## RESEARCH

**Open Access** 



# Primal-dual interior point QP-free algorithm for nonlinear constrained optimization

Jinbao Jian<sup>1</sup>, Hanjun Zeng<sup>2</sup>, Guodong Ma<sup>1\*</sup> and Zhibin Zhu<sup>3</sup>

\*Correspondence: mgd2006@163.com \*School of Mathematics and Statistics, Guangxi Colleges and Universities Key Laboratory of Complex System Optimization and Big Data Processing, Yulin Normal University, Yulin, China Full list of author information is available at the end of the article

## Abstract

In this paper, a class of nonlinear constrained optimization problems with both inequality and equality constraints is discussed. Based on a simple and effective penalty parameter and the idea of primal-dual interior point methods, a QP-free algorithm for solving the discussed problems is presented. At each iteration, the algorithm needs to solve two or three reduced systems of linear equations with a common coefficient matrix, where a slightly new working set technique for judging the active set is used to construct the coefficient matrix, and the positive definiteness restriction on the Lagrangian Hessian estimate is relaxed. Under reasonable conditions, the proposed algorithm is globally and superlinearly convergent. During the numerical experiments, by modifying the technique in Section 5 of (SIAM J. Optim. 14(1): 173-199, 2003), we introduce a slightly new computation measure for the Lagrangian Hessian estimate based on second order derivative information, which can satisfy the associated assumptions. Then, the proposed algorithm is tested and compared on 59 typical test problems, which shows that the proposed algorithm is promising.

MSC: 90C30; 49M37; 65K05

**Keywords:** inequality and equality constraints; optimization; primal-dual interior method; working set; global and superlinear convergence

## **1** Introduction

In this paper, we consider nonlinear constrained optimization problems with inequality and equality constraints

(P) min 
$$f(x)$$
, s.t.  $g_i(x) = 0$ ,  $i \in I^{\ell}$ ;  $g_j(x) \le 0$ ,  $j \in I^i$ , (1)

where  $I^{\ell} = \{1, 2, ..., m_{\ell}\}, I^{i} = \{m_{\ell} + 1, m_{\ell} + 2, ..., m_{\ell} + m_{i}\}$ , the functions f and  $g_{j} : \mathbb{R}^{n} \to \mathbb{R}$ . It is known that the nonlinear equality constraints are difficult to be dealt with in designing algorithms for (P), especially, in designing the *methods of feasible directions* (MFD). In 1976, Mayne and Polak [2] proposed a simple scheme to convert (P) to a sequence of inequality smoothing constrained optimization

$$(\mathbf{P}_{\rho}) \quad \min f_{\rho}(x) := f(x) - \rho \sum_{j \in I^{\ell}} g_j(x), \quad \text{s.t.} \quad g_j(x) \le 0, \quad j \in I^{\ell} \cup I^{\iota},$$
(2)

© The Author(s) 2017. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



where  $\rho > 0$  is a penalty parameter. Under suitable *constraint qualifications* (CQ), e.g., linear independence, it has been shown that (P<sub> $\rho$ </sub>) is equivalent to (P) when  $\rho$  is large enough. So, based on (P<sub> $\rho$ </sub>), one can study and present effective algorithms for the original problem (P), e.g., Refs. [1, 3–6].

In addition, with the help of inequality constrained non-smoothing optimization

$$\min f(x) + \sum_{j \in I^{\ell}} c_j |g_j(x)|, \quad \text{s.t.} \quad g_j(x) \le 0, \quad j \in I^{\ell} \cup I^{\iota},$$

one can also design an algorithm for solving the original problem (P), e.g., [7], where  $c_j > 0$  is the penalty parameter that needs to be updated.

It is known that the *sequential quadratic programming* (SQP) method is one of the efficient methods for constrained optimization due to its fast convergence, and it has been widely studied by many authors, see Refs. [8–17]. However, the *quadratic program* (QP) subproblems solved in the SQP methods may be inconsistent, and the computational cost for the QPs is high. Therefore, motivated by the KKT condition of the QPs and/or the quasi-Newton method, QP-free methods are put forward, in which the QPs are replaced by suitable *systems of linear equations* (SLEs), see Refs. [18–26].

Now we review briefly the study on the *primal-dual interior point* (PDIP) QP-free algorithms associated with our work. First, for problem (P) with no equality constraints, i.e.,  $I^{\ell} = \emptyset$ , in 1987, Panier et al. [22] presented a QP-free algorithm denoted by PTH, at iterate *k*, two SLEs are solved to yield a master search direction. Then a *least squares problem* (LSP) needs to be solved to avoid the so-called Maratos effect [27]. However, the SLEs solved in [22] may become ill-conditioned, and the PTH algorithm may be instable. Furthermore, the initial point must lie on the strict interior of the feasible set, and an additional assumption that 'the number of stationary points is finite' is used to ensure the global convergence. Later, under the assumption that the multiplier approximation sequence remains bounded, the PTH algorithm was improved by Gao et al. [3] by solving an extra SLE. The PTH algorithm was also improved by Qi and Qi [23], Zhu [26] and Cai [28].

To improve the PTH algorithm [22], by using the idea of PDIP and choosing different barrier parameters for each constraint, Bakthiari and Tits [18] proposed a new PDIP QPfree algorithm. The algorithm can start from a feasible point at the boundary of the feasible set, and it possesses global convergence without both the additional assumption of isolatedness of the stationary points and the positive definite restriction on matrix  $H_k$ . Almost at the same time, Tits et al. [1] extended and improved the PTH algorithm to problem (P) with both inequality and equality constraints. The algorithm [1] possesses two remarkable characters. One is that a new and simple rule to update the penalty parameter  $\rho$  in (P<sub> $\rho$ </sub>) is derived, the other is that, same as in [18], the uniformly positive definite restriction on the Lagrangian Hessian estimate is relaxed.

More recently, for inequality constrained optimization, Jian et al. [21] proposed a strongly sub-feasible primal-dual quasi interior-point algorithm with superlinear convergence, where the initial point can be chosen arbitrarily, the number of feasible constraints is nondecreasing, and the iteration points all enter into the interior of the feasible region after finite iterations; a new kind of working set was introduced, which further reduced the computational cost; the uniformly positive definite restriction on the sequence  $\{H_k\}$  was

relaxed; at each iteration, only two or three SLEs with the same coefficient matrix needed to be solved.

However, there are still some problems worthy of research on the PDIP-type algorithms [1, 18, 22]. First, the coefficient matrix of the Karush-Kuhn-Tucker (KKT) system of the LSP is not the same as the two previous SLEs, and this further increases the computational cost. Second, the coefficient matrices of the SLEs include all the constraints and their gradients, and this leads to a large increase in the scale of the SLEs. Third, the global convergence of the two algorithms [1, 18] relies on an additional assumption that the stationary points are finite or isolated.

On the other hand, to design more effective algorithms with small computational cost for solving constrained optimization, Facchinei et al. [29] first introduced the active set identifying technique (also called working set technique). And then this technique has been popularized and applied in many works, e.g., [17, 24, 25, 30, 31]. Particularly, the algorithm [30] needs to solve four SLEs at each iteration.

The goal of this paper is to improve and extend the algorithms [18, 21] to nonlinear constrained optimization (P) and, at the same time, to overcome the three problems mentioned above. As a result, by means of problem ( $P_\rho$ ), we propose a PDIP-type algorithm for problem (P). Compared with the previous PDIP-type algorithms, the proposed algorithm possesses the following features.

- (a) A slightly new identifying technique for the active set different from [17, 25] is introduced. The multiplier yielded at the previous iteration is used to compute the working set, and no additional computational cost is needed, so the computational cost is expected to be reduced.
- (b) At each iteration, to yield the search directions, only two or three SLEs with the same coefficient matrix need to be solved. Furthermore, the coefficient matrix has smaller scale than the ones in [1, 18, 22].
- (c) For a strict interior point  $x^k$  of the feasible set of  $(P_\rho)$ , the iteration at  $x^k$  is well defined without any other constraint qualification (CQ).
- (d) Under suitable CQ and assumptions including a relaxed positive definite restriction on the Lagrangian Hessian estimate *H<sub>k</sub>*, *but without the isolatedness of the stationary points*, the proposed algorithm is globally and superlinearly convergent.
- (e) A slightly new computation technique for  $H_k$  based on second order derivative information is introduced, which is a modification of the one in [1], Section 5.1, and satisfies the relaxed positive definite restriction.

Throughout this paper, for simplicity, denote vector  $(x^T, y^T, z^T, ...)^T$  by (x, y, z, ...) for column vectors x, y and z, and  $\|\cdot\|$  denotes the Euclidean norm.

### 2 Construction of algorithm

To analyze our algorithm, the following notations are used:

$$\begin{split} I &= I^{\ell} \cup I^{i}, \qquad \hat{e} = \left(1, \dots, 1(m_{\ell} \operatorname{th}), 0, \dots, 0\left((m_{\ell} + m_{i}) \operatorname{th}\right)\right)^{T}, \\ X &= \left\{x \in \mathbb{R}^{n} : g_{i}(x) = 0, i \in I^{\ell}; g_{j}(x) \leq 0, j \in I^{i}\right\}, \qquad e_{J} = (1, \dots, 1)^{T} \in \mathbb{R}^{|J|}, \\ \tilde{X} &= \left\{x \in \mathbb{R}^{n} : g_{j}(x) \leq 0, j \in I\right\}, \qquad \tilde{X}_{0} = \left\{x \in \mathbb{R}^{n} : g_{j}(x) < 0, j \in I\right\}, \\ I^{\ell}(x) &= \left\{j \in I^{\ell} : g_{j}(x) = 0\right\}, \qquad I^{i}(x) = \left\{j \in I^{i} : g_{j}(x) = 0\right\}, \qquad I(x) = I^{\ell}(x) \cup I^{i}(x), \\ g_{\ell}(x) &= \left(g_{j}(x), j \in I^{\ell}\right), \qquad g_{i}(x) = \left(g_{j}(x), j \in I^{i}\right), \qquad g(x) = \left(g_{j}(x), j \in I\right), \end{split}$$

$$\begin{split} g_J(x) &= \left(g_j(x), j \in J \subset I\right), \qquad \nabla g_J(x) = \left(\nabla g_j(x), j \in J\right), \\ g_j^k &= g_j(x^k), \qquad g_J^k = g_J(x^k), \qquad \nabla g_j^k = \nabla g_j(x^k), \qquad \nabla g_j^{k^T} = \left(\nabla g_j^k\right)^T. \end{split}$$

First, the following basic hypothesis is necessary.

