# SUPERLINEAR CONVERGENCE OF AN SQP-TYPE METHOD FOR NONLINEAR SEMIDEFINITE PROGRAMMING

#### WENHAO FU AND ZHONGWEN CHEN\*

Abstract. In this paper, we study the rate of convergence of a sequential quadratic programming (SQP) method for nonlinear semidefinite programming (SDP) problems. Since the linear SDP constraints does not contribute to the Hessian of the Lagrangian, we propose a reduced SQP-type method, which solves an equivalent and reduced type of the nonlinear SDP problem near the optimal point. For the reduced SDP problem, the well-known and often mentioned " $\sigma$ -term" in the second order sufficient condition vanishes. We analyze the rate of local convergence of the reduced SQP-type method and give a sufficient and necessary condition for its superlinear convergence. Furthermore, we give a sufficient and necessary condition for superlinear convergence of the SQP-type method under the nondegeneracy condition, the second-order sufficient condition with  $\sigma$ -term and the strict complementarity condition.

Key words. Nonlinear semidefinite programming, SQP-type method, second order sufficient condition, constraint nondegeneracy, superlinear convergence.

## 1. Introduction

Consider the following nonlinear semidefinite programming (SDP) problem

(1) 
$$\begin{array}{l} \min_{\substack{x \in \mathcal{R}^n \\ \text{s.t.}}} & f(x) \\ \text{s.t.} & h(x) = 0, \\ & G(x) \succeq 0, \end{array}$$

where  $f : \mathbb{R}^n \to \mathbb{R}$ ,  $h : \mathbb{R}^n \to \mathbb{R}^l$  and  $G : \mathbb{R}^n \to \mathbb{S}^m$  are all smooth functions.  $\mathcal{S}^m$  denotes the linear space of *m*-order real symmetric matrices,  $\mathcal{S}^m_+$  and  $\mathcal{S}^m_{++}$  denote the linear space of *m*-order real symmetric positive semidefinite matrices and symmetric positive definite ones, respectively.  $\succeq$  and  $\succ$  denote positive semidefinite order and positive definite order, which means  $A \succeq B$  if and only if  $A - B \in \mathcal{S}^m_+$  and  $A \succ B$  if and only if  $A - B \in \mathcal{S}^m_{++}$ , respectively. In the past few years, basic theoretical issues of nonlinear semidefinite programming have been studied, such as optimality conditions ([9, 15]), duality theory ([6]), stability analysis ([1, 7, 10]) and so forth.

There are various methods for solving nonlinear SDP problem, such as the penalty/barrier multiplier method ([13]), the Augmented Lagrangian method ([11, 12]), the primal-dual interior point method ([21]), sequential semidefinite programming (SSDP) method ([3, 22, 23]) and so forth. As one of effective methods for solving nonlinear SDP problem, the SSDP method is a development of the SQP-type method on semidefinite cone space. The main idea of such method is to generate steps by solving a sequence of quadratic semidefinite subproblems. At the current iterate  $x_k$ , the trial step  $d_k$  is obtained by solving the following quadratic

Received by the editors December 23, 2019 and, in revised form, March 25, 2020.

<sup>2000</sup> Mathematics Subject Classification. 90C22, 90C30, 65K05. \*Corresponding author.

semidefinite programming subproblem

(2) 
$$\min_{\substack{d \in \mathcal{R}^n \\ \text{s.t.}}} \nabla f(x_k)^T d + \frac{1}{2} d^T W_k d$$
$$\operatorname{s.t.} \quad h(x_k) + Dh(x_k) d = 0,$$
$$G(x_k) + DG(x_k) d \succeq 0,$$

where  $\nabla f(x)$  denotes the gradient of the objective function f(x),  $W_k$  denotes the Hessian matrix of the Lagrangian function of (1) (see (3) for its definition) or its approximate symmetric matrix, Dh(x) denotes the Jacobian matrix of h(x) and

$$Dh(x)^T = (\nabla h_1(x), \nabla h_2(x), \cdots, \nabla h_l(x)).$$

Linear operator DG(x) denotes by

$$DG(x) = \left(\frac{\partial G(x)}{\partial x_1}, \frac{\partial G(x)}{\partial x_2}, \cdots, \frac{\partial G(x)}{\partial x_n}\right)$$

which satisfies

$$DG(x)d = \sum_{i=1}^{n} d_i \frac{\partial G(x)}{\partial x_i}, \quad \forall d \in \mathcal{R}^n.$$

Hence, the new iteration is updated by a line search  $x_{k+1} = x_k + \alpha_k d_k$ , where  $\alpha_k \in (0, 1]$  is a step size.

Under suitable assumptions, the sequence generated by the algorithm above converges globally to a KKT point of the problem (1), e.g., see [3, 23]. A fast local rate of convergence can be arrived if  $W_k$  is a good approximation of the Hessian matrix. Fares et al. ([8]) proved the local quadratic rate of convergence under the maximal rank condition and the second order sufficient condition without  $\sigma$ -term. Under the nondegeneracy condition and the second-order sufficient condition with  $\sigma$ -term, Wang et al. ([19]) made a further analysis on the local property of convergence for SSDP method and also proved that the algorithm has a local quadratic convergence rate when  $W_k$  is chosen as the Hessian matrix of the Lagrangian function. Zhao and Chen ([23]) gave a globally convergent algorithm based on the references [3, 8, 19]and proved the step size  $\alpha_k$  of the algorithm is always equal to 1 for k sufficiently large under the nondegeneracy condition and the second-order sufficient condition with  $\sigma$ -term. Thus the algorithm is convergent superlinearly. The results on the local rate of convergence of some other methods, such as the primal-dual interior point method, the augmented Lagrangian method etc., for nonlinear SDP can be found in [14, 16, 18, 20].

It's worth noting that the results on the local convergence mentioned above all include multiplier term. There are few researches on the convergent rate of the sequence without multiplier term which is generated by SQP-type method for nonlinear SDP, while Boggs et al. ([2]) has already proposed an sufficient and necessary condition in the case of nonlinear programming in the 1980s. Another point of attention is that the Hessian matrix of the Lagrangian function at the optimal point is not necessarily positive definite on the critical cone (see the example showed in [4]). Therefore the subproblems in the SQP-type method near the optimal point may be nonconvex when  $W_k$  takes the Hessian matrix of the Lagrangian function of (1) or its approximate symmetric matrix, which may influence the local convergence properties of the method. In this paper, we analyze the convergent rate of the sequence without multiplier term. And then an equivalent and reduced type of the primal problem near the optimal point is analyzed and the conditions of superlinear convergence are discussed. Finally, a sufficient and necessary condition for superlinear convergence of the algorithm is given under the nondegeneracy condition, the second-order sufficient condition with  $\sigma\text{-term}$  and the strict complementarity condition.

The paper is organized as follows. In Section 2, we introduce the SSDP method and give some results on the local rate of convergence. In Section 3, we first reformulate the primal problem into an equivalent and reduced form, which is obtained by Schur-complement theorem. Then we analyze the conditions of superlinear convergence for the reduced SSDP method. In Section 4, we give an equivalent condition of superlinear convergence of the sequence generated by the SSDP method and conclude with final remarks in Section 5.

# 2. Local convergence of an SQP-type method

The Lagrangian function of the problem (1) is

(3) 
$$L(x,\mu,Y) = f(x) - \mu^T h(x) - \langle Y, G(x) \rangle,$$

where  $\mu \in \mathcal{R}^l$ ,  $Y \in \mathcal{S}^m$ ,  $\langle A, B \rangle = \operatorname{tr}(B^T A)$  denotes the inner product of  $A, B \in \mathcal{R}^{m \times n}$ ,  $\operatorname{tr}(X)$  denotes the trace of a matrix X. At the current iterate  $x_k$ , we solve the quadratic semidefinite programming subproblem (2). Suppose that the subproblem (2) has a solution  $d_k$  and that  $(\mu_{k+1}, Y_{k+1})$  is the Lagrange multiplier corresponding to the constraints. Let  $x_{k+1} = x_k + d_k$ , then the iterate sequence  $\{(x_k, \mu_k, Y_k)\}$  is generated. The detailed algorithm is described as follows.

Algorithm 2.1. Local SSDP Algorithm.

Initialization. Given an initial point  $(x_0, \mu_0, Y_0) \in \mathcal{R}^n \times \mathcal{R}^l \times \mathcal{S}^m_+, k := 0.$ 

Step 1. Solve (2) to get  $d_k$  and the corresponding Lagrange multiplier  $(\mu_{k+1}, Y_{k+1})$ . Step 2. If  $d_k = 0$ , then stop.

Step 3. Set  $x_{k+1} = x_k + d_k$ .

Step 4. Set k := k + 1 and go to Step 1.

Suppose that Algorithm 2.1 generates an infinite sequence  $\{(x_k, \mu_k, Y_k)\}$ . We give some basic assumptions in order to analyze the local convergence of the sequences  $\{(x_k, \mu_k, Y_k)\}$  and  $\{x_k\}$ .

# Assumption A

A1 f(x), h(x) and G(x) are twice continuously differentiable on  $\mathcal{R}^n$ . A2 The nondegeneracy condition holds at a feasible point  $x^*$ , i.e.,

$$\begin{pmatrix} Dh(x^*) \\ DG(x^*) \end{pmatrix} \mathcal{R}^n + \begin{pmatrix} \{0\} \\ \ln(T_{\mathcal{S}^m_+}(G(x^*))) \end{pmatrix} = \begin{pmatrix} \mathcal{R}^l \\ \mathcal{S}^m \end{pmatrix},$$

where  $\lim(T_{\mathcal{S}^m_+}(G(x^*)))$  is a linearity space of the critical cone  $T_{\mathcal{S}^m_+}(G(x^*))$ . A3 Second-order sufficient condition.

**A3.1** (Second-order sufficient condition with  $\sigma$ -term) Suppose that  $(x^*, \mu^*, Y^*) \in \mathcal{R}^n \times \mathcal{R}^l \times \mathcal{S}^m$  is a KKT triple of the problem (1), i.e.,

(4) 
$$\nabla f(x^*) - Dh(x^*)^T \mu^* - DG(x^*)^* Y^* = 0,$$

(5) 
$$h(x^*) = 0, \ G(x^*) \succeq 0, \ Y^* \succeq 0, \ \langle Y^*, G(x^*) \rangle = 0$$

where  $DG(x)^*$  denotes the adjoint operator of the linear operator DG(x) and satisfies

$$DG(x)^*Y = \left(\langle Y, \frac{\partial G(x)}{\partial x_1} \rangle, \langle Y, \frac{\partial G(x)}{\partial x_2} \rangle, \cdots, \langle Y, \frac{\partial G(x)}{\partial x_n} \rangle\right)^T, \ Y \in \mathcal{S}^m.$$

Moreover,

(6) 
$$d^T \nabla^2_{xx} L(x^*, \mu^*, Y^*) d + \Upsilon_{G(x^*)}(Y^*, DG(x^*)d) > 0 \text{ for all } d \in \mathcal{C}(x^*) \setminus \{0\},$$

where,

(7) 
$$\mathcal{C}(x^*) = \{ d \in \mathcal{R}^n \mid Dh(x^*)d = 0, DG(x^*)d \in T_{\mathcal{S}^m_+}(G(x^*)), \nabla f(x^*)^T d = 0 \}$$
  
is the critical cone, linear-quadratic function  $\Upsilon_B : \mathcal{S}^m \times \mathcal{S}^m \to \mathcal{R}$  is defined by

$$\Upsilon_B(D,A) = 2\langle D, AB^{\dagger}A \rangle, \ (D,A) \in \mathcal{S}^m \times \mathcal{S}^m,$$

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

**A3.2** (Second-order sufficient condition without  $\sigma$ -term) Suppose that  $(x^*, \mu^*, Y^*) \in \mathcal{R}^n \times \mathcal{R}^l \times \mathcal{S}^m$  is a KKT triple of the problem (1) and

(8) 
$$d^T \nabla^2_{xx} L(x^*, \mu^*, Y^*) d > 0 \text{ for all } d \in \mathcal{C}(x^*) \setminus \{0\}.$$

A4 The strict complementarity condition holds at  $x^*$ , i.e.,

$$\operatorname{rank}(G(x^*)) + \operatorname{rank}(Y^*) = m \text{ for all } Y^* \in \Omega^*,$$

where  $\Omega^*$  is the set of the multipliers corresponding to the constraint  $G(x) \succeq 0$  at  $x^*$ .

