Research Article

A Note on Convergence Analysis of an SQP-Type Method for Nonlinear Semidefinite Programming

Yun Wang,[1] Shaowu Zhang,[2] and Liwei Zhang[1]

[1] *Department of Applied Mathematics, Dalian University of Technology, Dalian 116024, China*
[2] *Department of Computer Science, Dalian University of Technology, Dalian 116024, China*

Correspondence should be addressed to Yun Wang, wangyun_3412@163.com

We reinvestigate the convergence properties of the SQP-type method for solving nonlinear semidefinite programming problems studied by Correa and Ramirez (2004). We prove, under the strong second-order sufficient condition with the sigma term, that the local SQP-type method is quadratically convergent and the line search SQP-type method is globally convergent.

1. Introduction

We consider the following nonlinear semidefinite programming:

$$\text{(SDP)} \quad \min f(x) \quad \text{s.t. } h(x) = 0, \ g(x) \in \mathcal{S}_+^p, \tag{1.1}$$

where $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$, and $g : \mathbb{R}^n \to \mathcal{S}^p$ are twice continuously differentiable functions, $\mathcal{S}^p$ is the linear space of all $p \times p$ real symmetric matrices, and $\mathcal{S}_+^p$ is the cone of all $p \times p$ symmetric positive semidefinite matrices.

Fares et al. (2002) [1] studied robust control problems via sequential semidefinite programming technique. They obtained the local quadratic convergence rate of the proposed SQP-type method and employed a partial augmented Lagrangian method to deal with the problems addressed there. Correa and Ramirez (2004) [2] systematically studied an SQP-type method for solving nonlinear SDP problems and analyzed the convergence properties, they obtained the global convergence and local quadratic convergence rate. Both papers used the same subproblems to generate search directions, but employed different merit functions for line search. The convergence analysis of both papers depends on a set of second-order conditions without sigma term, which is stronger than no gap second-order optimality condition with sigma term.

Comparing with the work by Correa and Ramirez (2004) [2], in this note, we make some modifications to the convergence analysis, and prove that all results in [2] still hold under the strong second-order sufficient condition with the sigma term.

It should be pointed out that the importance of exploring numerical methods for solving nonlinear semidefinite programming problems has been recognized in the optimization community. For instance, Kočvara and Stingl [3, 4] have developed PENNLP and PENBMI codes for nonlinear semidefinite programming and semidefinite programming with bilinear matrix inequality constraints, respectively. Recently, Sun et al. (2007) [5] considered the rate of convergence of the classical augmented Lagrangian method and Noll (2007) [6] investigated the convergence properties of a class of nonlinear Lagrangian methods.

In Section 2, we introduce preliminaries including differential properties of the metric projector onto $\mathcal{S}_+^p$ and optimality conditions for problem (1.1). In Section 3, we prove, under the strong second-order sufficient condition with the sigma term, that the local SQP-type method has the quadratic convergence rate and the global algorithm with line search is convergence.

2. Preliminaries

We use $\mathcal{R}^{m \times n}$ to denote the set of all the matrices of m rows and n columns. For A and B in $\mathcal{R}^{m \times n}$, we use the Frobenius inner product $\langle A, B \rangle = \mathrm{tr}(A^T B)$, and the Frobenius norm $\|A\|_F = \sqrt{\mathrm{tr}(A^T A)}$, where "tr" denotes the trace operation of a square matrix.

For a given matrix $A \in \mathcal{S}^p$, its spectral decomposition is

$$A = P\Lambda P^T = P \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_p \end{pmatrix} P^T, \tag{2.1}$$

where Λ is the diagonal matrix of eigenvalues of A and P is a corresponding orthogonal matrix. We can express Λ and P as

$$\Lambda = \begin{pmatrix} \Lambda_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Lambda_\gamma \end{pmatrix}, \qquad P = \begin{bmatrix} P_\alpha & P_\beta & P_\gamma \end{bmatrix}, \tag{2.2}$$

where α, β, γ are the index sets of positive, zero, negative eigenvalues of A, respectively.

2.1. Semismoothness of the metric projector

In this subsection, let X, Y, and Z be three arbitrary finite-dimensional real spaces with a scalar product $\langle \cdot, \cdot \rangle$ and its norm $\|\cdot\|$. We introduce some properties of the metric projector, especially its strong semismoothness.

The next lemma is about the generalized Jacobian for composite functions, proposed in [7].

Lemma 2.1. *Let $\Psi : X \to Y$ be a continuously differentiable function on an open neighborhood $\widehat{N}$ of $\overline{x}$ and let $\Xi : \mathcal{O} \subseteq Y \to Z$ be a locally Lipschitz continuous function on the open set $\mathcal{O}$ containing*

$\overline{y} := \Psi(\overline{x})$. *Suppose that Ξ is directionally differentiable at every point in O and that $\mathcal{J}_x\Psi(\overline{x}) : X \to Y$ is onto. Then it holds that*

$$\partial_B\Phi(\overline{x}) = \partial_B\Xi(\overline{y})\mathcal{J}_x\Psi(\overline{x}), \tag{2.3}$$

where $\Phi : \widehat{N} \to Z$ is defined by $\Phi(x) := \Xi(\Psi(x))$, $x \in \widehat{N}$.

