag{3.1}$$

Assume that $\bar{x} \in \mathcal{X}$ is a feasible solution to (3.1). Let $\mathcal{T}_K(g(\bar{x}))$ be the tangent cone of K and $\mathcal{N}_K(g(\bar{x}))$ be the normal cone of K at $g(\bar{x})$, respectively. We write $\mathrm{lin}(\mathcal{T}_K(g(\bar{x})))$ for the lineality space of $\mathcal{T}_K(g(\bar{x}))$. Then we can define the following nondegeneracy condition for problem (3.1).

DEFINITION 3.1. *We say that a feasible point $\bar{x}$ to problem* (3.1) *is constraint nondegenerate if*

$$g'(\bar{x})\mathcal{X} + \mathrm{lin}\big(\mathcal{T}_K(g(\bar{x}))\big) = \mathcal{Y}\,. \tag{3.2}$$

The concept of nondegeneracy for the abstract problem (3.1) first appeared in Robinson [32, 33]. The name "constraint nondegeneracy" was coined by Robinson in [34]. The nondegenerate constraint condition (3.2) including its various equivalent forms has been extensively used in [7, 39] for sensitivity and stability analysis in optimization and variational inequalities. If $\mathcal{Y}$ is the Euclidean space $\Re^m$ and $K = \{0\}^{m_1} \times \Re_+^{m_2}$ with $m_1 + m_2 = m$, then the constraint nondegenerate condition (3.2) is equivalent to the well known linear independence constraint qualification [32, 39]. Here we shall apply Definition 3.1 to both the SDP problem (1.1) and its dual (1.2) to define the primal constraint nondegeneracy and the dual constraint nondegeneracy, respectively.

DEFINITION 3.2. *We say that the primal constraint nondegeneracy holds at a feasible solution* $\overline{X} \in \mathcal{S}_+^n$ *to the SDP problem* (1.1) *if*

$$\begin{bmatrix} \mathcal{A} \\ \mathcal{I} \end{bmatrix} \mathcal{S}^n + \begin{bmatrix} \{0\} \\ \mathrm{lin}\big(\mathcal{T}_{\mathcal{S}_+^n}(\overline{X})\big) \end{bmatrix} = \begin{bmatrix} \Re^m \\ \mathcal{S}^n \end{bmatrix} \tag{3.3}$$

or equivalently

$$\mathcal{A}\,\mathrm{lin}\big(\mathcal{T}_{\mathcal{S}_+^n}(\overline{X})\big) = \Re^m\,, \tag{3.4}$$

where $\mathcal{I}$ *is the identity mapping from* $\mathcal{S}^n$ *to* $\mathcal{S}^n$. *Similarly, we say that the dual constraint nondegeneracy holds at a feasible solution* $(\bar{y}, \overline{S}) \in \Re^m \times \mathcal{S}_+^n$ *to the dual problem* (1.2) *if*

$$\begin{bmatrix} \mathcal{A}^* & \mathcal{I} \\ 0 & \mathcal{I} \end{bmatrix} \begin{pmatrix} \Re^m \\ \mathcal{S}^n \end{pmatrix} + \begin{bmatrix} \{0\} \\ \mathrm{lin}\big(\mathcal{T}_{\mathcal{S}_+^n}(\overline{S})\big) \end{bmatrix} = \begin{bmatrix} \mathcal{S}^n \\ \mathcal{S}^n \end{bmatrix} \tag{3.5}$$

or equivalently

$$\mathcal{A}^*\Re^m + \mathrm{lin}\big(\mathcal{T}_{\mathcal{S}_+^n}(\overline{S})\big) = \mathcal{S}^n\,. \tag{3.6}$$

Note that in the literature constraint nondegeneracy is called differently. Shapiro and Fan [40] and Shapiro [38] termed it transversality. Primal constraint nondegeneracy and dual constraint nondegeneracy are better known as primal nondegeneracy and dual nondegeneracy, respectively in the interior point methods community. See, for example, Alizadeh et al. [2]. To avoid potential confusions, we will stick to Robinson's terminology here and interpret different usages of constraint nondegeneracy in terms of Definition 3.2.

Let $\overline{Z} \equiv (\overline{X}, \bar{y}, \overline{S}) \in \mathcal{S}^n \times \Re^m \times \mathcal{S}^n$ be a KKT point satisfying the KKT conditions (1.3). Since $\mathcal{S}_+^n$ is a self-dual cone, from [13] we know that

$$\mathcal{S}_+^n \ni X \perp S \in \mathcal{S}_+^n \Longleftrightarrow -X \in \mathcal{N}_{\mathcal{S}_+^n}(S) \Longleftrightarrow S - \Pi_{\mathcal{S}_+^n}[\,S - X\,] = X - \Pi_{\mathcal{S}_+^n}[\,X - S\,] = 0\,. \tag{3.7}$$

Therefore, $(\overline{X}, \bar{y}, \overline{S}) \in \mathcal{S}^n \times \Re^m \times \mathcal{S}^n$ satisfies (1.3) if and only if $(\overline{X}, \bar{y}, \overline{S})$ is a solution to the following nonsmooth system of equations

$$F(X, y, S) \equiv \begin{bmatrix} C - \mathcal{A}^* y - S \\ \mathcal{A}X - b \\ S - \Pi_{\mathcal{S}_+^n}[\,S - X\,] \end{bmatrix} = \begin{bmatrix} C - \mathcal{A}^* y - S \\ \mathcal{A}X - b \\ X - \Pi_{\mathcal{S}_+^n}[\,X - S\,] \end{bmatrix} = 0\,, \tag{3.8}$$

where $(X, y, S) \in \mathcal{S}^n \times \Re^m \times \mathcal{S}^n$.

Note that both the KKT conditions (1.3) and the nonsmooth system (3.8) can be written as the following special generalized equation

$$0 \in \begin{bmatrix} C - \mathcal{A}^* y - S \\ \mathcal{A}X - b \\ X \end{bmatrix} + \begin{bmatrix} \mathcal{N}_{\mathcal{S}^n}(X) \\ \mathcal{N}_{\Re^m}(y) \\ \mathcal{N}_{\mathcal{S}_+^n}(S) \end{bmatrix}. \tag{3.9}$$

In [31], Robinson introduced an important concept called strong regularity for a solution of generalized equations. Here we only define the strong regularity for (3.9) rather than for the general problems.

Definition 3.3. *Let* $\mathcal{Z} \equiv \mathcal{S}^n \times \Re^m \times \mathcal{S}^n$. *We say that a KKT point* $\overline{Z} \equiv (\overline{X}, \bar{y}, \overline{S}) \in \mathcal{Z}$ *is a strongly regular solution of the generalized equation* (3.9) *if there exist neighborhoods* $\mathcal{B}$ *of the origin* $0 \in \mathcal{Z}$ *and* $\mathcal{V}$ *of* $\overline{Z}$ *such that for every* $\delta \in \mathcal{B}$, *the following generalized equation*

$$\delta \in \begin{bmatrix} C - \mathcal{A}^* y - S \\ \mathcal{A}X - b \\ X \end{bmatrix} + \begin{bmatrix} \mathcal{N}_{\mathcal{S}^n}(X) \\ \mathcal{N}_{\Re^m}(y) \\ \mathcal{N}_{\mathcal{S}^n_+}(S) \end{bmatrix} \tag{3.10}$$

has a unique solution in $\mathcal{V}$, *denoted by* $Z_{\mathcal{V}}(\delta)$, *and the mapping* $Z_{\mathcal{V}} : \mathcal{B} \to \mathcal{V}$ *is Lipschitz continuous.*

Recall that F is said to be a locally Lipschitz homeomorphism near $\overline{Z}$ if there exists an open neighborhood $\mathcal{V}$ of $\overline{Z}$ such that the restricted mapping $F\,|_{\mathcal{V}}: \mathcal{V} \to F(\mathcal{V})$ is Lipschitz continuous and bijective, and its inverse is also Lipschitz continuous. The following result, which holds in a more general framework, shows that F is locally Lipschitz homeomorphism near $\overline{Z}$ if and only if $\overline{Z}$ is a strongly regular solution of the generalized equation (3.9). This is almost intuitively true. For the sake of completeness, however, we include a short proof.

