Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 10, No. 2, pp. 285–303, April 2014

THE DIFFERENTIABLE EXACT PENALTY FUNCTION FOR NONLINEAR SEMIDEFINITE PROGRAMMING

LE HAN*

Abstract: This paper is concerned with a differentiable exact penalty function for the nonlinear semidefinite programming problem, which can be viewed as an extension of the one proposed by Fukuda, Silva and Fukushima (2012) for nonlinear second-order cone programming, which in turn is an extension of the one introduced by Di Pillo and Grippo (1989) for nonlinear programming. For the exact penalty function, we particularly show that its gradient has the positive definite Clarke generalized Jacobian at the KKT points if the constraint nondegeneracy condition and the strong second-order sufficient condition hold. This second-order property implies the locally superlinear (or quadratic) convergence when the semismooth Newton method is applied for the minimization of this penalty function.

Key words: *nonlinear semidefinite programming, differentiable exact penalty function, constraint nondegeneracy, strong second-order sufficient condition.*

Mathematics Subject Classification: *90C22, 90C26*

1 Introduction

Let $\mathbb{X}$ be a finite dimensional real vector space endowed with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$. Consider the nonlinear semidefinite programming

$$\min_{x \in \mathbb{X}} \left\{ f(x) : \ h(x) = 0, \ g(x) \in \mathbb{S}^n_+ \right\} \tag{1.1}$$

where $f\colon \mathbb{X} \to \mathbb{R}, h\colon \mathbb{X} \to \mathbb{R}^m$, $g\colon \mathbb{X} \to \mathbb{S}^n$ are twice continuously differentiable functions, $\mathbb{S}^n$ is the linear space of all $n \times n$ real symmetric matrices, and $\mathbb{S}^n_+$ is the cone of all $n \times n$ positive semidefinite matrices. This class of problems have an extensive applications in various fields such as control, finance, engineering design, robust optimization, combinatorial optimization, and so on (see, e.g., [4, 5, 32, 36]).

For the linear semidefinite programming problem, many effective algorithms have been developed in the past decade; for example, the interior point methods (see, e.g., [1, 17, 37, 33, 34]), the smoothing Newton methods (see, e.g., [9, 15, 30, 8, 13]), the multiplier methods [38, 16], and the regularization methods [19, 23]. However, for the nonlinear semidefinite programming problem, the theoretical and algorithmic study are still insufficient. For example, until recently Sun [29] established a verifiable characterization for the strong regularity

*The author's work is supported by the Fundamental Research Funds for the Central Universities (SCUT, 2013ZM0114, No. x2lxD2115730).

of locally optimal solutions via the strong second-order sufficient condition and the nonsingularity of Clarke's Jacobian of the KKT system, respectively. Later, Bi et al.[6] follow this line to provide another characterization for the strong regularity of locally optimal solutions via the nonsingularity study of the Clarke's Jacobian of the FB nonsmooth system. We also note that the modified barrier function methods were extended in [27] to solve the nonlinear semidefinite programming problem.

This paper is concerned with a class of differentiable exact penalty functions for the nonlinear semidefinite programming. It is well known that the exact penalty function methods inherit the advantage of the multiplier methods that only requires the penalty parameter to be over a certain threshold. Also, if the penalty parameter is appropriately chosen, the exact penalty methods may yield a desired solution of the original problem by solving a single penalty problem, instead of a sequence of penalty problems. For the polyhedral cone optimization problems, the differentiable and nondifferentiable exact penalty function methods are well studied (see, e.g., [2, 21, 22]). However, for the nonpolyhedral cone optimization problems, there are few works devoted to the exact penalty functions and the corresponding penalty function methods. Recently, Fukuda et al. [11] study the differentiable exact penalty function for the nonlinear second-order cone optimization problems, and propose a penalty function method with a superlinear (or quadratic) convergence rate but without the strict complementarity condition. Motivated by this, we focus on the differentiable exact penalty functions for the nonlinear semidefinite programming.

Specifically, for any given $x \in \mathbb{X}$, we first give an explicit estimation for the Lagrange multipliers via a strongly convex minimization problem with respect to multiplier variables, and then employ the multiplier estimation to construct a differentiable penalty function. We show that this penalty function is exact in the sense of Definition 3.1, and establish a desirable second-order property for this differentiable exact penalty function, i.e., the Clarke's Jacobian of its gradient function is positive definite at the KKT points if the constraint nondegeneracy condition and the strong second-order sufficient condition are satisfied. This property implies the locally superlinear (or quadratic) convergence when the semismooth Newton method is applied for the minimization of this penalty function. Note that the analysis techniques are different from that of [11] used for dealing with the nonlinear second-order cone optimization problems.

It is worthwhile to mention that Correa and Ramírez [10] gave two nondifferentiable exact penalty functions associated with (1.1). Their penalty functions are constructed by fixing the Lagrange multipliers involved in the proximal augmented Lagrangian. On the contrast, our exact penalty functions are differentiable and constructed by replacing the Lagrange multipliers involved in the augmented Lagrangian function with continuously differentiable multiplier functions with respect to the primal variables.

Throughout this paper, $\mathcal{I}$ represents an identity operator of appropriate dimension, $\mathbb{R}_{++}$ is the set of all positive real numbers, $\mathcal{T}_{\mathbb{S}^n_+}(X)$ denotes the tangent cone of $\mathbb{S}^n_+$ at a point $X \in \mathbb{S}^n_+$, and $\text{lin}(\mathcal{T}_{\mathbb{S}^n_+}(X))$ denotes the largest linear space in $\mathcal{T}_{\mathbb{S}^n_+}(X)$. From [7], the tangent cone $\mathcal{T}_{\mathbb{S}^n_+}(X)$ is defined as

$$\mathcal{T}_{\mathbb{S}^n_+}(X) := \left\{H \in \mathbb{S}^n : \text{dist}\left(X + tH, \mathbb{S}^n_+\right) = o(t),\ t \geq 0\right\}, \quad \forall X \in \mathbb{S}^n_+$$

where $\text{dist}(Y, \mathbb{S}^n_+) := \inf\left\{\|Y - Z\|_F : Z \in \mathbb{S}^n_+\right\}$ denotes the distance of Y from $\mathbb{S}^n_+$. With the directional derivative of projection operator $\Pi_{\mathbb{S}^n_+}(\cdot)$, it is easy to verify that

$$\mathcal{T}_{\mathbb{S}^n_+}(X) = \left\{H \in \mathbb{S}^n : H = \Pi'_{\mathbb{S}^n_+}(X; H)\right\}. \tag{1.2}$$

For any $n \times m$ real matrices A and B, $\langle A, B\rangle := \mathrm{tr}(A^T B)$ means the Frobenius inner product of A and B, and $\|A\|_F$ is the norm of A induced by the Frobenius inner product. For any given sets of indices α and β, we designate by $A_{\alpha\beta}$ the submatrix of A whose row indices belong to α and column indices belong to β. For a linear operator $\mathcal{A}$, we denote by $\mathcal{A}^*$ the adjoint of $\mathcal{A}$.

This paper is organized as follows. In Section 2, the smooth multiplier function and its properties are characterized in order to construct the differentiable exact penalty function for nonlinear semidefinite programming. In Section 3, we prove that the penalty function is an exact penalty function for nonlinear semidefinite programming. In Section 4, we show that the Clarke's Jacobian of the gradient of the exact penalty function at a KKT point is positive definite under the constraint nondegeneracy and the strong second-order sufficient condition. Specially, for linear semidefinite programming the strong second-order sufficient condition is weakened as the dual constraint nondegeneracy. Conclusions are given in Section 5.

2 Estimation of Multipliers

The differentiable exact penalty functions were early proposed in [22] for nonlinear programming problems, and were recently extended to the setting of the nonlinear second-order optimization by [11]. The basic idea to construct the differentiable exact penalty functions is to replace the multipliers involved in the augmented Lagrangian function with the continuously differentiable multiplier functions with respect to the primal variables. To construct the differentiable exact penalty function for the nonlinear semidefinite programming, we in this section present a characterization for the smooth multiplier function.

Let $L : \mathbb{X} \times \mathrm{I\!R}^m \times \mathbb{S}^n_+ \to \mathrm{I\!R}$ denote the Lagrangian function of problem (1.1), i.e.,

$$L(x, \mu, Y) := f(x) + \langle \mu, h(x)\rangle - \langle Y, g(x)\rangle \qquad \forall (x, \mu, Y) \in \mathbb{X} \times \mathrm{I\!R}^m \times \mathbb{S}^n_+.$$

For any given $X \in \mathbb{S}^n$, let $\mathcal{L}_X : \mathbb{S}^n \to \mathbb{S}^n$ be the Lyapunov operator associated with X :

$$\mathcal{L}_X(Y) := XY + YX \quad \forall Y \in \mathbb{S}^n.$$

Then, the Karush-Kuhn-Tucker (KKT) optimality conditions for problem (1.1) take the form of

$$\begin{cases} \nabla_x L(x, \mu, Y) = 0, \\ h(x) = 0, \\ g(x) \in \mathbb{S}^n_+, \ Y \in \mathbb{S}^n_+, \ \mathcal{L}_{g(x)}(Y) = 0, \end{cases} \tag{2.1}$$

where $\nabla_x L(x, \mu, Y)$ is the gradient of L at (x, μ, Y) with respect to x.

For any given $x \in \mathbb{X}$, we consider the multiplier estimation defined by the following minimization problem

$$(\mu(x), Y(x)) := \underset{\mu, Y}{\mathrm{argmin}} \ \|\nabla_x L(x, \mu, Y)\|^2 + \zeta_1^2 \|\mathcal{L}_{g(x)}(Y)\|_F^2 + \zeta_2^2 \alpha(x)(\|Y\|_F^2 + \|\mu\|^2), \tag{2.2}$$

where $\zeta_1, \zeta_2 > 0$ and

$$\alpha(x) := \frac{1}{2}\left(\|h(x)\|^2 + \|\Pi_{\mathbb{S}^n_+}(-g(x))\|_F^2\right).$$

To the best of our knowledge, the basic idea of multiplier estimation was given in [12] and the employment of the feasible measure-type function α was proposed in [18], moreover, Fukuda et al. [11] extended them to the multiplier estimation for nonlinear second-order cone programming. Noting that $\alpha(x) = 0$ if and only if x is feasible for (1.1), we use the feasible measure-type function α and the Lyapunov operator to construct the multiplier estimation (2.2) aiming to satisfy $\nabla_x L(x,\mu,Y) = 0, \mathcal{L}_{g(x)}(Y) = 0$, and $h(x) = 0, g(x) \in \mathbb{S}^n_+$ when $(\mu(x), Y(x)) \neq 0$, which force the KKT conditions (2.1) to hold. In addition, we must point out that the parameters ζ_1 and ζ_2 can not be omitted since their values may impact the robustness of the method, as we can see in [3] for the nonlinear programming case.