Fares et al. ([8]) and Correa and Ramirez ([3]) give the following result about the local convergence of Algorithm 2.1.

**Theorem 2.1.** ([3, 8]) Suppose that the assumptions A1 and A3.2 hold. Suppose that  $(Dh(x^*)^T, DG(x^*))$  has full rank and that

$$W_k - \nabla^2_{xx} L(x^*, \mu^*, Y^*) = o(1)$$

Then there is  $\delta > 0$  such that if  $||x_0 - x^*|| < \delta$ ,  $||\mu_0 - \mu^*|| < \delta$  and  $||Y_0 - Y^*|| < \delta$ , Algorithm 2.1 is well defined and the sequence  $\{(x_k, \mu_k, Y_k)\}$  generated by it converges superlinearly to  $(x^*, \mu^*, Y^*)$ . The convergence rate is quadratic especially when

(9) 
$$W_k - \nabla_{xx}^2 L(x^*, \mu^*, Y^*) = O(\|x_k - x^*\| + \|(\mu_k, Y_k) - (\mu^*, Y^*)\|)$$

and the second derivatives of f, h, G are locally Lipschitz continuous at  $x^*$ .

Though the sequence generated by Algorithm 2.1 under the assumptions of Theorem 2.1 is convergent quadratically, we need to suppose that  $(Dh(x^*)^T, DG(x^*))$ has full rank and that the assumption A3.2 holds. Zhao et al. ([23]) analyze the local convergence rate under the nondegeneracy condition and the second-order sufficient condition with  $\sigma$ -term.

**Theorem 2.2.** ([23]) Suppose that the assumptions A1, A2, A3.1 and A4 hold and that the second derivatives of f, h, G are locally Lipschitz continuous at  $x^*$ . The sequence  $\{(x_k, \mu_k, Y_k)\}$  is generated by Algorithm 2.1. Suppose that the matrix  $W_k$ is uniformly positive definite, uniformly bounded and satisfies

(10) 
$$(W_k - \nabla_{xx}^2 L(x^*, \mu^*, Y^*))d_k = o(||d_k||).$$

If the sequence  $\{x_k\}$  converges to  $x^*$ , then the sequence  $\{(x_k, \mu_k, Y_k)\}$  converges to  $(x^*, \mu^*, Y^*)$  superlinearly.

It's worth noting that it is not necessary to get the convergence rate of the sequence with multiplier term. Moreover, quadratic or superlinear convergence of the sequence  $\{(x_k, \mu_k, Y_k)\}$  above is not equivalent to the traditional one of the sequence  $\{x_k\}$ . Now we discuss the convergence rate of  $\{x_k\}$ . The following lemma is required.

**Lemma 2.3.** Suppose that the assumptions A1, A2, A3.1 hold and that the second derivatives of f, h, G are locally Lipschitz continuous at  $x^*$ . The sequence  $\{(x_k, \mu_k, Y_k)\}$  is generated by Algorithm 2.1. Suppose that the matrix  $W_k$  is uniformly positive definite, uniformly bounded and satisfies (9). Then there is  $\delta > 0$  such that if  $||x_k - x^*|| < \delta$ ,  $||\mu_k - \mu^*|| < \delta$  and  $||Y_k - Y^*|| < \delta$ , then

$$||d_k|| + ||\mu_{k+1} - \mu^*|| + ||Y_{k+1} - Y^*|| = O(||x_k - x^*|| + ||(\mu_k, Y_k) - (\mu^*, Y^*)||).$$

*Proof.* The proof is finished by combining Theorem 3.2 in [19] with Lemma 3.8 in [23].  $\Box$ 

**Theorem 2.4.** Suppose that the assumptions A1, A2 and A3.1 hold and that the second derivatives of f, h, G are locally Lipschitz continuous at  $x^*$ . The sequence  $\{(x_k, \mu_k, Y_k)\}$  is generated by Algorithm 2.1. Suppose that the matrix  $W_k$  is uniformly positive definite, uniformly bounded and satisfies (9). Then there is  $\delta > 0$ such that if  $||x_0 - x^*|| < \delta$ ,  $||\mu_0 - \mu^*|| < \delta$  and  $||Y_0 - Y^*|| < \delta$ , then Algorithm 2.1 is well defined and the sequence  $\{(x_k, \mu_k, Y_k)\}$  converges quadratically to  $(x^*, \mu^*, Y^*)$ and the sequence  $\{x_k\}$  converges to  $x^*$  superlinearly.

Proof. By Lemma 2.3 and

$$\frac{1}{2}(a+b)^2 \le a^2 + b^2 \le (a+b)^2, \ \forall a, b \ge 0, \ a, b \in \mathcal{R},$$

we have that

$$\begin{aligned} \|d_k\| &= O(\epsilon_k), \ \|Y_{k+1} - Y^*\| = O(\epsilon_k), \end{aligned}$$
  
where  $\epsilon_k = \|(x_k, \mu_k, Y_k) - (x^*, \mu^*, Y^*)\|.$  By the assumption A1,  
 $0 &= h(x_k) + Dh(x_k)d_k \\ &= h(x^*) + Dh(x^*)(x_k - x^*) + O(\|x_k - x^*\|^2) \\ &+ Dh(x^*)d_k + O(\|x_k - x^*\|\|d_k\|) \end{aligned}$   
(11)  $= Dh(x^*)(x_{k+1} - x^*) + O(\epsilon_k^2).$ 

By the definition and the property of the projection operator  $\Pi_{\mathcal{S}^m_+}(\cdot)$  ([17]), that is, let  $A \in \mathcal{S}^m$  has the following spectral decomposition,

$$A = Q\Lambda Q^T,$$

where  $\Lambda$  is the diagonal matrix of eigenvalues of A and Q is a corresponding orthogonal matrix of orthonormal eigenvectors. Then

$$\Pi_{\mathcal{S}^m_{\perp}}(A) = Q\Lambda_+ Q^T$$

where  $\Lambda_+$  is the diagonal matrix whose diagonal entries are the nonnegative parts of the respective diagonal entries of  $\Lambda$ . Thus, the complementarity condition of (1) and one of (2) can be reformulated as

(12) 
$$\Pi_{\mathcal{S}^m_{\perp}}(G(x^*) - Y^*) = G(x^*)$$

and

(13) 
$$\Pi_{\mathcal{S}^m_+}(G(x_k) + DG(x_k)d_k - Y_{k+1}) = G(x_k) + DG(x_k)d_k$$

respectively. Since the projection operator  $\Pi_{\mathcal{S}^m_+}(\cdot)$  is strongly semi-smooth, it follows that there is an operator  $M^* \in \partial \Pi_{\mathcal{S}^m_+}$   $(G(x^*) - Y^*)$  such that

$$\Pi_{\mathcal{S}^m_+}(G(x^*) - Y^*)$$

$$= \Pi_{\mathcal{S}^m_+}(G(x_k) + DG(x_k)d_k - Y_{k+1}) + M^*(G(x^*) - G(x_k) - DG(x_k)d_k$$

$$(14) + Y_{k+1} - Y^*) + O(||G(x^*) - G(x_k) - DG(x_k)d_k + Y_{k+1} - Y^*||^2).$$

Similar to (11), we have that

(15) 
$$G(x_k) + DG(x_k)d_k - G(x^*) = DG(x^*)(x_{k+1} - x^*) + O(\epsilon_k^2).$$
  
Thus,

(16)  $||G(x^*) - G(x_k) - DG(x_k)d_k + Y_{k+1} - Y^*|| = O(\epsilon_{k+1}) + O(\epsilon_k^2).$ Therefore, it follows from (12)–(16) that (17)  $-(M^* - I)DG(x^*)(x_{k+1} - x^*) + M^*(Y_{k+1} - Y^*) = O(\epsilon_k^2) + O(\epsilon_{k+1}^2).$ 

$$(W_k - \nabla_{xx}^2 L(x^*, \mu^*, Y^*))d_k = O(||d_k||\epsilon_k) = O(\epsilon_k^2).$$

Since  $(d_k, \mu_{k+1}, Y_{k+1})$  is a KKT triple of the subproblem (2), we have that

$$0 = \nabla f(x_k) + W_k d_k - Dh(x_k)^T \mu_{k+1} - DG(x_k)^* Y_{k+1}$$
  
=  $\nabla f(x_k) - Dh(x_k)^T \mu^* - DG(x_k)^* Y^* + \nabla^2_{xx} L(x^*, \mu^*, Y^*) d_k + O(\epsilon_k^2)$   
 $-Dh(x_k)^T (\mu_{k+1} - \mu^*) - DG(x_k)^* (Y_{k+1} - Y^*)$ 

$$\stackrel{(4)}{=} \nabla_x L(x_k, \mu^*, Y^*) - \nabla_x L(x^*, \mu^*, Y^*) + \nabla_{xx}^2 L(x^*, \mu^*, Y^*) d_k + O(\epsilon_k^2) - Dh(x_k)^T (\mu_{k+1} - \mu^*) - DG(x_k)^* (Y_{k+1} - Y^*) = \nabla_{xx}^2 L(x^*, \mu^*, Y^*) (x_k + d_k - x^*) + O(||x_k - x^*||^2) - Dh(x_k)^T (\mu_{k+1} - \mu^*)$$

$$= \nabla_{xx}^{2} L(x^{*}, \mu^{*}, Y^{*})(x_{k} + d_{k} - x^{*}) + O(||x_{k} - x^{*}||^{2}) - Dh(x_{k})^{*}(\mu_{k+1} - \mu^{*}) -DG(x_{k})^{*}(Y_{k+1} - Y^{*}) + O(\epsilon_{k}^{2}).$$

Noting that

$$Dh(x_k)^T(\mu_{k+1} - \mu^*) + DG(x_k)^*(Y_{k+1} - Y^*)$$
  
=  $Dh(x^*)^T(\mu_{k+1} - \mu^*) + DG(x^*)^*(Y_{k+1} - Y^*) + O(\epsilon_k \epsilon_{k+1}).$ 

Therefore,

(18)  

$$\nabla_{xx}^{2}L(x^{*},\mu^{*},Y^{*})(x_{k+1}-x^{*}) - Dh(x^{*})^{T}(\mu_{k+1}-\mu^{*}) - DG(x^{*})^{*}(Y_{k+1}-Y^{*}) = O(\epsilon_{k}^{2}) + O(\epsilon_{k}\epsilon_{k+1}).$$

By (18), (11) and (17), we have that

$$\begin{pmatrix} \nabla^2_{xx}L(x^*,\mu^*,Y^*) & -Dh(x^*)^T & -DG(x^*)^* \\ -Dh(x^*) & O & O \\ -(M^*-I)DG(x^*) & O & M^* \end{pmatrix} \begin{pmatrix} x_{k+1}-x^* \\ \mu_{k+1}-\mu^* \\ Y_{k+1}-Y^* \end{pmatrix}$$
  
9) =  $O(\epsilon_k^2) + o(\epsilon_{k+1}).$ 

It follows from the assumptions A2, which means that  $\Omega^*$  is a singleton, i.e.,  $\Omega^* = \{Y^*\}$ , and from A3.1 that the left operator of (19) at  $x^*$  is nonsingular (see [17], Proposition 3.2). So  $\epsilon_{k+1} = O(\epsilon_k^2)$ , i.e., the sequence  $\{(x_k, \mu_k, Y_k)\}$  converges to  $(x^*, \mu^*, Y^*)$  quadratically.

Now we will prove that the sequence  $\{x_k\}$  converges to  $x^*$  superlinearly. If  $\epsilon_{k+1}$  satisfies

(20) 
$$\epsilon_{k+1} = O(||x_k - x^*||\epsilon_k),$$

then

(1

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \frac{O(\epsilon_{k+1})}{\|x_k - x^*\|} = O(\epsilon_k) \to 0$$

i.e., the sequence  $\{x_k\}$  converges to  $x^*$  superlinearly. Therefore, it is enough to prove that (20) holds.