**H1** The inner set  $\tilde{X}_0$  is nonempty, and the functions f and  $g_j$  ( $j \in I$ ) are all continuously differentiable.

**Remark 1** Note that if there exists a point belonging to the set  $\tilde{X}$ , namely,  $\hat{x} \in \tilde{X}$ , and the active constraint gradient vectors  $\{\nabla g_j(\hat{x}), j \in I(\hat{x})\}$  are linearly independent, then one can yield a point  $x^0 \in \tilde{X}_0$  by simple computation, e.g., execute line search on g starting with  $\hat{x}$  along direction  $\hat{d} = -\hat{N}(\hat{N}^T\hat{N})^{-1}e$ , where  $\hat{N} = \nabla g_{I(\hat{x})}(\hat{x})$  and  $e = (1, ..., 1)^T$ .

Before proposing our algorithm, we give a proposition to show the equivalences between (P) and  $(P_{\rho})$ .

**Proposition 1** If  $(x, \lambda)$  is a KKT pair for problem  $(P_{\rho})$  and  $g_{\ell}(x) = 0$ , then  $(x, \lambda_{\rho})$  with multiplier  $\lambda_{\rho} = \lambda - \rho \hat{e}$  is a KKT pair for the original problem (P).

Based on Proposition 1, it is known that if one can construct an effective algorithm for problem ( $P_{\rho}$ ) and adjust parameter  $\rho$  to force the iterate to asymptotically satisfy  $g_{\ell}(x) = 0$ , then the solution to (P) can be yielded.

Now, refer to [29] and [24], we introduce optimal identification functions  $\Phi$  and  $\delta$  as follows:

$$\Phi(x,\lambda) = \begin{pmatrix} \nabla_x L(x,\lambda) \\ g_\ell(x) \\ \min\{-g_\ell(x),\lambda_\ell\} \end{pmatrix}, \qquad \delta(x,\lambda) = \|\Phi(x,\lambda)\|^r,$$
(3)

where  $\lambda = (\lambda_{\ell}, \lambda_{\iota})$ , parameter  $r \in (0, 1)$ , and the Lagrangian function

$$L(x,\lambda) = f(x) + \sum_{j \in I} \lambda_j g_j(x).$$
(4)

It is clear that  $(x^*, \lambda)$  is a KKT pair of (P) if and only if  $\delta(x^*, \lambda) = 0$ . Particularly, from [29] or/and [24], Definition 4.1, Theorems 4.1, 4.2 and 4.3, one can see that  $\{j \in I : g_j(x) + \delta(x, \lambda) \ge 0\}$  is an exact identification set for active constraint set  $I(x^*)$  if  $(x, \lambda)$  converges to a KKT pair  $(x^*, \lambda')$  of problem (P), and the Mangasarian-Fromovotz constraint qualification (MFCQ) and the second order sufficient conditions are satisfied at  $(x^*, \lambda')$ .

In this paper, similarly to the techniques in [21, 30], for the current iterate  $x^k \in \tilde{X}_0$ , we yield the corresponding multiplier vector  $\lambda^k = (\lambda_\ell^k, \lambda_\ell^k)$  in (3)-(4) as follows:

$$\lambda^{0} = z^{0}, \qquad \lambda^{k} = \bar{\lambda}^{k-1} - \rho_{k-1}\hat{e}, \quad k > 0,$$
(5)

where  $z^0 > 0$ , and  $(\bar{\lambda}^{k-1}, \rho_{k-1})$  is computed in the previous iteration (k-1)th. Then, similarly to [29], we structure our working set by

$$I_k^i = \left\{ j \in I^i : g_j(x^k) + \delta(x^k, \lambda^k) \ge 0 \right\}, \qquad I_k = I^\ell \cup I_k^i.$$
(6)

The reason why one does not compute  $I_k^{\ell}$  as  $I_k^{\iota}$  is to force  $g_{\ell}(x^k) \to 0$ , see the analysis of Theorem 1 in Section 3. The set  $I_k^{\iota}$  equals the exact active set  $I^{\iota}(x^*)$  when  $(x^k, \lambda^k)$  is sufficiently close to a KKT pair  $(x^*, \lambda')$  of (P) and the second order sufficient conditions as well as the MFCQ hold at  $(x^*, \lambda')$ . This important property allows us to construct the direction finding subproblems only considering the constraints in the working set  $I_k$ .

Taking into account that the iterates always execute within the feasible set  $\tilde{X}$ , let us consider the first order condition of optimality (KKT condition) for problem ( $P_{\rho_k}$ ) nearby the current iterate  $x^k$ :

$$abla f_{\rho_k}(x) + \sum_{j \in I_k} \lambda_j \nabla g_j(x) = 0, \qquad \lambda_j g_j(x) = 0, \quad j \in I_k, \lambda_{I_k} \ge 0.$$

Furthermore, if we ignore the non-negativity request ' $\lambda_{I_k} \geq 0$ ' and simultaneously introduce a suitable perturbation  $((1 - \zeta_k) \nabla f_{\rho_k}(x^k), \mu^k) \in \mathbb{R}^{(n+|I_k|)}$  in the right-hand side of the above system, then it can be reduced as a system of nonlinear equations with variables  $(x, \lambda_{I_k})$ 

$$\begin{pmatrix} \nabla f_{\rho_k}(x) + \sum_{j \in I_k} \lambda_j \nabla g_j(x) \\ \lambda_j g_j(x), j \in I_k \end{pmatrix} = \begin{pmatrix} (1 - \zeta_k) \nabla f_{\rho_k}(x^k) \\ \mu^k \end{pmatrix}.$$
 (7)

Applying the Newton method to system (7) starting with the current iterate ( $x^k$ ,  $\lambda_{I_k}^k$ ), it yields a SLE as follows:

$$\begin{pmatrix} \nabla_{xx}^2 L_{\rho_k}(x^k, \lambda_{I_k}^k) & \nabla g_{I_k}(x^k) \\ \Lambda_k \nabla g_{I_k}(x^k)^T & \text{diag}(g_{I_k}^k) \end{pmatrix} \begin{pmatrix} x - x^k \\ \lambda_{I_k} \end{pmatrix} = \begin{pmatrix} -\zeta_k \nabla f_{\rho_k}(x^k) \\ \mu^k \end{pmatrix},$$
(8)

where diagonal matrix  $\Lambda_k = \text{diag}(\lambda_{I_k}^k)$ , and the Lagrangian Hessian

$$\nabla^2_{xx}L_{\rho_k}(x^k,\lambda^k_{I_k}) = \nabla^2 f_{\rho_k}(x^k) + \sum_{j\in I_k}\lambda^k_j \nabla^2 g_j(x^k).$$

Subsequently, to make the coefficient matrix in SLE (8) possess nice property and low computational cost, we consider its optimization and modification as follows. First, replace the Lagrangian Hessian by a suitable approximate symmetric matrix  $H_k$ , and denote  $x - x^k$  by direction d. Second, replace the diagonal matrix  $\Lambda_k$  by positive diagonal matrix  $Z_k = \text{diag}(z_{l_k}^k)$ , where vector  $z_{l_k}^k$  is an approximation of  $\lambda^k$ .

As a result, from system (8), the coefficient matrix and the form of the SLEs that need to be solved in our algorithm are as follows:

$$V_k := \begin{pmatrix} H_k & \nabla g_{I_k}(x^k) \\ Z_k \nabla g_{I_k}(x^k)^T & \text{diag}(g_{I_k}^k) \end{pmatrix}, \tag{9}$$

$$SLE(V_k; \zeta_k, \mu^k): \quad V_k \begin{pmatrix} d \\ \lambda_{I_k} \end{pmatrix} = \begin{pmatrix} -\zeta_k \nabla f_{\rho_k}(x^k) \\ \mu^k \end{pmatrix}.$$
(10)

To yield improved search directions with superlinear convergence, our algorithm will solve two or three SLEs with the form of (10) with different perturbation vectors ( $\zeta_k$ ,  $\mu^k$ ). Subsequently, it is necessary to analyze the singularities of the coefficient matrix  $V_k$  above, i.e., the solvability of SLE (10).

**Lemma 1** For iterate  $x^k \in \tilde{X}_0$  and  $z_{I_k}^k > 0$ , if the matrix  $H_k$  satisfies

$$H_k \succ \sum_{j \in I_k} \frac{z_j^k}{g_j^k} \nabla g_j^k \nabla g_j^{k^T}, \tag{11}$$

then the coefficient matrix  $V_k$  defined by (9) is invertible, where matrix order  $A \succ B$  means (A - B) is positive definite on  $\mathbb{R}^n$ .

*Proof* One knows that it is sufficient to show that SLE  $V_k u = 0$  has a unique solution zero, and this is elementary and omitted here.

**Remark 2** Obviously, the positive definiteness request (11) on  $H_k$  is weaker than the positive definiteness of  $H_k$  itself on  $\mathbb{R}^n$ . But it is stronger than the positive definiteness of  $H_k$ on the null space of the gradients of approximate active constraints, i.e., on  $\Omega_k := \{d \in \mathbb{R}^n : \nabla g_{I_k}(x^k)^T d = 0\}$ . However, the latter cannot ensure the invertibility of  $V_k$ .