The following concept of semismoothness was first introduced by Mifflin [8] for functionals and was extended by Qi and Sun in [9] to vector valued functions.

Definition 2.2. Let $\Phi : \mathcal{O} \subseteq X \to Y$ be a locally Lipschitz continuous function on the open set $\mathcal{O}$. One says that Φ is semismooth at a point $x \in \mathcal{O}$ if

 (i) Φ is directionally differentiable at x;
 (ii) for any $\Delta x \in X$ and $V \in \partial\Phi(x + \Delta x)$ with $\Delta x \to 0$,

$$\Phi(x + \Delta x) - \Phi(x) - V(\Delta x) = o(\|\Delta x\|). \tag{2.4}$$

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

$$\Phi(x + \Delta x) - \Phi(x) - V(\Delta x) = O(\|\Delta x\|^2). \tag{2.5}$$

Let D be a closed convex set in a Banach space Z, and let $\Pi_D : Z \to Z$ be the metric projector over D. It is well known in [10, 11] that $\Pi_D(\cdot)$ is F-differentiable almost everywhere in Z and for any $y \in Z$, $\partial\Pi_D(y)$ is well defined.

Suppose $A \in \mathcal{S}^p$, then it has the spectral decomposition as (2.1), then the merit projector of A to $\mathcal{S}^p_+$ is denoted by $\Pi_{\mathcal{S}^p_+}(A)$ and

$$\Pi_{\mathcal{S}^p_+}(A) = P \begin{pmatrix} [\lambda_1]_+ & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & [\lambda_p]_+ \end{pmatrix} P^T, \tag{2.6}$$

where $[\lambda_i]_+ = \max\{0, \lambda_i\}, i = 1, \ldots, p$.

For our discussion, we know from [12] that the projection operator $\Pi_{\mathcal{S}^p_+}(\cdot)$ is directionally differentiable everywhere in $\mathcal{S}^p_+$ and is a strongly semismooth matrix-valued function. In fact, for any $A \in \mathcal{S}^p$, $H \in \mathcal{S}^p_+$, there exists $V \in \partial\Pi_{\mathcal{S}^p_+}(A + H)$, satisfying

$$\Pi_{\mathcal{S}^p_+}(A + H) = \Pi_{\mathcal{S}^p_+}(A) + V(H) + O(\|H\|^2). \tag{2.7}$$

2.2. Optimality conditions

Let the Lagrangian function of (1.1) be

$$L(x, \lambda, \mu) = f(x) + \langle \lambda, h(x) \rangle + \langle \mu, g(x) \rangle. \tag{2.8}$$

Robinson's constraint qualification(CQ) is said to hold at a feasible point x if

$$0 \in \text{int} \left\{ \begin{pmatrix} h(x) \\ g(x) \end{pmatrix} + \begin{pmatrix} \partial h(x) \\ \partial g(x) \end{pmatrix} \mathbf{R}^n - \begin{pmatrix} \{0\} \\ \mathcal{S}_+^p \end{pmatrix} \right\}. \tag{2.9}$$

If $\overline{x}$ is a locally optimal solution to (1.1) and Robinson's CQ holds at $\overline{x}$, then there exist Lagrangian multipliers $(\lambda, \mu) \in \mathbb{R}^l \times \mathcal{S}^p$, such that the following KKT condition holds:

$$0 = \nabla_x L(\overline{x}, \lambda, \mu) = \nabla f(\overline{x}) + \partial h(\overline{x})^* \lambda + \partial g(\overline{x})^* \mu, \quad 0 = h(\overline{x}),$$
$$g(\overline{x}) = \Pi_{\mathcal{S}_+^p}(g(\overline{x}) + \mu), \tag{2.10}$$

which is equivalent to $F(\overline{x}, \lambda, \mu) = 0$, where

$$F(x, \lambda, \mu) := \begin{pmatrix} \nabla f(x) + \partial h(x)^* \lambda + \partial g(x)^* \mu \\ h(x) \\ g(x) - \Pi_{\mathcal{S}_+^p}(g(x) + \mu) \end{pmatrix}. \tag{2.11}$$