Since 
$$||x_k - x^*|| = O(\epsilon_k)$$
 and  $||d_k|| = O(\epsilon_k)$ , similar to (11) and (15), we have  

$$0 = h(x^*) + Dh(x^*)(x_k - x^*) + O(||x_k - x^*||^2) + Dh(x^*)d_k + O(||x_k - x^*||d_k||)$$

$$= h(x^*) + Dh(x^*)(x_k - x^*) + Dh(x^*)d_k + O(||x_k - x^*||\epsilon_k)$$
(21)
$$= Dh(x^*)(x_{k+1} - x^*) + O(||x_k - x^*||\epsilon_k).$$

and

$$G(x_k) + DG(x_k)d_k - G(x^*) = DG(x^*)(x_{k+1} - x^*) + O(||x_k - x^*||\epsilon_k).$$

Therefore,

$$||G(x^*) - G(x_k) - DG(x_k)d_k + Y_{k+1} - Y^*|| = O(||x_k - x^*||\epsilon_k) + O(\epsilon_{k+1}).$$

Similar to (17), we have that

(22) 
$$-(M^* - I)DG(x^*)(x_{k+1} - x^*) + M^*(Y_{k+1} - Y^*) = O(||x_k - x^*||\epsilon_k) + O(\epsilon_{k+1}^2)$$
  
It follows from (9) that

It follows from (9) that

$$(W_k - \nabla_{xx}^2 L(x^*, \mu^*, Y^*))d_k$$
  
=  $O(||d_k||\epsilon_k)$   $(d_k = x_{k+1} - x^* + x^* - x_k)$   
=  $O(||x_k - x^*||\epsilon_k) + O(||x_{k+1} - x^*||\epsilon_k)$   
=  $O(||x_k - x^*||\epsilon_k) + O(\epsilon_{k+1}).$ 

Since  $(d_k, \mu_{k+1}, Y_{k+1})$  is a KKT triple of the subproblem (2), we have that

(23)  

$$\nabla_{xx}^{2}L(x^{*},\mu^{*},Y^{*})(x_{k+1}-x^{*}) \\
-Dh(x^{*})^{T}(\mu_{k+1}-\mu^{*}) - DG(x^{*})^{*}(Y_{k+1}-Y^{*}) \\
= O(||x_{k}-x^{*}||\epsilon_{k}) + o(\epsilon_{k+1}).$$

By (23), (21) and (22),

$$\begin{pmatrix} \nabla_{xx}^2 L(x^*, \mu^*, Y^*) & -Dh(x^*)^T & -DG(x^*)^* \\ -Dh(x^*) & O & O \\ -(M^* - I)DG(x^*) & O & M^* \end{pmatrix} \begin{pmatrix} x_{k+1} - x^* \\ \mu_{k+1} - \mu^* \\ Y_{k+1} - Y^* \end{pmatrix}$$

$$(24) = O(||x_k - x^*||\epsilon_k) + o(\epsilon_{k+1}).$$

Since the left operator of (24) is nonsingular at  $x^*$ , it follows that  $\epsilon_{k+1} = O(||x_k - C||)$  $x^* \| \epsilon_k$ ). Thus, the result is proved. 

Under the assumptions of Theorem 2.2 or Theorem 2.4, the sequence  $\{(x_k, \mu_k, Y_k)\}$ converges superlinearly when  $W_k$  is an approximate symmetric matrix of the Hessian matrix of the Lagrangian function. However, the Hessian matrix may not be positive definite (even negative definite) even if the second-order sufficient condition with  $\sigma$ -term holds. In this case, it is unsuitable to approximate the Hessian matrix by a positive definite matrix. Actually, Diehl et al. ([4]) show that the classic SQP-type method under the assumptions A1-A3.1 may only have a linear rate of convergence if the assumption A3.2 does not hold and  $W_k$  is any positive definite and bounded matrix. Therefore, we consider an equivalent and reduced problem near the optimal point where the Hessian matrix of the Lagrangian function of the reduced problem is always positive definite on its critical cone, i.e., the assumption A3.2 holds.

#### SUPERLINEAR CONVERGENCE OF AN SQP-TYPE METHOD

# 3. Local convergence of a reduced SQP-type method

Motivated by Dorsch et al ([5]), we assume, without loss of generality, that G(x) has the following form near the optimal solution  $x^*$  of (1).

$$G(x) = \begin{pmatrix} A(x) & B(x) \\ B(x)^T & C(x) \end{pmatrix},$$

where  $A(x^*)$  is nonsingular and  $\operatorname{rank}(G(x^*)) = \operatorname{rank}(A(x^*)) = r$ . In fact, this matrix partition holds after a reordering of the variables or, equivalently, by considering the matrix  $P^T G(x)P$  instead of G(x), where P is a permutation matrix. By continuity, there is a neighbourhood  $\mathcal{N}_1(x^*)$  of  $x^*$  such that  $\operatorname{rank}(A(x)) = r$  for all  $x \in \mathcal{N}_1(x^*)$ . Let

$$S(x) = C(x) - B(x)^T A(x)^{-1} B(x) \in \mathcal{S}^{m-r}, \ x \in \mathcal{N}_1(x^*).$$

Then, for all  $x \in \mathcal{N}_1(x^*)$ , we have that

$$G(x) = \begin{pmatrix} I_r & A(x)^{-1}B(x) \\ O & I_{m-r} \end{pmatrix}^T \begin{pmatrix} A(x) & O \\ O & S(x) \end{pmatrix} \begin{pmatrix} I_r & A(x)^{-1}B(x) \\ O & I_{m-r} \end{pmatrix},$$

where  $I_r$  and  $I_{m-r}$  are r-order and (m-r)-order identity matrices, respectively. Therefore, for all  $x \in \mathcal{N}_1(x^*)$ ,  $G(x) \succeq 0$  if and only if  $S(x) \succeq 0$ . Thus, we get an equivalent problem of (1) near the optimal point  $x^*$ 

(25) 
$$\min_{\substack{x \in X, \\ x \in Y, \\$$

We refer to the equivalent problem (25) as a reduced problem of (1). Denote the Lagrangian function of the reduced problem (25) by

(26) 
$$\tilde{L}(x,\tilde{\mu},\tilde{Y}) = f(x) - \tilde{\mu}^T h(x) - \langle \tilde{Y}, S(x) \rangle,$$

where  $\tilde{\mu} \in \mathcal{R}^l$ ,  $\tilde{Y} \in \mathcal{S}^{m-r}$ . Under certain constraint qualifications there is a multiplier pair  $(\tilde{\mu}^*, \tilde{Y}^*) \in \mathcal{R}^l \times \mathcal{S}^{m-r}$  associated with the optimal point  $x^*$  of (1) such that

(27) 
$$\nabla f(x^*) - Dh(x^*)^T \tilde{\mu}^* - DS(x^*)^* \tilde{Y}^* = 0,$$

(28) 
$$h(x^*) = 0, \ S(x^*) \succeq 0, \ \tilde{Y}^* \succeq 0, \ \langle \tilde{Y}^*, S(x^*) \rangle = 0$$

At the current iterate  $x_k$ , we have the reduced quadratic semidefinite programming subproblem as follows

(29) 
$$\min \quad \nabla f(x_k)^T d + \frac{1}{2} d^T \tilde{W}_k d$$
  
s.t. 
$$h(x_k) + Dh(x_k) d = 0,$$
  
$$S(x_k) + DS(x_k) d \succeq 0,$$

where  $\tilde{W}_k$  is the Hessian matrix of the Lagrangian function of (25) or its approximate symmetric matrix. It follows from the equation (28) in [5] that, for all  $d \in \mathcal{R}^n$ ,

$$DS(x_k)d = \begin{pmatrix} -A(x_k)^{-1}B(x_k) \\ I_{m-r} \end{pmatrix}^T DG(x_k)d \begin{pmatrix} -A(x_k)^{-1}B(x_k) \\ I_{m-r} \end{pmatrix}.$$

Denote the solution of (29) by  $\tilde{d}_k$  and the corresponding Lagrange multiplier by  $(\tilde{\mu}_{k+1}, \tilde{Y}_{k+1})$ . Set  $x_{k+1} = x_k + \tilde{d}_k$ . Thus, we obtain the following reduced sequential semidefinite programming (RSSDP) algorithm conceptually.

Algorithm 3.1. Local RSSDP Algorithm.

Initialization. Given an initial point  $(x_0, \tilde{\mu}_0, \tilde{Y}_0) \in \mathcal{R}^n \times \mathcal{R}^l \times \mathcal{S}^{m-r}_+, k := 0.$ 

Step 1. Solve (29) to get  $\tilde{d}_k$  and the corresponding Lagrange multiplier  $(\tilde{\mu}_{k+1}, \tilde{Y}_{k+1})$ . Step 2. If  $\tilde{d}_k = 0$ , stop.

Step 3. Set  $x_{k+1} = x_k + \tilde{d}_k$ .

Step 4. Set k := k + 1, go to Step 1.

Let the sequence  $\{(x_k, \tilde{\mu}_k, \tilde{Y}_k)\}$  be generated by Algorithm 3.1. By the definition of S(x), the  $\sigma$ -term in the second-order sufficient condition of the reduced problem (25) vanishes. By Lemma 5.a in [5], the multiplier pairs of the primal problem (1) and the reduced problem (25) are both determined uniquely when the assumptions A1 and A2 hold. Furthermore, it follows from Lemma 4 in [5] that the multiplier pairs associated with the optimal point  $x^*$  of the above two problems have the following relations

$$\tilde{\mu}^* = \mu^*, \ \tilde{Y}^* = \begin{pmatrix} -A(x^*)^{-1}B(x^*) \\ I_{m-r} \end{pmatrix}^T Y^* \begin{pmatrix} -A(x^*)^{-1}B(x^*) \\ I_{m-r} \end{pmatrix}.$$

**Lemma 3.1.** Suppose that the assumptions A1, A2 and A3.1 hold. The critical cone  $C(x^*)$  of the primal problem (1) at the optimal point  $x^*$  is equal to that of the reduced problem (25) and

$$d^{T} \nabla^{2}_{xx} L(x^{*}, \mu^{*}, Y^{*}) d + \Upsilon_{G(x^{*})}(Y^{*}, DG(x^{*})d) = d^{T} \nabla^{2}_{xx} \tilde{L}(x^{*}, \tilde{\mu}^{*}, \tilde{Y}^{*}) d$$

for all  $d \in \mathcal{C}(x^*)$ , where  $(\mu^*, Y^*)$  and  $(\tilde{\mu}^*, \tilde{Y}^*)$  are the Lagrange multipliers of (1) and (25) associated with  $x^*$ , respectively.

Proof. We first prove the following conclusion  $\begin{aligned} \mathcal{C}(x^*) &= \tilde{\mathcal{C}}(x^*) \triangleq \{ d \in \mathcal{R}^n \mid Dh(x^*)d = 0, \ DS(x^*)d \in T_{\mathcal{S}^{m-r}_+}(S(x^*)), \ \nabla f(x^*)^T d = 0 \}. \end{aligned}$ Denote  $G(x^*)$  by  $G^*$ . Let  $\mathcal{N}_2(G^*) = \{ X \in \mathcal{S}^m \mid X = G(x), \ x \in \mathcal{N}_1(x^*) \}. \end{aligned}$  Define  $F(X) = \begin{pmatrix} -((I_r, O)X(I_r, O)^T)^{-1}((I_r, O)X(O, I_{m-r})^T) \\ I_{m-r} \end{pmatrix}, \ X \in \mathcal{N}_2(G^*). \end{aligned}$ It is obvious that, for all  $x \in \mathcal{N}_1(x^*),$ 

$$(I_r, O)G(x)(I_r, O)^T = A(x), \ (I_r, O)G(x)(O, I_{m-r})^T = B(x),$$
$$F(G(x)) = \begin{pmatrix} -A(x)^{-1}B(x) \\ I_{m-r} \end{pmatrix},$$
$$F(G(x))^T DG(x)dF(G(x)) = DS(x)d, \ for \ all \ d \in \mathcal{R}^n.$$

Since  $F(\cdot)$  is twice continuously differentiable on  $\mathcal{N}_2(G^*)$ , so is  $F(G(\cdot))$  on  $\mathcal{N}_1(x^*)$ . Moreover, the columns of  $F(G(x^*))$  span the kernel space of  $G(x^*)$  and the kernel

space of 
$$S(x^*)$$
 is  $S^{m-r}$ . By the definition of tangent cone,  

$$\{d \in \mathcal{R}^n \mid DG(x^*)d \in T_{\mathcal{S}^m_+}(G(x^*))\}$$

$$= \{d \in \mathcal{R}^n \mid F(G(x^*))^T DG(x^*)dF(G(x^*)) \succeq 0\}$$

$$= \{d \in \mathcal{R}^n \mid DS(x^*)d \succeq 0\}$$

$$= \{ d \in \mathcal{R}^n \mid DS(x^*)d \in T_{\mathcal{S}^{m-r}_+}(S(x^*)) \}.$$

Therefore,  $\mathcal{C}(x^*) = \tilde{\mathcal{C}}(x^*)$ .