Lemma 3.4. *Let* $\mathcal{Z} \equiv \mathcal{S}^n \times \Re^m \times \mathcal{S}^n$. *Let* $F : \mathcal{Z} \to \mathcal{Z}$ *be defined by* (3.8) *and* $\overline{Z}$ *be a KKT point of the SDP problem. Then, it holds that* F *is locally Lipschitz homeomorphism near* $\overline{Z}$ *if and only if* $\overline{Z}$ *is a strongly regular solution of the generalized equation* (3.9).

Proof. " $\Longrightarrow$ " Assume that F is locally Lipschitz homeomorphism near $\overline{Z}$. Then, there exists an open neighborhood $\mathcal{V}$ of $\overline{Z}$ such that $F(\mathcal{V})$ is an open neighborhood of the origin $0 \in \mathcal{Z}$ and for any $\hat{\delta} \in F(\mathcal{V})$, the equation $F(Z) = \hat{\delta}$ has a unique solution $\widehat{Z}_{\mathcal{V}}(\hat{\delta})$ in $\mathcal{V}$ and $\widehat{Z}_{\mathcal{V}} : F(\mathcal{V}) \to \mathcal{V}$ is Lipschitz continuous.

For any $\delta = (\delta^1, \delta^2, \delta^3) \in \mathcal{B} \equiv \frac{1}{2}F(\mathcal{V})$, let $Z(\delta) = (X(\delta), y(\delta), S(\delta))$ be a solution, if exists, to (3.10). Write $\delta \equiv (\delta^1, \delta^2, \delta^3) \in \mathcal{S}^n \times \Re^m \times \mathcal{S}^n$. Then we have

$$\begin{bmatrix} C - \mathcal{A}^* y(\delta) - S(\delta) \\ \mathcal{A}X(\delta) - b \\ (S(\delta) + \delta^3) - \Pi_{\mathcal{S}^n_+}[\,(S(\delta) + \delta^3) - X(\delta)\,] \end{bmatrix} = \begin{bmatrix} \delta^1 \\ \delta^2 \\ \delta^3 \end{bmatrix},$$

i.e.,

$$F(X(\delta), y(\delta), S(\delta) + \delta^3) = \begin{bmatrix} \delta^1 - \delta^3 \\ \delta^2 \\ \delta^3 \end{bmatrix}.$$

Then $Z(\delta)$ uniquely exists in $\mathcal{V}$ and $Z(\delta) = \widehat{Z}_{\mathcal{V}}(\delta^1 - \delta^3, \delta^2, \delta^3) - \begin{bmatrix} 0 \\ 0 \\ \delta^3 \end{bmatrix}$. Hence, $Z(\cdot)$ is Lipschitz coninuous on $\mathcal{B}$.

" $\Longleftarrow$ " Assume that $\overline{Z}$ is a strongly regular solution of the generalized equation (3.9). Then, there exist neighborhoods $\mathcal{B}$ of the origin $0 \in \mathcal{Z}$ and $\mathcal{V}$ of $\overline{Z}$, and a locally Lipschitz function $Z_{\mathcal{V}} : \mathcal{B} \to \mathcal{V}$ such that for any $\delta \in \mathcal{B}$, $Z_{\mathcal{V}}(\delta)$ is the unique solution in $\mathcal{V}$ to (3.10). By reversing the arguments in the first part of the proof, we can conclude that for any $\hat{\delta} \equiv (\hat{\delta}^1, \hat{\delta}^2, \hat{\delta}^3) \in (\frac{1}{2}\mathcal{B}) \cap (\,\mathcal{S}^n \times \Re^m \times \mathcal{S}^n\,)$, $F(Z) = \hat{\delta}$ has a unique solution $\widehat{Z}(\hat{\delta}) \in \mathcal{V}$ given by $\widehat{Z}(\hat{\delta}) = Z_{\mathcal{V}}(\hat{\delta}^1 + \hat{\delta}^3, \hat{\delta}^2, \hat{\delta}^3) + \begin{bmatrix} 0 \\ 0 \\ \hat{\delta}^3 \end{bmatrix}$, which implies that $\widehat{Z}(\cdot)$ is Lipschitz continuous on $\frac{1}{2}\mathcal{B}$. Thus, F is Lipschitz homeomorphism near $\overline{Z}$. □

The concept of strong regularity for general nonlinear semidefinite programming is closely related to another concept called the strong second order sufficient condition as shown by Sun in [41]. Here we will only present the strong second order sufficient condition in terms of the SDP problem (1.1). First, for any $B \in \mathcal{S}^n$, we define a linear-quadratic function $\Upsilon_B : \mathcal{S}^n \times \mathcal{S}^n \to \Re$.

DEFINITION 3.5. [41, Definition 2.1] *For any given $B \in \mathcal{S}^n$, define the linear-quadratic function $\Upsilon_B : S^n \times \mathcal{S}^n \to \Re$, which is linear in the first argument and quadratic in the second argument, by*

$$\Upsilon_B(S, H) := 2\langle S, HB^{\dagger}H\rangle, \quad (S, H) \in \mathcal{S}^n \times \mathcal{S}^n,$$

where $B^{\dagger}$ is the Moore-Penrose pseudo-inverse of B.

Let $\overline{X} \in \mathcal{S}^n_+$ be an optimal solution to the SDP problem (1.1). Denote $\mathcal{M}(\overline{X})$ by the set of points $(y, S) \in \Re^m \times \mathcal{S}^n$ such that $(\overline{X}, y, S)$ is a KKT point, i.e., for any $(y, S) \in \mathcal{M}(\overline{X})$, $(\overline{X}, y, S)$ satisfies the KKT conditions (1.3). Let $(\bar{y}, \overline{S}) \in \mathcal{M}(\overline{X})$. Write $\overline{A} \equiv \overline{X} - \overline{S}$. By using the fact $\mathcal{S}^n_+ \ni \overline{X} \perp \overline{S} \in \mathcal{S}^n_+$, we may assume that $\overline{A}$ has the spectral decomposition as in (2.4) by replacing A with $\overline{A}$ and

$$\overline{A} = P \begin{bmatrix} \Lambda_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Lambda_\gamma \end{bmatrix} P^T, \ \overline{X} = P \begin{bmatrix} \Lambda_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} P^T, \ \overline{S} = P \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\Lambda_\gamma \end{bmatrix} P^T. \tag{3.11}$$

Write $P = [\ P_\alpha \quad P_\beta \quad P_\gamma\ ]$. Then, according to (2.7) and (2.9), we have

$$\operatorname{lin}\left(\mathcal{T}_{\mathcal{S}^n_+}(\overline{X})\right) = \{B \in \mathcal{S}^n \ : \ P_\beta^T B P_\beta = 0, \quad P_\beta^T B P_\gamma = 0, \quad P_\gamma^T B P_\gamma = 0\}, \tag{3.12}$$

$$\operatorname{lin}\left(\mathcal{T}_{\mathcal{S}^n_+}(\overline{S})\right) = \{B \in \mathcal{S}^n \ : \ P_\alpha^T B P_\alpha = 0, \quad P_\alpha^T B P_\beta = 0, \quad P_\beta^T B P_\beta = 0\}, \tag{3.13}$$

and

$$\operatorname{aff}\left(\mathcal{C}(\overline{A}; \mathcal{S}^n_+)\right) = \left\{B \in \mathcal{S}^n \ : \ P_\beta^T B P_\gamma = 0, \ P_\gamma^T B P_\gamma = 0\right\}.$$

Define

$$\begin{aligned} \operatorname{app}(\bar{y}, \overline{S}) \ &:= \ \left\{B \in \mathcal{S}^n \ : \ \mathcal{A}B = 0, \quad B \in \operatorname{aff}\left(\mathcal{C}(\overline{A}; \mathcal{S}^n_+)\right)\right\} \\ &= \ \{B \in \mathcal{S}^n \ : \ \mathcal{A}B = 0, \quad P_\beta^T B P_\gamma = 0, \quad P_\gamma^T B P_\gamma = 0\}. \end{aligned} \tag{3.14}$$

Then we can state the strong second order sufficient condition for the SDP problem tailored from Sun [41] for the general nonlinear SDP problem.