Based on the above analysis and preparation, now we can describe the steps of our algorithm solving (P) as follows.

#### Algorithm A

*Parameters:*  $\alpha \in (0, \frac{1}{2}), \sigma, \beta, \theta, r \in (0, 1), \xi \in (2, 3), \nu > 2, \vartheta > 1, M, p > 0$ ; suitable small positive parameters  $\gamma_1, \gamma$  and  $\gamma_3$ ; sufficiently small lower bound  $\underline{\varepsilon} > 0$  and sufficiently large upper bound  $\overline{\varepsilon} > 0$ ; termination accuracy  $\epsilon > 0$ .

*Data*:  $x^0 \in \tilde{X}_0, \rho_0 > 0$ , vectors  $z^0$  with weights  $z_i^0 \in [\underline{\varepsilon}, \overline{\varepsilon}], j \in I$ . Set k := 0.

**Step 1** *Compute working set.* Compute  $\lambda^k$  by (5),  $\Phi(x^k, \lambda^k)$  and  $\delta(x^k, \lambda^k)$  by (3)-(4). If  $\Phi(x^k, \lambda^k) \leq \epsilon$  or other suitable termination rule is satisfied, then  $(x^k, \lambda^k)$  is an approximate KKT pair of problem (P) and stop; otherwise, generate the working sets  $I_k^i$  and  $I_k$  by (6).

**Step 2** *Yield matrix*  $H_k$ . Yield matrix  $H_k$  such that it approximates to the Hessian of the Lagrangian associated with ( $P_{o_k}$ ) and satisfies request (11).

Step 3 Compute the main search directions.

(i) Compute  $(\bar{d}^k, \bar{\lambda}_{I_k}^k)$  by solving SLE $(V_k; 1, 0)$ , see (10), then set  $\bar{\lambda}^k = (\bar{\lambda}_{I_k}^k, 0_{I \setminus I_k}) = (\bar{\lambda}_{\ell}^k, \bar{\lambda}_{\iota}^k)$  with  $\bar{\lambda}_{\iota}^k = (\bar{\lambda}_{I_{\iota}}^k, 0_{I \setminus I_{\iota}^k})$ .

(ii) Check conditions: (a)  $\|\vec{d^k}\| \le \gamma_1$ , (b)  $\bar{\lambda}^k \ge -\gamma_2 e_I$ , (c)  $\bar{\lambda}^k_\ell \ne \gamma_3 e_{I^\ell}$ . If all the three conditions above hold, then increase penalty parameter  $\rho$  by  $\rho_{k+1} = \vartheta \rho_k$ , set  $x^{k+1} = x^k, z^{k+1} = z^k, H_{k+1} = H_k, I'_{k+1} = I'_k, I_{k+1} = I_k, k := k+1$ , and go back to Step 3(i). Otherwise, set  $\rho_{k+1} = \rho_k$ , proceed to Step 3(ii) as follows.

(iii) Yield the weights of vector  $\phi^k$  by

$$\phi_{j}^{k} = \min\{0, -(\max\{-\bar{\lambda}_{j}^{k}, 0\})^{p} - Mg_{j}^{k}\}, \quad j \in I_{k}.$$
(12)

Then compute

$$\xi_k = \nabla f_{\rho_k} \left( x^k \right)^T \vec{d}^k - \sum_{j \in I_k} \frac{\bar{\lambda}_j^k \phi_j^k}{z_j^k},\tag{13}$$

$$b_{k} = \left( \left\| \bar{d}^{k} \right\|^{\nu} + \left\| \phi^{k} \right\| \right) \left( \sum_{j \in I_{k}} \bar{\lambda}_{j}^{k} \right) + \sum_{j \in I_{k}} \frac{\bar{\lambda}_{j}^{k}}{z_{j}^{k}} \phi_{j}^{k},$$
(14)

$$\varphi_{k} = \begin{cases} 1, & \text{if } b_{k} \leq 0; \\ \min\{\frac{(1-\theta)|\xi_{k}|}{b_{k}}, 1\}, & \text{if } b_{k} > 0, \end{cases}$$
(15)

and yield perturbation vectors via convex combinations

$$\mu^{k} = (1 - \varphi_{k})\phi^{k} + \varphi_{k} \left( - \|\bar{d}^{k}\|^{\nu} - \|\phi^{k}\| \right) z_{I_{k}}^{k}.$$
(16)

(iv) Compute  $(d^k, \lambda_{I_k}^k)$  by solving SLE $(V_k; 1, \mu^k)$ , see (10), then set  $\lambda^k = (\lambda_{I_k}^k, 0_{I \setminus I_k}) =$  $(\lambda_{\ell}^k, \lambda_{\iota}^k)$  with  $\lambda_{\iota}^k = (\lambda_{I_{\iota}^k}^k, \mathbf{0}_{I^{\iota} \setminus I_k^{\iota}})$ .

Step 4 Trial of unit step. If

$$f_{\rho_k}(x^k + d^k) \leq f_{\rho_k}(x^k) + \alpha \nabla f_{\rho_k}(x^k)^T d^k, \qquad g_j(x^k + d^k) < 0, \quad \forall j \in I,$$

then let the step size  $t_k = 1$ , the high order correction direction  $\tilde{d}^k = 0$ , and enter Step 7. Otherwise, proceed to Step 5.

**Step 5** Generate high order correction direction. Compute  $(\tilde{d}^k, \tilde{\lambda}_{I_k}^k)$  by solving SLE( $V_k$ ; 0,  $\tilde{\mu}^k$ ), where

$$\tilde{\mu}^k = -\omega_k e_{I_k} - Z_k g_{I_k} \left( x^k + d^k \right), \tag{17}$$

$$\omega_{k} = \max\left\{ \left\| d^{k} \right\|^{\xi}; \left\| d^{k} \right\|^{2} \max\left\{ \left| 1 - \frac{z_{j}^{\kappa}}{\lambda_{j}^{k}} \right|^{\sigma}, j \in I_{k}, \lambda_{j}^{k} \neq 0 \right\} \right\}.$$
(18)

If  $\|\tilde{d}^k\| > \|d^k\|$ , reset  $\tilde{d}^k = 0$ .

**Step 6** *Perform arc search.* Compute the step size  $t_k$ , the maximum number t of sequence  $\{1, \beta, \beta^2, \ldots\}$  satisfying

$$f_{\rho_k}\left(x^k + td^k + t^2\tilde{d}^k\right) \le f_{\rho_k}\left(x^k\right) + \alpha t\nabla f_{\rho_k}\left(x^k\right)^T d^k,\tag{19}$$

$$g_j(x^k + td^k + t^2\tilde{d}^k) < 0, \quad j \in I.$$

$$\tag{20}$$

**Step 7** *Update.* Yield a new iterate by  $x^{k+1} = x^k + t_k d^k + t_k^2 \tilde{d}^k$  and compute

$$z_{j}^{k+1} = \min\{\max\{\|d^{k}\|^{2} + \underline{\varepsilon}, \lambda_{j}^{k}\}, \overline{\varepsilon}\}, \quad j \in I.$$

$$(21)$$

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

Subsequently, we analyze and describe some properties of Algorithm A by the following lemma and several remarks. For convenience of writing, denote matrix

$$Q_k := H_k - \sum_{j \in I_k} \frac{z_j^k}{g_j^k} \nabla g_j^k \nabla g_j^{k^T}.$$
(22)

Then request (11) implies that matrix  $Q_k$  is positive definite.

**Lemma 2** For the directions  $\overline{d}^k$  and  $d^k$  yielded in Step 3(i), (iv), the following two relations hold:

$$\nabla f_{\rho_k}(x^k)^T \bar{d}^k = -(\bar{d}^k)^T Q_k \bar{d}^k \le 0, \quad \forall k \ge 0,$$

$$(23)$$

$$\nabla f_{\rho_k} (x^k)^T d^k \le \theta \xi_k \le 0, \quad \forall k \ge 0.$$
(24)

Furthermore, when the iterative process goes into Step 3(iii), (iv), one has  $\overline{d}_k \neq 0$  and  $\xi_k < 0$ , so  $d^k$  is a feasible direction of descent of problem ( $P_{\rho_k}$ ) at point  $x^k$  and the arc search in Step 6 can be finished by finite calculations. Therefore, Algorithm A is well defined.

*Proof* First, from (9) and  $SLE(V_k; 1, 0)$  (10), we have

$$\nabla f_{\rho_k} (x^k)^T \bar{d}^k = -(\bar{d}^k)^T \left( H_k \bar{d}^k + \sum_{j \in I_k} \nabla g_j^k \bar{\lambda}_j^k \right)$$
$$= -(\bar{d}^k)^T \left( H_k - \sum_{j \in I_k} \frac{z_j^k}{g_j^k} \nabla g_j^k \nabla g_j^{kT} \right) \bar{d}^k$$
$$= -(\bar{d}^k)^T Q_k \bar{d}^k \le 0.$$

So, conclusion (23) is at hand. Second, from (12)-(13), one gets

$$\phi_j^k \bar{\lambda}_j^k \ge 0, \quad \forall j \in I_k; \qquad \xi_k \le \nabla f_{\rho_k} \left( x^k \right)^T \bar{d}^k \le 0.$$
(25)

On the other hand, taking into account SLE( $V_k$ ; 1, 0) and SLE( $V_k$ ; 1,  $\mu^k$ ) as well as (13)-(14), it is not difficult to show that

$$\nabla f_{\rho_k}(x^k)^T d^k = \nabla f_{\rho_k}(x^k)^T \bar{d}^k - \sum_{j \in I_k} \frac{\bar{\lambda}_j^k \mu_j^k}{z_j^k} = \xi_k + \varphi_k b_k.$$
(26)

Again, in view of (15), it follows that  $\varphi_k b_k = b_k \le 0$  if  $b_k \le 0$ , hence, the relations  $\xi_k + \varphi_k b_k \le \xi_k \le \theta \xi_k$  hold since  $\xi_k \le 0$ . If  $b_k > 0$ , then  $\xi_k + \varphi_k b_k \le \xi_k + (\theta - 1)\xi_k = \theta \xi_k$ . In all, one gets  $\xi_k + \varphi_k b_k \le \theta \xi_k$ . This, together with (26) and (25), shows that  $\nabla f_{\rho_k} (x^k)^T d^k \le \theta \xi_k \le 0$ .