Define map  $\phi : \mathcal{N}_2(G^*) \to \mathcal{S}^{m-r}$  satisfying

$$\phi(X) = F(X)^T X F(X).$$

Obviously,  $\phi(\cdot)$  is twice continuously differentiable on  $\mathcal{N}_2(G^*)$  as well. It follows from chain rule and

$$G^*F(G^*) = \begin{pmatrix} O \\ S(x^*) \end{pmatrix} = 0$$

that for all  $Y \in \mathcal{S}^m$ , we have that ((())) - -

-

$$D_X \phi(G^*) Y$$
  
=  $(D_X F(G^*) Y)^T G^* F(G^*) + F(G^*)^T Y F(G^*) + F(G^*)^T G^* (D_X F(G^*) Y)$   
=  $F(G^*)^T Y F(G^*) \in S^{m-r}.$ 

As  $F(G^*)$  has full column rank, for all  $\tilde{Y} \in \mathcal{S}^{m-r}$ , there is a matrix

$$Y_0 = F(G^*)(F(G^*)^T F(G^*))^{-1} \tilde{Y}(F(G^*)^T F(G^*))^{-1} F(G^*)^T$$

such that

$$D_X\phi(G^*)Y_0=\tilde{Y}$$

Therefore,  $D_X \phi(G^*)$  is onto. Next we can prove that  $\phi(\cdot)$  satisfies

$$\mathcal{S}^m_+ \cap \mathcal{N}_2(G^*) = \{ X \in \mathcal{S}^m \mid \phi(X) \in \mathcal{S}^{m-r}_+ \}.$$

In fact, it is obvious that  $\phi(X_0) \in \mathcal{S}_+^{m-r}$  for all  $X_0 \in \mathcal{S}_+^m \cap \mathcal{N}_2(G^*)$ . On the contrary, the domain of  $\phi(\cdot)$  implies that  $X_0 \in \mathcal{N}_2(G^*)$  for all  $X_0 \in \{X \in \mathcal{S}^m \mid \phi(X) \in \mathcal{S}_+^{m-r}\}$ . Hence,  $\phi(X_0) \in \mathcal{S}_+^{m-r}$  is equivalent to  $X_0 \in \mathcal{S}_+^m$ . So  $\mathcal{S}_+^m$  is  $\mathcal{C}^2$  reducible to  $\mathcal{S}^{m-r}_+$  at  $G^*$ . The remainder of the proof follows from (3.272) in [1]. 

Similar to Theorem 2.4, we have the following result.

Theorem 3.2. Suppose that the assumptions A1, A2 and A3.1 hold and that the second derivatives of f, h, G are locally Lipschitz continuous at  $x^*$ . The sequence  $\{(x_k, \tilde{\mu}_k, Y_k)\}$  is generated by Algorithm 3.1. Suppose that the matrix  $\tilde{W}_k$ is uniformly positive definite, uniformly bounded and satisfies

(30) 
$$\tilde{W}_k - \nabla^2_{xx} \tilde{L}(x^*, \tilde{\mu}^*, \tilde{Y}^*) = O(\|x_k - x^*\| + \|(\tilde{\mu}_k, \tilde{Y}_k) - (\tilde{\mu}^*, \tilde{Y}^*)\|).$$

Then there is  $\delta > 0$  such that if  $||x_0 - x^*|| < \delta$ ,  $||\tilde{\mu}_0 - \tilde{\mu}^*|| < \delta$  and  $||\tilde{Y}_0 - \tilde{Y}^*|| < \delta$ , then Algorithm 3.1 is well defined and the sequence  $\{(x_k, \tilde{\mu}_k, \tilde{Y}_k)\}$  converges to  $(x^*, \tilde{\mu}^*, \tilde{Y}^*)$  quadratically and the sequence  $\{x_k\}$  converges to  $x^*$  superlinearly.

*Proof.* It follows from the assumption A1 and the definition of S(x) that S(x) is twice continuously differentiable in some neighbourhood  $\mathcal{N}_1(x^*)$  of  $x^*$ . Since G(x)is locally Lipschitz continuous at  $x^*$ , so is S(x) at  $x^*$ . By Lemma 5.a in [5], the nondegeneracy condition of the reduced problem (25) holds at  $x^*$ . The rest of the proof runs as in Theorem 2.4. 

Under the assumptions of Theorem 3.2, the condition (30) ensures that the whole sequence  $\{(x_k, \tilde{\mu}_k, \tilde{Y}_k)\}$  converges quadratically and then  $\{x_k\}$  converges superlinearly, which is unnecessary for superlinear convergence of the sequence  $\{x_k\}$ . Now we discuss the equivalent condition of superlinear convergence of the sequence  $\{x_k\}$ . We first have the following lemma.

Lemma 3.3. Suppose that the assumptions A1, A2, A3.1 and A4 hold. The sequence  $\{(x_k, \tilde{\mu}_k, Y_k)\}$  is generated by Algorithm 3.1.  $\{\tilde{d}_k\}$  is a solution to the subproblem (29). If the sequence  $\{x_k\}$  converges to  $x^*$ , then

$$S(x_k) + DS(x_k)\dot{d}_k = 0$$

holds for all k sufficiently large.

*Proof.* Suppose, by contradiction, that there exists an infinity index set  $\mathcal{K}_1$  such that

$$S(x_k) + DS(x_k)d_k \in \mathcal{S}^{m-r}_+ \setminus \{0\}, \ \forall k \in \mathcal{K}_1,$$

which implies that

$$\lambda_1(S(x_k) + DS(x_k)d_k) > 0, \ \forall k \in \mathcal{K}_1,$$

where  $\lambda_1(\cdot)$  denotes the maximal eigenvalue function of a square matrix. By the complementarity condition of the subproblem (29), we have that

$$\lambda_1(-Y_{k+1}) = 0, \ \forall k \in \mathcal{K}_1.$$

It follows from the continuous property of the maximal eigenvalue function that

$$\lambda_1(-\tilde{Y}^*) = \lambda_1(\lim_{k \in \mathcal{K}_1} -\tilde{Y}_{k+1}) = 0.$$

By A4 and Lemma 5.b in [5], the strict complementarity condition of the reduced problem (25) holds, which implies by  $S(x^*) = 0$  that  $\tilde{Y}^* \succ 0$ . That is a contradiction. Thus, the result is true.

For convenience, we define an  $n \times \frac{1}{2}(m-r)(m-r+1)$ -order matrix as

$$A_S(x) = \left(\operatorname{svec}\left(\frac{\partial S(x)}{\partial x_1}\right), \operatorname{svec}\left(\frac{\partial S(x)}{\partial x_2}\right), \cdots, \operatorname{svec}\left(\frac{\partial S(x)}{\partial x_n}\right)\right)^T$$

where, for all  $Z = (z_{ij}) \in \mathcal{S}^m$ , the operator  $\operatorname{svec}(\cdot)$  is defined by

$$\operatorname{svec}(Z) = (z_{11}, \sqrt{2}z_{12}, z_{22}, \sqrt{2}z_{13}, \sqrt{2}z_{23}, z_{33}, \cdots, z_{mm})^T \in \mathcal{R}^{\frac{1}{2}m(m+1)}$$

Obviously,  $A_S(x)^T d = \operatorname{svec}(DS(x)d)$  for all  $d \in \mathbb{R}^n$ . It follows from the strict complementarity condition that

(31) 
$$\mathcal{C}(x^*) = \tilde{\mathcal{C}}(x^*) = \{ d \in \mathcal{R}^n \mid Dh(x^*)d = 0, \ A_S(x^*)^T d = 0 \}.$$

Moreover, for all  $\tilde{Y} \in \mathcal{S}^{m-r}$ , we have that

 $(\tilde{\mathbf{x}}_{z})$ 

. . .

$$A_{S}(x)\operatorname{svec}(Y) = \left( \left( \operatorname{svec}\left(\frac{\partial S(x)}{\partial x_{1}}\right) \right)^{T} \operatorname{svec}(\tilde{Y}), \cdots, \left( \operatorname{svec}\left(\frac{\partial S(x)}{\partial x_{n}}\right) \right)^{T} \operatorname{svec}(\tilde{Y}) \right)^{T} = \left( \left\langle \tilde{Y}, \left(\frac{\partial S(x)}{\partial x_{1}}\right) \right\rangle, \cdots, \left\langle \tilde{Y}, \left(\frac{\partial S(x)}{\partial x_{n}}\right) \right\rangle \right)^{T} = DS(x)^{*} \tilde{Y}.$$

Therefore, the nondegeneracy condition of (25) is equal to that  $(Dh(x^*)^T, A_S(x^*))$  having full column rank. Now, we define a matrix near the optimal point  $x^*$  as

$$\tilde{P}(x_k) := I - \left(Dh(x_k)^T, A_S(x_k)\right) \left[ \left( \begin{array}{c} Dh(x_k) \\ A_S(x_k)^T \end{array} \right) \left(Dh(x_k)^T, A_S(x_k)\right) \right]^{-1} \left( \begin{array}{c} Dh(x_k) \\ A_S(x_k)^T \end{array} \right).$$

Clearly,  $\tilde{P}(x_k)$  is an orthogonal projection matrix from  $\mathcal{R}^n$  to  $\operatorname{Ker}((Dh(x_k)^T, A_S(x_k))^T)$ . For abbreviation, we denote  $\tilde{P}(x_k)$  by  $\tilde{P}_k$ .