To study the properties of $(\mu(x), Y(x))$, we first recall the definition of nondegeneracy.

Definition 2.1 ([7])**.** A feasible point $\overline{x}$ of the nonlinear semidefinite programming is called constraint nondegenerate if

$$\begin{pmatrix} \mathcal{J}h(\overline{x}) \\ \mathcal{J}g(\overline{x}) \end{pmatrix} \mathbb{X} + \begin{pmatrix} \{0\} \\ \mathrm{lin}(\mathcal{T}_{\mathbb{S}^n_+}(g(\overline{x}))) \end{pmatrix} = \begin{pmatrix} \mathbb{R}^m \\ \mathbb{S}^n \end{pmatrix}, \tag{2.3}$$

where $\mathcal{J}h(\overline{x})$ and $\mathcal{J}g(\overline{x})$ denote the Jacobian of h and g at $\overline{x}$, respectively.

In addition, we also need the linear operator $\mathcal{N}(x)\colon \mathbb{S}^n \times \mathbb{R}^m \to \mathbb{S}^n \times \mathbb{R}^m$ defined by

$$\mathcal{N}(x) := \begin{pmatrix} \mathcal{J}g(x)\mathcal{J}g(x)^T + \zeta_1^2 \mathcal{L}^*_{g(x)}\mathcal{L}_{g(x)} & -\mathcal{J}g(x)\mathcal{J}h(x)^T \\ -\mathcal{J}h(x)\mathcal{J}g(x)^T & \mathcal{J}h(x)\mathcal{J}h(x)^T \end{pmatrix} + \zeta_2^2\alpha(x) \begin{pmatrix} \mathcal{I}_n & 0 \\ 0 & \mathcal{I}_m \end{pmatrix}. \tag{2.4}$$

Lemma 2.2. *For any given $x \in \mathbb{X}$, if x is an infeasible point or a nondegenerate feasible point of problem (1.1), then the operator $\mathcal{N}(x)$ in (2.4) is symmetric and positive definite.*

Proof. The symmetry of $\mathcal{N}(x)$ is clear by its definition (2.4). We next prove that $\mathcal{N}(x)$ is positive definite. Let (Z, ξ) be an arbitrary nonzero element from $\mathbb{S}^n \times \mathbb{R}^m$. Then,

$$\begin{aligned} \langle (Z,\xi), \mathcal{N}(x)(Z,\xi)\rangle &= \left\langle Z, \mathcal{J}g(x)\mathcal{J}g(x)^T Z + \zeta_1^2 \mathcal{L}^*_{g(x)}\mathcal{L}_{g(x)}(Z) + \zeta_2^2\alpha(x)Z - \mathcal{J}g(x)\mathcal{J}h(x)^T\xi \right\rangle \\ &\quad + \left\langle \xi, -\mathcal{J}h(x)\mathcal{J}g(x)^T Z + \mathcal{J}h(x)\mathcal{J}h(x)^T\xi + \zeta_2^2\alpha(x)\xi \right\rangle \\ &= \|\mathcal{J}g(x)^T Z - \mathcal{J}h(x)^T\xi\|^2 + \zeta_1^2|\mathcal{L}_{g(x)}(Z)\|_F^2 + \zeta_2^2\alpha(x)(\|Z\|_F^2 + \|\xi\|^2). \end{aligned}$$

If x is infeasible, then $\langle (Z,\xi), \mathcal{N}(x)(Z,\xi)\rangle > 0$ since $\alpha(x) \neq 0$. Let x be a nondegenerate feasible point. Noting that $\alpha(x) = 0$, we only need to prove that

$$\|\mathcal{J}g(x)^T Z - \mathcal{J}h(x)^T\xi\|^2 + \zeta_1^2\|\mathcal{L}_{g(x)}(Z)\|_F^2 > 0.$$

If not, then we have $\mathcal{J}g(x)^T Z - \mathcal{J}h(x)^T\xi = 0$ and $\mathcal{L}_{g(x)}(Z) = 0$ for some nonzero $(Z,\xi) \in \mathbb{S}^n \times \mathbb{R}^m$. From $\mathcal{J}g(x)^T Z - \mathcal{J}h(x)^T\xi = 0$, it follows that $\langle \mathcal{J}g(x)^T Z + \mathcal{J}h(x)^T(-\xi), z\rangle = 0$ for any $z \in \mathbb{X}$. Thus,

$$0 = \langle \mathcal{J}g(x)^T Z, z\rangle + \langle \mathcal{J}h(x)^T(-\xi), z\rangle = \langle Z, \mathcal{J}g(x)z\rangle + \langle -\xi, \mathcal{J}h(x)z\rangle,$$

which in turn implies that

$$\begin{pmatrix} -\xi \\ Z \end{pmatrix} \in \left(\begin{pmatrix} \mathcal{J}h(x) \\ \mathcal{J}g(x) \end{pmatrix} \mathbb{X} \right)^{\perp},$$

where

$$\left(\begin{pmatrix} \mathcal{J}h(x) \\ \mathcal{J}g(x) \end{pmatrix} \mathbb{X} \right)^{\perp} := \left\{ \begin{pmatrix} \mu \\ Y \end{pmatrix} \in \begin{pmatrix} \mathbb{R}^m \\ \mathbb{S}^n \end{pmatrix} : \left\langle \begin{pmatrix} \mu \\ Y \end{pmatrix}, \begin{pmatrix} \mathcal{J}h(x) \\ \mathcal{J}g(x) \end{pmatrix} \mathbb{X} \right\rangle = 0 \right\}.$$

From $\mathcal{L}_{g(x)}(Z) = 0$, we have $g(x)Z + Zg(x) = 0$. Without loss of generality, assume that $g(x)$ has the spectral decomposition $g(x) = P\Lambda P^T = P\text{diag}(\lambda_1, \dots, \lambda_n)P^T$, where $\text{diag}(\lambda_1, \dots, \lambda_n)$ is the diagonal matrix of eigenvalues of $g(x)$ and P is a corresponding orthogonal matrix of orthogonal eigenvectors. Noting that $g(x) \in \mathbb{S}^n_+$, we define

$$\kappa := \{i : \lambda_i > 0\} \quad \text{and} \quad \beta := \{i : \lambda_i = 0\}.$$

Then, by permuting the rows and columns of $g(x)$ if necessary, we may assume that

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

with $P_\kappa \in \mathbb{R}^{n\times|\kappa|}$, $P_\beta \in \mathbb{R}^{n\times|\beta|}$. Then, $g(x)Z + Zg(x) = 0$ can be equivalently written as

$$0 = \Lambda\widetilde{Z} + \widetilde{Z}\Lambda = \begin{bmatrix} \Lambda_\kappa\widetilde{Z}_{\kappa\kappa} + \widetilde{Z}_{\kappa\kappa}\Lambda_\kappa & \Lambda_\kappa\widetilde{Z}_{\kappa\beta} \\ \widetilde{Z}_{\beta\kappa}\Lambda_\kappa & 0 \end{bmatrix}$$

with $\widetilde{Z} = P^T ZP$, which in turn implies that

$$P_\beta^T ZP_\kappa = 0,\ P_\kappa^T ZP_\beta = 0 \text{ and } P_\kappa^T ZP_\kappa = 0. \tag{2.5}$$

In addition, using the spectral decomposition of $g(x)$ and (1.2), it is not hard by [29] to deduce that

$$\text{lin}\left(\mathcal{T}_{\mathbb{S}^n_+}(g(x))\right) = \left\{B \in \mathbb{S}^n : \ P_\beta^T BP_\beta = 0\right\}. \tag{2.6}$$

Then, from (2.5) and (2.6) we have that for any $B \in \text{lin}\left(\mathcal{T}_{\mathbb{S}^n_+}(g(x))\right)$ it holds that

$$\langle Z, B\rangle = \langle P^T ZP, P^T BP\rangle = \left\langle \begin{bmatrix} 0 & 0 \\ 0 & \widetilde{Z}_{\beta\beta} \end{bmatrix}, \begin{bmatrix} \widetilde{B}_{\kappa\kappa} & \widetilde{B}_{\kappa\beta} \\ \widetilde{B}_{\beta\kappa} & 0 \end{bmatrix} \right\rangle = 0,$$

which implies that $\begin{pmatrix} -\xi \\ Z \end{pmatrix} \in \begin{pmatrix} \{0\} \\ \text{lin}\left(\mathcal{T}_{\mathbb{S}^n_+}(g(x))\right) \end{pmatrix}^\perp$. The above arguments show that the nonzero vector $\begin{pmatrix} -\xi \\ Z \end{pmatrix}$ satisfies

$$\begin{pmatrix} -\xi \\ Z \end{pmatrix} \in \left(\begin{pmatrix} \mathcal{J}h(x) \\ \mathcal{J}g(x) \end{pmatrix}\mathbb{X}\right)^\perp \bigcap \begin{pmatrix} \{0\} \\ \text{lin}\left(\mathcal{T}_{\mathbb{S}^n_+}(g(x))\right) \end{pmatrix}^\perp,$$

which contradicts the fact that x is nondegenerate. The proof is completed. □

Proposition 2.3. *For any given $x \in \mathbb{X}$, if x is an infeasible point or a nondegenerate feasible point of problem (1.1), then the following statements hold.*

(a) *The problem (2.2) has a unique optimal solution $(\mu(x), Y(x))$ that takes the form of*

$$\begin{pmatrix} Y(x) \\ \mu(x) \end{pmatrix} = \mathcal{N}(x)^{-1} \begin{pmatrix} \mathcal{J}g(x) \\ -\mathcal{J}h(x) \end{pmatrix} \nabla f(x). \tag{2.7}$$

In particular, if $(x^*, \mu^*, Y^*) \in \mathbb{X} \times \mathbb{R}^m \times \mathbb{S}^n$ satisfies the KKT conditions (2.1), then $\mu(x^*) = \mu^*$ and $Y(x^*) = Y^*$.