DEFINITION 3.6. *Let $\overline{X} \in \mathcal{S}^n_+$ be an optimal solution to the SDP problem* (1.1). *We say that the strong second order sufficient condition holds at $\overline{X}$ if*

$$\sup_{(y,S)\in\mathcal{M}(\overline{X})} \left\{-\Upsilon_{\overline{X}}(-S, H)\right\} > 0 \quad \forall\, 0 \neq H \in \Big\{ \bigcap_{(y,S)\in\mathcal{M}(\overline{X})} \operatorname{app}(y, S) \Big\}. \tag{3.15}$$

The strong second order sufficient condition (3.15) may look very complicated. When $\mathcal{M}(\overline{X})$ is a singleton, the following result gives a very simple characterization.

LEMMA 3.7. *Let $\overline{X} \in \mathcal{S}^n_+$ be an optimal solution to the SDP problem* (1.1). *Assume that $\mathcal{M}(\overline{X}) = \{(\bar{y}, \overline{S})\}$. Let $\overline{X}$ and $\overline{S}$ have the spectral decompositions as in* (3.11). *Then the strong second order sufficient condition* (3.15) *holds at $\overline{X}$ if and only if, for any $H \in \mathcal{S}^n$, the following condition holds*

$$\text{(3.16)} \quad \mathcal{A}H = 0,\ P_\beta^T H P_\gamma = 0,\ P_\gamma^T H P_\gamma = 0,\ \text{and } P_\alpha^T H P_\gamma = 0 \implies H = 0\,.$$

Proof. For any $H \in \mathcal{S}^n$, write $\widetilde{H} = P^T H P$. Since $\mathcal{M}(\overline{X}) = \{(\bar{y}, \overline{S})\}$, the strong second order sufficient condition (3.15) becomes

$$-\Upsilon_{\overline{X}}(-\overline{S}, H) > 0 \quad \forall\, H \in \text{app}(\bar{y}, \overline{S})\backslash\{0\}\,,$$

which, by the definition of $\Upsilon_{\overline{X}}(-\overline{S}, H)$ and (3.11), is equivalent to

$$2 \sum_{i\in\alpha, j\in\gamma} \frac{-\lambda_j}{\lambda_i} (\widetilde{H}_{ij})^2 > 0 \qquad \forall\, H \in \text{app}(\bar{y}, \overline{S})\backslash\{0\}\,.$$

For details, see [41]. Then, by (3.14), the strong second order sufficient condition (3.15) holds at $\overline{X}$ if and only if

$$\mathcal{A}H = 0,\ \widetilde{H}_{\beta\gamma} = 0,\ \widetilde{H}_{\gamma\gamma} = 0,\ \text{and } H \neq 0 \implies \widetilde{H}_{\alpha\gamma} \neq 0 \quad \forall\, H \in \mathcal{S}^n\,,$$

which is equivalent to (3.16). This completes the proof. □

Next, we shall establish a link between the strong second order sufficient condition and the dual constraint nondegeneracy.[2]

PROPOSITION 3.8. *Let $\overline{X} \in \mathcal{S}^n_+$ be an optimal solution to the SDP problem* (1.1). *Under the assumption $\mathcal{M}(\overline{X}) = \{(\bar{y}, \overline{S})\}$, the following are equivalent:*
(i) *The strong second order sufficient condition* (3.15) *holds at $\overline{X}$.*
(ii) *The dual constraint nondegenerate condition* (3.6) *holds at $(\bar{y}, \overline{S})$.*

Proof. Let $\overline{X}$ and $\overline{S}$ have the spectral decompositions as in (3.11). For any $H \in \mathcal{S}^n$, let $\widetilde{H} = P^T H P$. We prove "(i) $\Longrightarrow$ (ii)" first. By Lemma 3.7, (i) holds if and only if we have the following implication:

$$\text{(3.17)} \qquad \mathcal{A}H = 0,\ \widetilde{H}_{\beta\gamma} = 0,\ \widetilde{H}_{\gamma\gamma} = 0,\ \text{and } \widetilde{H}_{\alpha\gamma} = 0 \Longrightarrow H = 0\ \forall\, H \in \mathcal{S}^n\,.$$

Suppose, for the sake of contradiction, that the dual constraint nondegenerate condition (3.6) does not holds at $(\bar{y}, \overline{S})$. Then, we have

$$\text{(3.18)} \qquad [\mathcal{A}^*\Re^m]^\perp \cap \left[\text{lin}\big(\mathcal{T}_{\mathcal{S}^n_+}(\overline{S})\big)\right]^\perp \neq \{0\}$$

Take an arbitrary $0 \neq \overline{H} \in [\mathcal{A}^*\Re^m]^\perp \cap \left[\text{lin}\big(\mathcal{T}_{\mathcal{S}^n_+}(\overline{S})\big)\right]^\perp$. We obtain from $\overline{H} \in [\mathcal{A}^*\Re^m]^\perp$ that

$$\text{(3.19)} \qquad \langle \overline{H}, \mathcal{A}^* y\rangle = 0 \ \ \forall\, y \in \Re^m \Longrightarrow \langle \mathcal{A}\overline{H}, y\rangle = 0 \ \ \forall\, y \in \Re^m \Longrightarrow \mathcal{A}\overline{H} = 0$$

and from $\overline{H} \in \left[\text{lin}\big(\mathcal{T}_{\mathcal{S}^n_+}(\overline{S})\big)\right]^\perp$ that

$$\langle P^T \overline{H} P, P^T B P\rangle = \langle \overline{H}, B\rangle = 0 \ \ \forall\, B \in \text{lin}\big(\mathcal{T}_{\mathcal{S}^n_+}(\overline{S})\big)\,,$$

[2] A similar statement for the dual SDP problem (1.2) also holds. We omit it here for brevity.

which, together with (3.13), implies

$$P_\alpha^T \overline{H} P_\gamma = 0, \quad P_\beta^T \overline{H} P_\gamma = 0, \quad \text{and} \quad P_\gamma^T \overline{H} P_\gamma = 0\,. \tag{3.20}$$

By making use of (3.17), (3.19), and (3.20), we obtain $\overline{H} = 0$, which contradicts the choice of $\overline{H}$. This contradiction shows that (ii) holds.

Next, we show "(ii) $\Longrightarrow$ (i)". Since the dual constraint nondegenerate condition (3.6) holds at $(\overline{y}, \overline{S})$, for any $H \in \mathcal{S}^n$ such that $\mathcal{A} H = 0$, $\widetilde{H}_{\beta\gamma} = 0$, $\widetilde{H}_{\gamma\gamma} = 0$, and $\widetilde{H}_{\alpha\gamma} = 0$, there exist $y \in \Re^m$ and $S \in \mathrm{lin}\big(\mathcal{T}_{\mathcal{S}^n_+}(\overline{S})\big)$ such that

$$H = \mathcal{A}^* y + S\,,$$

which, together with (3.13), implies

$$\begin{aligned} \langle H, H\rangle &= \langle H, \mathcal{A}^* y + S\rangle = \langle \mathcal{A} H, y\rangle + \langle H, S\rangle = 0 + \langle P^T H P, P^T S P\rangle \\ &= \left\langle \begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & 0 \\ \widetilde{H}_{\alpha\beta}^T & \widetilde{H}_{\beta\beta} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & P_\alpha^T S P_\gamma \\ 0 & 0 & P_\beta^T S P_\gamma \\ P_\gamma^T S P_\alpha & P_\gamma^T S P_\beta & P_\gamma^T S P_\gamma \end{bmatrix} \right\rangle = 0\,. \end{aligned}$$

Therefore, by Lemma 3.7, it follows that (i) holds. □

Let $(\overline{X}, \overline{y}, \overline{S}) \in \mathcal{S}^n \times \Re^m \times \mathcal{S}^n$ be a KKT point satisfying the KKT conditions (1.3) and F be defined by (3.8). As we menioned in the introduction, by combining the two papers [14] and [18], we know that if the primal constraint nondegeneracy holds at $\overline{X}$ and the dual constraint nondegeneracy holds at $(\overline{y}, \overline{S})$, then every element in $\partial_B F(\overline{X}, \overline{y}, \overline{S})$ is nonsingular. Actually, Proposition 3.8 and [41, Proposition 3.2] allow us to prove even the nonsingularity of Clarke's generalized Jacobian $\partial F(\overline{X}, \overline{y}, \overline{S})$ under the same primal and dual constraint nondegenerate conditions.