Third, if  $\bar{d}^k = 0$ , then, from SLE( $V_k$ ; 1, 0) (10),  $g(x^k) < 0$  and (9), it follows that  $\bar{\lambda}_{I_k}^k = 0$ . So, by the structure of Step 3, the iterate k does not go into Step 3(iii), (iv). Thus,  $\bar{d}^k \neq 0$  when the iterative process goes into Step 3(iii), (iv).

Finally,  $\xi_k < 0$  follows from (25), (23) and  $d_k \neq 0$ . The remaining claims in Lemma 2 are at hand by  $\xi_k < 0$  and  $g(x^k) < 0$ .

As an end of this section, to help the readers understand our algorithm, we further analyze the steps/structure of Algorithm A with three remarks below.

### **Remark 3** (Analysis for Step 3)

(i) The role of solving SLE( $V_k$ ;1,0) with no perturbation in Step 3(i) is to check whether the current iterate  $x^k$  is an approximate KKT point of ( $P_{\rho_k}$ ) and yield an 'improved' direction  $\bar{d}^k$  to a certain extent.

- (ii) If conditions (a) and (b) in Step 3(ii) are satisfied, and the parameters γ<sub>1</sub> and γ<sub>2</sub> are small enough, then SLE(V<sub>k</sub>;1,0) implies that x<sup>k</sup> is an approximate KKT point of (P<sub>ρk</sub>). However, if case (c) is also satisfied, one cannot estimate ||g<sub>ℓ</sub>(x<sup>k</sup>)||. So, we increase the penalty parameter ρ. In practical computation, if conditions (a) and (b) are satisfied and ||g<sub>ℓ</sub>(x<sup>k</sup>)|| is small enough, we can terminate the algorithm.
- (iii) From result (23), one knows that  $\overline{d}^k$  is a descent direction of the merit function  $f_{\rho_k}(x)$  at  $x^k$  when  $\overline{d}^k \neq 0$ . However, the primal feasibility and dual feasibility are relaxed to a large extent in SLE( $V_k$ ; 1, 0),  $\overline{d}^k$  cannot be used as an effective search direction. So, generally, the first direction  $\overline{d}^k$  should be corrected by another SLE. For this goal, refer to [21], we construct and solve SLE( $V_k$ ; 1,  $\mu^k$ ) in Step 3(iii), (iv). Lemma 2 and the global convergence analysis in the next section show that the algorithm with search direction  $d^k$  is well defined and globally convergent.

**Remark 4** (Explanation for Steps 4 and 5) Usually, search direction  $d^k$  cannot avoid the Maratos effect, i.e., unit step cannot be accepted by the associated line search for all sufficiently large iterates k. So, to overcome the Maratos effect and obtain superlinear convergence, one needs to compute an additional high order correction direction. Here, we generate it by solving  $SLE(V_k; 0, \tilde{\mu}^k)$  in Step 5. Obviously, solving  $SLE(V_k; 0, \tilde{\mu}^k)$  should add computational cost more or less. On the other hand, numerical testing shows that  $d^k$  can still avoid the Maratos effect at some iterates. Therefore, to save computational cost as much as possible, the trial of unit step in Step 4 is added.

**Remark 5** With the help of the working set technique, the three SLEs solved in Algorithm A have a common coefficient matrix  $V_k$ , which can save the cost of computation and is different from those in Refs. [18, 26], etc. Furthermore, due to being interior point type and the constructing technique for  $V_k$ , Algorithm A is well defined at each iterate without any other CQ except the strict inner  $\tilde{X}_0 \neq \emptyset$ , see Lemmas 1 and 2. In many existing QP-free type algorithms, see Refs. [1, 3, 21–24], the linearly independent constraint qualification (LICQ) is necessary to ensure the iterate itself is well defined. Of course, as we see in Assumption H3, to obtain the global and superlinear convergence of Algorithm A, a suitable CQ on the boundary of  $\tilde{X}$  is still necessary.

#### 3 Analysis of global convergence

In this section, we assume that the proposed algorithm (Algorithm A) generates an infinite iteration sequence  $\{x^k\}$  of points. First, we show that the penalty parameter  $\rho_k$  can be fixed after finite iterates. And then, we prove that Algorithm A is globally convergent. For this goal, the following hypotheses are necessary.

**H2** Suppose that the sequences both  $\{x^k\}$  and  $\{H_k\}$  yielded by Algorithm A are bounded, and assume that there exists a positive constant *a* such that

$$d^{T}H_{k}d \geq a \|d\|^{2} - \sum_{j \in I_{k}} \frac{z_{j}^{k}}{|g_{j}^{k}|} \|\nabla g_{j}^{k^{T}}d\|^{2},$$
(27)

i.e.,  $d^T Q_k d \ge a ||d||^2$ ,  $\forall k, \forall d \in \mathbb{R}^n$ .

**H3** For each  $x \in \tilde{X}$ , suppose that

- (i) the gradient vectors  $\{\nabla g_j(x), j \in I(x)\}$  are linearly independent; and
- (ii) if  $x \notin X$ , i.e.,  $g_{\ell}(x) \neq 0$ , then there exist no scalars  $\lambda_j \ge 0, j \in I(x)$  such that  $\sum_{j \in I^{\ell}} \nabla g_j(x) = \sum_{j \in I(x)} \lambda_j \nabla g_j(x)$ .

**Remark 6** (Analysis for H2) The uniform 'positive-definiteness' request (27) on  $\{H_k\}$  is weaker than the usual uniform positive-definiteness of  $\{H_k\}$  itself on  $\mathbb{R}^n$ , namely,  $d^T H_k d \ge a \|d\|^2$ ,  $\forall k, \forall d \in \mathbb{R}^n$ . However, it is stronger than the uniform positive-definiteness of  $H_k$  on the null space  $\Omega_k$ . It is encouraging that, based on the Lagrangian Hessian, we can design an alternative computational technique for  $H_k$  such that  $\{H_k\}$  is bounded and satisfies request (27), which implies (11) whenever  $\{x^k\}$  is bounded, see formulas (52), (54) and (55) as well as Theorem 5 in Section 5.

## Remark 7 (Analysis for H3)

- (i) Hypothesis H3 was introduced by Tits et al. in [1], Assumption 3. In our work, it plays two roles in the convergence analysis of Algorithm A. One is to ensure the correction for the penalty parameter  $\rho$  can be finished in a finite number of iterations, the other is to assure that the sequence  $\{V_k\}$  of coefficient matrices is uniform invertible, see Lemmas 3 and 4. Furthermore, H3 is considerably milder than the linear independence of the gradients  $\{\nabla g_i(x), i \in I^\ell; \nabla g_j(x), j \in I^\iota(x)\}$ , a detailed analysis for this assumption can be seen in [1, 32].
- (ii) First, H3 automatically holds at each interior point x ∈ X
  <sub>0</sub>. Second, H3 can be reduced to each accumulation point x\* of the iterate sequence {x<sup>k</sup>}, which satisfies x\* ∉ X
  <sub>0</sub>. However, the latter is difficult to be verified.

**Lemma 3** Suppose that H1, H2 and H3 hold. Then the penalty parameter  $\rho_k$  in Algorithm A is increased at most finite times.

The proof of Lemma 3 is similar to the one of [1], Lemma 4.1, and omitted here. In what follows,  $\bar{\rho}$  denotes the final value of  $\rho_k$ , i.e.,  $\rho_k \equiv \bar{\rho}$  when k is sufficiently large.

Lemma 4 Suppose that H1, H2 and H3 hold. Then

- (i) the sequence {V<sub>k</sub>} of coefficient matrices is unified invertible, i.e., there exists a positive constant M
   such that ||V<sub>k</sub><sup>-1</sup>|| ≤ M
   ,∀k ≥ 0, and
- (ii) both sequences  $\{(\bar{d}^k, \bar{\lambda}^k)\}$  and  $\{(d^k, \lambda^k)\}$  are bounded.

*Proof* (i) By contradiction, suppose that there exists an infinite subset K such that  $||V_k^{-1}|| \xrightarrow{K} \infty$ . In view of the boundedness of  $\{x^k\}$  and  $\{H_k\}$ , Step 6 and the finite choice of  $I_k^r$ , without loss of generality, for  $k \in K$ , assume that

$$I_k^{\prime} \equiv I^{\prime}, \qquad x^k \to x^*, \qquad H_k \to H_*, \qquad z^k \to z^* \geq \underline{\varepsilon} e_I > 0.$$

Denote  $\hat{I} = I^{\ell} \cup I'$  and  $Z_* = \text{diag}(z_{\hat{i}}^*)$ , then

$$V_k \xrightarrow{K} V_* := \begin{pmatrix} H_* & \nabla g_{\hat{I}}(x^*) \\ Z_* \nabla g_{\hat{I}}(x^*)^T & \operatorname{diag}(g_{\hat{I}}(x^*)) \end{pmatrix}.$$
(28)

Consequently, under H1-H3, refer to the proof of [21], Lemma 3.1(i), one can show that  $V_*$  is nonsingular. So  $\|V_k^{-1}\| \stackrel{K}{\to} \|V_*^{-1}\| < \infty$ , which contradicts  $\|V_k^{-1}\| \stackrel{K}{\to} \infty$ .