**Theorem 3.4.** Suppose that the assumptions A1, A2, A3.1 and A4 hold and that the second derivatives of f, h, G are locally Lipschitz continuous at  $x^*$ . The sequence  $\{(x_k, \tilde{\mu}_k, \tilde{Y}_k)\}$  is generated by Algorithm 3.1. If the sequence  $\{x_k\}$  converges to  $x^*$ , then the step  $\tilde{d}_k$  in Algorithm 3.1 is a superlinearly convergent one, i.e.,

$$\lim_{k \to +\infty} \frac{\|x_k + d_k - x^*\|}{\|x_k - x^*\|} = 0$$

if and only if the matrix  $\tilde{W}_k$  satisfies

(32) 
$$\tilde{P}_k(\tilde{W}_k - \nabla_{xx}^2 \tilde{L}(x^*, \tilde{\mu}^*, \tilde{Y}^*))\tilde{d}_k = o(\|\tilde{d}_k\|).$$

*Proof.* By the definition of the projection matrix  $\tilde{P}_k$ ,

$$\begin{split} \tilde{P}_k(Dh(x_k)^T \tilde{\mu}_{k+1} + DS(x_k)^* \tilde{Y}_{k+1}) \\ &= \tilde{P}_k \left( Dh(x_k)^T, A_S(x_k) \right) \left( \begin{array}{c} \tilde{\mu}_{k+1} \\ \operatorname{svec}(\tilde{Y}_{k+1}) \end{array} \right) \\ &= O \\ &= \tilde{P}_k \left( Dh(x_k)^T, A_S(x_k) \right) \left( \begin{array}{c} \tilde{\mu}^* \\ \operatorname{svec}(\tilde{Y}^*) \end{array} \right) \\ &= \tilde{P}_k(Dh(x_k)^T \tilde{\mu}^* + DS(x_k)^* \tilde{Y}^*). \end{split}$$

Since  $(\tilde{d}_k, \tilde{\mu}_{k+1}\tilde{Y}_{k+1})$  is a KKT triple of the subproblem (29), we have that

$$\tilde{W}_k \tilde{d}_k = -\nabla f(x_k) + Dh(x_k)^T \tilde{\mu}_{k+1} + DS(x_k)^* \tilde{Y}_{k+1}.$$

Abbreviate  $\tilde{P}(x^*)$  to  $\tilde{P}^*$ , then,  $\tilde{P}(\tilde{W} = \nabla^2 \tilde{L}(x^*, \tilde{x}^*, \tilde{V}^*))\tilde{L}$ 

$$P_{k}(W_{k} - \nabla_{xx}^{2}L(x^{*}, \tilde{\mu}^{*}, Y^{*}))d_{k}$$

$$= \tilde{P}_{k}(-\nabla f(x_{k}) + Dh(x_{k})^{T}\tilde{\mu}_{k+1} + DS(x_{k})^{*}\tilde{Y}_{k+1} - \nabla_{xx}^{2}\tilde{L}(x^{*}, \tilde{\mu}^{*}, \tilde{Y}^{*})\tilde{d}_{k})$$

$$= -\tilde{P}_{k}(\nabla f(x_{k}) - Dh(x_{k})^{T}\tilde{\mu}^{*} - DS(x_{k})^{*}\tilde{Y}^{*} + \nabla_{xx}^{2}\tilde{L}(x^{*}, \tilde{\mu}^{*}, \tilde{Y}^{*})\tilde{d}_{k})$$

$$= -\tilde{P}_{k}(\nabla_{x}\tilde{L}(x_{k}, \tilde{\mu}^{*}, \tilde{Y}^{*}) - \nabla_{x}\tilde{L}(x^{*}, \tilde{\mu}^{*}, \tilde{Y}^{*}) + \nabla_{xx}^{2}\tilde{L}(x^{*}, \tilde{\mu}^{*}, \tilde{Y}^{*})\tilde{d}_{k})$$

$$(33) = -\tilde{P}_{k}\nabla_{xx}^{2}\tilde{L}(x^{*}, \tilde{\mu}^{*}, \tilde{Y}^{*})(x_{k} + \tilde{d}_{k} - x^{*}) + o(||x_{k} - x^{*}||).$$

Again by  $P(x^*) = P(x_k) + o(1)$  that

$$\tilde{P}_{k}(\tilde{W}_{k} - \nabla^{2}_{xx}\tilde{L}(x^{*}, \tilde{\mu}^{*}, \tilde{Y}^{*}))\tilde{d}_{k} = -\tilde{P}^{*}\nabla^{2}_{xx}\tilde{L}(x^{*}, \tilde{\mu}^{*}, \tilde{Y}^{*})(x_{k} + \tilde{d}_{k} - x^{*}) + o(\|x_{k+1} - x^{*}\|) + o(\|x_{k} - x^{*}\|).$$

$$(34)$$

By (11),

(35) 
$$-Dh(x^*)(x_{k+1} - x^*) = o(||x_k - x^*||)$$

Similarly, by Lemma 3.3, it holds for k sufficiently large that

(36) 
$$A_S(x^*)^T(x_{k+1} - x^*) = o(||x_k - x^*||).$$

By (34), (35) and (36),

(37)  
$$\begin{pmatrix} \tilde{P}^* \nabla^2_{xx} \tilde{L}(x^*, \tilde{\mu}^*, \tilde{Y}^*) \\ Dh(x^*) \\ A_S(x^*)^T \end{pmatrix} (x_{k+1} - x^*) \\ = \begin{pmatrix} -\tilde{P}_k(\tilde{W}_k - \nabla^2_{xx} \tilde{L}(x^*, \tilde{\mu}^*, \tilde{Y}^*)) \tilde{d}_k \\ O \\ O \end{pmatrix} \\ +o(\|x_{k+1} - x^*\|) + o(\|x_k - x^*\|).$$

We can prove that the left coefficient matrix in (37) has full column rank. Suppose, by contraction, that there exists a nonzero vector  $d \in \mathbb{R}^n$  such that

$$\tilde{P}^* \nabla^2_{xx} \tilde{L}(x^*, \tilde{\mu}^*, \tilde{Y}^*) d = 0, \quad (Dh(x^*)^T, A_S(x^*))^T d = 0.$$

By the second part of the equation above,  $\tilde{P}^*d = d$ . Hence,

$$d^{T} \nabla^{2}_{xx} \tilde{L}(x^{*}, \tilde{\mu}^{*}, \tilde{Y}^{*}) d = d^{T} \tilde{P}^{*} \nabla^{2}_{xx} \tilde{L}(x^{*}, \tilde{\mu}^{*}, \tilde{Y}^{*}) d = 0.$$

By Lemma 3.1 and (31),  $\nabla^2_{xx} \tilde{L}(x^*, \tilde{\mu}^*, \tilde{Y}^*)$  is positive definite on  $\operatorname{Ker}(Dh(x^*)^T, A_S(x^*)) \setminus \{0\}$ . Thus, d = 0, which is a contradiction.

If the sequence  $\{x_k\}$  converges to  $x^*$  superlinearly, then

$$\lim_{k \to +\infty} \|\tilde{d}_k\| / \|x_k - x^*\| = 1$$

By (33),

 $\tilde{P}_k(\tilde{W}_k - \nabla_{xx}^2 \tilde{L}(x^*, \tilde{\mu}^*, \tilde{Y}^*))\tilde{d}_k = O(\|x_{k+1} - x^*\|) + o(\|x_k - x^*\|)$ (38)and thus

$$\frac{\|\tilde{P}_{k}(\tilde{W}_{k}-\nabla_{xx}^{2}\tilde{L}(x^{*},\tilde{\mu}^{*},\tilde{Y}^{*}))\tilde{d}_{k}\|}{\|\tilde{d}_{k}\|} = \frac{\|\tilde{P}_{k}(\tilde{W}_{k}-\nabla_{xx}^{2}\tilde{L}(x^{*},\tilde{\mu}^{*},\tilde{Y}^{*}))\tilde{d}_{k}\|}{\|x_{k}-x^{*}\|} \cdot \frac{\|x_{k}-x^{*}\|}{\|\tilde{d}_{k}\|} \\ \stackrel{(38)}{=} \left(\frac{O(\|x_{k+1}-x^{*}\|)}{\|x_{k}-x^{*}\|} + \frac{O(\|x_{k}-x^{*}\|)}{\|x_{k}-x^{*}\|}\right) \frac{\|x_{k}-x^{*}\|}{\|\tilde{d}_{k}\|} \to 0.$$

i.e., (32) holds.

On the contrary, if (32) holds, then, it follows from (37) and its left coefficient matrix having full column rank that

$$\begin{aligned} \|x_{k+1} - x^*\| &= O(\|\tilde{P}_k(\tilde{W}_k - \nabla^2_{xx}\tilde{L}(x^*, \tilde{\mu}^*, \tilde{Y}^*))\tilde{d}_k\| \\ &+ o(\|x_{k+1} - x^*\|) + o(\|x_k - x^*\|) \\ \stackrel{(32)}{=} o(\|x_{k+1} - x^*\|) + o(\|x_k - x^*\|) + o(\|\tilde{d}_k\|). \end{aligned}$$

By ||d|lt is proved.

The preceding theorem guarantees a sufficient and necessary condition for the step  $d_k$  being a superlinear convergent one, which is also equivalent to the fact that the primal sequence  $\{x_k\}$  generated by Algorithm 3.1 converges superlinearly when the unit step size is always accepted. Inspired by the technology in Section 3, where a conceptual algorithm 3.1 is applied and an unspecific reduced problem (25) is solved, we next analyze the equivalent condition for the primal sequence  $\{x_k\}$  generated by Algorithm 2.1 being superlinearly convergent.

# 4. Superlinear convergence of the SSDP algorithm

In this section, we will construct a projection matrix to obtain an equivalent condition for superlinear convergence of the standard sequential semidefinite programming algorithm.

At the optimal point  $x^*$ , we suppose that  $\operatorname{rank}(G(x^*)) = r$  and that G(x) has a spectral decomposition as follows

$$G(x) = Q(x)^T \operatorname{Diag}(\lambda_1(x), \lambda_2(x), \cdots, \lambda_m(x))Q(x),$$

where  $Q(x) = (q_1(x), q_2(x), \dots, q_m(x))$  is an orthogonal matrix,  $\lambda_1(x), \lambda_2(x), \dots, \lambda_n(x)$  $\lambda_m(x)$  are the eigenvalues of G(x) in decreasing order. Denote that  $Q^{\alpha}(x) =$  $(q_1(x), q_2(x), \cdots, q_r(x)), Q^{\gamma}(x) = (q_{r+1}(x), \cdots, q_m(x)).$ 

Let  $\mathcal{L}(G(x))$  be the linear space generated by the column vectors of  $Q^{\gamma}(x)$ . Then the column vectors of  $Q_*^{\gamma} = Q^{\gamma}(x^*)$  span the linear space  $\mathcal{L}(G(x^*))$ . Denote the orthogonal projection matrix from  $\mathcal{R}^n$  to  $\mathcal{L}(G(x))$  by  $P_{\gamma}(G(x))$ . Note that  $P_{\gamma}(G(x))$  is continuously differentiable in a neighbourhood of  $G_* = G(x^*)$  about G (see Example 3.140 in [1]). Therefore,  $F_{\gamma}(G(x)) = P_{\gamma}(G(x))Q_{\gamma}^{\gamma}$  is also continuously differentiable in a neighbourhood of  $G_* = G(x^*)$  about G. Especially,  $F_{\gamma}(G_*) = Q_*^{\gamma}$ . The rank of  $F_{\gamma}(G(x))$  is equal to m-r for x close to  $x^*$  sufficiently, i.e., its column

vectors are linear independent. Let U(G(x)) be a matrix generated by the column vectors of the matrix  $F_{\gamma}(G(x))$  using Gram-Schmidt orthogonal method. In this case, the matrix U(G(x)) is well defined and is continuously differentiable about Gin a neighbourhood of  $G_*$  and  $U(G(x^*)) = Q_*^{\gamma}$ . Moreover, the linear space  $\mathcal{L}(G(x))$ generated by the column vectors of  $Q^{\gamma}(x)$  coincides with that of U(G(x)), and we have  $U(G(x))^T U(G(x)) = I_{m-r}$ . Thus,  $U(G(x))^T G(x) U(G(x)) \succeq 0$  is equivalent to  $G(x) \succeq 0$  in a neighbourhood of  $x^*$ .