Let $\Lambda(\overline{x})$ be the set of all the Lagrangian multipliers satisfying (2.10). Then $\Lambda(\overline{x})$ is a nonempty, compact convex set of $\mathbb{R}^l \times \mathcal{S}^p$ if and only if Robinson's CQ holds at $\overline{x}$, see [13]. Moreover, it follows from [13] that the constraint nondegeneracy condition is a sufficient condition for Robinson constraint qualification. In the setting of the problem (1.1), the constraint nondegeneracy condition holding at a feasible point $\overline{x}$ can be expressed as

$$\begin{pmatrix} \partial h(\overline{x}) \\ \partial g(\overline{x}) \end{pmatrix} \mathbb{R}^n + \begin{pmatrix} \{0\} \\ \text{lin}(T_{\mathcal{S}_+^p}(g(\overline{x}))) \end{pmatrix} = \begin{pmatrix} \mathbb{R}^l \\ \mathcal{S}^p \end{pmatrix}, \tag{2.12}$$

where $\text{lin}(T_{\mathcal{S}_+^p}(g(\overline{x})))$ is the lineality space of the tangent cone of $\mathcal{S}_+^p$ at $g(\overline{x})$. If $\overline{x}$, a locally optimal solution to (1.1), is nondegenerate, then $\Lambda(\overline{x})$ is a singleton.

For a KKT point $(\overline{x}, \lambda, \mu)$ of (1.1), without loss of generality, we assume that $g(\overline{x})$ and μ have the spectral decomposition forms

$$g(\overline{x}) = P \begin{pmatrix} \Lambda_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} P^T, \qquad \mu = P \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Lambda_\gamma \end{pmatrix} P^T. \tag{2.13}$$

We state the strong second-order sufficient condition (SSOSC) coming from [7].

Definition 2.3. Let $\overline{x}$ be a stationary point of (1.1) such that (2.12) holds at $\overline{x}$. One says that the strong second-order sufficient condition holds at $\overline{x}$ if

$$\langle d, \nabla_{xx}^2 L(\overline{x}, \overline{\lambda}, \overline{\mu}) d \rangle - \Upsilon_{g(\overline{x})}(\overline{\mu}, \partial g(\overline{x}) d) > 0, \quad \forall d \in \text{aff } C(\overline{x}) \setminus \{0\}, \tag{2.14}$$

where $\{(\overline{\lambda}, \overline{\mu})\} = \Lambda(\overline{x}) \subset \mathbb{R}^l \times \mathcal{S}^p$, aff $C(\overline{x})$ is the affine hull of the critical cone $C(\overline{x})$:

$$\text{aff } C(\overline{x}) = \{d : \partial h(\overline{x}) d = 0, \ P_\beta^T (\partial g(\overline{x}) d) P_\gamma = 0, \ P_\gamma^T (\partial g(\overline{x}) d) P_\gamma = 0\}. \tag{2.15}$$

And the linear-quadratic function $\Upsilon_B : \mathcal{S}^p \times \mathcal{S}^p \to \mathbb{R}$ is defined by

$$\Upsilon_B(D, A) := 2\langle D, A B^\dagger A \rangle, \qquad (D, A) \in \mathcal{S}^p \times \mathcal{S}^p, \tag{2.16}$$

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

The next proposition relates the SSOSC and nondegeneracy condition to nonsingularity of Clarke's Jacobian of the mapping F defined by (2.11). The details of this proof can be found in [7].

Proposition 2.4. *Let* $(\overline{x}, \overline{\lambda}, \overline{\mu})$ *be a KKT point of* (1.1). *If nondegeneracy condition* (2.12) *and SSOSC* (2.14) *hold at* $\overline{x}$, *then any element in* $\partial F(\overline{x}, \overline{\lambda}, \overline{\mu})$ *is nonsingular, where* F *is defined by* (2.11).

3. Convergence analysis of the SQP-type method

In this section, we analyze the local quadratic convergence rate of an SQP-type method and then prove that the SQP-type method proposed in [2] is globally convergent. The analysis is based on the strong second-order sufficient condition, which is weaker than the conditions used in [1, 2].

3.1. Local convergence rate

Linearizing (1.1) at the current point (x^k, λ^k, μ^k), we obtain the following tangent quadratic problem:

$$\min{}_{\Delta x} \nabla f(x^k)^T \Delta x + \frac{1}{2} \Delta x^T \nabla_{xx}^2 L(x^k, \lambda^k, \mu^k) \Delta x,$$
$$\text{s.t. } h(x^k) + \partial h(x^k) \Delta x = 0, \ g(x^k) + \partial g(x^k) \Delta x \in \mathcal{S}_+^p, \tag{3.1}$$

where $\nabla_{xx}^2 L(x^k, \lambda^k, \mu^k) = \partial_x(\nabla_x L)(x^k, \lambda^k, \mu^k)$. Let $(\Delta x^k, \lambda_{\mathrm{QP}}^k, \mu_{\mathrm{QP}}^k)$ be a KKT point of (3.1), then we have $\widehat{F}(\Delta x^k, \lambda_{\mathrm{QP}}^k, \mu_{\mathrm{QP}}^k; x^k, \lambda^k, \mu^k) = 0$, where

$$\widehat{F}(\zeta, \eta, \xi; x^k, \lambda^k, \mu^k) := \begin{pmatrix} \nabla f(x^k) + \nabla_{xx}^2 L(x^k, \lambda^k, \mu^k)\zeta + \partial h(x^k)^* \eta + \partial g(x^k)^* \xi \\ h(x^k) + \partial h(x^k)\zeta \\ g(x^k) + \partial g(x^k)\zeta - \Pi_{\mathcal{S}_+^p}(g(x^k) + \partial g(x^k)\zeta + \xi) \end{pmatrix}. \tag{3.2}$$