(ii) First, the boundedness of  $\{(\bar{d}^k, \bar{\lambda}^k)\}$  follows from SLE( $V_k$ ; 1, 0) and conclusion (i) as well as  $\rho_k \equiv \bar{\rho}$ . Second, the boundedness of  $\{\mu^k\}$  follows from formulas (12)-(16) and the boundedness of  $\{(\bar{d}^k, \bar{\lambda}^k)\}$  as well as the positive boundary below of  $\{z^k\}$ . Therefore, the boundedness of  $\{(d^k, \lambda^k)\}$  is also at hand by SLE( $V_k$ ; 1,  $\mu^k$ ).

**Lemma 5** Suppose that H1, H2 and H3 hold. Let  $x^*$  be an accumulation point of the sequence  $\{x^k\}$  generated by Algorithm A, and suppose that  $\{x^k\}_K \to x^*$  for some infinite index set K. If  $\{\xi_k\}_K \to 0$ , then  $x^*$  is a KKT point of problem  $(P_{\bar{\rho}})$ , and both  $\{\bar{\lambda}^k\}_K$  and  $\{\lambda^k\}_K$  converge to the unique multiplier vector  $\lambda^*$  associated with  $x^*$ .

*Proof* Let  $(\bar{\lambda}^*; \hat{\lambda})$  be any given limit point of  $\{(\bar{\lambda}^k; \lambda^k)\}_K$ . We first show that  $(x^*, \bar{\lambda}^*)$  is a KKT pair of  $(P_{\bar{\rho}})$ . In view of H2, Lemma 4 and the finite choice of  $I_k^i$ , we know that there is an infinite index  $K' \subseteq K$  such that

$$I_{k}^{i} \equiv I', \qquad (\bar{\lambda}^{k}; \lambda^{k}) \to (\bar{\lambda}^{*}; \hat{\lambda}),$$

$$H_{k} \to H_{*}, \qquad \bar{d}^{k} \to \bar{d}^{*}, \qquad z^{k} \to z^{*} \geq \underline{\varepsilon} e_{I}, \quad k \in K'.$$
(29)

Therefore, from (25), (23) and H2, one can easily get  $\bar{d}^* = 0$  by  $\{\xi_k\}_K \to 0$ . Further, taking the limit in SLE( $V_k$ ; 1, 0) for  $k \in K'$ , we have, here  $\hat{I} = I^{\ell} \cup I'$ ,

$$\nabla f_{\bar{\rho}}(x^*) + \sum_{j \in \hat{I}} \bar{\lambda}_j^* \nabla g_j(x^*) = 0; \qquad \bar{\lambda}_j^* g_j(x^*) = 0, \quad \forall j \in \hat{I}.$$

$$(30)$$

Next, divert our attention to showing that  $\bar{\lambda}^* \geq 0$ . It is obvious that  $\bar{\lambda}_j^* = 0$  follows from  $\bar{\lambda}_j^* g_j(x^*) = 0$  for  $j \in \hat{I} \setminus I(x^*)$ . Moreover, from the definition of  $\xi_k$ , i.e., (13), and  $(\xi_k, \bar{d}^k) \xrightarrow{K'} (0, 0)$ , we can deduce that  $\sum_{j \in \hat{I}} \frac{\bar{\lambda}_j^k \phi_j^k}{z_j^k} \to 0, k \in K'$ . Further, in view of (25), we know that each term  $\frac{\bar{\lambda}_j^k \phi_j^k}{z_j^k} \leq 0$ , which together with (29) implies that  $\bar{\lambda}_j^k \phi_j^k \xrightarrow{K'} 0$ . This, plus (12), shows that  $\bar{\lambda}_j^* \min\{0, -(\max\{-\bar{\lambda}_j^*, 0\})^p - Mg_j(x^*)\} = 0$  for  $j \in \hat{I}$ , and this includes  $\bar{\lambda}_j^* \geq 0$  for  $j \in \hat{I} \cap I(x^*)$ . Therefore,  $\bar{\lambda}_{\hat{I}}^* \geq 0$ . Obviously,  $\bar{\lambda}_{I\setminus\hat{I}}^* = 0$ . So  $\bar{\lambda}^* \geq 0$  is at hand.

Hence, taking into account  $x^* \in \tilde{X}$ , we can conclude from (30) that  $(x^*, \bar{\lambda}^*)$  is a KKT pair and  $x^*$  is a KKT point for  $(P_{\bar{\rho}})$ . Furthermore, the analysis above further shows that the sequence  $\{\bar{\lambda}^k\}_K$  possesses a unique limit point, i.e., the unique KKT multiplier vector  $\lambda^*$ . So  $\lim_{k \in K} \bar{\lambda}^k = \lambda^*$ .

Finally, taking into account  $(\bar{d}^k, \bar{\lambda}^k) \xrightarrow{K'} (0, \lambda^*) \ge 0$ , from (12) and (16), we have  $(\phi^k, \mu^k) \xrightarrow{K'} 0$ . Therefore, SLE $(V_k; 1, \mu^k)$  minus SLE $(V_k; 1, 0)$  gives

$$V_k \begin{pmatrix} d^k - \bar{d}^k \\ \lambda_{\hat{I}}^k - \bar{\lambda}_{\hat{I}}^k \end{pmatrix} = \begin{pmatrix} 0 \\ \mu^k \end{pmatrix} \xrightarrow{K'} \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$
(31)

This, along with Lemma 4(i), shows that  $\hat{\lambda} = \lim_{k \in K'} \lambda^k = \lim_{k \in K'} \overline{\lambda}^k = \lambda^*$ .

**Theorem 1** Suppose that H1, H2 and H3 hold. Then each accumulation point  $x^*$  of the sequence  $\{x^k\}$  generated by Algorithm A is a KKT point of the original problem (P), i.e. problem (1).

*Proof* First, there exists an infinite index set K' such that  $x^k \to x^*, k \in K'$ , and relation (29) holds. By contradiction, suppose that  $x^*$  is not a KKT point of (P). Then, from Lemma 4, without loss of generality, one can suppose that  $\lambda^k = \bar{\lambda}^{k-1} - \bar{\rho}\hat{e} \to \bar{\lambda}', k \in K'$ . Therefore, it follows that  $(x^*, \bar{\lambda}')$  is not a KKT pair of (P), which further implies that  $\delta(x^*, \bar{\lambda}') > 0$  and  $I'(x^*) \subseteq I'_k, k \in K'$  large enough. There are two cases as follows to be considered.

*Case I*: Assume that  $x^*$  is a KKT point of  $(P_{\bar{\rho}})$ . Then there exists a multiplier  $\bar{\lambda}'' \ge 0$  such that the KKT condition of  $(P_{\bar{\rho}})$  is satisfied at  $(x^*, \bar{\lambda}'')$ . In view of  $I^i(x^*) \subseteq I^i_k \equiv I'$  holds for  $k \in K'$  large enough, it is easy to know, from the KKT condition of  $(P_{\bar{\rho}})$ , that  $(0, \bar{\lambda}''_{I^*})$  is a solution to SLE in (u, v)

$$V_* \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} -\nabla f_{\bar{\rho}}(x^*) \\ 0 \end{pmatrix},\tag{32}$$

where matrix  $V_*$  is defined by (28). On the other hand, passing to the limit in SLE( $V_k; 1, 0$ ) for  $k \in K'$  and  $k \to \infty$ , one knows that  $(\bar{d}^*, \bar{\lambda}_{I^*}^*)$  also solves system (32) above. Taking into account the nonsingularity of matrix  $V_*$  (by Lemma 4(i)), one knows that the solution of (32) is unique. So  $\bar{d}^* = 0$  and  $\bar{\lambda}_{I^*}^* = \bar{\lambda}_{I^*}' \ge 0$ , which implies  $\bar{\lambda}^* = \bar{\lambda}'' \ge 0$ . Thus, conditions (a) and (b) in Step 3(ii) are always satisfied for  $k \in K'$  large enough. Therefore, in view of  $\rho_k \equiv \bar{\rho} < \infty$  for k large enough, Step 3(ii) implies  $\bar{\lambda}_{I^\ell}^k > \gamma_3 e_{I^\ell}$  for  $k \in K'$  large enough, which further implies that  $\bar{\lambda}_{I^\ell}^* \ge \gamma_3 e_{I^\ell} > 0$ . Hence, it follows from the complementary slackness at KKT pair  $(x^*, \bar{\lambda}'')$ ,  $0 = \bar{\lambda}_{J}''g_j(x^*) = \bar{\lambda}_{J}^*g_j(x^*) = 0$  ( $j \in I^\ell$ ). So  $g_\ell(x^*) = 0$ , which together with Proposition 1 implies that  $x^*$  is also a KKT point of (P), which contradicts the assumption that  $x^*$  is not a KKT point of (P).

*Case II*: Suppose that  $x^*$  is not a KKT point of  $(P_{\bar{\rho}})$ . And, by Lemma 5 and  $\xi_k \leq 0$ , one can deduce that  $\xi_k \rightarrow \bar{\xi} < 0, k \in K'$ . Further, this along with (13) and (23) as well as H2, shows that  $\lim_{k \in K'} (\|\bar{d}^k\|^{\nu} + \|\phi^k\|) > 0$ . So there exist a subset  $K'' \subseteq K'$  and a positive constant  $\varpi$  such that

$$\xi_k \leq \bar{\xi}/2 < 0, \qquad (\|\bar{d}^k\|^{\nu} + \|\phi^k\|) \geq \varpi > 0, \quad k \in K''.$$

The remaining proof is divided into two steps.

Step A: Show that there exists a constant  $\overline{t} > 0$  such that the step-length  $t_k \ge \overline{t}$  holds for all  $k \in K''$ .