(b) *The Jacobians of the multiplier functions $Y(\cdot)$ and $\mu(\cdot)$ are given by*

$$\begin{pmatrix} \mathcal{J}Y(x) \\ \mathcal{J}\mu(x) \end{pmatrix} = \mathcal{N}(x)^{-1} \begin{pmatrix} R_1(x) \\ -R_2(x) \end{pmatrix}$$

with

$$\begin{aligned} R_1(x) &:= \mathcal{J}^2 g(x)\nabla_x L(x,\mu(x),Y(x)) + \mathcal{J}g(x)\nabla^2_{xx}L(x,\mu(x),Y(x)) \\ &\quad -\zeta_1^2\mathcal{J}(\mathcal{L}^*_{g(x)}\mathcal{L}_{g(x)}(Y(x))) - \zeta_2^2\mathcal{J}\alpha(x)Y(x) \\ R_2(x) &:= \mathcal{J}^2 h(x)\nabla_x L(x,\mu(x),Y(x)) + \mathcal{J}h(x)\nabla^2_{xx}L(x,\mu(x),Y(x)) + \zeta_2^2\mathcal{J}\alpha(x)\mu(x), \end{aligned}$$

where $\mathcal{J}^2 g(x)$ and $\mathcal{J}^2 h(x)$ denote the second-order Jacobian of g and h at x, respectively, $\mathcal{J}\alpha(x)$ denotes the Jacobian of α at x and

$$\begin{aligned} \nabla_x L(x,\mu(x),Y(x)) &:= \nabla_x L(x,\mu,Y)|_{\mu=\mu(x),Y=Y(x)}, \\ \nabla^2_{xx} L(x,\mu(x),Y(x)) &:= \nabla^2_{xx} L(x,\mu,Y)|_{\mu=\mu(x),Y=Y(x)}. \end{aligned}$$

Proof. (a) Note that the minimization problem (2.2) can be equivalently written as

$$(\mu(x),Y(x)) := \operatorname*{argmin}_{\mu,Y} \left\| \mathcal{A}(x) \begin{pmatrix} Y \\ \mu \end{pmatrix} - \mathcal{B}(x) \right\|^2$$

with

$$\mathcal{A}(x) := \begin{pmatrix} -\mathcal{J}g(x)^T & \mathcal{J}h(x)^T \\ \zeta_1\mathcal{L}_{g(x)} & 0 \\ \zeta_2\alpha(x)^{1/2}\mathcal{I}_n & 0 \\ 0 & \zeta_2\alpha(x)^{1/2}\mathcal{I}_m \end{pmatrix} \quad \text{and} \quad \mathcal{B}(x) := \begin{pmatrix} -\nabla f(x) \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

By the definition of $\mathcal{N}(x)$, it is not hard to verify that $\mathcal{N}(x) = \mathcal{A}^*(x)\mathcal{A}(x)$. This, together with Lemma 2.2, implies that the minimization problem (2.2) is strictly convex, and consequently it has the unique optimal solution $(\mu(x),Y(x))$. An elementary calculation yields (2.7). Now assume that $(x^*,\mu^*,Y^*) \in \mathbb{X}\times\mathbb{R}^m\times\mathbb{S}^n$ satisfies the KKT conditions, that is, $\nabla_x L(x^*,\mu^*,Y^*) = 0$, $\mathcal{L}_{g(x^*)}(Y^*) = 0$ and $\alpha(x^*) = 0$. Then, the objective function of (2.2) with x replaced by x^* is equal to zero. Since this objective function is always nonnegative, (μ^*,Y^*) is a solution to the problem. This proves part (a).

(b) From part (a), it is immediate to have the following equalities

$$-\mathcal{J}g(x)\nabla_x L(x,\mu(x),Y(x)) + \zeta_1^2\mathcal{L}^*_{g(x)}\mathcal{L}_{g(x)}(Y(x)) + \zeta_2^2\alpha(x)Y(x) = 0 \tag{2.8}$$

$$\mathcal{J}h(x)\nabla_x L(x,\mu(x),Y(x)) + \zeta_2^2\alpha(x)\mu(x) = 0. \tag{2.9}$$

Making differentiation to (2.8) and (2.9) with respect to x yields the desired result. □

Note that if x is a feasible point for nonlinear semidefinite programming, constraint nondegeneracy is a sufficient condition for Proposition 2.3. Then, in the subsequent analysis, we always assume constraint nondegeneracy is satisfied at the feasible point x.

3 Exact Penalty Function

It is known that the augmented Lagrangian function of (1.1), given by Rockafellar and Wets [26], Shapiro and Sun [28] and Sun et al. [31], takes the form of

$$L_c(x,\mu,Y) = f(x) + \langle h(x),\mu\rangle + \frac{c}{2}\|h(x)\|^2 + \frac{1}{2c}(\|\Pi_{\mathbb{S}^n_+}(Y - cg(x))\|_F^2 - \|Y\|_F^2),$$

where $c > 0$ is the penalty parameter. Now replacing (μ, Y) by the multiplier function $(\mu(x), Y(x))$ defined in (2.7), we get a function only involving the primal variable x:

$$\Phi_c(x) := f(x) + \langle h(x),\mu(x)\rangle + \frac{c}{2}\|h(x)\|^2 + \frac{1}{2c}\left(\|\Pi_{\mathbb{S}^n_+}(Y(x) - cg(x))\|_F^2 - \|Y(x)\|_F^2\right).$$

This section is devoted to proving that $\Phi_c\colon \mathbb{X} \to \mathrm{I\!R}$ is an exact penalty function of (1.1) in the sense of Definition 3.1 below. The structure of the proofs is based on [2, 11, 22] and the Definition 3.1 comes from [21] without the extra compact set. For the convenience of discussions, in the rest of this paper, we denote by $\mathcal{G}^*$ and $\mathcal{S}^*$ the set of global minimizers and the set of local minimizers of (1.1), respectively, and denote by $\mathcal{G}(\Phi_c)$ and $\mathcal{S}(\Phi_c)$ the set of global minimizers and the set of local minimizers of Φ_c, respectively.

Definition 3.1. The function $\Phi_c : \mathbb{X} \to \mathrm{I\!R}$ is called an exact penalty function of (1.1) if there exists $\widehat{c} > 0$ such that $\mathcal{G}^* = \mathcal{G}(\Phi_c)$ and $\mathcal{S}(\Phi_c) \subseteq \mathcal{S}^*$ for all $c > \widehat{c}$.

We first study the relation between the set of stationary points of (1.1) and that of Φ_c. Here we call $x \in \mathbb{X}$ a stationary point of (1.1) if there exists $(\mu, Y) \in \mathrm{I\!R}^m \times \mathbb{S}^n_+$ such that (x,μ,Y) is a KKT triple of (1.1). Note that Φ_c is continuously differentiable everywhere in $\mathbb{X}$. Therefore, the set of stationary points of Φ_c is given by $\{x \in \mathbb{X} : \nabla\Phi_c(x) = 0\}$ where, for any $x \in \mathbb{X}$,

$$\nabla\Phi_c(x) = \nabla_x L(x,\mu(x),Y(x)) + [\mathcal{J}\mu(x) + c\mathcal{J}h(x)]^T h(x) + [\mathcal{J}Y(x) - c\mathcal{J}g(x)]^T y_c(x),$$

where

$$y_c(x) := c^{-1}\left[\Pi_{\mathbb{S}^n_+}(Y(x) - cg(x)) - Y(x)\right]. \tag{3.1}$$

The following lemma gives a characterization for the zeros of y_c, which will be used later.

Lemma 3.2. *For any $x \in \mathbb{X}$, $y_c(x) = 0$ if and only if $\mathbb{S}^n_+ \ni Y(x)\perp g(x) \in \mathbb{S}^n_+$.*

Proof. From [35, Lemma 2.1(b)], we know that $Y(x) - \Pi_{\mathbb{S}^n_+}(Y(x) - cg(x)) = 0$ is equivalent to $\mathbb{S}^n_+ \ni Y(x)\perp cg(x) \in \mathbb{S}^n_+$. Noticing that c is greater than 0, we complete this proof. □

Let x^* be a stationary point of (1.1). Then, there exists $(\mu^*, Y^*) \in \mathrm{I\!R}^m \times \mathbb{S}^n_+$ such that (x^*,μ^*,Y^*) is a KKT triple of (1.1). By Proposition 2.3 (a), it is not hard to obtain that $\mathbb{S}^n_+ \ni Y(x^*)\perp g(x^*) \in \mathbb{S}^n_+$. From Lemma 3.2, it follows that $y_c(x^*) = 0$. This, together with the fact that $\mu(x^*) = \mu^*$,$Y(x^*) = Y^*$ and (x^*,μ^*,Y^*) is a KKT triple, implies that x^* is a stationary point of Φ_c. This shows that the set of stationary points of (1.1) is included in that of Φ_c associated to any $c > 0$. However, the converse does not hold generally. The following proposition states that under a certain condition the stationary point of Φ_c associated with a sufficiently large penalty parameter $c > 0$ will be that of (1.1).

Proposition 3.3. *Let $\widehat{x}$ be a feasible point of (1.1). Suppose that there exist sequences $\{x^k\} \subset \mathbb{X}$ with $x^k \to \widehat{x}$ and $\{c_k\} \subset \mathrm{I\!R}_{++}$ with $c_k \to \infty$ such that $\nabla\Phi_{c_k}(x^k) = 0$ for all k. Then, there exists $\widehat{k} > 0$ such that for any $k > \widehat{k}$, x^k is a stationary point of (1.1).*

Proof. First, for any given $c > 0$ and $x \in \mathbb{X}$, we present a simplified expression for $-c^{-1}\mathcal{J}g(x)\nabla\Phi_c(x)$ and $c^{-1}\mathcal{J}h(x)\nabla\Phi_c(x)$. To this end, let $\widetilde{y}_c(x) := y_c(x) + g(x)$. By the definition of $y_c(x)$, it is easy to obtain that $\widetilde{y}_c(x) = c^{-1}\Pi_{\mathbb{S}^n_+}(cg(x) - Y(x))$. Notice that

$$\langle \Pi_{\mathbb{S}^n_+}(cg(x) - Y(x)), \Pi_{\mathbb{S}^n_+}(Y(x) - cg(x))\rangle = 0.$$

Hence, we have that $\mathcal{L}_{\widetilde{y}_c(x)}\left(\Pi_{\mathbb{S}^n_+}(Y(x) - cg(x))\right) = 0$. Along with the definition of $y_c(x)$, we obtain $\mathcal{L}_{\widetilde{y}_c(x)}(y_c(x)) = -c^{-1}\mathcal{L}_{\widetilde{y}_c(x)}(Y(x))$. Thus, it holds that

$$\mathcal{L}^*_{g(x)}\mathcal{L}_{g(x)}(Y(x)) = \mathcal{L}^*_{g(x)}(\mathcal{L}_{\widetilde{y}_c(x)} - \mathcal{L}_{y_c(x)})(Y(x)) = -\mathcal{L}^*_{g(x)}\left(c\mathcal{L}_{\widetilde{y}_c(x)} + \mathcal{L}_{Y(x)}\right)(y_c(x)). \quad (3.2)$$

In addition, from the formula of $\nabla\Phi_c(x)$, it follows that

$$\begin{aligned}\mathcal{J}g(x)\nabla\Phi_c(x) \quad = \quad & \mathcal{J}g(x)\nabla_x L(x,\mu(x),Y(x)) + \mathcal{J}g(x)[\mathcal{J}\mu(x) + c\mathcal{J}h(x)]^T h(x)\\ & +\mathcal{J}g(x)[\mathcal{J}Y(x) - c\mathcal{J}g(x)]^T y_c(x).\end{aligned} \quad (3.3)$$