The following algorithm is an SQP-type algorithm for solving (1.1), which is based on computing at each iteration a primal-dual stationary point $(\Delta x^k, \lambda_k^{\mathrm{QP}}, \mu_k^{\mathrm{QP}})$ of (3.1).

Algorithm 3.1
Step 1. Given an initial iterate point (x^1, λ^1, μ^1). Compute $h(x^1)$, $g(x^1)$, $\nabla f(x^1)$, $\partial h(x^1)$ and $\partial g(x^1)$. Set $k := 1$.
Step 2. If $\nabla_x L(x^k, \lambda^k, \mu^k) = 0$, $h(x^k) = 0$, $g(x^k) \in \mathcal{S}_+^P$, stop.
Step 3. Compute $\nabla_{xx}^2 L(x^k, \lambda^k, \mu^k)$, and find a solution $(\Delta x^k, \lambda_{\mathrm{QP}}^k, \mu_{\mathrm{QP}}^k)$ to (3.1).
Step 4. Set $x^{k+1} := x^k + \Delta x^k$, $\lambda^{k+1} := \lambda_{\mathrm{QP}}^k$, $\mu^{k+1} := \mu_{\mathrm{QP}}^k$.
Step 5. Compute $h(x^{k+1})$, $g(x^{k+1})$, $\nabla f(x^{k+1})$, $\partial h(x^{k+1})$ and $\partial g(x^{k+1})$. Set $k := k + 1$ and go to step 2.

From item (f) of [7, Theorem 4.1], we obtain the error between $(\Delta x^k, \lambda_k^{\mathrm{QP}}, \mu_k^{\mathrm{QP}})$ and $(\overline{x}, \overline{\lambda}, \overline{\mu})$ directly.

Theorem 3.2. *Suppose that* f, h, g *are twice continuously differentiable and their derivatives are locally Lipschitz in a neighborhood of a local solution* $\overline{x}$ *to* (1.1). *Suppose nondegeneracy condition* (2.12)

and SSOSC (2.14) *hold at* $\overline{x}$. *Then there exists a neighborhood* U *of* $(\overline{x}, \overline{\lambda}, \overline{\mu})$ *such that if* (x^k, λ^k, μ^k) *in* U, (3.1) *has a local solution* Δx^k *together with corresponding Lagrangian multiplies* $(\lambda_{QP}^k, \mu_{QP}^k)$ *satisfying*

$$\|\Delta x^k\| + \|\lambda_{QP}^k - \overline{\lambda}\| + \|\mu_{QP}^k - \overline{\mu}\| = O(\|(x^k, \lambda^k, \mu^k) - (\overline{x}, \overline{\lambda}, \overline{\mu})\|). \tag{3.3}$$

Now we are in a position to state that the sequence of primal-dual points generated by Algorithm 3.1 has quadratic convergence rate.

Theorem 3.3. *Suppose that* f, h, g *are twice continuously differentiable and their derivatives are locally Lipschitz in a neighborhood of a local solution* $\overline{x}$ *to* (1.1). *Suppose nondegeneracy condition* (2.12) *and SSOSC* (2.14) *hold at* $\overline{x}$. *Consider Algorithm 3.1, in which* Δx^k *is a minimum norm stationary point of the tangential quadratic problem* (3.1). *Then there exists a neighborhood* U *of* $(\overline{x}, \overline{\lambda}, \overline{\mu})$ *such that, if* $(x^1, \lambda^1, \mu^1) \in U$, *Algorithm 3.1 is well defined and the sequence* $\{(x^k, \lambda^k, \mu^k)\}$ *converges quadratically to* $(\overline{x}, \overline{\lambda}, \overline{\mu})$.