PROPOSITION 3.9. *Let $(\overline{X}, \overline{y}, \overline{S}) \in \mathcal{S}^n \times \Re^m \times \mathcal{S}^n$ be a KKT point. Assume that the primal constraint nondegenerate condition* (3.4) *holds at $\overline{X}$ and the dual constraint nondegenerate condition* (3.6) *holds at $(\overline{y}, \overline{S})$, respectively. Then, every element in $\partial F(\overline{X}, \overline{y}, \overline{S})$ is nonsingular.*

Proof. Since the primal constraint nondegenerate condition (3.4) implies that $\mathcal{M}(\overline{X}) = \{(\overline{y}, \overline{S})\}$, we know from Proposition 3.8 that the strong second order sufficient condition (3.15) holds at $\overline{X}$. Consequently, by [41, Proposition 3.2], every element in $\partial F(\overline{X}, \overline{y}, \overline{S})$ is nonsingular. □

Proposition 3.9 says that the primal and dual constraint nondegenerate conditions are sufficient for the nonsingularity of all elements in $\partial F(\overline{X}, \overline{y}, \overline{S})$. Next, we shall show that the nonsingularity of only two elements in $\partial_B F(\overline{X}, \overline{y}, \overline{S})$ will imply both the primal and dual constraint nondegenerate conditions.

From Lemma 2.1, we know that $W \in \partial_B F(\overline{X}, \overline{y}, \overline{S})$ if and only if there exists a $V \in \partial_B \Pi_{\mathcal{S}^n_+}(\overline{A})$ such that

$$W(\Delta X, \Delta y, \Delta S) = \begin{bmatrix} -\mathcal{A}^*(\Delta y) - \Delta S \\ \mathcal{A}(\Delta X) \\ \Delta X - V(\Delta X - \Delta S) \end{bmatrix} \tag{3.21}$$

for all $(\Delta X, \Delta y, \Delta S) \in \mathcal{S}^n \times \Re^m \times \mathcal{S}^n$, where $\overline{A} \equiv \overline{X} - \overline{S}$. Let $\mathrm{ex}\,(\partial_B \Pi_{\mathcal{S}^n_+}(\overline{A}))$ be defined by (2.15). For $V^0, V^{\mathcal{I}} \in \mathrm{ex}\,(\partial_B \Pi_{\mathcal{S}^n_+}(\overline{A}))$, let W^0 and $W^{\mathcal{I}}$ be defined by (3.21), respectively. Denote

$$\mathrm{ex}\,(\partial_B F(\overline{X}, \overline{y}, \overline{S})) := \left\{ W^0, W^{\mathcal{I}} \right\} \subseteq \partial_B F(\overline{X}, \overline{y}, \overline{S})\,. \tag{3.22}$$

PROPOSITION 3.10. *Let* $(\overline{X}, \bar{y}, \overline{S}) \in \mathcal{S}^n \times \Re^m \times \mathcal{S}^n$ *be a KKT point. If both* W^0 *and* $W^{\mathcal{I}}$ *in* $\mathrm{ex}\,(\partial_B F(\overline{X}, \bar{y}, \overline{S}))$ *are nonsingular, then the primal constraint nondegenerate condition* (3.4) *holds at* $\overline{X}$ *and the dual constraint nondegenerate condition* (3.6) *holds at* $(\bar{y}, \overline{S})$*, respectively.*

Proof. First we show that the nonsingularity of W^0 implies the primal constraint nondegenerate condition (3.4). Assume on the contrary that (3.4) does not hold. Since, equivalently, (3.3) fails to hold too, we have

$$\left\{ \begin{bmatrix} \mathcal{A} \\ I \end{bmatrix} \mathcal{S}^n \right\}^{\perp} \cap \begin{bmatrix} 0 \\ \mathrm{lin}\left(\mathcal{T}_{\mathcal{S}^n_+}(\overline{X})\right) \end{bmatrix}^{\perp} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} \Re^m \\ \mathcal{S}^n \end{bmatrix},$$

which implies that there exists $0 \neq (\Delta y, \Delta S) \in \left\{ \begin{bmatrix} \mathcal{A} \\ I \end{bmatrix} \mathcal{S}^n \right\}^{\perp} \cap \begin{bmatrix} 0 \\ \mathrm{lin}\left(\mathcal{T}_{\mathcal{S}^n_+}(\overline{X})\right) \end{bmatrix}^{\perp}$.

We obtain from $(\Delta y, \Delta S) \in \left\{ \begin{bmatrix} \mathcal{A} \\ I \end{bmatrix} \mathcal{S}^n \right\}^{\perp}$ that

$$\big\langle (\Delta y, \Delta S), (\mathcal{A}H, H) \big\rangle = 0 \quad \forall\, H \in \mathcal{S}^n \Longrightarrow \mathcal{A}^*(\Delta y) + \Delta S = 0 \tag{3.23}$$

and from $(\Delta y, \Delta S) \in \begin{bmatrix} 0 \\ \mathrm{lin}\left(\mathcal{T}_{\mathcal{S}^n_+}(\overline{X})\right) \end{bmatrix}^{\perp}$ we obtain that

$$\big\langle P^T(\Delta S)P, P^T H P \big\rangle = \big\langle \Delta S, H \big\rangle = 0 \quad \forall\, H \in \mathrm{lin}\left(\mathcal{T}_{\mathcal{S}^n_+}(\overline{X})\right),$$

which, together with (3.12), implies

$$P_\alpha^T(\Delta S)P_\alpha = 0, \quad P_\alpha^T(\Delta S)P_\beta = 0, \quad \text{and} \quad P_\alpha^T(\Delta S)P_\gamma = 0\,. \tag{3.24}$$

Let $U \in \mathcal{S}^n$ be defined by (2.5). Recall from Proposition 2.2 that for $V^0 \in \mathrm{ex}\,(\partial_B \Pi_{\mathcal{S}^n_+}(\overline{A}))$, it holds that

$$V^0(\Delta S) = P \begin{bmatrix} P_\alpha^T(\Delta S)P_\alpha & P_\alpha^T(\Delta S)P_\beta & U_{\alpha\gamma} \circ (P_\alpha^T(\Delta S)P_\gamma) \\ \left(P_\alpha^T(\Delta S)P_\beta\right)^T & 0 & 0 \\ \left(P_\alpha^T(\Delta S)P_\gamma\right)^T \circ U_{\alpha\gamma}^T & 0 & 0 \end{bmatrix} P^T,$$

which, together with (3.24), implies $V^0(\Delta S) = 0 \in \mathcal{S}^n$. Therefore, by (3.21) and (3.23), we have for $\Delta X \equiv 0$ that

$$W^0(\Delta X, \Delta y, \Delta S) = \begin{bmatrix} -\mathcal{A}^*(\Delta y) - \Delta S \\ \mathcal{A}(\Delta X) \\ \Delta X - V^0(\Delta X - \Delta S) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ V^0(\Delta S) \end{bmatrix} = 0\,,$$

which, implies that W^0 is singular. This contradiction shows that the primal constraint nondegenerate condition (3.4) holds at $\overline{X}$.