(A1) Analyze inequality (20). First, for  $j \notin I(x^*), g_j(x^*) < 0$ , from the boundedness of  $\{(d^k, \tilde{d}^k)\}_{K''}$  and the continuity of  $g_j$ , one gets that  $g_j(x^k + td^k + t^2\tilde{d}^k) < 0$  holds for  $k \in K''$  large enough and t > 0 sufficiently small. Second, consider index  $j \in I(x^*)$ , i.e.,  $g_j(x^*) = 0$ . In view of  $I^t(x^*) \subseteq I_k^t$ , which implies  $j \in I_k$ , from Taylor expansion, formulas (9), (16) and SLE( $V_k; 1, \mu^k$ ) as well as  $\|\tilde{d}^k\| \le \|\bar{d}^k\|$ , for t > 0 small enough, we obtain that

$$g_{j}(x^{k} + td^{k} + t^{2}\tilde{d}^{k}) = g_{j}^{k} + t\nabla g_{j}^{k^{T}}d^{k} + o(t) = g_{j}^{k} + t\frac{\mu_{j}^{k} - \lambda_{j}^{k}g_{j}^{k}}{z_{j}^{k}} + o(t)$$
$$= \left(1 - t\frac{\lambda_{j}^{k}}{z_{j}^{k}}\right)g_{j}^{k} + t\frac{1 - \varphi_{k}}{z_{j}^{k}}\phi_{j}^{k} - t\varphi_{k}(\|\bar{d}^{k}\|^{\nu} + \|\phi^{k}\|) + o(t)$$
$$\leq -t\varphi_{k}(\|\bar{d}^{k}\|^{\nu} + \|\phi^{k}\|) + o(t),$$

where the last inequality follows from Lemma 4(ii),  $z_i^k \ge \underline{\varepsilon}$ ,  $\varphi_k \le 1$  and  $\phi_i^k \le 0$ .

On the other hand, taking into account  $\xi_k \leq \overline{\xi}/2 < 0$  and the boundedness of  $b_k$  (14) (by Lemma 4) as well as (15), we know that there exists a constant  $\varphi > 0$  such that  $\varphi_k \geq \varphi > 0$ ,  $k \in K''$ . So  $g_j(x^k + td^k + t^2\overline{d}^k) \leq -\varphi \varpi t + o(t) < 0$  holds for  $k \in K''$  large enough and t > 0 sufficiently small. Therefore, inequality (20) holds for t > 0 sufficiently small and  $k \in K''$  large enough.

(A2) Analyze inequality (19). From Taylor expansion and (24), one gets

$$\begin{aligned} f_{\bar{\rho}}(x^{k} + td^{k} + t^{2}\tilde{d}^{k}) - f_{\bar{\rho}}(x^{k}) - \alpha t \nabla f_{\bar{\rho}}(x^{k})^{T}d^{k} &= (1 - \alpha)t \nabla f_{\bar{\rho}}(x^{k})^{T}d^{k} + o(t) \\ &\leq (1 - \alpha)t\theta\xi_{k} + o(t) \\ &\leq (1 - \alpha)t\theta\bar{\xi}/2 + o(t) \\ &< 0. \end{aligned}$$

Hence, inequality (19) holds for  $k \in K''$  large enough and t > 0 sufficiently small. Up to now, one can conclude that there exists a constant  $\overline{t} > 0$  such that  $t_k \ge \overline{t}$  for each  $k \in K''$ .

Step B: Use  $t_k \ge \overline{t} > 0$  ( $k \in K''$ ) to bring a contradiction. Because of  $\lim_{k \in K''} f_{\bar{\rho}}(x^k) = f_{\bar{\rho}}(x^*)$ and the monotone property of { $f_{\bar{\rho}}(x^k)$ }, one knows that  $\lim_{k\to\infty} f_{\bar{\rho}}(x^k) = f_{\bar{\rho}}(x^*)$ . Further, in view of (19) and (24), it follows that for  $k \in K''$  large enough

$$f_{\bar{\rho}}(x^{k+1}) - f_{\bar{\rho}}(x^k) \leq \alpha t_k \nabla f_{\bar{\rho}}(x^k)^T d^k \leq \alpha t_k \theta \xi_k \leq \alpha \theta \bar{\xi} \bar{t}/2.$$

Passing to the limit for  $k \in K''$  and  $k \to \infty$  in the inequality above, we can bring a contradiction. Summarizing the discussions above, the whole proof of Theorem 1 is completed.

## 4 Analysis of strong and superlinear convergence

In this part, under some additional mild assumptions, we first show that the proposed algorithm is strongly convergent, that is, the whole sequence  $\{x^k\}$  is convergent. Then the unit step can be accepted and the Maratos effect can be avoided for all k large enough. At last, we prove that Algorithm A achieves superlinear convergence.

**H4** (i) The functions f(x) and g(x) are all twice continuously differentiable over  $\tilde{X}$ ; and

(ii) there exists an accumulation point  $x^*$  of the sequence  $\{x^k\}$  of iterative points with (unique) KKT multiplier  $\lambda'$  associated with (P) such that the second order sufficiency conditions (SOSC) and the strict complementarity hold, i.e., the KKT pair ( $x^*$ ,  $\lambda'$ ) of (P) satisfies  $\lambda'_{\mu(x^*)} > 0$  and

$$d^T \nabla^2_{xx} L(x^*, \lambda') d > 0, \quad \forall d \in \left\{ d \in \mathbb{R}^n : d \neq 0, \nabla g_{I(x^*)}(x^*)^T d = 0 \right\}.$$

**Remark 8** Denote the Lagrangian function of problem  $(P_{\bar{\rho}})$  by  $L_{\bar{\rho}}(x, \lambda) = f_{\bar{\rho}}(x) + \sum_{j \in I} \lambda_j \times g_j(x)$ . Then, with relation  $\lambda_{\bar{\rho}} = \lambda - \bar{\rho}\hat{e}$ , we have  $L(x, \lambda_{\bar{\rho}}) = L_{\bar{\rho}}(x, \lambda)$ . Therefore, taking into account Lemma 6(iv), it is readily checked that the SOSC with the strict complementarity for  $(P_{\bar{\rho}})$  is identical with that for (P).

**Lemma 6** Suppose that  $\tilde{X} \neq \emptyset$  and assumptions H2, H3 and H4 are satisfied (by Remark 1,  $\tilde{X} \neq \emptyset$  plus H3(i) implies  $\tilde{X}_0 \neq \emptyset$ ). Then, for any subset K such that  $\{x^k\}_K$  converges to the limit point  $x^*$  stated in H4, there exists an infinite subset  $K' \subseteq K$  such that

- (i)  $I^{i}(x^{*}) \subseteq I^{i}_{k}$  for  $k \in K'$  sufficiently large;
- (ii)  $x^*$  is a KKT point of problem ( $P_{\bar{o}}$ );
- (iii)  $\{(\bar{d}^k, \bar{\lambda}^k)\}_{K'} \to (0, \lambda^*) \text{ and } \{(d^k, \lambda^k)\}_{K'} \to (0, \lambda^*), \text{ where } \lambda^* \text{ together with } x^* \text{ is a } KKT pair of problem (P_{\bar{o}}); \text{ and}$
- (iv) the KKT multiplier  $\lambda'$  of (P) and  $\lambda^*$  of (P<sub> $\bar{\rho}$ </sub>) associated with the KKT point  $x^*$  satisfy  $\lambda' = \lambda^* \bar{\rho}\hat{e}, \lambda^*_{I(x^*)} > 0.$

*Proof* (i) From Lemma 4(ii), there exists an infinite subset  $K' \subseteq K$  such that

$$x^k \to x^*$$
,  $\lambda^k = (\bar{\lambda}^{k-1} - \bar{\rho}\hat{e}) \to \bar{\lambda}'$ ,  $k \in K'$ .

If  $(x^*, \overline{\lambda}')$  is a KKT pair of (P), then  $\overline{\lambda}' = \overline{\lambda}^*$ . Further, under H4, by [24, 29], one knows that  $I_k^i \equiv I^i(x^*)$  for  $k \in K'$  large enough. Otherwise, we have  $0 < \delta(x^*, \overline{\lambda}') \stackrel{K'}{\leftarrow} \delta(x^k, \lambda^k)$ . So, from (6),  $I^i(x^*) \subseteq I_k^i$  also holds for  $k \in K'$  large enough.

(ii) By contradiction, suppose that  $x^*$  is not a KKT point of  $(P_{\bar{\rho}})$ . Then, taking into account conclusion  $I^i(x^*) \subseteq I'_k$  ( $k \in K'$  large enough), by Case II of the proof of Theorem 1, we can bring a contradiction.

(iii) To show  $\{(\bar{d}^k, \bar{\lambda}^k)\}_{K'} \to (0, \lambda^*)$ , it is sufficient to show that  $(0, \lambda^*)$  is a unique accumulation point of  $\{(\bar{d}^k, \bar{\lambda}^k)\}_{K'}$ . Let  $(\bar{d}, \bar{\lambda})$  be any given accumulation point of  $\{(\bar{d}^k, \bar{\lambda}^k)\}_{K'}$ . Since the sequences  $\{(\bar{d}^k, \bar{\lambda}^k)\}$  and  $\{z^k\}$  are all bounded, in view of H2, H3 and  $I'_k \subseteq I'$ , there exists an infinite subset  $K'' \subseteq K'$  such that

$$I_k^{l} \equiv I', \qquad H_k \to H_*, \qquad \left(\bar{d}^k, \bar{\lambda}^k\right) \to \left(\bar{d}, \bar{\lambda}\right), \qquad z^k \to z^*, \quad k \in K''.$$
(33)

Now, passing to the limit for  $k \in K''$  and  $k \to \infty$  in SLE( $V_k$ ;1,0), we deduce that  $(\bar{d}, \bar{\lambda}_{\hat{l}})$  $(\hat{I} := I^{\ell} \cup I')$  solves SLE (32). Further, it follows from Lemma 4(i) that the coefficient matrix of SLE (32) is nonsingular. Thus the solution of (32) is unique. On the other hand, in view of  $I'(x^*) \subseteq I', I^{\ell}(x^*) = I^{\ell}$  and  $(x^*, \lambda^*)$  being a KKT pair of  $(P_{\bar{\rho}})$ , we know that  $(0, \lambda_{\hat{l}}^*)$  is also a solution to system (32). Therefore  $(\bar{d}, \bar{\lambda}_{\hat{l}}) = (0, \lambda_{\hat{l}}^*)$ , this further implies that  $(\bar{d}, \bar{\lambda}) = (0, \lambda^*)$  and  $(0, \lambda^*)$  is a unique limit point of  $\{(\bar{d}^k, \bar{\lambda}^k)\}_{K'}$ .