Proof. By Theorem 3.2, we know Algorithm 3.1 is well defined. Let

$$\delta_k := \|(x^k, \lambda^k, \mu^k) - (\overline{x}, \overline{\lambda}, \overline{\mu})\|, \tag{3.4}$$

then

$$\Delta x^k = O(\delta_k), \qquad \lambda^{k+1} - \overline{\lambda} = O(\delta_k), \qquad \mu^{k+1} - \overline{\mu} = O(\delta_k), \tag{3.5}$$

where Δx^k is the minimum norm solution to (3.1), and $\lambda^{k+1} = \lambda_{QP}^k$, $\mu^{k+1} = \mu_{QP}^k$ are the associated multipliers. Using Taylor expansion of (3.2) at $(\overline{x}, \overline{\lambda}, \overline{\mu})$, noting that $\nabla_x L(\overline{x}, \overline{\lambda}, \overline{\mu}) = 0$, $x^{k+1} = x^k + \Delta x^k$, and (3.5), we obtain

$$\nabla_{xx}^2 L(\overline{x}, \overline{\lambda}, \overline{\mu})(x^{k+1} - \overline{x}) + \partial h(\overline{x})^*(\lambda^{k+1} - \overline{\lambda}) + \partial g(\overline{x})^*(\mu^{k+1} - \overline{\mu}) = O(\delta_k^2),$$
$$\partial h(\overline{x})(x^{k+1} - \overline{x}) = O(\delta_k^2). \tag{3.6}$$

As the projection operator $\Pi_{\mathcal{S}_+^p}(\cdot)$ is strongly semismooth, we have that there exists $V \in \partial \Pi_{\mathcal{S}_+^p}(g(\overline{x}) + \overline{\mu})$ such that

$$\Pi_{\mathcal{S}_+^p}(g(\overline{x}) + \overline{\mu}) = \Pi_{\mathcal{S}_+^p}(g(x^k) + \partial g(x^k)\Delta x^k + \mu_{QP}^k)$$
$$+ V(g(\overline{x}) + \overline{\mu} - g(x^k) - \partial g(x^k)\Delta x^k - \mu_{QP}^k) \tag{3.7}$$
$$+ O(\|g(\overline{x}) + \overline{\mu} - g(x^k) - \partial g(x^k)\Delta x^k - \mu_{QP}^k\|^2).$$

Since

$$g(\overline{x}) + \overline{\mu} - g(x^k) - \partial g(x^k)\Delta x^k - \mu_{QP}^k = \partial g(x^k)(\overline{x} - x^{k+1}) + (\overline{\mu} - \mu_{QP}^k) + O(\delta_k^2), \tag{3.8}$$

we have

$$\Pi_{\mathcal{S}_+^p}(g(x^k) + \partial g(x^k)\Delta x^k + \mu_{QP}^k)$$
$$= \Pi_{\mathcal{S}_+^p}(g(\overline{x}) + \overline{\mu}) - V(\partial g(x^k)(\overline{x} - x^{k+1}) + (\overline{\mu} - \mu_{QP}^k)) + O(\delta_k^2). \tag{3.9}$$

Noting the fact that $g(\overline{x}) = \Pi_{\mathcal{S}_+^p}(g(\overline{x}) + \overline{\mu})$, by Taylor expansion of the third equation of (3.2) at $(\overline{x}, \overline{\mu})$, we obtain

$$(V - I)\mathcal{D}g(\overline{x})(x^{k+1} - \overline{x}) + V(\mu^{k+1} - \overline{\mu}) = O(\delta_k^2). \tag{3.10}$$

Therefore, we can conclude that

$$\begin{pmatrix} \nabla_{xx}^2 L(\overline{x}, \overline{\lambda}, \overline{\mu}) & \mathcal{D}h(\overline{x})^* & \mathcal{D}g(\overline{x})^* \\ \mathcal{D}h(\overline{x}) & 0 & 0 \\ -\mathcal{D}g(\overline{x}) + V\mathcal{D}g(\overline{x}) & 0 & V \end{pmatrix} \begin{pmatrix} x^{k+1} - \overline{x} \\ \lambda^{k+1} - \overline{\lambda} \\ \mu^{k+1} - \overline{\mu} \end{pmatrix} = O(\delta_k^2). \tag{3.11}$$

Since the nondegeneracy condition (2.12) and SSOSC (2.14) hold, we have from Proposition 2.4 that (3.11) implies the quadratic convergence of the sequence $\{(x^k, \lambda^k, \mu^k)\}$. $\qquad\square$

3.2. The global convergence

The tangential quadratic problem constrained here is slightly more general than (3.1) in the sense that the Hessian of the Lagrangian $\nabla_{xx}^2 L(x^k, \lambda^k, \mu^k)$ is replaced by some positive definite matrix M_k. Thus the tangential quadratic problem in Δx now becomes

$$\min_{\Delta x} \nabla f(x^k)^T \Delta x + \frac{1}{2} \Delta x^T M_k \Delta x, \qquad \text{s.t. } h(x^k) + \mathcal{D}h(x^k)\Delta x = 0,$$
$$g(x^k) + \mathcal{D}g(x^k)\Delta x \in \mathcal{S}_+^p. \tag{3.12}$$

The KKT system of (3.12) is

$$\nabla f(x^k) + M_k \Delta x^k + \mathcal{D}h(x^k)^* \lambda_{QP}^k + \mathcal{D}g(x^k)^* \mu_{QP}^k = 0, \qquad h(x^k) + \mathcal{D}h(x^k)\Delta x^k = 0,$$
$$g(x^k) + \mathcal{D}g(x^k)\Delta x^k - \Pi_{\mathcal{S}_+^p}(g(x^k) + \mathcal{D}g(x^k)\Delta x^k + \mu_{QP}^k) = 0. \tag{3.13}$$