Next, we show that the nonsingularity of $W^{\mathcal{I}}$ implies the dual constraint nondegenerate condition (3.6). Suppose not. Then,

$$[\mathcal{A}^*\Re^m]^{\perp} \cap \left[\mathrm{lin}\left(\mathcal{T}_{\mathcal{S}^n_+}(\overline{S})\right)\right]^{\perp} \neq \{0\}\,.$$

Let $0 \neq \Delta X \in [\mathcal{A}^*\Re^m]^{\perp} \cap \left[\mathrm{lin}\left(\mathcal{T}_{\mathcal{S}^n_+}(\overline{S})\right)\right]^{\perp}$. We obtain from $\Delta X \in [\mathcal{A}^*\Re^m]^{\perp}$ that

$$\langle \Delta X, \mathcal{A}^* y \rangle = 0 \quad \forall\, y \in \Re^m \Longrightarrow \mathcal{A}(\Delta X) = 0 \tag{3.25}$$

and from $\Delta X \in \left[\,\mathrm{lin}\left(\mathcal{T}_{\mathcal{S}^n_+}(\overline{S})\right)\right]^{\perp}$ that

$$\left\langle P^T(\Delta X)P, P^T SP\right\rangle = \left\langle \Delta X, S\right\rangle = 0 \quad \forall\, S \in \mathrm{lin}\left(\mathcal{T}_{\mathcal{S}^n_+}(\overline{S})\right),$$

which, together with (3.13), implies

$$P_\alpha^T(\Delta X)P_\gamma = 0, \quad P_\beta^T(\Delta X)P_\gamma = 0, \quad \text{and} \quad P_\gamma^T(\Delta X)P_\gamma = 0\,. \tag{3.26}$$

From Proposition 2.2 that for $V^{\mathcal{I}} \in \mathrm{ex}\,(\partial_B \Pi_{\mathcal{S}^n_+}(\overline{A}))$, it holds that

$$V^{\mathcal{I}}(\Delta X) = P\begin{bmatrix} P_\alpha^T(\Delta X)P_\alpha & P_\alpha^T(\Delta X)P_\beta & U_{\alpha\gamma}\circ(P_\alpha^T(\Delta X)P_\gamma) \\ \left(P_\alpha^T(\Delta X)P_\beta\right)^T & P_\beta^T(\Delta X)P_\beta & 0 \\ \left(P_\alpha^T(\Delta X)P_\gamma\right)^T \circ U_{\alpha\gamma}^T & 0 & 0 \end{bmatrix} P^T,$$

which, together with (3.26), implies $V^{\mathcal{I}}(\Delta X) = \Delta X$. Therefore, by (3.21) and (3.25), we have for $(\Delta y, \Delta S) \equiv (0,0) \in \Re^m \times \mathcal{S}^n$ that

$$W^{\mathcal{I}}(\Delta X, \Delta y, \Delta S) = \begin{bmatrix} -\mathcal{A}^*(\Delta y) - \Delta S \\ \mathcal{A}(\Delta X) \\ \Delta X - V^{\mathcal{I}}(\Delta X - \Delta S) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \Delta X - V^{\mathcal{I}}(\Delta X) \end{bmatrix} = 0\,,$$

which, implies that $W^{\mathcal{I}}$ is singular. This contradiction shows that the dual constraint nondegenerate condition (3.6) holds at $(\overline{y}, \overline{S})$. This completes the proof. ◻

Now, we are ready to state our main result of this paper.

THEOREM 3.11. *Let* $(\overline{X}, \overline{y}, \overline{S}) \in \mathcal{S}^n \times \Re^m \times \mathcal{S}^n$ *be a KKT point satisfying the KKT conditions* (1.3) *and* F *be defined by* (3.8)*. Then, the following are all equivalent:*

(i) *The KKT point* $(\overline{X}, \overline{y}, \overline{S})$ *is a strongly regular solution of the generalized equation* (3.9)*.*
(ii) *The function* F *is locally Lipschitz homeomorphism near* $(\overline{X}, \overline{y}, \overline{S})$*.*
(iii) *The primal constraint nondegenerate condition* (3.4) *holds at* $\overline{X}$ *and the dual constraint nondegenerate condition* (3.6) *holds at* $(\overline{y}, \overline{S})$*.*
(iv) *Every element in* $\partial F(\overline{X}, \overline{y}, \overline{S})$ *is nonsingular.*
(v) *Every element in* $\partial_B F(\overline{X}, \overline{y}, \overline{S})$ *is nonsingular.*
(vi) *The two elements in* $\mathrm{ex}\left(\partial_B F(\overline{X}, \overline{y}, \overline{S})\right)$ *are nonsingular.*

Proof. We have already known from Lemma 3.4 that (i) $\Longleftrightarrow$ (ii) and from Propositions 3.9 and 3.10 that (iii) $\Longleftrightarrow$ (iv) $\Longleftrightarrow$ (v) $\Longleftrightarrow$ (vi). Furthermore, Clarke's inverse function theorem for Lipschitz functions [11, 12] implies that (iv) $\Longrightarrow$ (ii). The proof of this theorem will be complete if one can show (ii) $\Longrightarrow$ (v). However, the latter has been known to be true since [20] (Gowda [15] even obtained a stronger conclusion than this by employing the degree theory). ◻

REMARK 3.12. *Note that the relations* (i) $\Longleftrightarrow$ (ii) $\Longleftrightarrow$ (iv) *even hold for the general nonlinear semidefinite programming case* [41, Theorem 4.1]*, whose proof further relies on a number of important results achieved by* Bonnans *and* Shapiro *in their excellent monograph* [7] *on sensitivity analysis in optimization and variational inequalities. Here, the structure displayed uniquely by the SDP problem* (1.1) *allows us to derive them directly by avoiding the detour employed in* [41] *for the nonlinear SDP problem. An SDP example satisfying* (iii) *but with the strict complementary condition failing to hold can be found in* [2]*. See also* [19]*.*

4. Quadratic Convergence of Smoothing Newton Methods. In this section, we shall show how the theoretical results obtained in Sections 2 and 3 can be used to provide a quadratic convergence analysis on smoothing Newton methods for solving the nonsmooth equation $F(X, y, S) = 0$, where F is defined by (3.8). Let $\Phi : \Re \times \mathcal{S}^n \to \mathcal{S}^n$ be defined by (2.16). We then introduce the following smoothing function for F:

$$G(\varepsilon, X, y, S) \equiv \begin{bmatrix} C - \mathcal{A}^* y - S \\ \mathcal{A}X - b \\ S - \Phi(\varepsilon, S - X) \end{bmatrix} = \begin{bmatrix} C - \mathcal{A}^* y - S \\ \mathcal{A}X - b \\ X - \Phi(\varepsilon, X - S) \end{bmatrix}, \tag{4.1}$$

where $(\varepsilon, X, y, S) \in \Re \times \mathcal{S}^n \times \Re^m \times \mathcal{S}^n$. The above function G is continuously differentiable around any $(\varepsilon, X, y, S) \in \Re \times \mathcal{S}^n \times \Re^m \times \mathcal{S}^n$ when $\varepsilon \neq 0$ and has been used by several authors [9, 10, 19, 45] to design smoothing Newton methods for solving SDP problems (1.1) and (1.2).

Define $E : \Re \times \mathcal{S}^n \times \Re^m \times \mathcal{S}^n \to \Re \times \mathcal{S}^n \times \Re^m \times \mathcal{S}^n$ by

$$E(\varepsilon, X, y, S) \equiv \begin{bmatrix} \varepsilon \\ G(\varepsilon, X, y, S) \end{bmatrix}, \quad (\varepsilon, X, y, S) \in \Re \times \mathcal{S}^n \times \Re^m \times \mathcal{S}^n. \tag{4.2}$$

Then we have

$$F(X, y, S) = 0 \Longleftrightarrow E(\varepsilon, X, y, S) = 0 \quad \forall\, (\varepsilon, X, y, S) \in \Re \times \mathcal{S}^n \times \Re^m \times \mathcal{S}^n.$$

Let $(\overline{X}, \bar{y}, \overline{S}) \in \mathcal{S}^n \times \Re^m \times \mathcal{S}^n$ be a KKT point satisfying the KKT conditions (1.3). Then

$$E(0, \overline{X}, \bar{y}, \overline{S}) = 0.$$

Write $\overline{A} \equiv \overline{X} - \overline{S}$. Let $\overline{A}$, $\overline{X}$, and $\overline{S}$ have the spectral decompositions as in (3.11). Let the linear-quadratic function $\Upsilon_{\overline{X}}(\cdot, \cdot)$ be defined as in Definition 3.5. Then, we have the following result, which will play a key role in our analysis on quadratic convergence of smoothing Newton methods.