Suppose that the sequence  $\{(x_k, \mu_k, Y_k)\}$  is generated by Algorithm 2.1 and that  $d_k$  is a solution to the subproblem (2). Let

$$\hat{S}(x) = U(G(x))^T G(x) U(G(x))$$

and  $U_k = U(G(x_k)), \hat{S}_k = \hat{S}(x_k), \hat{Y}_k = U_k^T Y_k U_k.$ 

**Lemma 4.1.** Suppose that the assumptions A1, A2 and A3.1 hold.  $\{(x_k, \mu_k, Y_k)\}$  generated by Algorithm 2.1 is an infinite sequence.  $\{d_k\}$  is a solution to (2). If  $\{x_k\}$  converges to  $x^*$ , then

$$DG(x_k)^* Y_k - D\hat{S}(x_k)^* \hat{Y}_k = O(||d_{k-1}||^2).$$

*Proof.* By the definition of  $(\hat{S}_k, \hat{Y}_k)$  and  $U_k^T U_k = I_{m-r}$ ,

$$\left\langle \frac{\partial \hat{S}(x_{k})}{\partial x_{i}}, \hat{Y}_{k} \right\rangle$$

$$= \left\langle \frac{\partial U(G(x_{k}))^{T}G(x_{k})U(G(x_{k}))}{\partial x_{i}}, U_{k}^{T}Y_{k}U_{k} \right\rangle$$

$$= \left\langle \frac{\partial U(G(x_{k}))^{T}}{\partial x_{i}}G(x_{k})U_{k} + U_{k}^{T}\frac{\partial G(x_{k})}{\partial x_{i}}U_{k} + U_{k}^{T}G(x_{k})\frac{\partial U(G(x_{k}))}{\partial x_{i}}, U_{k}^{T}Y_{k}U_{k} \right\rangle$$

$$= \left\langle U_{k}\frac{\partial U(G(x_{k}))^{T}}{\partial x_{i}}G(x_{k}) + \frac{\partial G(x_{k})}{\partial x_{i}} + G(x_{k})\frac{\partial U(G(x_{k}))}{\partial x_{i}}U_{k}^{T}, U_{k}U_{k}^{T}Y_{k}U_{k}U_{k}^{T} \right\rangle$$

$$= \left\langle U_{k}\frac{\partial U(G(x_{k}))^{T}}{\partial x_{i}}G(x_{k}), U_{k}U_{k}^{T}Y_{k}U_{k}U_{k}^{T} \right\rangle + \left\langle \frac{\partial G(x_{k})}{\partial x_{i}}, U_{k}U_{k}^{T}Y_{k}U_{k}U_{k}^{T} \right\rangle$$

$$(39) + \left\langle G(x_{k})\frac{\partial U(G(x_{k}))}{\partial x_{i}}U_{k}^{T}, U_{k}U_{k}^{T}Y_{k}U_{k}U_{k}^{T} \right\rangle.$$

It follows from  $Q_k = (Q_k^{\alpha}, Q_k^{\gamma})$  and  $Q_k Q_k^T = I_m$  that

$$Q_k^{\gamma}(Q_k^{\gamma})^T + Q_k^{\alpha}(Q_k^{\alpha})^T = I_m, \ (Q_k^{\gamma})^T Q_k^{\alpha} = 0.$$

Since both  $U_k$  and  $Q_k^{\gamma}$  have full column rank and the linear spaces spanned by them respectively coincide with each other,  $U_k^T Q_k^{\gamma}$  is invertible and  $U_k^T Q_k^{\alpha} = 0$ . Therefore,

$$(U_k^T Q_k^{\gamma})^{-1} U_k^T = (U_k^T Q_k^{\gamma})^{-1} U_k^T Q_k^{\gamma} (Q_k^{\gamma})^T + (U_k^T Q_k^{\gamma})^{-1} U_k^T Q_k^{\alpha} (Q_k^{\alpha})^T$$

$$(40) = (Q_k^{\gamma})^T.$$

Similarly, let  $U_k^{\perp}$  be the matrix constructed by the standard orthogonal basis of the orthogonal complement space of the linear space  $\mathcal{L}(G(x_k))$ . Then,

$$Q_k^{\gamma} (U_k^T Q_k^{\gamma})^{-1} = (U_k U_k^T + U_k^{\perp} (U_k^{\perp})^T) Q_k^{\gamma} (U_k^T Q_k^{\gamma})^{-1} = U_k.$$

Therefore,

(41) 
$$U_k U_k^T = Q_k^{\gamma} (U_k^T Q_k^{\gamma})^{-1} U_k^T \stackrel{(40)}{=} Q_k^{\gamma} (Q_k^{\gamma})^T.$$

Set the eigenvalues of  $G(x_{k-1}) + DG(x_{k-1})d_{k-1}$  in decreasing order and let  $\tilde{Q}_k^{\gamma}$  be the matrix whose column vectors are the last m-r eigenvectors. By the definition of the orthogonal projection operator  $P_{\gamma}(\cdot)$  (also see Example 3.140 in [1]),

$$P_{\gamma}(G(x_{k-1}) + DG(x_{k-1})d_{k-1}) = \tilde{Q}_k^{\gamma}(\tilde{Q}_k^{\gamma})^T.$$

Since  $P_{\gamma}(G(x_k)) = Q_k^{\gamma}(Q_k^{\gamma})^T$  and  $P_{\gamma}(G)$  is continuously differentiable in a neighbourhood of  $G_*$  about G, we have that

(42) 
$$Q_k^{\gamma}(Q_k^{\gamma})^T = \tilde{Q}_k^{\gamma}(\tilde{Q}_k^{\gamma})^T + O(\|G(x_k) - G(x_{k-1}) - DG(x_{k-1})d_{k-1}\|)$$
$$= \tilde{Q}_k^{\gamma}(\tilde{Q}_k^{\gamma})^T + O(\|d_{k-1}\|^2).$$

By  $\langle Y_k, G(x_{k-1}) + DG(x_{k-1})d_{k-1} \rangle = 0$ ,

(43) 
$$Y_k = \tilde{Q}_k^{\gamma} \tilde{\Gamma}_k (\tilde{Q}_k^{\gamma})^T,$$

where  $\tilde{\Gamma}_k$  is a diagonal matrix formed by the eigenvalues of  $Y_k$ . Therefore, it follows from (41), (42) and (43) that

$$U_{k}U_{k}^{T}Y_{k}U_{k}U_{k}^{T} = Q_{k}^{\gamma}(Q_{k}^{\gamma})^{T}Y_{k}Q_{k}^{\gamma}(Q_{k}^{\gamma})^{T} 
= \tilde{Q}_{k}^{\gamma}(\tilde{Q}_{k}^{\gamma})^{T}Y_{k}\tilde{Q}_{k}^{\gamma}(\tilde{Q}_{k}^{\gamma})^{T} + O(\|d_{k-1}\|^{2}) 
= \tilde{Q}_{k}^{\gamma}(\tilde{Q}_{k}^{\gamma})^{T}\tilde{Q}_{k}^{\gamma}\tilde{\Gamma}_{k}(\tilde{Q}_{k}^{\gamma})^{T}\tilde{Q}_{k}^{\gamma}(\tilde{Q}_{k}^{\gamma})^{T} + O(\|d_{k-1}\|^{2}) 
= \tilde{Q}_{k}^{\gamma}\tilde{\Gamma}_{k}(\tilde{Q}_{k}^{\gamma})^{T} + O(\|d_{k-1}\|^{2}) 
= Y_{k} + O(\|d_{k-1}\|^{2}).$$
(44)

By the complementarity condition of the subproblem (2),

$$Y_k G(x_k) = Y_k (G(x_{k-1}) + DG(x_{k-1})d_{k-1}) + O(||d_{k-1}||^2) = O(||d_{k-1}||^2).$$

Hence,

(45) 
$$\langle U_k \frac{\partial U(G(x_k))^T}{\partial x_i} G(x_k), Y_k \rangle = \langle G(x_k) \frac{\partial U(G(x_k))}{\partial x_i} U_k^T, Y_k \rangle = O(||d_{k-1}||^2).$$

It follows from (39), (44) and (45) that

$$(D\hat{S}(x_k)^*\hat{Y}_k)_i$$

$$= \langle U_k \frac{\partial U(G(x_k))^T}{\partial x_i} G(x_k), Y_k \rangle + \langle \frac{\partial G(x_k)}{\partial x_i}, Y_k \rangle$$

$$+ \langle G(x_k) \frac{\partial U(G(x_k))}{\partial x_i} U_k^T, Y_k \rangle + O(||d_{k-1}||^2)$$

$$= (DG(x_k)^*Y_k)_i + O(||d_{k-1}||^2),$$

which proves the lemma.

Since  $U^* = Q^{\gamma}_*$  at the optimal point  $x^*$ , we have that

(46) 
$$\hat{S}(x^*) = (Q_*^{\gamma})^T G(x^*) Q_*^{\gamma}, \quad \hat{Y}^* = (Q_*^{\gamma})^T Y^* Q_*^{\gamma}$$

and  $(x^*, \mu^*, \hat{Y}^*)$  is a KKT triple of the following SDP problem

(47) 
$$\begin{array}{l} \min \quad f(x) \\ \text{s.t.} \quad h(x) = 0, \\ \hat{S}(x) \succeq 0. \end{array}$$

Denote the Lagrangian function of (47) by

$$\hat{L}(x,\hat{\mu},\hat{Y}) = f(x) - \hat{\mu}^T h(x) - \langle \hat{Y}, \hat{S}(x) \rangle, \ \hat{\mu} \in \mathcal{R}^l, \ \hat{Y} \in \mathcal{S}^{m-r}.$$

606

### SUPERLINEAR CONVERGENCE OF AN SQP-TYPE METHOD

Then the second-order sufficient condition with  $\sigma$ -term (6) is equivalent to

(48) 
$$d^T \nabla^2_{xx} \hat{L}(x^*, \mu^*, \hat{Y}^*) d > 0, \ \forall d \in \mathcal{C}(x^*) \setminus \{0\},$$

where the critical cone  $\mathcal{C}(x^*)$  is as follows

$$\mathcal{C}(x^*) = \{ d \in \mathcal{R}^n \mid Dh(x^*)d = 0, \ (Q_*^{\gamma})^T DG(x^*) dQ_*^{\gamma} \succeq 0, \ \nabla f(x^*)^T d = 0 \} \\ = \{ d \in \mathcal{R}^n \mid Dh(x^*)d = 0, \ (U^*)^T DG(x^*) dU^* = 0 \}.$$

Denote  $U(G(x)) = (u_1(x), u_2(x), \dots, u_{m-r}(x))$ . For  $1 \le i \le j \le m - r$ , we define

$$v_{ij}(x) = \left(\frac{\partial(u_i(x)^T G(x) u_j(x))}{\partial x_1}, \cdots, \frac{\partial(u_i(x)^T G(x) u_j(x))}{\partial x_n}\right)^T$$

 $V(x_k) = (v_{11}^k, \sqrt{2}v_{12}^k, v_{22}^k, \cdots, \sqrt{2}v_{1,m-r}^k, \cdots, v_{m-r,m-r}^k) \in \mathcal{R}^{n \times \frac{1}{2}(m-r)(m-r+1)},$ where  $v_{ij}^k = v_{ij}(x_k)(1 \le i \le j \le m-r).$ 

**Lemma 4.2.** It holds for all  $d \in \mathbb{R}^n$ ,  $\hat{Y} \in \mathcal{S}^{m-r}$  that

(49) 
$$V(x_k)^T d = \operatorname{svec}(D\hat{S}(x_k)d),$$

(50) 
$$V(x_k)\operatorname{svec}(\hat{Y}) = D\hat{S}(x_k)^*\hat{Y}.$$

*Proof.* For all  $d \in \mathbb{R}^n$ , by the definition of  $V_k$ ,

$$V(x_k)^T d$$

$$= ((v_{11}^k)^T d, \sqrt{2}(v_{12}^k)^T d, (v_{22}^k)^T d, \cdots, \sqrt{2}(v_{1,m-r}^k)^T d, \cdots, (v_{m-r,m-r}^k)^T d)^T$$

$$= \operatorname{svec} \left( \begin{pmatrix} (v_{11}^k)^T d & (v_{12}^k)^T d & \cdots & (v_{1,m-r}^k)^T d \\ \vdots & \vdots & \ddots & \vdots \\ (v_{m-r,1}^k)^T d & (v_{m-r,2}^k)^T d & \cdots & (v_{m-r,m-r}^k)^T d \end{pmatrix} \right).$$

Moreover,

$$(v_{ij}^k)^T d$$

$$= \frac{\partial (u_i(x_k)^T G(x_k) u_j(x_k))}{\partial x_1} d_1 + \dots + \frac{\partial (u_i(x_k)^T G(x_k) u_j(x_k))}{\partial x_n} d_n$$

$$= (D\hat{S}(x_k) d)_{ij}.$$

So (49) holds.