To obtain the global convergence, we use the Han penalty function given by [14], as a merit function and Armijo line search. For problem (1.1), the Han penalty function is defined by

$$\Theta_\sigma(x) = f(x) + \sigma\|h(x)\| - \sigma\lambda_{\min}(g(x))_-, \tag{3.14}$$

where $\lambda_{\min}(g(x))$ is the smallest eigenvalue of $g(x)$, $(\cdot)_-$ denote min $\{\cdot, 0\}$ and $\sigma > 0$ is a positive constant.

The following proposition comes from [2] directly.

Proposition 3.4. (i) *If f, h, g have a directional derivative at x in the direction $d \in \mathbf{R}^n$, then Θ_σ has also a directional derivative at x in the direction d. If, in addition, x is feasible for (1.1), we have*

$$\Theta_\sigma'(x; d) = f'(x; d) + \sigma\|h'(x; d)\| - \sigma\lambda_{\min}(N^T \mathcal{D}g(x)dN), \tag{3.15}$$

where $N = [\nu_1, \dots, \nu_r]$ is the matrix whose columns ν_i form an orthonormal basis of $\mathrm{Ker}g(x)$.

(ii) *If $\overline{x}$ is a feasible point of (1.1) and Θ_σ has a local minimum at $\overline{x}$, then $\overline{x}$ is the local solution to (1.1). Furthermore, if f, h, g are differentiable at $\overline{x}$ and nondegeneracy condition (2.12) holds at $\overline{x}$, then $\sigma \geq \max\{\|\overline{\lambda}\|, \mathrm{tr}(-\overline{\mu})\}$.*

(iii) *If $\mu < 0$ and $\sigma \geq \max\{\|\lambda\|, \mathrm{tr}(-\mu)\}$, then $L(\cdot, \lambda, \mu) \leq \Theta_\sigma(\cdot)$.*

To discuss the conditions ensuring the exactness of Θ_σ, we need the following lemma from (3.10).

Lemma 3.5. *Suppose nondegeneracy condition (2.12) and SSOSC (2.14) hold at $\overline{x}$. Then there exists $c_0 > 0$, such that for any $c > c_0$ there exist a neighborhood V of $\overline{x}$ and a neighborhood U of $(\overline{\lambda}, \overline{\mu})$, for any $(\lambda, \mu) \in U$, the problem*

$$\min L_c(x, \lambda, \mu) \quad \text{s.t. } x \in V \tag{3.16}$$

has a unique solution denote $x_c(\lambda, \mu)$. The function $x_c(\cdot, \cdot)$ is locally Lipschitz continuous and semismooth on U. Furthermore, there exists $\rho > 0$, for any $(\lambda, \mu) \in U$,

$$\|\overline{x} - x_c(\lambda, \mu)\| \le \rho \|(\lambda, \mu) - (\overline{\lambda}, \overline{\mu})\| / c, \tag{3.17}$$

where

$$L_c(x, \lambda, \mu) := f(x) + \langle h(x), \lambda \rangle + \frac{c}{2}\|h(x)\|^2 + \frac{1}{2c}\left[\|\Pi_{\mathcal{S}_+^p}(-\mu - cg(x))\|_F^2 - \|\mu\|_F^2\right] \tag{3.18}$$

is the augmented Lagrangian function with the penalty parameter c for (1.1).

Theorem 3.6. *Suppose that f, h, g are twice differentiable around a local solution $\overline{x}$ to (1.1), at which nondegeneracy condition (2.12) and SSOSC (2.14) hold. If $\sigma > \max\{\|\overline{\lambda}\|, \operatorname{tr}(-\overline{\mu})\}$, then Θ_σ has a strict local minimum at $\overline{x}$.*

Proof. For the definition of the projection operator $\Pi_{\mathcal{S}_+^p}(\cdot)$, we have

$$\Pi_{\mathcal{S}_+^p}(-\mu - cg(x)) = -\mu - cg(x) + \Pi_{\mathcal{S}_+^p}(cg(x) + \mu), \tag{3.19}$$

and for any $W \in \mathcal{S}_+^p$, $c > 0$,

$$\|\Pi_{\mathcal{S}_+^p}(cg(x) + \mu) - (cg(x) + \mu)\|_F^2 \le \|W - (cg(x) + \mu)\|_F^2. \tag{3.20}$$