PROPOSITION 4.1. *Let $V \in \partial\Phi(0, \overline{A})$. Then, for any ΔX and ΔS in $\mathcal{S}^n$ such that $\Delta X = V(0, \Delta X - \Delta S)$, it holds that*

$$\langle \Delta X, \Delta S \rangle \leq \Upsilon_{\overline{X}}(-\overline{S}, \Delta X). \tag{4.3}$$

Proof. Let ΔX and ΔS be in $\mathcal{S}^n$ such that $\Delta X = V(0, \Delta X - \Delta S)$. Write $\Delta\widetilde{X} \equiv P^T(\Delta X)P$ and $\Delta\widetilde{S} \equiv P^T(\Delta S)P$. Let $\Phi_{|\beta|}$ be defined by (2.21). Then, by Proposition 2.5, there exists $V_{|\beta|} \in \partial\Phi_{|\beta|}(0, 0)$ such that

$$V(0, \Delta X - \Delta S) = P \begin{bmatrix} \Delta\widetilde{H}_{\alpha\alpha} & \Delta\widetilde{H}_{\alpha\beta} & U_{\alpha\gamma} \circ \Delta\widetilde{H}_{\alpha\gamma} \\ (\Delta\widetilde{H}_{\alpha\beta})^T & V_{|\beta|}(0, \Delta\widetilde{H}_{\beta\beta}) & 0 \\ (\Delta\widetilde{H}_{\alpha\gamma})^T \circ U_{\alpha\gamma}^T & 0 & 0 \end{bmatrix} P^T,$$

where $\Delta\widetilde{H} \equiv \Delta\widetilde{X} - \Delta\widetilde{S}$ and $U \in \mathcal{S}^n$ is defined by (2.5). Thus, by using $\Delta X = V(0, \Delta X - \Delta S)$, we obtain

$$\Delta\widetilde{S}_{\alpha\alpha} = 0, \quad \Delta\widetilde{S}_{\alpha\beta} = 0, \quad \Delta\widetilde{X}_{\beta\gamma} = 0, \quad \Delta\widetilde{X}_{\gamma\gamma} = 0, \tag{4.4}$$

$$\Delta\widetilde{X}_{\beta\beta} = V_{|\beta|}(0, \Delta\widetilde{X}_{\beta\beta} - \Delta\widetilde{S}_{\beta\beta})\,, \tag{4.5}$$

and

$$\Delta\widetilde{X}_{\alpha\gamma} - U_{\alpha\gamma}\circ\Delta\widetilde{X}_{\alpha\gamma} = -U_{\alpha\gamma}\circ\Delta\widetilde{S}_{\alpha\gamma}\,. \tag{4.6}$$

By applying Proposition 2.7 to $\Phi_{|\beta|}$ and using (4.5), we obtain

$$\begin{aligned} &\langle \Delta\widetilde{X}_{\beta\beta}, -\Delta\widetilde{S}_{\beta\beta}\rangle \\ &= \big\langle V_{|\beta|}(0, \Delta\widetilde{X}_{\beta\beta} - \Delta\widetilde{S}_{\beta\beta}), (\Delta\widetilde{X}_{\beta\beta} - \Delta\widetilde{S}_{\beta\beta}) - V_{|\beta|}(0, \Delta\widetilde{X}_{\beta\beta} - \Delta\widetilde{S}_{\beta\beta})\big\rangle \geq 0\,, \end{aligned} \tag{4.7}$$

Therefore, from (4.4), (4.6), and (4.7), we have

$$\begin{aligned} \langle \Delta X, \Delta S\rangle &= \langle \Delta\widetilde{X}, \Delta\widetilde{S}\rangle \\ &= \langle \Delta\widetilde{X}_{\beta\beta}, \Delta\widetilde{S}_{\beta\beta}\rangle + 2\langle \Delta\widetilde{X}_{\alpha\gamma}, \Delta\widetilde{S}_{\alpha\gamma}\rangle \\ &\leq 2\langle \Delta\widetilde{X}_{\alpha\gamma}, \Delta\widetilde{S}_{\alpha\gamma}\rangle \\ &= 2\sum_{i\in\alpha, j\in\gamma} \frac{\lambda_j}{\lambda_i}((\Delta\widetilde{X})_{ij})^2\,, \end{aligned}$$

which, together with the fact that

$$\Upsilon_{\overline{X}}(-\overline{S}, \Delta X) = 2\sum_{i\in\alpha, j\in\gamma} \frac{\lambda_j}{\lambda_i}((\Delta\widetilde{X})_{ij})^2\,,$$

shows that (4.3) holds. □

The following result relates the nonsingularity of $\partial_B E(0, \overline{X}, \bar{y}, \overline{S})$ and $\partial E(0, \overline{X}, \bar{y}, \overline{S})$ to both the primal constraint nondegeneracy and the dual constraint nondegeneracy.

PROPOSITION 4.2. *Let $(\overline{X}, \bar{y}, \overline{S}) \in \mathcal{S}^n \times \Re^m \times \mathcal{S}^n$ be a KKT point satisfying the KKT conditions* (1.3) *and E be defined by* (4.2)*. Then the following are equivalent*

(i) *The primal constraint nondegenerate condition* (3.4) *holds at $\overline{X}$ and the dual constraint nondegenerate condition* (3.6) *holds at $(\bar{y}, \overline{S})$.*

(ii) *Every element in $\partial_B E(0, \overline{X}, \bar{y}, \overline{S})$ is nonsingular.*

(iii) *Every element in $\partial E(0, \overline{X}, \bar{y}, \overline{S})$ is nonsingular.*

Proof. Since "(iii) $\Longrightarrow$ (ii)" holds trivially and "(ii) $\Longrightarrow$ (i)" follows from Proposition 2.6 and Theorem 3.11 directly, we only need to show "(i) $\Longrightarrow$ (iii)". So in the remaining part of our proof we always assume that part (i) holds.

Let W be an arbitrary element in $\partial E(0, \overline{X}, \bar{y}, \overline{S})$. We need to show that W is nonsingular. Let $(\Delta\varepsilon, \Delta X, \Delta y, \Delta S) \in \Re\times\mathcal{S}^n\times\Re^m\times\Re^n$ be such that

$$W(\Delta\varepsilon, \Delta X, \Delta y, \Delta S) = 0.$$

Then, by Lemma 2.1, there exists $V \in \partial\Phi(0, \overline{A})$ such that

$$W(\Delta\varepsilon, \Delta X, \Delta y, \Delta S) = \begin{bmatrix} \Delta\varepsilon \\ -\mathcal{A}^*(\Delta y) - \Delta S \\ \mathcal{A}(\Delta X) \\ \Delta X - V(\Delta\varepsilon, \Delta X - \Delta S) \end{bmatrix} = 0\,,$$

which, implies that $\Delta\varepsilon = 0$. Thus, we have

$$W(0, \Delta X, \Delta y, \Delta S) = \begin{bmatrix} 0 \\ -\mathcal{A}^*(\Delta y) - \Delta S \\ \mathcal{A}(\Delta X) \\ \Delta X - V(0, \Delta X - \Delta S) \end{bmatrix} = 0\,. \tag{4.8}$$

Since the primal constraint nondegenerate condition (3.4) implies $\mathcal{M}(\overline{X}) = \{(\bar{y}, \overline{S})\}$, we know from Proposition 3.8 that the strong second order sufficient condition (3.15) holds at $\overline{X}$ and takes the following form

$$-\Upsilon_{\overline{X}}(-\overline{S}, H) > 0 \quad \forall\, 0 \neq H \in \mathrm{app}(\bar{y}, \overline{S})\,, \tag{4.9}$$

where the set $\mathrm{app}(\bar{y}, \overline{S})$ is defined by (3.14). From Proposition 2.5, (4.8), and (3.14), we know that

$$\Delta X \in \mathrm{app}(\bar{y}, \overline{S})\,. \tag{4.10}$$

By the second and the third equations of (4.8), we obtain that

$$0 = \langle \Delta X, -\mathcal{A}^*(\Delta y) - \Delta S\rangle + \langle \Delta y, \mathcal{A}(\Delta X)\rangle = \langle \Delta X, -\Delta S\rangle\,,$$

which, together with Proposition 4.1 and the last equation of (4.8), implies that

$$\Upsilon_{\overline{X}}(-\overline{S}, \Delta X) \geq 0\,.$$

Hence, from (4.9) and (4.10), we can conclude that

$$\Delta X = 0\,.$$

Thus, from (4.8), we get

$$\begin{bmatrix} \mathcal{A}^*(\Delta y) + \Delta S \\ V(0, -\Delta S) \end{bmatrix} = 0\,, \tag{4.11}$$

which, by Proposition 2.5, gives rise to

$$P_\alpha^T(\Delta S)P_\alpha = 0, \quad P_\alpha^T(\Delta S)P_\beta = 0, \text{ and } P_\alpha^T(\Delta S)P_\gamma = 0\,. \tag{4.12}$$