Combining the equality in (3.3) with equations (2.8) and (3.2), we have that

$$\mathcal{J}g(x)\nabla\Phi_c(x) = -c\widetilde{\mathcal{N}}_c(x)y_c(x) + \mathcal{J}g(x)[\mathcal{J}\mu(x) + c\mathcal{J}h(x)]^T h(x) + \zeta_2^2\alpha(x)Y(x), \quad (3.4)$$

where

$$\widetilde{\mathcal{N}}_c(x) := \mathcal{J}g(x)\left(\mathcal{J}g(x) - c^{-1}\mathcal{J}Y(x)\right)^T + \zeta_1^2\mathcal{L}^*_{g(x)}\left(\mathcal{L}_{\widetilde{y}_c(x)} + c^{-1}\mathcal{L}_{Y(x)}\right).$$

Similarly, combining $\mathcal{J}h(x)\nabla\Phi_c(x)$ with the equality in (2.9), we obtain

$$\begin{aligned}\mathcal{J}h(x)\nabla\Phi_c(x) = {} & \mathcal{J}h(x)[\mathcal{J}Y(x) - c\mathcal{J}g(x)]^T y_c(x)\\ & +\mathcal{J}h(x)[\mathcal{J}\mu(x) + c\mathcal{J}h(x)]^T h(x) - \zeta_2^2\alpha(x)\mu(x).\end{aligned} \quad (3.5)$$

The equations (3.4) and (3.5) can be equivalently written as

$$\frac{1}{c}\begin{pmatrix} -\mathcal{J}g(x) \\ \mathcal{J}h(x)\end{pmatrix}\nabla\Phi_c(x) = \mathcal{N}_c(x)\begin{pmatrix} y_c(x) \\ h(x)\end{pmatrix} - \frac{1}{c}\zeta_2^2\alpha(x)\begin{pmatrix} Y(x) \\ \mu(x)\end{pmatrix}, \quad (3.6)$$

where

$$\mathcal{N}_c(x) := \begin{pmatrix} \widetilde{\mathcal{N}}_c(x) & -\mathcal{J}g(x)(c^{-1}\mathcal{J}\mu(x) + \mathcal{J}h(x))^T \\ \mathcal{J}h(x)(c^{-1}\mathcal{J}Y(x) - \mathcal{J}g(x))^T & \mathcal{J}h(x)(c^{-1}\mathcal{J}\mu(x) + \mathcal{J}h(x))^T\end{pmatrix}.$$

Using equation (3.6) and noting that $\alpha(x) \leq \frac{1}{2}(\|h(x)\|^2 + \|y_c(x)\|_F^2)$, which can be obtained by the definition (3.1) and the inequality $\|\Pi_{\mathbb{S}^n_+}(Y(x) - cg(x)) - Y(x)\|_F \geq \|\Pi_{\mathbb{S}^n_+}(-cg(x))\|_F$, we have that

$$\begin{aligned}&\frac{1}{c^2}\left\|\begin{pmatrix} -\mathcal{J}g(x) \\ \mathcal{J}h(x)\end{pmatrix}\nabla\Phi_c(x)\right\|^2\\ \geq \quad & \frac{1}{2}\left\|\mathcal{N}_c(x)\begin{pmatrix} y_c(x) \\ h(x)\end{pmatrix}\right\|^2 - \frac{\zeta_2^4}{c^2}\alpha(x)^2\left(\|Y(x)\|_F^2 + \|\mu(x)\|^2\right)\\ \geq \quad & \frac{1}{2}\left\|\mathcal{N}_c(x)\begin{pmatrix} y_c(x) \\ h(x)\end{pmatrix}\right\|^2 - \frac{\zeta_2^4}{2c^2}\alpha(x)(\|h(x)\|^2 + \|y_c(x)\|_F^2)\left(\|Y(x)\|_F^2 + \|\mu(x)\|^2\right).\end{aligned}$$

Noting that $g(\widehat{x}) \in \mathbb{S}^n_+$, we have that $\widetilde{y}_{c_k}(\widehat{x}) \to g(\widehat{x})$ as $k \to \infty$, which by the definition of $\mathcal{N}_c(x)$ and $\alpha(\widehat{x}) = 0$ implies that $\mathcal{N}_{c_k}(x^k) \to \mathcal{N}(\widehat{x})$. Since $\mathcal{N}(\widehat{x})$ is nonsingular by Lemma 2.2, $\mathcal{N}_{c_k}(x^k)$ is also nonsingular for sufficiently large k. That is, there exists $\widehat{k} > 0$ such that for all $k > \widehat{k}$, $\mathcal{N}_{c_k}(x^k)$ is nonsingular, and consequently, $\mathcal{N}_{c_k}(x^k)^*\mathcal{N}_{c_k}(x^k)$ is positive definite. Thus, we have the following inequality

$$\left\| \mathcal{N}_{c^k}(x^k) \begin{pmatrix} y_{c^k}(x^k) \\ h(x^k) \end{pmatrix} \right\|^2 \geq \sigma^k_{\min} \left(\|y_{c^k}(x^k)\|_F^2 + \|h(x^k)\|^2 \right),$$

where $\sigma^k_{\min}$ means the smallest eigenvalue of the operator $\mathcal{N}_{c_k}(x^k)^*\mathcal{N}_{c_k}(x^k)$. Note that $\mathcal{N}(\widehat{x})^*\mathcal{N}(\widehat{x})$ is positive definite by Lemma 2.2. Hence, it holds that $\sigma^k_{\min} \to \sigma^* > 0$ as $k \to \infty$. This means that there exists $\widehat{\rho} > 0$ such that for all $k > \widehat{k}$ (by increasing $\widehat{k}$ if necessary),

$$\left[\sigma^k_{\min} - \frac{\zeta_2^4}{c_k^2} \alpha(x^k)(\|Y(x^k)\|_F^2 + \|\mu(x^k)\|^2) \right] \geq \widehat{\rho}.$$

Using the last three inequalities and $\nabla\Phi_{c_k}(x^k) = 0$, it is easy to obtain that for all $k > \widehat{k}$,

$$\begin{aligned} 0 &= \frac{1}{c_k^2} \left\| \begin{pmatrix} -\mathcal{J}g(x^k) \\ \mathcal{J}h(x^k) \end{pmatrix} \nabla\Phi_{c_k}(x^k) \right\|^2 \\ &\geq \frac{1}{2} \left[\sigma^k_{\min} - \frac{\zeta_2^4}{c_k^2} \alpha(x^k)(\|Y(x^k)\|_F^2 + \|\mu(x^k)\|^2) \right] \left(\|y_{c^k}(x^k)\|_F^2 + \|h(x^k)\|^2 \right) \\ &\geq \frac{\widehat{\rho}}{2} \left(\|y_{c^k}(x^k)\|_F^2 + \|h(x^k)\|^2 \right), \end{aligned}$$

which in turn implies that $y_{c_k}(x^k) = 0$ and $h(x^k) = 0$ for all $k > \widehat{k}$. Substituting $y_{c_k}(x^k) = 0$ and $h(x^k) = 0$ into $\nabla\Phi_{c_k}(x^k)$ we get $\nabla_x L(x^k, \mu(x^k), Y(x^k)) = 0$. Moreover, by Lemma 3.2, the complementarity condition $\mathbb{S}^n_+ \ni Y(x^k) \perp g(x^k) \in \mathbb{S}^n_+$ holds. Thus, the proof is completed. □

To get the result of this section, we also require the relation between f and Φ_c.

Lemma 3.4. *If x is a feasible point of (1.1), then $\Phi_c(x) \leq f(x)$ for all $c > 0$. In particular, if x is a stationary point of (1.1), then $\Phi_c(x) = f(x)$ for all $c > 0$.*

Proof. Since x is a feasible point, we have $h(x) = 0$. From the expression of $\Phi_c(x)$ we only need to prove that $\|\Pi_{\mathbb{S}^n_+}(Y(x) - cg(x))\|_F \leq \|Y(x)\|_F$. Recalling that the metric projector $\Pi_{\mathbb{S}^n_+}(\cdot)$ is globally Lipschitz continuous with modulus 1, we have

$$\|\Pi_{\mathbb{S}^n_+}(Y(x) - cg(x)) - \Pi_{\mathbb{S}^n_+}(-cg(x))\|_F \leq \|Y(x) - cg(x) + cg(x)\|_F = \|Y(x)\|_F.$$

Noting that $\Pi_{\mathbb{S}^n_+}(-cg(x)) = 0$, we have that $\|\Pi_{\mathbb{S}^n_+}(Y(x) - cg(x))\|_F \leq \|Y(x)\|_F$.

If x is a stationary point of (1.1), then there exists a multiplier $(\mu, Y) \in \mathbb{R}^m \times \mathbb{S}^n_+$ such that (x, μ, Y) is a KKT triple. This, along with Proposition 2.3 (a), implies that $h(x) = 0$. and $g(x), Y(x)$ satisfy the complementarity condition. Then for any $c > 0$, we obtain $Y(x) = \Pi_{\mathbb{S}^n_+}(Y(x) - cg(x))$, which implies that $\Phi_c(x) = f(x)$. □

Now we can give the main result, the function Φ_c is an exact penalty function for (1.1), presented in the two theorems below.

Theorem 3.5. *Assume that $\mathcal{G}^* \neq \emptyset$ and that there exists $\widetilde{c} > 0$ such that $\bigcup_{c\geq\widetilde{c}} \mathcal{G}(\Phi_c)$ is bounded. Then, there exists $\widehat{c} > 0$ such that for all $c > \widehat{c}$, $\mathcal{G}(\Phi_c) = \mathcal{G}^*$.*

Proof. First, we show that there exists $\widehat{c} > 0$ such that, for all $c > \widehat{c}$ we have $\mathcal{G}(\Phi_c) \subseteq \mathcal{G}^*$. Suppose that this assertion is false. Then, for any positive integer k there must exist a $c_k > k$ and a $x^k \in \mathcal{G}(\Phi_{c_k})$ such that x^k is not a global solution of problem (1.1). Note that there exists $\widetilde{c} > 0$ such that $\bigcup_{c\geq\widetilde{c}} \mathcal{G}(\Phi_c)$ is bounded. Then, we only consider the bounded sequence $\{x^k\}$ with $k > \widetilde{c}$. It is obvious that the sequence has a convergent subsequence, without loss of generality, we denote this subsequence by $\{x^k\}$ and assume that $x^k \to \widehat{x}$ as $k \to \infty$. We next show that there exists $\widehat{k} > 0$ such that $x^k \in \mathcal{G}^*$ for all $k > \widehat{k}$, which contradicts our hypothesis. Let $\widetilde{x}$ be an arbitrary point in $\mathcal{G}^*$. Then, $\widetilde{x}$ is a stationary point of (1.1), which by Lemma 3.4 implies that $\Phi_c(\widetilde{x}) = f(\widetilde{x})$ for all $c > 0$. Note that $\Phi_{c_k}(x^k) \leq \Phi_{c_k}(\widetilde{x}) = f(\widetilde{x})$ for all k. Taking the supremum limit to this inequality and using the expression of $\Phi_{c_k}(x^k)$, it is not hard to obtain $h(\widehat{x}) = 0$ and $\Pi_{\mathbb{S}^n_+}(-g(\widehat{x})) = 0$. This shows that $\widehat{x}$ is a feasible point of (1.1). Then, by Proposition 3.3, there exists $\widehat{k} > 0$ such that x^k is a stationary point of (1.1) for all $k > \widehat{k}$. Using Lemma 3.4 again, we obtain $f(x^k) = \Phi_{c_k}(x^k) \leq f(\widetilde{x})$ for all $k > \hat{k}$, which means that $x^k \in \mathcal{G}^*$ for all $k > \widehat{k}$. This shows that there exists $\widehat{c} > 0$ such that, for all $c > \widehat{c}$ we have $\mathcal{G}(\Phi_c) \subseteq \mathcal{G}^*$.