For all 
$$\hat{Y} \in \mathcal{S}^{m-r}$$
, by the definition of  $V(x_k)$ ,  

$$V(x_k)\operatorname{svec}(\hat{Y}) = (v_{11}^k, \sqrt{2}v_{12}^k, v_{22}^k, \cdots, \sqrt{2}v_{1,m-r}^k, \cdots, v_{m-r,m-r}^k)\operatorname{svec}(\hat{Y})$$

$$= \begin{pmatrix} \left(\operatorname{svec}\left(\frac{\partial\left(U(G(x_k))^T G(x_k)U(G(x_k))\right)}{\partial x_1}\right)\right)^T \\ \vdots \\ \left(\operatorname{svec}\left(\frac{\partial\left(U(G(x_k))^T G(x_k)U(G(x_k))\right)}{\partial x_n}\right)\right)^T \end{pmatrix} \operatorname{svec}(\hat{Y})$$

$$= \begin{pmatrix} \left\langle\frac{\partial \hat{S}(x_k)}{\partial x_1}, \hat{Y}\right\rangle & \cdots & \left\langle\frac{\partial \hat{S}(x_k)}{\partial x_n}, \hat{Y}\right\rangle \end{pmatrix}^T$$

$$= D\hat{S}(x_k)^*\hat{Y},$$

which means that (50) holds.

Since the assumption A2 is equivalent to  $(Dh(x^*)^T, V(x^*))$  having full column rank, the critical cone can be further reformed as

(51) 
$$\mathcal{C}(x^*) = \{ d \in \mathcal{R}^n \mid Dh(x^*)d = 0, \ V(x^*)^T d = 0 \}.$$

We next construct a matrix  $P(x_k)$  according to  $(Dh(x_k)^T, V(x_k))$ . Let

$$P(x_k) = I - \left(Dh(x_k)^T, V(x_k)\right) \left[ \left(\begin{array}{c} Dh(x_k) \\ V(x_k)^T \end{array}\right) \left(Dh(x_k)^T, V(x_k)\right) \right]^{-1} \left(\begin{array}{c} Dh(x_k) \\ V(x_k)^T \end{array}\right).$$

It is easy to check that  $P(x_k)$  is an orthogonal projection matrix from  $\mathcal{R}^n$  to the Kernel space  $\operatorname{Ker}((Dh(x_k)^T, V(x_k))^T)$ .

Now, we give an equivalent condition for superlinear convergence of the SSDP algorithm by abbreviating  $P_k = P(x_k)$ .

**Theorem 4.3.** Suppose that the assumptions A1, A2, A3.1 and A4 hold and that the second derivatives of f, h, G are locally Lipschitz continuous at  $x^*$ . The sequence  $\{(x_k, \mu_k, Y_k)\}$  is generated by Algorithm 2.1 and assume that  $x_k$  converges to  $x^*$ , rank $(G(x^*)) = r$ . Then the iterate  $x_k$  is a superlinearly convergent one, i.e.,

$$\lim_{k \to +\infty} \frac{\|x_k + d_k - x^*\|}{\|x_k - x^*\|} = 0$$

if and only if the matrix  $W_k$  satisfies

(52) 
$$P_k(W_k - \nabla_{xx}^2 L(x^*, \mu^*, Y^*))d_k = o(||d_k||).$$

*Proof.* By Lemma 3.1 in [23],  $\lim_{k\to\infty} (d_k, \mu_{k+1}, Y_{k+1}) = (0, \mu^*, Y^*)$ . According to the assumptions, we have that

$$\nabla_x L(x^*, \mu^*, Y^*) = \nabla_x \hat{L}(x^*, \mu^*, \hat{Y}^*) = 0,$$
  
$$\nabla^2_{xx} \hat{L}(x^*, \mu^*, \hat{Y}^*) - \nabla^2_{xx} L(x^*, \mu^*, Y^*) = \nabla^2_{xx} \langle Y^*, G(x^*) \rangle - \nabla^2_{xx} \langle \hat{Y}^*, \hat{S}^* \rangle,$$

where  $\hat{Y}^*$  is defined by (46). By the definition of  $P_k$  and (50), for all  $(\mu, \hat{Y}) \in \mathcal{R}^l \times \mathcal{S}^{m-r}$ ,

(53) 
$$P_k(Dh(x_k)^T \mu + D\hat{S}(x_k)^* \hat{Y}) = P_k(Dh(x_k)^T \mu + V(x_k) \operatorname{svec}(\hat{Y})) = 0.$$

Therefore, it follows from  $(d_k, \mu_{k+1}, Y_{k+1})$  being a KKT triple of (2) that

$$\begin{split} & P_{k}(W_{k} - \nabla_{xx}^{2}L(x^{*}, \mu^{*}, Y^{*}))d_{k} \\ = & P_{k}\left(-\nabla f(x_{k}) + Dh(x_{k})^{T}\mu_{k+1} + DG(x_{k})^{*}Y_{k+1} - \nabla_{xx}^{2}L(x^{*}, \mu^{*}, Y^{*})d_{k}\right) \\ \stackrel{(53)}{=} & P_{k}\left(-\nabla f(x_{k}) + Dh(x_{k})^{T}\mu^{*} + D\hat{S}(x_{k})^{*}\hat{Y}^{*} - \nabla_{xx}^{2}\hat{L}(x^{*}, \mu^{*}, \hat{Y}^{*})d_{k} \\ & + Dh(x_{k})^{T}\mu_{k+1} + DG(x_{k})^{*}Y_{k+1} - \nabla_{xx}^{2}L(x^{*}, \mu^{*}, Y^{*})d_{k} + \nabla_{xx}^{2}\hat{L}(x^{*}, \mu^{*}, \hat{Y}^{*})d_{k}\right) \\ & = -P_{k}\left(\nabla_{x}\hat{L}(x_{k}, \mu^{*}, \hat{Y}^{*}) - \nabla_{x}\hat{L}(x^{*}, \mu^{*}, \hat{Y}^{*}) + \nabla_{xx}^{2}\hat{L}(x^{*}, \mu^{*}, \hat{Y}^{*})d_{k}\right) \\ & + P_{k}\left(Dh(x_{k})^{T}\mu_{k+1} + DG(x_{k})^{*}Y_{k+1} + \left(\nabla_{xx}^{2}\langle Y^{*}, G(x^{*})\rangle - \nabla_{xx}^{2}\langle \hat{Y}^{*}, \hat{S}(x^{*})\rangle\right)d_{k}\right) \\ \stackrel{(53)}{=} P_{k}\left(DG(x_{k})^{*}Y_{k+1} - D\hat{S}(x_{k})^{*}\hat{Y}_{k+1} + \left(\nabla_{xx}^{2}\langle Y^{*}, G(x^{*})\rangle - \nabla_{xx}^{2}\langle \hat{Y}^{*}, \hat{S}(x^{*})\rangle\right)d_{k}\right) \\ (54) \\ & - P_{k}\nabla_{xx}^{2}\hat{L}(x^{*}, \mu^{*}, \hat{Y}^{*})(x_{k+1} - x^{*}) + o(||x_{k} - x^{*}||). \end{split}$$

By the assumption A1 and  $Y_{k+1} - Y^* = o(1)$ , one has

$$\begin{aligned} \nabla_{xx}^{2} \langle Y_{k+1}, G(x_{k}) \rangle &= \nabla_{xx}^{2} \langle Y_{k+1}, G(x^{*}) \rangle + O(\|x_{k} - x^{*}\|) \\ &= \nabla_{xx}^{2} \langle Y^{*}, G(x^{*}) \rangle + O(\|x_{k} - x^{*}\|) + o(1) \\ &= \nabla_{xx}^{2} \langle Y^{*}, G(x^{*}) \rangle + o(1), \end{aligned}$$

which implies that

$$DG(x_{k+1})^*Y_{k+1} = DG(x_k)^*Y_{k+1} + \nabla^2_{xx}\langle Y_{k+1}, G(x_k)\rangle d_k + o(||d_k||) = DG(x_k)^*Y_{k+1} + \nabla^2_{xx}\langle Y^*, G(x^*)\rangle d_k + o(||d_k||), D\hat{S}(x_{k+1})^*\hat{Y}_{k+1} = D\hat{S}(x_k)^*\hat{Y}_{k+1} + \nabla^2_{xx}\langle \hat{Y}^*, \hat{S}^*\rangle d_k + o(||d_k||).$$

Therefore, it follows from Lemma 4.1 that

$$DG(x_k)^* Y_{k+1} - D\hat{S}(x_k)^* \hat{Y}_{k+1} + \left(\nabla^2_{xx} \langle Y^*, G(x^*) \rangle - \nabla^2_{xx} \langle \hat{Y}^*, \hat{S}^* \rangle \right) d_k$$
  
=  $DG(x_{k+1})^* Y_{k+1} - D\hat{S}(x_{k+1})^* \hat{Y}_{k+1} + o(||d_k||)$   
(55) =  $O(||d_k||^2) + o(||d_k||) = o(||d_k||).$ 

Thus, by (54), (55) and  $P(x_k) = P(x^*) + o(1)$ ,

$$P^* \nabla^2_{xx} \hat{L}(x^*, \mu^*, \hat{Y}^*)(x_{k+1} - x^*) = -P_k(W_k - \nabla^2_{xx} L(x^*, \mu^*, Y^*))d_k$$
  
(56) 
$$+ o(\|x_{k+1} - x^*\|) + o(\|x_k - x^*\|) + o(\|d_k\|).$$

By (21),

(57) 
$$Dh(x^*)(x_{k+1} - x^*) = o(||x_k - x^*||)$$

Then, similar to (14), there exists an  $\tilde{M}^* \in \partial \Pi_{\mathcal{S}^m_+} \left( U^* (U^*)^T (G(x^*) - Y^*) U^* (U^*)^T \right)$  such that

$$\Pi_{\mathcal{S}^{m}_{+}} \left( U^{*}(U^{*})^{T} (G(x^{*}) - Y^{*}) U^{*}(U^{*})^{T} \right)$$

$$= \Pi_{\mathcal{S}^{m}_{+}} \left( U^{*}(U^{*})^{T} (G(x_{k}) + DG(x_{k})d_{k} - Y_{k+1}) U^{*}(U^{*})^{T} \right)$$

$$+ \tilde{M}^{*} \left( U^{*}(U^{*})^{T} (G(x^{*}) - G(x_{k}) - DG(x_{k})d_{k} + Y_{k+1} - Y^{*}) U^{*}(U^{*})^{T} \right)$$

$$(58) \qquad + O \left( \| U^{*}(U^{*})^{T} (G(x^{*}) - G(x_{k}) - DG(x_{k})d_{k} + Y_{k+1} - Y^{*}) U^{*}(U^{*})^{T} \|^{2} \right).$$

Moreover,

(59) 
$$\Pi_{\mathcal{S}^m_+} \left( U^* (U^*)^T (G(x^*) - Y^*) U^* (U^*)^T \right) = U^* (U^*)^T G(x^*) U^* (U^*)^T,$$

(60) 
$$\Pi_{\mathcal{S}^{m}_{+}} \left( U^{*} (U^{*})^{T} (G(x_{k}) + DG(x_{k})d_{k} - Y_{k+1})U^{*} (U^{*})^{T} \right)$$
$$= U^{*} (U^{*})^{T} (G(x_{k}) + DG(x_{k})d_{k})U^{*} (U^{*})^{T},$$

(61) 
$$G(x_k) + DG(x_k)d_k - G(x^*) = DG(x^*)(x_{k+1} - x^*) + o(||x_k - x^*||).$$