From (3.3), which is equivalent to the primal constraint nondegenerate condition (3.4), we know that there exist $X \in \mathcal{S}^n$ and $S \in \mathrm{lin}\left(\mathcal{T}_{\mathcal{S}^n_+}(\overline{X})\right)$ such that

$$\mathcal{A}\,X = \Delta y \quad \text{and} \quad X + S = \Delta S\,,$$

which, together with (4.12), (3.13), and the first equation of (4.11), imply

$$\begin{aligned} \langle \Delta y, \Delta y\rangle + \langle \Delta S, \Delta S\rangle &= \langle \mathcal{A}\,X, \Delta y\rangle + \langle X + S, \Delta S\rangle \\ &= \langle \mathcal{A}\,X, \Delta y\rangle + \langle X, -\mathcal{A}^*(\Delta y)\rangle + \langle S, \Delta S\rangle \\ &= \langle S, \Delta S\rangle = \langle P^T SP, P^T(\Delta S)P\rangle = 0\,. \end{aligned}$$

Thus, $\Delta y = 0$ and $\Delta S = 0$, which together with $\Delta\varepsilon = 0$ and $\Delta X = 0$, imply the following

$$W(\Delta\varepsilon, \Delta X, \Delta y, \Delta S) = 0 \Longrightarrow (\Delta\varepsilon, \Delta X, \Delta y, \Delta S) = 0\,.$$

This shows that W is nonsingular. So, the proof is completed. □

The significance of Proposition 4.2 is that it allows us to offer a quadratic convergence analysis on several globally convergent smoothing Newton methods presented in [9, 10, 19, 45] for solving the SDP problem even when the strict complementarity condition is not satisfied, i.e., when the condition $\overline{X}+\overline{S}\succ 0$ fails to hold. Instead of working on these different smoothing Newton methods one by one (with some necessary modifications), for simplicity we only use the smoothing Newton method presented in [45] as an example to show how this objective can be achieved.

For any $(\varepsilon, X, y, S)\in \Re\times\mathcal{S}^n\times\Re^m\times\mathcal{S}^n$, write $Z\equiv(X,y,S)$ and define $f(\varepsilon,Z):=\|E(\varepsilon,Z)\|^2$ and $\theta(\varepsilon,Z) := r\min\{1, f(\varepsilon,Z)\}$. Let $\bar\varepsilon\in(0,\infty)$ and $r\in(0,1)$ be such that $r\bar\varepsilon<1$. The smoothing Newton method presented in [45] can then be stated as follows.

Algorithm I (A Squared Smoothing Newton Method)

Step 0. Select constants $\delta\in(0,1)$ and $\sigma\in(0,1/2)$. Let $\varepsilon_0:=\bar\varepsilon$, $Z^0\in\mathcal{S}^n\times\Re^m\times\mathcal{S}^n$ be an arbitrary point, and $k:=0$.

Step 1. If $E(\varepsilon_k, Z^k)=0$, then stop. Otherwise, let $\theta_k:=\theta(\varepsilon_k,Z^k)$.

Step 2. Compute $(\Delta\varepsilon_k,\Delta Z^k)$ by

$$E(\varepsilon_k,Z^k) + E'(\varepsilon_k,Z^k)(\Delta\varepsilon_k,\Delta Z^k) = \theta_k\begin{bmatrix}\bar\varepsilon\\ 0\end{bmatrix}. \tag{4.13}$$

Step 3. Let l_k be the smallest nonnegative integer l satisfying

$$f(\varepsilon_k+\delta^l\Delta\varepsilon_k, Z^k+\delta^l\Delta Z^k) \le [\,1-2\sigma(1-r\bar\varepsilon)\delta^l\,]f(\varepsilon_k,Z^k)\,. \tag{4.14}$$

Define $(\varepsilon_{k+1},Z^{k+1}):=(\varepsilon_k+\delta^{l_k}\Delta\varepsilon_k, Z^k+\delta^{l_k}\Delta Z^k)$.

Step 4. Replace k by $k+1$ and go to Step 1.

The well posedness of Algorithm I hinges on the nonsingularity of $E'(\varepsilon,Z)$ for any $\varepsilon>0$, which is equivalent to the surjectivity of the linear operator $\mathcal{A}:\mathcal{S}^n\to\Re^m$ [45]. The two conditions required for quadratic convergence of Algorithm I are: (i) the strong semismoothness of the smoothing function E and (ii) the nonsingularity of all $W\in\partial_B E(0,Z^*)$ (or all $W\in\partial E(0,Z^*)$). However, (i) has been proven in [45] and (ii) can be derived from Proposition 4.2 under both the primal constraint nondegeneracy and the dual constraint nondegeneracy. Thus, by employing the standard convergence analysis laid out in [29] for the vector version of the squared smoothing Newton method, we have the following convergence theorem. For more explanations on this, see [45].

THEOREM 4.3. *Assume that $\mathcal{A}:\mathcal{S}^n\to\Re^m$ is onto. Then an infinite sequence $\{(\varepsilon_k,Z^k)\}$ is generated by* Algorithm I *and each accumulation point $(0,\overline{Z})$ of $\{(\varepsilon_k,Z^k)\}$ is a solution of $E(\varepsilon,Z)=0$. Let $\overline{Z}=(\overline{X},\bar y,\overline{S})\in\mathcal{S}^n\times\Re^m\times\mathcal{S}^n$. If the primal constraint nondegenerate condition* (3.4) *holds at $\overline{X}$ and the dual constraint nondegenerate condition* (3.6) *holds at $(\bar y,\overline{S})$, then the whole sequence $\{(\varepsilon_k,Z^k)\}$ converges to $(0,\overline{Z})$,*

$$\|\,(\varepsilon_{k+1},Z^{k+1})-(0,\overline{Z})\,\| = O(\|\,(\varepsilon_k,Z^k)-(0,\overline{Z})\,\|^2) \tag{4.15}$$

and

$$\varepsilon_{k+1}=O(\varepsilon_k^2)\,. \tag{4.16}$$

Note that in Theorem 4.3, the quadratic convergence does not rely on the strict complementarity, one common condition that was assumed in all known smoothing Newton methods for solving the SDP problem (1.1) and its dual, as far as we know. The smoothing function G can certainly take other forms. For example, in order to improve the global convergence of Algorithm I, one may consider Tikhonov-type regularized smoothing functions such as

$$G(\varepsilon, X, y, S) := \begin{bmatrix} C - \mathcal{A}^* y - S + \varepsilon X \\ \mathcal{A}X - b + \varepsilon y \\ S - \Phi\big(\varepsilon, S - (X + \varepsilon S)\big) \end{bmatrix} = \begin{bmatrix} C - \mathcal{A}^* y - S + \varepsilon X \\ \mathcal{A}X - b + \varepsilon y \\ X - \Phi\big((X + \varepsilon S) - S\big) + \varepsilon S \end{bmatrix}. \tag{4.17}$$

The quadratic convergence of Algorithm I will not be affected because, by Lemma 2.1, the set $\partial_B E(0, X, S, Y)$ is still kept the same for any $(X, y, S) \in \mathcal{S}^n \times \Re^m \times \mathcal{S}^n$ if one replaces the smoothing function G in (4.2) by the one given in (4.17).

5. Conclusions. In this paper, we presented several equivalent links among the primal and dual constraint nondegenerate conditions, the strong regularity, and the nonsingularity of both the B-subdifferential and Clarke's generalized Jacobian of a nonsmooth system at a KKT point in the context of linear semidefinite programming. These links were further used to derive for the first time a quadratic convergence analysis of globally convergent smoothing Newton methods without assuming the strict complementarity. Variational analysis on the metric projector over the cone of positive semidefinite matrices and its smoothed counterpart plays a fundamental role in achieving these. Given the fact that the metric projector over the more general symmetric cone behaves quite similar to the metric projector over the cone of positive semidefinite matrices [44], one is tempted to wonder if the results obtained in this paper can be extended to linear symmetric cone programming. We leave this interesting question as our future research topic.