Secondly, we prove that for all $c > \widehat{c}$ we have $\mathcal{G}(\Phi_c) \supseteq \mathcal{G}^*$. Let c be an arbitrary constant satisfying $c > \widehat{c}$. Take an arbitrary element $\widehat{x}$ from $\mathcal{G}^*$. Then, $\widehat{x}$ must be a stationary point of problem (1.1), which by Lemma 3.4 implies $\Phi_c(\widehat{x}) = f(\widehat{x})$. Now take a point $\widetilde{x} \in \mathcal{G}(\Phi_c)$. Since we have proved that $\mathcal{G}(\Phi_c) \subseteq \mathcal{G}^*$, we have $\widetilde{x} \in \mathcal{G}^*$ which, together with Lemma 3.4, implies that $f(\widetilde{x}) = \Phi_c(\widetilde{x}) = \Phi_c(\widehat{x}) = f(\widehat{x})$. This shows that $\widehat{x} \in \mathcal{G}(\Phi_c)$. Thus, we prove that $\mathcal{G}(\Phi_c) \supseteq \mathcal{G}^*$. The proof is completed. □

Theorem 3.6. *Suppose that there exist sequences $\{c_k\} \subset \mathbb{R}_{++}$ with $c_k \to \infty$ and $\{x^k\} \subset \mathbb{X}$ with $x^k \to \widehat{x}$ such that $x^k \in \mathcal{S}(\Phi_{c_k})$ for all k. If $\widehat{x}$ is a feasible point of (1.1), then there exists $\widehat{k} > 0$ such that $x^k \in \mathcal{S}^*$ for all $k > \widehat{k}$; if $\widehat{x}$ is an infeasible point of (1.1), then*

$$\nabla\alpha(\widehat{x}) = \mathcal{J}h(\widehat{x})^T h(\widehat{x}) - \mathcal{J}g(\widehat{x})^T \Pi_{\mathbb{S}^n_+}(-g(\widehat{x})) = 0.$$

Proof. Assume that $\widehat{x}$ is a feasible point of (1.1). Then, from Proposition 3.3, there exists $\widehat{k} > 0$ such that x^k is a stationary point of (1.1) for all $k > \widehat{k}$. Now let k be an arbitrary but fixed $k > \widehat{k}$. Since $x^k \in \mathcal{S}(\Phi_{c_k})$, there exists a neighborhood $\mathcal{U}(x^k)$ of x^k such that

$$f(x^k) = \Phi_{c_k}(x^k) \leq \Phi_{c_k}(x) \quad \text{for all } x \in \mathcal{U}(x^k). \tag{3.7}$$

Now take an arbitrary $x \in \mathcal{U}(x^k)$ that is feasible to (1.1). From Lemma 3.4 it then follows that $\Phi_{c_k}(x) \leq f(x)$, which, together with (3.7), shows that $f(x^k) \leq f(x)$. Since $x \in \mathcal{U}(x^k)$ is arbitrary, we prove that $x^k \in \mathcal{S}^*$.

If $\widehat{x}$ is infeasible to (1.1), then by the expression of $\nabla\Phi_c(x)$ we have that, for all k,

$$\begin{aligned} &\frac{\nabla_x L(x^k, \mu(x^k), Y(x^k))}{c_k} + \left(\frac{\mathcal{J}\mu(x^k)}{c_k} + \mathcal{J}h(x^k)\right)^T h(x^k) \\ &+ \left(\frac{\mathcal{J}Y(x^k)}{c_k} - \mathcal{J}g(x^k)\right)^T y_{c_k}(x^k) = 0. \end{aligned}$$

Taking the limit $k \to \infty$ and taking into account the definition of $y_c(x)$, we obtain that $\nabla\alpha(\widehat{x}) = \mathcal{J}h(\widehat{x})^T h(\widehat{x}) - \mathcal{J}g(\widehat{x})^T \Pi_{\mathbb{S}^n_+}(-g(\widehat{x})) = 0$. The proof is then completed. □

Theorem 3.5 and Theorem 3.6 show that if the sequence $\{x^k\}$ converges to some point $\widehat{x} \in \mathbb{X}$, then there exists $\widehat{c} > 0$ such that $\mathcal{G}^* = \mathcal{G}(\Phi_c)$ for all $c > \widehat{c}$, especially, if $\widehat{x}$ is feasible, we have $\mathcal{S}^* = \mathcal{S}(\Phi_c)$ for all $c > \widehat{c}$, otherwise, $\widehat{x}$ is a stationary point of $\alpha(x)$. In short, the function $\Phi_c(x)$ is a continuously differentiable exact penalty function for nonlinear semidefinite programming. However, we must point out that the construction of $\Phi_c(x)$ is based on Proposition 2.3, which assumes x is constraint nondegenerate if x is feasible. So this section is established on this assumption.

4 Second-order Property of Φ_c

Since f, g, h are twice continuously differentiable functions and $\Pi_{\mathbb{S}^n_+}(\cdot)$ is globally Lipschitz continuous, the mapping $\nabla\Phi_c(\cdot)$ is Lipschitz continuous on $\mathbb{X}$. Thus, the Clarke's Jacobian of $\nabla\Phi_c(x)$ at any $x \in \mathbb{X}$ is well defined. In this section, we will show that the Clarke's Jacobian of $\nabla\Phi_c$ at the feasible points satisfying the constraint nondegeneracy and the strong second-order sufficient condition is positive definite. Then, the semismooth Newton method [20, 24, 25] can be applied to solve this problem with superlinear or quadratic convergence under mild conditions.

First, we introduce some notations that will be used in this section. For any given $A \in \mathbb{S}^n$, assume that A has the spectral decomposition

$$A = P\Lambda P^T, \tag{4.1}$$

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 orthogonal eigenvectors. Then $\Pi_{\mathbb{S}^n_+}(A) = P\Pi_{\mathbb{S}^n_+}(\Lambda)P^T$. Define three index sets associated to the 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

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

with $P_\alpha \in \mathbb{R}^{n\times|\alpha|}, P_\beta \in \mathbb{R}^{n\times|\beta|}$ and $P_\gamma \in \mathbb{R}^{n\times|\gamma|}$. Let $\overline{x} \in \mathbb{X}$ be an optimal solution to (1.1). Denote $\mathcal{M}(\overline{x})$ by the set of points $(\mu, Y) \in \mathbb{R}^m \times \mathbb{S}^n$ such that $(\overline{x}, \mu, Y)$ is a KKT triple. Let $(\overline{\mu}, \overline{Y}) \in \mathcal{M}(\overline{x})$ and $A := g(\overline{x}) - \overline{Y}$. Denote A_+ by $\Pi_{\mathbb{S}^n_+}(A)$. The critical cone of $\mathbb{S}^n_+$ at A is defined as

$$\mathcal{C}(A; \mathbb{S}^n_+) := \mathcal{T}_{\mathbb{S}^n_+}(A_+) \cap (A_+ - A)^\perp,$$

where $(A_+ - A)^\perp := \{B \in \mathbb{S}^n : \langle B, A_+ - A\rangle = 0\}$. Denote aff $\big(\mathcal{C}(A; \mathbb{S}^n_+)\big)$ by the affine hull of $\mathcal{C}(A; \mathbb{S}^n_+)$. Define

$$\text{app}(\overline{\mu}, \overline{Y}) := \left\{\xi : \ \mathcal{J}h(\overline{x})\xi = 0, \ \ \mathcal{J}g(\overline{x})\xi \in \text{aff}\left(\mathcal{C}(A; \mathbb{S}^n_+)\right)\right\}.$$

Noting that $\mathbb{S}^n_+ \ni g(\overline{x}) \perp \overline{Y} \in \mathbb{S}^n_+$, by [29] it is not hard to deduce that

$$\text{app}(\overline{\mu}, \overline{Y}) = \left\{\xi : \mathcal{J}h(\overline{x})\xi = 0, P_\beta^T(\mathcal{J}g(\overline{x})\xi)P_\gamma = 0, P_\gamma^T(\mathcal{J}g(\overline{x})\xi)P_\gamma = 0\right\}. \tag{4.2}$$

Next, we recall the strong second-order sufficient condition introduced by Sun [29].

Definition 4.1. Let $\overline{x}$ be a stationary point of (1.1). We say that the strong second-order sufficient condition holds at $\overline{x}$ if for any nonzero $\xi \in \bigcap_{(\mu,Y)\in\mathcal{M}(\overline{x})} \text{app}(\mu, Y)$,

$$\sup_{(\mu,Y)\in\mathcal{M}(\overline{x})} \left\{ \langle \xi, \nabla^2_{xx} L(\overline{x}, \mu, Y)\xi\rangle - \Upsilon_{g(\overline{x})}(-Y, \mathcal{J}g(\overline{x})\xi) \right\} > 0, \tag{4.3}$$

where, for any given $B \in \mathbb{S}^n$, $\Upsilon_B : \mathbb{S}^n \times \mathbb{S}^n \to \mathbb{R}$ is the linear-quadratic function

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

with $B^\dagger$ denoting the Moore-Penrose pseudoinverse of B.

Let x^* be an arbitrary stationary point of (1.1). For any given $c > 0$, the following proposition characterizes the expression of $\partial\nabla\Phi_c(x^*)$.