Since U(G(x)) is continuously differentiable, by (41) and (42), it holds that

$$U^{*}(U^{*})^{T} = U_{k+1}U_{k+1}^{T} + O(||x_{k+1} - x^{*}||)$$
  
=  $Q_{k+1}^{\gamma}(Q_{k+1}^{\gamma})^{T} + O(||x_{k+1} - x^{*}||)$   
=  $\tilde{Q}_{k+1}^{\gamma}(\tilde{Q}_{k+1}^{\gamma})^{T} + O(||x_{k+1} - x^{*}||) + O(||d_{k}||^{2})$ 

and

610

$$U^{*}(U^{*})^{T}(G(x^{*}) - G(x_{k}) - DG(x_{k})d_{k} + Y_{k+1} - Y^{*})U^{*}(U^{*})^{T}$$

$$= U^{*}(U^{*})^{T}(G(x^{*}) - Y^{*})U^{*}(U^{*})^{T}$$

$$+ \tilde{Q}_{k+1}^{\gamma}(\tilde{Q}_{k+1}^{\gamma})^{T}(G(x_{k}) + DG(x_{k})d_{k} - Y_{k+1})\tilde{Q}_{k+1}^{\gamma}(\tilde{Q}_{k+1}^{\gamma})^{T}$$

$$+ O(||x_{k+1} - x^{*}||) + O(||d_{k}||^{2})$$

$$= U^{*}(U^{*})^{T}G(x^{*})U^{*}(U^{*})^{T} - \tilde{Q}_{k+1}^{\gamma}(\tilde{Q}_{k+1}^{\gamma})^{T}(G(x_{k}) + DG(x_{k})d_{k})\tilde{Q}_{k+1}^{\gamma}(\tilde{Q}_{k+1}^{\gamma})^{T}$$

$$+ O(||x_{k+1} - x^{*}||) + O(||d_{k}||^{2})$$

$$= U^{*}(U^{*})^{T}G(x^{*})U^{*}(U^{*})^{T} - U^{*}(U^{*})^{T}(G(x_{k}) + DG(x_{k})d_{k})U^{*}(U^{*})^{T}$$

$$+ O(||x_{k+1} - x^{*}||) + O(||d_{k}||^{2})$$

$$= U^{*}(U^{*})^{T}(G(x^{*}) - G(x_{k}) - DG(x_{k})d_{k})U^{*}(U^{*})^{T} + O(||x_{k+1} - x^{*}||) + O(||d_{k}||^{2})$$

$$(62) = O(||x_{k+1} - x^{*}||) + O(||d_{k}||^{2}).$$

Let

$$H_{k} = U^{*}(U^{*})^{T}(G(x^{*}) - G(x_{k}) - DG(x_{k})d_{k} + Y_{k+1} - Y^{*})U^{*}(U^{*})^{T} \in \mathcal{S}^{m}.$$
  
Then,  $H_{k} = O(||x_{k+1} - x^{*}||) + O(||d_{k}||^{2}).$  From (41) and (58)-(62),  
 $Q_{*}^{\gamma}(Q_{*}^{\gamma})^{T}DG(x^{*})(x_{k+1} - x^{*})Q_{*}^{\gamma}(Q_{*}^{\gamma})^{T} = U^{*}(U^{*})^{T}DG(x^{*})(x_{k+1} - x^{*})U^{*}(U^{*})^{T}$   
 $= \tilde{M}^{*}(H_{k}) + O(||H_{k}||^{2})$   
(63)  
 $\tilde{M}^{*}(H_{k}) + o(||x_{k+1} - x^{*}||) + o(||d_{k}||).$ 

Since  $G(x^*) \succeq 0, Y^* \succeq 0$  and  $\langle G(x^*), Y^* \rangle = 0$ , we assume, without loss of generality, that

$$G(x^*) - Y^* = Q_* \Lambda_* Q_*^T,$$

where  $\Lambda_*$  is a diagonal matrix whose diagonal elements are in decreasing order. Moreover, the first r diagonal elements of  $\Lambda_*$  are the positive eigenvalues of  $G(x^*)$ as well as the last m - r diagonal elements being the negative eigenvalues of  $-Y^*$ . Let

$$\hat{H}_k = Q_*^T H_k Q_* = \begin{pmatrix} \hat{H}_k^{\alpha \alpha} & \hat{H}_k^{\alpha \gamma} \\ \hat{H}_k^{\gamma \alpha} & \hat{H}_k^{\gamma \gamma} \end{pmatrix} = \begin{pmatrix} (Q_*^{\alpha})^T H_k Q_*^{\alpha} & (Q_*^{\alpha})^T H_k Q_*^{\gamma} \\ (Q_*^{\gamma})^T H_k Q_*^{\alpha} & (Q_*^{\gamma})^T H_k Q_*^{\gamma} \end{pmatrix},$$

where  $Q_*^{\alpha}$  is formed by the first r columns of  $Q_*$  and  $Q_*^{\gamma}$  by the last m-r ones. Since

$$\Pi_{\mathcal{S}^m_+} \left( U^* (U^*)^T (G(x^*) - Y^*) U^* (U^*)^T \right) = \Pi_{\mathcal{S}^m_+} (-Y^*),$$

by Proposition 2.2 in [17], there exists an  $\hat{M} \in \partial \Pi_{\mathcal{S}^{r}_{+}}(0)$  such that

$$\tilde{M}^*(H_k) = Q_* \begin{pmatrix} \hat{M}(\hat{H}_k^{\alpha\alpha}) & O \\ O & O \end{pmatrix} Q_*^T,$$

Thus, we have that

$$(Q_*^{\gamma})^T \tilde{M}^*(H_k) Q_*^{\gamma} = \begin{pmatrix} O & I_{m-r} \end{pmatrix} \begin{pmatrix} \hat{M}(\hat{H}_k^{\alpha\alpha}) & O \\ O & O \end{pmatrix} \begin{pmatrix} O \\ I_{m-r} \end{pmatrix} = 0.$$

Multiplying  $Q_*^{\gamma}$  to the both sides of (63),

$$(Q_*^{\gamma})^T DG(x^*)(x_{k+1} - x^*) Q_*^{\gamma} = o(||x_{k+1} - x^*||) + o(||x_k - x^*||).$$

Applying  $\operatorname{svec}(\cdot)$  to the both sides above again, we conclude from

$$D\hat{S}(x^*)(x_{k+1} - x^*) = (Q_*^{\gamma})^T DG(x^*)(x_{k+1} - x^*)Q_*^{\gamma}$$

and Lemma 4.2 that

(64)  

$$V(x^*)^T(x_{k+1} - x^*) = \operatorname{svec}(D\hat{S}(x^*)(x_{k+1} - x^*))$$

$$= \operatorname{svec}((Q_*^{\gamma})^T DG(x^*)(x_{k+1} - x^*)Q_*^{\gamma})$$

$$= o(||x_{k+1} - x^*||) + o(||x_k - x^*||).$$

By (56), (57) and (64),

$$\begin{pmatrix} P^* \nabla_{xx}^2 \hat{L}(x^*, \mu^*, \hat{Y}^*) \\ Dh(x^*) \\ V(x^*)^T \end{pmatrix} (x_{k+1} - x^*) \\ = \begin{pmatrix} -P_k(W_k - \nabla_{xx}^2 L(x^*, \mu^*, Y^*))d_k \\ O \\ O \end{pmatrix} \\ + o(\|x_{k+1} - x^*\|) + o(\|x_k - x^*\|) + o(\|d_k\|). \end{cases}$$

The rest of the proof is similar to that in Theorem 3.4.

### 5. Conclusions

In this paper, we introduced a sequential semidefinite programming (SSDP) local method for solving nonlinear semidefinite programming problems, which is inspired by the classic sequential quadratic programming method. We first give the sufficient conditions for superlinear convergence of  $\{x_k\}$ . Since the curvature of the SDP constraints does not contribute to the Lagrangian Hessian matrix when they are linear, Dorsch et al (5) consider a reduced SDP, where the positive semidefiniteness of a symmetric matrix G(x), depending continuously on x, is locally equivalent to the fact that a certain Schur complement S(x) of G(x) is positive semidefinite. For the reduced SDP problem, the well-known and often mentioned " $\sigma$ -term" in the second order sufficient condition vanishes. A reduced sequential semidefinite programming (RSSDP) method is proposed for solving the reduced SDP problem. Under the nondegeneracy condition, the second-order sufficient condition with  $\sigma$ term and the strict complementarity condition, we made an analysis on the local convergence rate of the RSSDP algorithm and proposed an equivalent condition for its superlinear convergence. Finally, we proposed an equivalent condition for superlinear convergence of the sequence  $\{x_k\}$  generated by SSDP method.

# Acknowledgment

This research was supported by Chinese NSF grant 11871362.

## References

- J.F. Bonnans and A. Shapiro, Perturbation Analysis of Optimization Problems, Springer, New York, 2000.
- [2] P.T. Boggs, J.W. Tolle, and P. Wang, On the local convergence of quasi-Newton methods for constrained optimization, SIAM J. Control Optim., 20(2)(1982), 161-171.
- [3] R. Correa and H.C. Ramirez, A global algorithm for nonlinear semidefinite programming, SIAM J. Optim., 15(1)(2004), 303-318.
- [4] M. Diehl, F. Jarre, and C.H. Vogelbusch, Loss of superlinear convergence for an SQP-type method with conic constraints, SIAM J. Optim., 16(4)(2006), 1201-1210.
- [5] D. Dorsch, W. Gómez, and V. Shikhman, Sufficient optimality conditions hold for almost all nonlinear semidefinite programs, Math. Program., 158(1-2)(2016), 77-97.
- [6] J.Y. Fan, Duality theories in nonlinear semidefinite programming, Appl. Math. Lett., 18(9)(2005), 1068-1073.

611

- [7] R.W. Freund, F. Jarre, and C.H. Vogelbusch, Nonlinear semidefinite programming: sensitivity, convergence, and an application in passive reduced-order modeling, Math. Program., 109(2-3)(2007), 581-611.
- [8] B. Fares, D. Noll, and P. Apkarian, Robust control via sequential semidefinite programming, SIAM J. Control Optim., 40(6)(2002), 1791-1820.
- [9] A. Forsgren, Optimality conditions for nonconvex semidefinite programming, Math. Program., 88(1)(2008), 105-128.
- [10] R. Garcés, W. Gómez and F. Jarre, A sensitivity result for quadratic semidefinite program with an application to a sequential quadratic semidefinite programming algorithm, Comput. Appl. Math., 31(1)(2012), 205-218.
- [11] X. X. Huang, K. L. Teo, X. Q. Yang, Approximate augmented Lagrangian functions and nonlinear semidefinite programs, Acta Math. Sin. (Engl. Ser.) 22(5)(2006), 1283-1296.
- [12] X. X. Huang, X. Q. Yang, K. L. Teo, Lower-order penalization approach to nonlinear semidefinite programming, J. Optim. Theory Appl. 132(1)(2007), 1-20.
- [13] L. Mosheyev and M. Zibulevsky, Penalty/barrier multiplier algorithm for semidefinite programming, Optim. Method Softw., 13(4)(2000), 235-261.
- [14] D. Noll, Local convergence of an augmented Lagrangian method for matrix inequality constrained programming, Optim. Method Softw., 22(5)(2007), 777-802.
- [15] A. Shapiro, First and second order analysis of nonlinear semidefinite programs, Math. Program., 77(1)(1997), 301-320.
- [16] D. Sun, J. Sun, and L. Zhang, The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming, Math. Program., 114(2008), 349-391.
- [17] D. Sun, The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications, Math. Oper. Res., 31(4)(2006), 761-776.
- [18] H. Wolkowicz, R. Saigal, and L. Vandenberghe, Handbook of Semidefinite Programming: Theory, Algorithms, and Applications, Springer Science & Business Media, 2012.
- [19] Y. Wang, S. Zhang, and L. Zhang, A note on convergence analysis of an SQP-type method for nonlinear semidefinite programming, J. Inequal. App., https:// doi:10.1155/2008/218345, 2008.
- [20] H. Yamashita and H. Yabe, Local and superlinear convergence of a primal-dual interior point method for nonlinear semidefinite programming, Math. Program., 132(2012), 1-30.
- [21] H. Yamashita, H. Yabe, and K. Harada, A primal-dual interior point method for nonlinear semidefinite programming, Math. Program., 135(1-2)(2012), 89-121.
- [22] Q. Zhao and Z.W. Chen, On the superlinear local convergence of a penalty-free method for nonlinear semidefinite programming, J. Comput. Appl. Math., 308(2016), 1-19.
- [23] Q. Zhao and Z.W. Chen, An SQP-type method with superlinear convergence for nonlinear semidefinite programming, Asia Pac. J. Oper. Res., 35(3)(2018), 1850009.

School of Mathematical Sciences, Soochow University, Suzhou, PR China. E-mail: wenhfu@163.com

School of Mathematical Sciences, Soochow University, Suzhou, 215006, PR China.E-mail:zwchen@suda.edu.cn