Acknowledgments. The authors are grateful to the referees and the associate editor for their helpful suggestions on improving the quality of this paper. Defeng Sun would also like to thank Jong-Shi Pang and Alexander Shapiro for discussions on the constraint nondegeneracy when they visited National University of Singapore during 2002-2005.

REFERENCES

[1] F. Alizadeh, *Interior point methods in semidefinite programming with applications to combinatorial optimization*, SIAM Journal on Optimization, 5 (1995), pp. 13–51.
[2] F. Alizadeh, J. -P. A. Haeberly, and O. L. Overton, *Complementarity and nondegeneracy in semidefinite programming*, Mathematical Programming, 77 (1997), pp. 111–128.
[3] V. I. Arnold, *On matrices depending on parameters*, Russian Mathematical Surveys, 26 (1971), pp. 29–43.
[4] R. Bhatia, *Matrix Analysis*, Springer-Verlag, New York, 1997.
[5] J. F. Bonnans, R. Cominetti, and A. Shapiro, *Sensitivity analysis of optimization problems under second order regularity constraints*, Mathematics of Operations Research, 23 (1998), pp. 803–832.
[6] J. F. Bonnans, R. Cominetti, and A. Shapiro, *Second order optimality conditions based on parabolic second order tangent sets*, SIAM Journal on Optimization 9 (1999), pp. 466–493.
[7] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.
[8] S. Chandrasekaran and I. C. F. Ipsen, *Backward errors for eigenvalue and singular value decompositions*, Numerische Mathematik, 68 (1994), pp. 215–223.

[9] X. Chen, H. -D. Qi, and P. Tseng, *Analysis of nonsmooth symmetric matrix valued functions with applications to semidefinite complementarity problems*, SIAM Journal on Optimization, 13 (2003), pp. 960–985.

[10] X. Chen and P. Tseng, *Non-interior continuation methods for solving semidefinite complementarity problems*, Mathematical Programming, 95 (2003), pp. 431–474.

[11] F. H. Clarke, *On the inverse function theorem*, Pacific Journal of Mathematics, 64 (1976), pp. 97–102.

[12] F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley and Sons, New York, 1983.

[13] B. C. Eaves, *On the basic theorem of complementarity*, Mathematical Programming, 1 (1971), pp. 68–75.

[14] M. L. Flegel and C. Kanzow, *Equivalence of two nondegeneracy conditions for semidefinite programs*, Journal of Optimization Theory and Applications, to appear.

[15] M. S. Gowda, *Inverse and implicit function theorems for H-differentiable and semismooth functions*, Optimization Methods and Software, 19 (2004), pp. 443-461.

[16] N. J. Higham, *Computing a nearest symmetric positive semidefinite matrix*, Linear Algebra and Its Applications, 103 (1988), pp. 103–118.

[17] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.

[18] C. Kanzow and C. Nagel, *Semidefinite programs: new search directions, smoothing-type methods, and numerical results*, SIAM Journal on Optimization, 13 (2002), pp. 1–23. (Corrigendum in SIAM Journal on Optimization, 14 (2003), pp. 936–937).

[19] C. Kanzow and C. Nagel, *Quadratic convergence of a nonsmooth Newton-type method for semidefinite programs without strict complementarity*, SIAM Journal on Optimization, 15 (2005), pp. 654–672.

[20] B. Kummer, *Lipschitzian inverse functions, directional derivatives, and applications in $C^{1,1}$-optimization*, Journal of Optimization Theory and Applications, 70 (1991), pp. 559–580.

[21] K. Löwner, *Über monotone matrixfunktionen*, Mathematische Zeitschrift, 38 (1934), pp. 177–216.

[22] J. Malick and H. S. Sendov, *Clarke generalized Jacobian of the projection onto the cone of positive semidefinite matrices*, Set-Valued Analysis, 14 (2006), pp. 273–293.

[23] T. Matsumoto, *An algebraic condition equivalent to strong stability of stationary solutions of nonlinear positive semidefinite programms*, SIAM Journal on Optimization, 16 (2005), pp. 452–470.

[24] F. Meng, D. F. Sun, and G. Zhao, *Semismoothness of solutions to generalized equations and the Moreau-Yosida regularization*, Mathematical Programming, 104 (2005), pp. 561–581.

[25] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM Journal on Control and Optimization, 15 (1977), pp. 957–972.

[26] Yu. Nesterov and A. Nemirovskii, *Interior-point polynomial algorithms in convex programming*, SIAM Studies in Applied Mathematics 13, Philadelphia, 1994.

[27] J. S. Pang, D. F. Sun, and J. Sun, *Semismooth homeomorphisms and strong stability of semidefinite and Lorentz complementarity problems*, Mathematics of Operations Research, 28 (2003), pp. 39–63.

[28] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Mathematics of Operations Research, 18 (1993), pp. 227–244.

[29] L. Qi, D. F. Sun, and G. Zhou, *A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities*, Mathematical Programming, 87 (2000), pp. 1-35.

[30] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Mathematical Programming, 58 (1993), pp. 353–367.

[31] S. M. Robinson, *Strongly regular generalized equations*, Mathematics of Operations Research, 5 (1980), pp. 43–62.

[32] S. M. Robinson, *Local structure of feasible sets in nonlinear programming, Part II: Nondegeneracy*, Mathematical Programming Study, 22 (1984), pp. 217–230.

[33] S. M. Robinson, *Local structure of feasible sets in nonlinear programming, Part III: Stability and sensitivity*, Mathematical Programming Study, 30 (1987), pp. 45–66.

[34] S. M. Robinson, *Constraint nondegeneracy in variational analysis*, Mathematics of Operations Research, 28 (2003), pp. 201–232.

[35] R. T. Rockafellar, *Convex Analysis*, Princeton, New Jersey, 1970.

[36] R. T. Rockafellar and R. J. -B. Wets, *Variational Analysis*, Springer, New York, 1998.

[37] N. C. Schwertman and D. M. Allen, *Smoothing an indefinite variance-covariance matrix*, Journal of Statistical Computation and Simulation, 9 (1979), pp. 183–194.

[38] A. Shapiro, *First and second order analysis of nonlinear semidefinite programs*, Mathematical Programming, 77 (1997), pp. 301–320.

[39] A. Shapiro, *Sensitivity analysis of generalized equations*, Journal of Mathematical Sciences, 115 (2003), pp. 2554–2565.

[40] A. Shapiro and M. K. H. Fan, *On eigenvalue optimization*, SIAM Journal on Optimization, 5 (1995), pp. 552–569.

[41] D. F. Sun, *The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications*, Mathematics of Operations Research, 31 (2006), pp. 761–776.

[42] D. F. Sun and J. Sun, *Semismooth matrix valued functions*, Mathematics of Operations Research, 27 (2002), pp. 150–169.

[43] D. F. Sun and J. Sun, *Strong semismoothness of Fischer-Burmeister SDC and SOC functions*, Mathematical Programming, 103 (2005), pp. 575–581.

[44] D. F. Sun and J. Sun, *Löwner's operator and spectral functions in Euclidean Jordan algebras*, Mathematics of Operations Research, 33 (2008), to appear.

[45] J. Sun, D. F. Sun, and L. Qi, *A squared smoothing Newton method for nonsmooth matrix equations and its applications in semidefinite optimization problems*, SIAM Journal on Optimization, 14 (2003), pp. 783–806.

[46] M. J. Todd, *Semidefinite optimization*, Acta Numerica, 10 (2001), pp. 515–560.

[47] P. Tseng, *Merit functions for semi-definite complementarity problems*, Mathematical Programming, 83 (1998), pp. 159–185.

[48] E. H. Zarantonello, *Projections on convex sets in Hilbert space and spectral theory I and II*, in Contributions to Nonlinear Functional Analysis, E. H. Zarantonello, eds., Academic Press, New York, 1971, pp. 237–424.