Proposition 4.2. *Let (x^*, μ^*, Y^*) be a KKT triple of (1.1). Then, for any given $c > 0$ and $V \in \partial_B\nabla\Phi_c(x^*)$ (respectively, $V \in \partial\nabla\Phi_c(x^*)$), there exists $R \in \partial_B\Pi_{\mathbb{S}^n_+}(Y^* - cg(x^*))$ (respectively, $R \in \partial\Pi_{\mathbb{S}^n_+}(Y^* - cg(x^*))$ such that*

$$\begin{aligned} V \;=\; & \nabla^2_{xx}L(x^*, \mu^*, Y^*) + \mathcal{J}h(x^*)^T(\mathcal{J}\mu(x^*) + c\mathcal{J}h(x^*)) + \mathcal{J}\mu(x^*)^T\mathcal{J}h(x^*) \\ & -c^{-1}\mathcal{J}Y(x^*)^T\mathcal{J}Y(x^*) + c^{-1}(\mathcal{J}Y(x^*) - c\mathcal{J}g(x^*))^T R(\mathcal{J}Y(x^*) - c\mathcal{J}g(x^*)). \end{aligned}$$

Proof. It suffices to argue that the result holds for $V \in \partial_B\nabla\Phi_c(x^*)$. By the expression of Φ_c, it is easy to see that $\nabla\Phi_c(x)$ is locally Lipschitz continuous on $\mathbb{X}$, and hence is almost everywhere Frechet-differentiable in $\mathbb{X}$. After an elementary computation, we can obtain that for any $V \in \partial_B\nabla\Phi_c(x)$, there exists $R \in \partial_B\Pi_{\mathbb{S}^n_+}(Y(x) - cg(x))$ such that

$$\begin{aligned} V \;=\; & \nabla^2_{xx}L(x, \mu(x), Y(x)) + \mathcal{J}h(x)^T(\mathcal{J}\mu(x) + c\mathcal{J}h(x)) + \mathcal{J}\mu(x)^T\mathcal{J}h(x) \\ & -c^{-1}\mathcal{J}Y(x)^T\mathcal{J}Y(x) + c^{-1}(\mathcal{J}Y(x) - c\mathcal{J}g(x))^T R(\mathcal{J}Y(x) - c\mathcal{J}g(x)) + \varphi_c(x), \end{aligned}$$

where

$$\varphi_c(x) := [\mathcal{J}^2Y(x) - c\mathcal{J}^2g(x)]^T y_c(x) + [\mathcal{J}^2\mu(x) + c\mathcal{J}^2h(x)]^T h(x).$$

For the stationary point x^*, we have $h(x^*) = 0$, and $y_c(x^*) = 0$ by Lemma 3.2. Substituting the two equalities into the last equality, we get the desired result. □

Let (x^*, μ^*, Y^*) be a KKT triple of (1.1). By Proposition 4.2, the expression of the elements in $\partial\nabla\Phi_c(x^*)$ can be obtained from that of the elements in $\partial\Pi_{\mathbb{S}^n_+}(Y^* - cg(x^*))$. We next present the explicit expression of the elements in $\partial\Pi_{\mathbb{S}^n_+}(Y^* - cg(x^*))$. For this purpose, we assume that $cg(x^*) - Y^*$ has the spectral decomposition as in (4.1). Then, $g(x^*)$ and $-Y^*$ take the form of

$$g(x^*) = P\begin{bmatrix} D_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}P^T, \qquad -Y^* = P\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & D_\gamma \end{bmatrix}P^T$$

where D_α is the diagonal matrix of eigenvalues $d_1 \geq \cdots \geq d_\alpha > 0$ of $g(x^*)$ and D_γ is the diagonal matrix of eigenvalues $0 > d_{\alpha+\beta+1} \geq \cdots \geq d_n$ of $-Y^*$. Obviously, $\Lambda_\alpha = cD_\alpha$ and $\Lambda_\gamma = D_\gamma$. Then $\partial\Pi_{\mathbb{S}^n_+}(Y^* - cg(x^*))$ can be characterized as follows.

Proposition 4.3. *Let (x^*, μ^*, Y^*) be a KKT triple of (1.1) and $A = cg(x^*) - Y^*$ for any given $c > 0$. Suppose that A has the spectral decomposition as in (4.1). Then, for any $R \in \partial_B \Pi_{\mathbb{S}^n_+}(-A)$ (respectively, $R \in \partial \Pi_{\mathbb{S}^n_+}(-A)$), there exists a $R_{|\beta|} \in \partial_B \Pi_{\mathbb{S}^{|\beta|}_+}(0)$ (respectively, $R_{|\beta|} \in \partial \Pi_{\mathbb{S}^{|\beta|}_+}(0)$) such that*

$$R(H) = P \begin{pmatrix} 0 & 0 & U_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} \\ 0 & R_{|\beta|}(\widetilde{H}_{\beta\beta}) & \widetilde{H}_{\beta\gamma} \\ \widetilde{H}^T_{\alpha\gamma} \circ U^T_{\alpha\gamma} & \widetilde{H}^T_{\beta\gamma} & \widetilde{H}_{\gamma\gamma} \end{pmatrix} P^T, \quad \forall H \in \mathbb{S}^n.$$

where $\widetilde{H} := P^T HP$, "$\circ$" denotes the Hadamard product and $U_{\alpha\gamma}$ is a $\alpha \times \gamma$ real matrix with entries

$$u_{ij} = \frac{\max\{-cd_i, 0\} + \max\{-d_j, 0\}}{|cd_i| + |d_j|}, \quad i = 1, \dots, \alpha, j = \alpha + \beta + 1, \dots, n,$$

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

Proof. It is straightforward from [29]. □

Definition 4.1, Proposition 4.2 and 4.3 allow us to give the following important result.

Theorem 4.4. *Let $x^* \in \mathbb{X}$ be a stationary point of (1.1). Suppose that the strong second-order sufficient condition (4.3) holds at x^* and x^* is constraint nondegenerate. Then, for sufficiently large $c > 0$, any element in $\partial \nabla \Phi_c(x^*)$ is positive definite.*

Proof. Suppose on the contrary that the conclusion does not hold. Then there exist sequences $\{c_k\} \subset \mathbb{R}_{++}$ with $c_k \to +\infty$ and $\{V_k\}$ with $V_k \in \partial \nabla \Phi_{c_k}(x^*)$ such that V_k is not positive definite for all k. So, there exists $\{\xi^k\} \subset \mathbb{X}$ with $\|\xi^k\| = 1$ such that $\langle V_k \xi^k, \xi^k \rangle \leq 0$ for all k. Notice that $\{V_k\}$ and $\{\xi^k\}$ are bounded. Then both of them have convergent subsequences, without loss of generality, we denote the two subsequences by $\{V_k\}$ and $\{\xi^k\}$, respectively. Moreover, suppose that $V_k \to V$ and $\xi^k \to \xi$ with $\|\xi\| = 1$. Thus, $\langle V_k \xi^k, \xi^k \rangle \to \langle V\xi, \xi \rangle \leq 0$. Next, we prove that $\langle V\xi, \xi \rangle > 0$ which will be a contradiction to $\langle V\xi, \xi \rangle \leq 0$. To this end, we first claim that $\xi \in \mathrm{app}(\mu^*, Y^*)$ characterized below.

Since $x^* \in \mathbb{X}$ is a stationary point of (1.1), there exists $(\mu^*, Y^*) \in \mathbb{R}^m \times \mathbb{S}^n_+$ such that (x^*, μ^*, Y^*) is a KKT triple of (1.1). Let $A := g(x^*) - Y^*$. Without loss of generality, we assume that A has the spectral decomposition as in (4.1),

$$g(x^*) = P \begin{bmatrix} \Lambda_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} P^T, \quad \text{and} \quad -Y^* = P \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Lambda_\gamma \end{bmatrix} P^T.$$

By (4.2), we have

$$\mathrm{app}(\mu^*, Y^*) = \left\{ \xi : \mathcal{J}h(x^*)\xi = 0, P^T_\beta(\mathcal{J}g(x^*)\xi)P_\gamma = 0, P^T_\gamma(\mathcal{J}g(x^*)\xi)P_\gamma = 0 \right\}.$$

For each k, from Proposition 4.2 by replacing V, c with V_k, c_k respectively, it follows that

$$\begin{aligned} \langle V_k \xi^k, \xi^k \rangle = {} & \langle \nabla^2_{xx} L(x^*, \mu^*, Y^*)\xi^k, \xi^k \rangle + 2\langle \mathcal{J}\mu(x^*)\xi^k, \mathcal{J}h(x^*)\xi^k \rangle + c_k \|\mathcal{J}h(x^*)\xi^k\|^2 \\ & + c_k^{-1} \left(\langle R_k(\mathcal{J}Y(x^*)\xi^k), \mathcal{J}Y(x^*)\xi^k \rangle - \|\mathcal{J}Y(x^*)\xi^k\|^2 \right) \\ & - 2\langle R_k(\mathcal{J}Y(x^*)\xi^k), \mathcal{J}g(x^*)\xi^k \rangle + c_k \langle R_k(\mathcal{J}g(x^*)\xi^k), \mathcal{J}g(x^*)\xi^k \rangle, \end{aligned} \tag{4.4}$$

where $R_k \in \partial\Pi_{\mathbb{S}^n_+}(Y^* - c_k g(x^*))$. Dividing the equation by c_k and taking the limit $k \to \infty$, we conclude that $\mathcal{J}h(x^*)\xi = 0$ and $\langle R_k(\mathcal{J}g(x^*)\xi^k), \mathcal{J}g(x^*)\xi^k\rangle \to 0$. Furthermore, noting that $c_k > 0$ and $\mathbb{S}^n_+ \ni g(x^*) \perp Y^* \in \mathbb{S}^n_+$, we get that $c_k g(x^*) - Y^*$ has the spectral decomposition

$$c_k g(x^*) - Y^* = P \begin{bmatrix} c_k\Lambda_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Lambda_\gamma \end{bmatrix} P^T.$$