Then

$$\begin{aligned}
&\|\Pi_{\mathcal{S}_+^p}(cg(x) + \mu) - cg(x)\|_F^2 - 2\langle \mu, \Pi_{\mathcal{S}_+^p}(cg(x) + \mu) - cg(x)\rangle \\
&\quad \le -2\langle \mu, W - cg(x)\rangle + \|W - cg(x)\|_F^2
\end{aligned} \tag{3.21}$$

holds for any $W \in \mathcal{S}_+^p$. So taking $\mu = \overline{\mu}$ and $W = c\Pi_{\mathcal{S}_+^p}(g(x))$, we obtain that

$$\begin{aligned}
&\|\Pi_{\mathcal{S}_+^p}(cg(x) + \overline{\mu}) - cg(x)\|_F^2 - 2\langle \overline{\mu}, \Pi_{\mathcal{S}_+^p}(cg(x) + \overline{\mu}) - cg(x)\rangle \\
&\quad \le -2c\langle \overline{\mu}, \Pi_{\mathcal{S}_+^p}(-g(x))\rangle + c^2 \|\Pi_{\mathcal{S}_+^p}(-g(x))\|_F^2,
\end{aligned} \tag{3.22}$$

which implies

$$\begin{aligned}
L_c(x, \overline{\lambda}, \overline{\mu}) &\le f(x) + \langle \overline{\lambda}, h(x)\rangle + \frac{c}{2}\|h(x)\|^2 - \langle \overline{\mu}, \Pi_{\mathcal{S}_+^p}(-g(x))\rangle + \frac{c}{2}\|\Pi_{\mathcal{S}_+^p}(-g(x))\|_F^2 \\
&\le f(x) + \|h(x)\|\left(\|\overline{\lambda}\| + \frac{c}{2}\|h(x)\|\right) \\
&\quad + \lambda_{\max}(\Pi_{\mathcal{S}_+^p}(-g(x)))\left[\operatorname{tr}(-\overline{\mu}) + \frac{c}{2}\sum_{i=1}^p \lambda_i(\Pi_{\mathcal{S}_+^p}(-g(x)))\right].
\end{aligned} \tag{3.23}$$

Since $\sigma > \max\{\|\overline{\lambda}\|, \operatorname{tr}(-\overline{\mu})\}$, for any fixed $c > 0$, there exists a neighborhood V_c of $\overline{x}$ such that

$$L_c(x, \overline{\lambda}, \overline{\mu}) \leq f(x) + \sigma\|h(x)\| + \sigma\lambda_{\max}\left(\Pi_{\mathcal{S}_+^p}(-g(x))\right) = \Theta_\sigma(x), \quad \forall x \in V_c. \tag{3.24}$$

From Lemma 3.5, we know that there exist an $r > c_0$ and a neighborhood V_r of $\overline{x}$ where $\overline{x}$ is a strict minimum of $L_r(\cdot, \overline{\lambda}, \overline{\mu})$. So we can conclude that $\overline{x}$ is a strict minimum of Θ_σ on $V_c \cap V_r$. $\qquad\square$

Let us outline the line-search SQP-type algorithm that uses the merit function $\Theta_\sigma(\cdot)$ defined in (3.14) and the parameter updating scheme from [14], which is a generalized version to the algorithm in [2].

Algorithm 3.7
Step 1. Given a positive number $\overline{\sigma} > 0$, $\varrho \in (0, 1/2)$, $\beta \in (0, 1/2)$. Choose an initial iterate $(x^1, \lambda^1, \mu^1) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathcal{S}^p$. Compute $f(x^1), h(x^1), g(x^1), \nabla f(x^1), \mathcal{D}h(x^1)$ and $\mathcal{D}g(x^1)$. Set $k := 1, \sigma_1 = \overline{\sigma}$.
Step 2. If $\nabla_x L(x^k, \lambda^k, \mu^k) = 0$, $h(x^k) = 0$, $g(x^k) \in \mathcal{S}_+^p$, stop.
Step 3. Compute a symmetric matrix M_k and find a solution $(\Delta x^k, \lambda_{\mathrm{QP}}^k, \mu_{\mathrm{QP}}^k)$ to (3.12).
Step 4. Adapt σ_k.

if $\sigma_{k-1} \geq \max\{\operatorname{tr}(-\mu^{k+1}), \|\lambda^{k+1}\|\} + \overline{\sigma}$

then $\sigma_k = \sigma_{k-1}$

else $\sigma_k = \max\{1.5\sigma_{k-1}, \max\{\operatorname{tr}(-\mu^{k+1}), \|\lambda^{k+1}\|\} + \overline{\sigma}\}$

Step 5. Compute

$$w_k := -\langle \Delta x^k, M_k \Delta x^k \rangle + \langle \mu_{\mathrm{QP}}^k, g(x^k) \rangle + \langle \lambda_{\mathrm{QP}}^k, h(x^k) \rangle - \sigma_k\|h(x^k)\| + \sigma_k\lambda_{\min}(g(x^k))_-. \tag{3.25}$$