Finally, conclusion  $\{(d^k, \lambda^k)\}_{K'} \to (0, \lambda^*)$  follows from  $\{(\bar{d}^k, \bar{\lambda}^k)\}_{K'} \to (0, \lambda^*)$  and (31).

(iv) By Proposition 1 and  $g_{\ell}(x^*) = 0$ , we have  $\lambda' = \lambda^* - \bar{\rho}\hat{e}$ , and  $\lambda^*_{l'(x^*)} = \lambda'_{l'(x^*)} > 0$  by H4(ii). Further, in view of  $\bar{d}^k \to 0$ ,  $\bar{\lambda}^k \to \lambda^* \ge 0$ ,  $k \in K'$ , one knows that conditions (a) and (b) in Step 3(ii) hold for k large enough. Therefore, taking into account  $\rho_k \equiv \bar{\rho}$  for k large enough, it follows that  $\bar{\lambda}^k > \gamma e_{l^\ell}$  by Step 3(ii), so  $\lambda^*_{\ell} \ge \gamma_3 e_{l^\ell} > 0$ . Therefore  $\lambda^*_{l(x^*)} > 0$  holds.

**Remark 9** In view of  $I^{\ell}(x^*) = I^{\ell}$ , from H3, H4 and Lemma 6(ii), (iv), the following conclusion holds: The LICQ, SOSC and strict complementarity of problem (P) and problem (P<sub> $\bar{\rho}$ </sub>) are satisfied at their KKT pair ( $x^*, \lambda'$ ) and ( $x^*, \lambda^*$ ), respectively.

In view of Remark 9, similarly to the proof of [21], Theorem 4.1, Lemma 4.2, we have the following result.

**Theorem 2** Suppose that  $\tilde{X} \neq \emptyset$  and assumptions H2, H3 and H4 are satisfied. Then

- (i)  $x^k \rightarrow x^*$ , *i.e.*, Algorithm A is strongly convergent;
- (ii)  $(\overline{d}^k, \overline{\lambda}^k) \to (0, \lambda^*), (d^k, \lambda^k) \to (0, \lambda^*), z^k \to \min\{\max\{\underline{\varepsilon}e_I, \lambda^*\}, \overline{\varepsilon}e_I\}, and$
- (iii)  $\phi^k = 0, \mu^k = -\varphi_k \|\bar{d}^k\|^v z_{I_k}^k, I_k^i \equiv I^i(x^*) \text{ and } I_k \equiv I_* := I(x^*) \text{ if } k \text{ is sufficiently large.}$

**Lemma** 7 Suppose that the hypotheses in Lemma 6 hold, and assume that the boundary parameters  $\underline{\varepsilon}$  and  $\overline{\varepsilon}$  satisfy

$$\underline{\varepsilon} \le \min\{\lambda_j^*, j \in I_*\}, \qquad \overline{\varepsilon} \ge \max\{\lambda_j^*, j \in I_*\}.$$
(34)

Then

$$z_{I_*}^k \to \lambda_{I_*}^*, \qquad \omega_k = o(\|d^k\|^2), \tag{35}$$

and the solution  $(\tilde{d}^k, \tilde{\lambda}_{I_*}^k)$  of SLE $(V_k; 0, \tilde{\mu}^k)$  satisfies

$$\|\left(\tilde{d}^k, \tilde{\lambda}_{I_*}^k\right)\| = O(\hat{\omega}_k) = o\left(\left\|d^k\right\|\right), \qquad \hat{\omega}_k \left\|d^k\right\| = o(\omega_k), \tag{36}$$

where

$$\hat{\omega}_{k} = \max\{|z_{j}^{k}/\lambda_{j}^{k} - 1| \cdot \|d^{k}\|, j \in I_{*}; \|d^{k}\|^{2}\}.$$
(37)

Furthermore, the correction direction  $\tilde{d}^k$  in Step 6 is always yielded by the solution of  $SLE(V_k; 0, \tilde{\mu}^k)$ .

*Proof* First, from the given conditions and Theorem 2(iv), relation  $z_{I_*}^k \to \lambda_{I_*}^*$  is at hand. Further, this, together with Theorem 2(ii), shows that  $z_j^k/\lambda_j^k \to 1$  for  $j \in I_*$ . So, it follows that  $\omega_k = o(||d^k||^2)$  from (18).

Second, we prove relation (36). From Theorem 2(ii), (iii), we know that  $\mu^k = -\varphi_k \| \overline{d}^k \|^{\nu} \times z_{I_*}^k \to 0, k \to \infty$ . This, along with SLE( $V_k$ ; 1, 0) and SLE( $V_k$ ; 1,  $\mu^k$ ) as well as Lemma 4(i), implies that there exists a positive constant *c* such that

$$\|d^{k} - \bar{d}^{k}\| \le c \|\bar{d}^{k}\|^{\nu}, \qquad \|\lambda^{k} - \bar{\lambda}^{k}\| \le c \|\bar{d}^{k}\|^{\nu}, \qquad \|d^{k}\| \sim \|\bar{d}^{k}\|.$$
 (38)

Therefore, from definition (17) of  $\tilde{\mu}^k$ , Taylor expansion and  $SLE(V_k; 1, \mu^k)$ , one has for  $j \in I_*$ 

$$\begin{split} \tilde{\mu}_{j}^{k} &= -\omega_{k} - z_{j}^{k} g_{j} \left( x^{k} + d^{k} \right) \\ &= -\omega_{k} - z_{j}^{k} \left( g_{j}^{k} + \nabla g_{j}^{k^{T}} d^{k} \right) + O(\|d^{k}\|^{2}) \\ &= -\omega_{k} - z_{j}^{k} \left( 1 - z_{j}^{k} / \lambda_{j}^{k} \right) \nabla g_{j}^{k^{T}} d^{k} + O(\|d^{k}\|^{2}) \\ &= -\omega_{k} + O(\max\{|z_{i}^{k} / \lambda_{i}^{k} - 1| \cdot \|d^{k}\|, i \in I_{*}\}) + O(\|d^{k}\|^{2}) \\ &= -\omega_{k} + O(\hat{\omega}_{k}) + O(\|d^{k}\|^{2}) \\ &= o(\|d^{k}\|^{2}) + O(\hat{\omega}_{k}) + O(\|d^{k}\|^{2}). \end{split}$$

Obviously, definition (37) implies  $||d^k||^2 = O(\hat{\omega}_k)$ . Thus  $||\tilde{\mu}^k|| = O(\hat{\omega}_k) = o(||d^k||)$ . Therefore, from SLE( $V_k$ ; 0,  $\tilde{\mu}^k$ ) and Lemma 4, it is clear that the first relation of (36) holds. Finally, relation  $\hat{\omega}_k ||d^k|| = o(\omega_k)$  follows from definitions (37) of  $\hat{\omega}_k$  and (18) of  $\omega_k$ .

To ensure the step size  $t_k \equiv 1$  for k large enough, which is necessary to obtain superlinear convergence, similarly or refer to [8, 9], the following second order approximate condition is necessary.

**H5** Assume that the relation  $||P_k(\nabla_{xx}^2 L_{\bar{\rho}}(x^k, \lambda^k) - H_k)P_k\bar{d}^k|| = o(||\bar{d}^k||)$  holds, where the projective matrix  $P_k$  is defined by  $P_k = E_n - N_k(N_k^T N_k)^{-1}N_k^T$  with  $N_k = \nabla g_{I_k}(x^k)$  and *n*-order unit matrix  $E_n$ .

#### Remark 10 (About H5)

- (i) Due to  $I_* = I(x^*) = I^{\ell} \cup I^{\ell}(x^*)$ , one knows from H3(i) that matrix  $N_k \to \nabla g_{I_*}(x^*)$  which is column full rank, and matrix  $P_k$  is well defined when k is large enough.
- (ii) The 2-sided projection second order approximation H5 above, also used in [1, 8, 18, 22], is milder than the 1-sided projection second order approximation:

**H5**<sup>+</sup>  $||P_k(\nabla_{xx}^2 L_{\bar{\rho}}(x^k, \lambda^k) - H_k)\bar{d}^k|| = o(||\bar{d}^k||).$ 

Both the two can ensure the step unit is achieved. However, the associated algorithms can attain (one-step) q-superlinear convergence under the latter, and only two-step superlinear convergence under the former.

(iii) In view of relation (38), assumptions H5 and H5<sup>+</sup> are equivalent to  $\|P_k(\nabla^2_{xx}L_{\bar{\rho}}(x^k,\lambda^k) - H_k)P_kd^k\| = o(\|d^k\|) \text{ and } \|P_k(\nabla^2_{xx}L_{\bar{\rho}}(x^k,\lambda^k) - H_k)d^k\| = o(\|d^k\|),$ respectively.

**Theorem 3** Suppose that  $\tilde{X} \neq \emptyset$  and hypotheses H2-H5 hold, and assume that the boundary parameters  $\underline{\varepsilon}$  and  $\overline{\varepsilon}$  satisfy (34). Then the step size  $t_k$  of Algorithm A always equals one, *i.e.*,  $t_k \equiv 1$  for k large enough.