Write $H^k = \mathcal{J}g(x^*)\xi^k$, by Proposition 4.3, there exists $R^k_{|\beta|} \in \partial\Pi_{\mathbb{S}^{|\beta|}_+}(0)$ such that

$$R_k(H^k) = P \begin{pmatrix} 0 & 0 & U^k_{\alpha\gamma} \circ \widetilde{H}^k_{\alpha\gamma} \\ 0 & R^k_{|\beta|}(\widetilde{H}^k_{\beta\beta}) & \widetilde{H}^k_{\beta\gamma} \\ (\widetilde{H}^k_{\alpha\gamma})^T \circ (U^k_{\alpha\gamma})^T & (\widetilde{H}^k_{\beta\gamma})^T & \widetilde{H}^k_{\gamma\gamma} \end{pmatrix} P^T,$$

where $\widetilde{H}^k := P^T H^k P$ and $U^k_{\alpha\gamma}$ is a $\alpha \times \gamma$ matrix with entries $u^k_{ij} = \frac{\max\{-c_k\lambda_i, 0\} + \max\{-\lambda_j, 0\}}{|c_k\lambda_i| + |\lambda_j|}$, $i = 1, \dots, \alpha, j = \alpha + \beta + 1, \dots, n$. Thus,

$$\begin{aligned} \langle R_k(\mathcal{J}g(x^*)\xi^k), \mathcal{J}g(x^*)\xi^k\rangle &= 2\mathrm{tr}\left(U^k_{\alpha\gamma} \circ \widetilde{H}^k_{\alpha\gamma}(\widetilde{H}^k_{\alpha\gamma})^T\right) + \mathrm{tr}\left(R^k_{|\beta|}(\widetilde{H}^k_{\beta\beta})\widetilde{H}^k_{\beta\beta}\right) \\ &\quad + 2\mathrm{tr}\left(\widetilde{H}^k_{\beta\gamma}(\widetilde{H}^k_{\beta\gamma})^T\right) + \mathrm{tr}\left(\widetilde{H}^k_{\gamma\gamma}\widetilde{H}^k_{\gamma\gamma}\right). \end{aligned} \tag{4.5}$$

Note that every term on the right-hand side is nonnegative in (4.5). Recalling that $\langle R_k(\mathcal{J}g(x^*)\xi^k), \mathcal{J}g(x^*)\xi^k\rangle \to 0$, it follows that $P^T_\beta\left(\mathcal{J}g(x^*)\xi\right)P_\gamma = 0$, $P^T_\gamma\left(\mathcal{J}g(x^*)\xi\right)P_\gamma = 0$. So, we have $\xi \in \mathrm{app}(\mu^*, Y^*)$.

Now we claim that $\langle V\xi, \xi\rangle > 0$. In a similar way to (4.4) and recalling that $\mathcal{J}h(x^*)\xi = 0$, we can write

$$\begin{aligned} \langle V_k\xi, \xi\rangle &= \langle \xi, \nabla^2_{xx}L(x^*, \mu^*, Y^*)\xi\rangle - \Upsilon_{g(x^*)}(-Y^*, \mathcal{J}g(x^*)\xi) - c_k^{-1}\|\mathcal{J}Y(x^*)\xi\|^2 \\ &\quad + c_k^{-1}\langle R_k(\mathcal{J}Y(x^*)\xi), \mathcal{J}Y(x^*)\xi\rangle - 2\langle R_k(\mathcal{J}Y(x^*)\xi), \mathcal{J}g(x^*)\xi\rangle \\ &\quad + c_k\langle R_k(\mathcal{J}g(x^*)\xi), \mathcal{J}g(x^*)\xi\rangle + \Upsilon_{g(x^*)}(-Y^*, \mathcal{J}g(x^*)\xi). \end{aligned} \tag{4.6}$$

For the first two terms in (4.6), note that

$$\mathcal{M}(x^*) = \{(\mu^*, Y^*)\}$$

since the constraint nondegeneracy condition (2.3) is assumed to hold at x^* then, we get $\xi \in \bigcap_{(\mu,Y)\in\mathcal{M}(x^*)} \mathrm{app}(\mu, Y) \setminus \{0\}$, in addition,

$$\langle \xi, \nabla^2_{xx}L(x^*, \mu^*, Y^*)\xi\rangle - \Upsilon_{g(x^*)}(-Y^*, \mathcal{J}g(x^*)\xi) > 0$$

from the strong second-order sufficient condition (4.3).

For the last three terms on the right-hand side in (4.6), let $H := \mathcal{J}g(x^*)\xi$, $G := \mathcal{J}Y(x^*)\xi$, $\widetilde{H} := P^T HP$ with entries $\widetilde{h}_{ij}, i, j = 1, \cdots, n$ and $\widetilde{G} := P^T GP$ with entries $\widetilde{g}_{ij}, i, j = 1, \dots, n$, we have that

$$\begin{aligned} -2\langle R_k(\mathcal{J}Y(x^*)\xi), \mathcal{J}g(x^*)\xi\rangle &= -2\langle \mathcal{J}Y(x^*)\xi, R_k(\mathcal{J}g(x^*)\xi)\rangle \\ &= -2\mathrm{tr}\left(\widetilde{G}_{\alpha\gamma}(\widetilde{H}^T_{\alpha\gamma} \circ (U^k_{\alpha\gamma})^T)\right) - 2\mathrm{tr}\left(\widetilde{G}_{\beta\beta}R^k_{|\beta|}(\widetilde{H}_{\beta\beta})\right) \\ &\quad -2\mathrm{tr}\left(\widetilde{G}^T_{\alpha\gamma}(U^k_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma})\right) \\ &= 4\Sigma^n_{j=\alpha+\beta+1}\Sigma^\alpha_{i=1}\frac{\lambda_j}{c_k\lambda_i - \lambda_j}\widetilde{h}_{ij}\widetilde{g}_{ij} - 2\langle \widetilde{G}_{\beta\beta}, R^k_{|\beta|}(\widetilde{H}_{\beta\beta})\rangle, \end{aligned}$$

$$
\begin{aligned}
c_k\langle R_k(\mathcal{J}g(x^*)\xi),\mathcal{J}g(x^*)\xi\rangle &= c_k\mathrm{tr}\left(\widetilde{H}_{\alpha\gamma}(\widetilde{H}_{\alpha\gamma}^T\circ(U_{\alpha\gamma}^k)^T)\right)+c_k\mathrm{tr}\left(\widetilde{H}_{\beta\beta}R_{|\beta|}^k(\widetilde{H}_{\beta\beta})\right)\\
&\quad +c_k\mathrm{tr}\left(\widetilde{H}_{\alpha\gamma}^T(U_{\alpha\gamma}^k\circ\widetilde{H}_{\alpha\gamma})\right)\\
&= 2\Sigma_{j=\alpha+\beta+1}^n\Sigma_{i=1}^\alpha\frac{-c_k\lambda_j}{c_k\lambda_i-\lambda_j}\widetilde{h}_{ij}^2+c_k\langle R_{|\beta|}^k(\widetilde{H}_{\beta\beta}),\widetilde{H}_{\beta\beta}\rangle,
\end{aligned}
$$

and

$$
\begin{aligned}
\Upsilon_{g(x^*)}(-Y^*,\mathcal{J}g(x^*)\xi) &= 2\langle -Y^*,\mathcal{J}g(x^*)\xi(g(x^*))^\dagger\mathcal{J}g(x^*)\xi\rangle\\
&= 2\mathrm{tr}\left(\Lambda_\gamma\widetilde{H}_{\alpha\gamma}^T(\Lambda_\alpha)^{-1}\widetilde{H}_{\alpha\gamma}\right)=2\Sigma_{j=\alpha+\beta+1}^n\Sigma_{i=1}^\alpha\frac{\lambda_j}{\lambda_i}\widetilde{h}_{ij}^2.
\end{aligned}
$$

Thus, the last three terms on the right-hand side in (4.6) can be simplified to be

$$
\begin{aligned}
&4\Sigma_{j=\alpha+\beta+1}^n\Sigma_{i=1}^\alpha\frac{\lambda_j}{c_k\lambda_i-\lambda_j}\widetilde{h}_{ij}\widetilde{g}_{ij}-2\Sigma_{j=\alpha+\beta+1}^n\Sigma_{i=1}^\alpha\frac{\lambda_j^2}{\lambda_i(c_k\lambda_i-\lambda_j)}\widetilde{h}_{ij}^2\\
&-c_k^{-1}\langle R_{|\beta|}^k(\widetilde{G}_{\beta\beta}^k),\widetilde{G}_{\beta\beta}^k\rangle+c_k\left[\langle R_{|\beta|}^k(\widetilde{H}_{\beta\beta}^k-c_k^{-1}\widetilde{G}_{\beta\beta}^k),(\widetilde{H}_{\beta\beta}^k-c_k^{-1}\widetilde{G}_{\beta\beta}^k)\rangle\right].
\end{aligned}
$$

Take the limit $k\to 0$ on above, it is obvious that the first three terms converge to zero and the last term is nonnegative since $R_{|\beta|}^k\in\mathbb{S}_+^{|\beta|}$.

For the rest terms in (4.6), it is easy to get

$$
\lim_{k\to+\infty}(c_k^{-1}\langle R_k(\mathcal{J}Y(x^*)\xi),\mathcal{J}Y(x^*)\xi\rangle-c_k^{-1}\|\mathcal{J}Y(x^*)\xi\|^2)=0.
$$

Now, we can summarize the discussion above to have

$$
\langle V_k\xi,\xi\rangle\to\langle V\xi,\xi\rangle>0\quad(k\to+\infty)
$$

which gives the desired contradiction. The proof is completed. □

The positive definite property of $\partial\nabla\Phi_c(x^*)$ entails the semismooth Newton algorithms based on this exact penalty function for solving nonlinear semidefinite programming problems. In addition, it deserves to be mentioned that the strong-second sufficient condition can be weakened as the dual constraint nondegeneracy when f and g are linear matrix functions (see. [8]), which will be discussed below.

4.1 Linear Semidefinite Programming

When f and g are linear matrix functions, the problem (1.1) becomes linear semidefinite programming

$$
\begin{aligned}
\min\ &\langle C,X\rangle\\
\text{s.t.}\ &\mathcal{A}X=b,\\
&X\in\mathbb{S}_+^n,
\end{aligned}
\tag{4.7}
$$

where $C\in\mathbb{S}^n,\mathcal{A}\colon\mathbb{S}^n\to\mathbb{R}^m$ is a linear operator and $b\in\mathbb{R}^m$. For this problem, the exact penalty function takes the form of

$$
\Phi_c(X)=\langle C,X\rangle+\langle\mathcal{A}X-b,\mu(X)\rangle+\frac{c}{2}\|\mathcal{A}X-b\|^2+\frac{1}{2c}(\|\Pi_{\mathbb{S}_+^n}(Y(X)-cX)\|_F^2-\|Y(X)\|_F^2),
$$

where $\mu : \mathbb{S}^n \to \mathbb{R}^m$ and $Y : \mathbb{S}^n \to \mathbb{S}^n$ are multiplier functions defined in (2.2). Obviously, Theorem 4.4 holds for the problem (4.7). However, due to the special structure of the problem (4.7) and its dual, the strong second-order sufficient condition can be weakened as the dual constraint nondegeneracy defined below. This subsection is devoted to proving this result.

For this purpose, we first recall from [1, 7, 8] the concept of primal and dual constraint nondegeneracies and from [8] the link between the strong second-order sufficient condition and the dual constraint nondegeneracy.