Using backtracking line search rule to compute the step length α_k:
Step 6. set $i = 0$, $\alpha_{k,0} = 1$;
Step 7. if

$$\Theta_{\sigma_k}(x^k + \alpha\Delta x^k) \leq \Theta_{\sigma_k}(x^k) + \varrho\alpha w_k \tag{3.26}$$

holds for $\alpha = \alpha_{k,i}$, then $\alpha_k = \alpha$ and stop the line search.
Step 8. else, choose $\alpha_{k,i+1} \in [\beta\alpha_{k,i}, (1-\beta)\beta\alpha_{k,i}]$;
Step 9. set $i := i + 1$, go to step 7
Step 10. Set $x^{k+1} := x^k + \alpha_k\Delta x^k$, $\lambda^{k+1} := \lambda_{\mathrm{QP}}^k$, $\mu^{k+1} := \mu_{\mathrm{QP}}^k$.
Step 11. Compute $f(x^{k+1}), h(x^{k+1}), g(x^{k+1}), \nabla f(x^{k+1}), \mathcal{D}h(x^{k+1})$ and $\mathcal{D}g(x^{k+1})$. Set $k := k + 1$ and go to step 2.

Now we are in a position to state the global convergence of the line search SQP Algorithm 3.7, whose proof can be found in [2].

Theorem 3.8. *Suppose that f, h, g are continuously differentiable and their derivatives are Lipschitz continuous. Consider Algorithm 3.7, if positive definite matrices M_k and M_k^{-1} are bounded, then one of the following situations occurs:*

(i) *the sequence $\{\sigma_k\}$ is unbounded, in which case $\{(\lambda^{k+1}, \mu^{k+1})\}$ is also unbounded;*

(ii) *there exists an index k_2 such that $\sigma_k = \sigma$ for any $k \geq k_2$, and one of the following situations occurs:*

    (a) $\Theta_\sigma(x^k) \to \infty$,

    (b) $\nabla_x L(x^k, \lambda^k, \mu^k) \to 0$, $h(x^k) \to 0$, $\lambda_{\min}(g(x^k))_- \to 0$, *and* $\langle \mu^{k+1}, g(x^k) \rangle \to 0$.

Acknowledgments

References

[1] B. Fares, D. Noll, and P. Apkarian, "Robust control via sequential semidefinite programming," *SIAM Journal on Control and Optimization*, vol. 40, no. 6, pp. 1791–1820, 2002.

[2] R. Correa and H. C. Ramirez, "A global algorithm for nonlinear semidefinite programming," *SIAM Journal on Optimization*, vol. 15, no. 1, pp. 303–318, 2004.

[3] M. Kočvara and M. Stingl, "Pennon: a code for convex nonlinear and semidefinite programming," *Optimization Methods & Software*, vol. 18, no. 3, pp. 317–333, 2003.

[4] M. Kočvara and M. Stingl, "Solving nonconvex SDP problems of structural optimization with stability control," *Optimization Methods & Software*, vol. 19, no. 5, pp. 595–609, 2004.

[5] D. Sun, J. Sun, and L. Zhang, "The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming," *Mathematical Programming*, 2008.

[6] D. Noll, "Local convergence of an augmented Lagrangian method for matrix inequality constrained programming," *Optimization Methods and Software*, vol. 22, no. 5, pp. 777–802, 2007.

[7] D. Sun, "The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications," *Mathematics of Operations Research*, vol. 31, no. 4, pp. 761–776, 2006.

[8] R. Mifflin, "Semismooth and semiconvex functions in constrained optimization," *SIAM Journal on Control and Optimization*, vol. 15, no. 6, pp. 959–972, 1977.

[9] L. Q. Qi and J. Sun, "A nonsmooth version of Newton's method," *Mathematical Programming*, vol. 58, no. 1–3, pp. 353–367, 1993.

[10] E. H. Zarantonello, "Projections on convex sets in Hilbert space and spectral theory—I: projections on convex sets," in *Contributions to Nonlinear Functional Analysis (Proceedings of a Symposium held at the Mathematics Research Center, University of Wisconsin, Madison, Wis., 1971)*, E. H. Zarantonello, Ed., pp. 237–341, Academic Press, New York, NY, USA, 1971.

[11] E. H. Zarantonello, "Projections on convex sets in Hilbert space and spectral theory—II: spectral theory," in *Contributions to Nonlinear Functional Analysis (Proceedings of a Symposium held at the Mathematics Research Center, University of Wisconsin, Madison, Wis., 1971)*, E. H. Zarantonello, Ed., pp. 343–424, Academic Press, New York, NY, USA, 1971.

[12] D. Sun and J. Sun, "Semismooth matrix-valued functions," *Mathematics of Operations Research*, vol. 27, no. 1, pp. 150–169, 2002.

[13] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer Series in Operations Research, Springer, New York, NY, USA, 2000.

[14] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal, *Numerical Optimization. Theoretical and Practical Aspects*, Universitext, Springer, Berlin, Germany, 2003.