Definition 4.5. We say that the primal constraint nondegeneracy holds at a feasible solution $\overline{X}$ to the linear semidefinite programming problem (4.7) if

$$\begin{pmatrix} \mathcal{A} \\ \mathcal{I} \end{pmatrix} \mathbb{S}^n + \begin{pmatrix} \{0\} \\ \mathrm{lin}\Big(\mathcal{T}_{\mathbb{S}^n_+}(\overline{X})\Big) \end{pmatrix} = \begin{pmatrix} \mathbb{R}^m \\ \mathbb{S}^n \end{pmatrix} \tag{4.8}$$

or, equivalently,

$$\mathcal{A}\ \mathrm{lin}\Big(\mathcal{T}_{\mathbb{S}^n_+}(\overline{X})\Big) = \mathbb{R}^m.$$

Similarly, we say that the dual constraint nondegeneracy holds at a feasible solution $(\overline{\mu}, \overline{Y})$ to the dual problem of (4.7) if

$$\begin{pmatrix} \mathcal{A}^* & \mathcal{I} \\ 0 & \mathcal{I} \end{pmatrix} \begin{pmatrix} \mathbb{R}^m \\ \mathbb{S}^n \end{pmatrix} + \begin{pmatrix} \{0\} \\ \mathrm{lin}\Big(\mathcal{T}_{\mathbb{S}^n_+}(\overline{Y})\Big) \end{pmatrix} = \begin{pmatrix} \mathbb{S}^n \\ \mathbb{S}^n \end{pmatrix} \tag{4.9}$$

or, equivalently,

$$\mathcal{A}^*\mathbb{R}^m + \mathrm{lin}\Big(\mathcal{T}_{\mathbb{S}^n_+}(\overline{Y})\Big) = \mathbb{S}^n.$$

Proposition 4.6 ([8])**.** *Let $\overline{X} \in \mathbb{S}^n_+$ be an optimal solution to problem (4.7). Under the assumption $\mathcal{M}(\overline{X}) = \{(\overline{\mu}, \overline{Y})\}$, the following are equivalent:*

(1) *The strong second-order sufficient condition holds at $\overline{X}$.*

(2) *The dual constraint nondegenerate condition holds at $(\overline{\mu}, \overline{Y})$.*

Comparing problem (1.1) with (4.7) and Definition 2.1 with 4.5, we can easily obtain that the constraint nodegenerate condition is equivalent to the primal constraint nondegeneracy for linear semidefinite programming. Then, combining Proposition 4.6, we have the following result.

Theorem 4.7. *Let $(X^*, \mu^*, Y^*) \in \mathbb{S}^n \times \mathbb{R}^m \times \mathbb{S}^n$ be a KKT triple of the problem (4.7). Assume that the primal constraint nondegenerate condition (4.8) holds at X^* and the dual constraint nondegenerate (4.9) holds at (μ^*, Y^*), respectively. Then, any element in $\partial\nabla\Phi_c(X^*)$ is symmetric and positive definite for any $c > 0$.*

Proof. For any given $c > 0$ and $V \in \partial\nabla\Phi_c(X^*)$, from Proposition 4.2, we get

$$\begin{aligned} V &= (\mathcal{A}^*\mathcal{J}\mu(X) + +(\mathcal{J}\mu(X))^*\mathcal{A}) + c\mathcal{A}^*\mathcal{A} \\ &\quad -\frac{1}{c}(\mathcal{J}Y(X))^*\mathcal{J}Y(X) + \frac{1}{c}(\mathcal{J}Y(X) - cI)^*R(\mathcal{J}Y(X) - cI), \end{aligned}$$

where $R \in \partial\Pi_{\mathbb{S}^n_+}(Y^* - cX^*)$. Then, it is easy to obtain that V is symmetric since R is self-adjoint. On the other hand, the primal constraint nondegeneracy implies the constrain nondegeneracy for the problem (4.7). Since $\mathcal{M}(X^*) = \{(\mu^*, Y^*)\}$, we have from Proposition 4.6, that the strong second-order sufficient condition holds at X^*. Thus, from Theorem 4.4, $\partial\nabla\Phi_c(X^*)$ is positive definite. The proof is completed. $\square$

5 Conclusions

In this paper, we have established a differentiable exact penalty function for the nonlinear semidefinite programming. Also, under the constraint nondegeneracy and the strong second-order sufficient condition, we showed that the Clarke generalized Jacobian of the gradient of this exact penalty function at the stationary point is positive definite. This property implies the rate of superlinear (or quadratic) convergence when the semismooth Newton method is applied for the minimization of this penalty function. The applications of this penalty function in the development of algorithms will be our future topic.

Acknowledgements

In addition, I would like to thanks one anonymous referee for some helpful suggestions to the revision of this paper.

References

[1] F. Alizadeh, J.-P. A. Haeberly and M. L. Overton, Primal-dual interior point methods for semidefinite programming: convergence rates, stability and numerical results, *SIAM J. Optim.* 8 (1998) 746–768.

[2] T.A. André and P.J.S. Silva, Exact penalties for variational inequalities with applications to nonlinear complementarity problems, *Comput. Optim. Appl.* 47 (2010) 401–429.

[3] R. Andreani, E.H. Fukuda and P.J.S. Silva, A Gauss-Newton approach for solving constrained optimization problems using differentiable exact penalties, *J. Optim. Theory Appl.* (2012) DOI: 10.1007/s10957-012-0114-6.

[4] A. Ben-Tal, F. Jarre, M. Kocvara, A. Nemirovski and J. Zowe, Optimal design of trusses under a nonconvex global buckling constraint, *Optim. Eng.* 1 (2000) 189–213.

[5] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*, MPSSIAM Series on Optimization, SIAM Philadelphia, 2001.

[6] S.-J. Bi, S. H. Pan and J. S. Chen, Nonsingularity conditions for FB system of nonlinear SDPs, *SIAM J. Optim.* 21 (2011) 1392–1417.

[7] J.-F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.

[8] Z.-X. Chan and D.F. Sun, Constraint nondegeneracy, strong regularity, and nonsingularity in semidefinite programming, *SIAM J. Optim.* 19 (2008) 370–396.

[9] X. Chen and P. Tseng, Non-interior continuation methods for solving semidefinite complementarity problems, *Math. Program.* 95 (2003) 431–474.

[10] R. Correa and H.Ramírez C., A global algorithm for nonlinear semidefinite programming, *SIAM J. Optim.* 15 (2004) 303–318.

[11] E.H. Fukuda, P.J.S. Silva and M. Fukushima, Differentiable exact penalty functions for nonlinear second-order cone programs, *SIAM J. Optim.* 22 (2012) 1607–1633.

[12] T. Glad and E. Polak, A multiplier method with automatic limitation of penalty growth, *Math. Program.* 17 (1979) 140–155.

[13] L. Han, S. Bi and S.H. Pan, Nonsingularity of FB system and constraint nondegeneracy in semidefinite programming, *Numer. Algorithms* DOI: 10.1007/s11075–012–9567–9, 2012.

[14] R.A. Horn and C.R. Johnson, *Topics of Matrix Analysis*, Cambridge University Press, Cambridge, 1991.

[15] C. Kanzow and C. Nagel, Semidefinite programs: new search directions, smoothing-type methods, and numerical results, *SIAM J. Optim.* 13 (2003) 1–23, Corrigendum in *SIAM J. Optim.* 14 (2004) 936–937.

[16] M. Kocvara and M. Stingl, On the solution of large-scale SDP problems by the modified barrier method using iterative solvers, *Math. Program.* 109 (2007) 413–444.

[17] M. Kojima, M. Shida and S. Shindoh, A predictor-corrector interior-point algorithms for the semidefinite linear complementarity problem using the Alizadeh-Haeberly-Overton search direction, *SIAM J. Optim.* 9 (1999) 444–465.

[18] S. Lucidi, New results on a continuously differentiable exact penalty function, *SIAM J. Optim.* 2 (1992) 558–574.

[19] J. Malick, J. Povh, F. Rendl and A. Wiegele, Regularization methods for semidefinite programming, *SIAM J. Optim.* 20 (2009) 336–356.

[20] J.S. Pang and L. Qi, Nonsmooth equations: Motivations and algorithms, *SIAM J. Optim.* 3 (1993) 443–465.

[21] G.Di Pillo and L. Grippo, A continuously differentiable exact penalty function for nonlinear programming problems with inequality constraints, *SIAM J. Control Optim.* 23 (1985) 72–84.

[22] G.Di Pillo and L. Grippo, Exact penalty functions in constrained optimization, *SIAM J. Control Optim.* 27 (1989) 1333–1360.

[23] J. Povh, F. Rendl, and A. Wiegele, A boundary point method to solve semidefinite programs, *Computing* 78 (2006) 277–286.

[24] L. Qi, Convergence analysis of some algorithms for solving nonsmooth equations, *Math. Oper. Res.* 18 (1993) 227–244.

[25] L. Qi and J. Sun, A nonsmooth version of Newton's method, *Math. Program.* 58 (1993) 353–367.

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

[27] M. Stingl, On the solution of nonlinear semidefinite programs by augmented Lagrangian methods, *Available at* `http://www2.am.uni-erlangen.de/~kocvara/pennon/michael_stingl_thesis.pdf`, 2005.

[28] A. Shapiro and J. Sun, Some properties of the augmented Lagrangian in cone constrained optimization, *Math. Oper. Res.* 29 (2001) 479–491.

[29] D.F. Sun, The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications, *Math. Oper. Res.* 31 (2006) 761–776.

[30] 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 J. Optim.* 14 (2004) 783–806.

[31] J. Sun, D.F. Sun and L. W. Zhang, The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming, *Math. Program.* 114 (2008) 349–391.

[32] M.J. Todd, Semidefinite optimization, *Acta Numer.* 10 (2001) 515–560.

[33] K.C. Toh, Solving large scale semidefinite programs via an iterative solver on the augmented systems, *SIAM J. Optim.* 14 (2003) 670–698.

[34] K.C. Toh, An inexact primal-dual path-following algorithm for convex quadratic SDP, *Math. Program.* 112 (2007) 221–254.

[35] P. Tseng, Merit functions for semidefinite complementarity problems, *Math. Program.* 83 (1998) 159–185.

[36] L. Vanderberghe, H. Wolkowocz and R. Safonov, *Handbook of Semidefinite Programming-Theory, Algorithm and Applications*, Kluwer Academic Publishers, Dordrecht, 2000.

[37] L. Vandenberghe and S. Boyd, Semidefinite programming, *SIAM Rev.* 38 (1996) 49–95.

[38] X.Y. Zhao, D.F. Sun and K.C. Toh, A Newton-CG augmented lagrangian method for semidefinite programming *SIAM J. Optim.* 20 (2010) 1737–1765.

Manuscript received 28 August 2012
revised 28 October 2012, 25 December 2012
accepted for publication 8 January 2013

LE HAN
Department of Mathematics
South China University of Technology
Guangzhou 510640, China
E-mail address: hanle@scut.edu.cn