*Proof* (i) Discuss (20). For  $j \notin I_* = I(x^*)$ ,  $g_j(x^*) < 0$ , using the continuity of  $g_j$  and  $(x^k, d^k, \tilde{d}^k) \to (x^*, 0, 0), k \to \infty$ , we know that (20) holds for t = 1 and k large enough.

For  $j \in I_* = I(x^*) = I_k$ , in view of SLE( $V_k; 1, \mu^k$ ), Theorem 2(iii) and  $||d^k|| \sim ||\bar{d}^k||$  as well as  $\lambda_i^k \to \lambda_i^* > 0$ , we have

$$z_{j}^{k} \nabla g_{j}^{k^{T}} d^{k} + g_{j}^{k} \lambda_{j}^{k} = \mu_{j}^{k} = o(\|d^{k}\|^{2}), \quad g_{j}^{k} = O(\|d^{k}\|).$$
(39)

Again, taking into account SLE( $V_k$ ; 0,  $\tilde{\mu}^k$ ), one has

$$z_j^k \nabla g_j^{k^T} \tilde{d}^k + z_j^k g_j (x^k + d^k) + \tilde{\lambda}_j^k g_j^k = -\omega_k$$

This, together with (39), (36) and  $\omega_k = o(||d^k||^2)$ , shows that

$$g_{j}(x^{k} + d^{k}) + \nabla g_{j}^{k^{T}} \tilde{d}^{k} = -\frac{\omega_{k}}{z_{j}^{k}} + \tilde{\lambda}_{j}^{k} O(\|d^{k}\|) = -\frac{\omega_{k}}{z_{j}^{k}} + O(\hat{\omega}_{k} \|d^{k}\|)$$
$$= -\frac{\omega_{k}}{z_{j}^{k}} + o(\omega_{k}) = O(\omega_{k}) = o(\|d^{k}\|^{2}).$$
(40)

Further, using Taylor expansion and (36), one has

$$g_{j}(x^{k} + d^{k} + \tilde{d}^{k}) = g_{j}(x^{k} + d^{k}) + \nabla g_{j}(x^{k} + d^{k})^{T}\tilde{d}^{k} + O(\|\tilde{d}^{k}\|^{2})$$
  
$$= g_{j}(x^{k} + d^{k}) + \nabla g_{j}^{k^{T}}\tilde{d}^{k} + O(\|d^{k}\| \cdot \|\tilde{d}^{k}\|) + O(\|\tilde{d}^{k}\|^{2})$$
  
$$= -\frac{\omega_{k}}{z_{j}^{k}} + o(\omega_{k}) + O(\hat{\omega}_{k}\|d^{k}\|) = -\frac{\omega_{k}}{z_{j}^{k}} + o(\omega_{k}) = o(\|d^{k}\|^{2}).$$

Hence, we can conclude from the fourth equality above that inequality (20) holds for  $j \in I_*$ , t = 1 and k large enough since  $z_j^k \to \lambda_j^* > 0$ .

(ii) Analyze (19). From Taylor expansion and (36), it follows that

$$w_{k} := f_{\bar{\rho}} \left( x^{k} + d^{k} + \tilde{d}^{k} \right) - f_{\bar{\rho}} \left( x^{k} \right) - \alpha \nabla f_{\bar{\rho}} \left( x^{k} \right)^{T} d^{k}$$
  
$$= \nabla f_{\bar{\rho}} \left( x^{k} \right)^{T} \left( d^{k} + \tilde{d}^{k} \right) + \frac{1}{2} \left( d^{k} \right)^{T} \nabla^{2} f_{\bar{\rho}} \left( x^{k} \right) d^{k} - \alpha \nabla f_{\bar{\rho}} \left( x^{k} \right)^{T} d^{k} + o \left( \left\| d^{k} \right\|^{2} \right).$$
(41)

On the other hand, from SLE( $V_k$ ; 1,  $\mu^k$ ), we have

$$H_k d^k + \nabla f_{\bar{\rho}} \left( x^k \right) + \sum_{j \in I_*} \lambda_j^k \nabla g_j^k = 0, \tag{42}$$

which, together with (36), gives

$$\nabla f_{\bar{\rho}} \left( x^k \right)^T d^k = - \left( d^k \right)^T H_k d^k - \sum_{j \in I_*} \lambda_j^k \nabla g_j^{k^T} d^k, \tag{43}$$

$$\nabla f_{\bar{\rho}}(x^{k})^{T}(d^{k}+\tilde{d}^{k}) = -(d^{k})^{T}H_{k}d^{k} - \sum_{j\in I_{*}}\lambda_{j}^{k}\nabla g_{j}^{k^{T}}(d^{k}+\tilde{d}^{k}) + o(\|d^{k}\|^{2}).$$
(44)

Therefore, by (40) and Taylor expansion for  $g_i(x^k + d^k)$  at point  $x^k$ , one yields

$$\nabla g_{j}^{k^{T}}(d^{k}+\tilde{d}^{k})=-g_{j}^{k}-\frac{1}{2}(d^{k})^{T}\nabla^{2}g_{j}(x^{k})d^{k}+o(\|d^{k}\|^{2}), \quad j\in I_{*}.$$

This, together with (44), shows that

$$\nabla f_{\bar{\rho}}(x^k)^T (d^k + \tilde{d}^k) = (d^k)^T \left( -H_k + \frac{1}{2} \sum_{j \in I_*} \lambda_j^k \nabla^2 g_j(x^k) \right) d^k$$
$$+ \sum_{j \in I_*} \lambda_j^k g_j^k + o(\|d^k\|^2).$$
(45)

On the other hand, the first relation of (39) gives

$$\lambda_{j}^{k} \nabla g_{j}^{k^{T}} d^{k} = -((\lambda_{j}^{k})^{2}/z_{j}^{k})g_{j}^{k} + o(||d^{k}||^{2}), \quad j \in I_{*}.$$

This, along with (43), shows that

$$(d^{k})^{T}H_{k}d^{k} = -\nabla f_{\bar{\rho}}(x^{k})^{T}d^{k} + \sum_{j\in I_{*}}\frac{(\lambda_{j}^{k})^{2}}{z_{j}^{k}}g_{j}^{k} + o(\|d^{k}\|^{2}).$$
(46)

Again, substituting (45) into (41), one has

$$w_{k} = \sum_{j \in I_{*}} \lambda_{j}^{k} g_{j}^{k} + \frac{1}{2} (d^{k})^{T} (\nabla_{xx}^{2} L_{\bar{\rho}}(x^{k}, \lambda^{k}) - H_{k}) d^{k} - \frac{1}{2} (d^{k})^{T} H_{k} d^{k} - \alpha \nabla f_{\bar{\rho}}(x^{k})^{T} d^{k} + o(||d^{k}||^{2}).$$

Therefore, substituting (46) into the relation above, we have

$$w_{k} = \left(\frac{1}{2} - \alpha\right) \nabla f_{\bar{\rho}} \left(x^{k}\right)^{T} d^{k} + \frac{1}{2} \left(d^{k}\right)^{T} \left(\nabla_{xx}^{2} L_{\bar{\rho}} \left(x^{k}, \lambda^{k}\right) - H_{k}\right) d^{k} + \sum_{j \in I_{*}} \lambda_{j}^{k} \left(1 - \frac{\lambda_{j}^{k}}{2z_{j}^{k}}\right) g_{j}^{k} + o\left(\left\|d^{k}\right\|^{2}\right).$$
(47)

On the other hand, from the definition of the projection matrix  $P_k$ , we get

$$d^{k} = P_{k}d^{k} + d_{0}^{k}, \quad d_{0}^{k} = N_{k}(N_{k}^{T}N_{k})^{-1}N_{k}^{T}d^{k}.$$

Furthermore, in view of SLE( $V_k$ ; 1,  $\mu^k$ ), Theorem 2(iii) and the above division, one has

$$N_{k}^{T}d^{k} = Z_{k}^{-1}(\mu^{k} - \operatorname{diag}(g_{I_{*}}^{k})\lambda_{I_{*}}^{k}), \qquad d_{0}^{k} = o(\|d^{k}\|^{2}) + O(\|g_{I_{*}}^{k}\|).$$
(48)

Thus, relation (47), together with the relations above and H5, implies that

$$w_{k} = \sum_{j \in I_{*}} \lambda_{j}^{k} \left(1 - \frac{\lambda_{j}^{k}}{2z_{j}^{k}}\right) g_{j}^{k} + o\left(\left\|d^{k}\right\|^{2}\right) + \left(\frac{1}{2} - \alpha\right) \nabla f_{\bar{\rho}}\left(x^{k}\right)^{T} d^{k} + \frac{1}{2} \left(d_{0}^{k} + P_{k} d^{k}\right)^{T} \left(\nabla_{xx}^{2} L_{\bar{\rho}}\left(x^{k}, \lambda^{k}\right) - H_{k}\right) \left(d_{0}^{k} + P_{k} d^{k}\right) = \left(\frac{1}{2} - \alpha\right) \nabla f_{\bar{\rho}}\left(x^{k}\right)^{T} d^{k} + \sum_{j \in I_{*}} \lambda_{j}^{k} \left(1 - \frac{\lambda_{j}^{k}}{2z_{j}^{k}}\right) g_{j}^{k} + O\left(\left\|g_{I_{*}}^{k}\right\|\right) + o\left(\left\|d^{k}\right\|^{2}\right).$$
(49)

On the other hand, taking into account Lemma 6(ii), Lemma 7 and Theorem 2, one has (when  $k \to \infty$ )

$$\lambda_j^k \to \lambda_j^* > 0, \qquad \lambda_j^k \left( 1 - \frac{\lambda_j^k}{2z_j^k} \right) \to \frac{\lambda_j^*}{2} > 0, \quad \forall j \in I_*.$$
 (50)

Further, relations (24), (13), (25), (23) and  $\|d^k\| \sim \|\overline{d}^k\|$  as well as H2 yield

$$\nabla f_{\bar{\rho}}(x^{k})^{T}d^{k} \leq \theta \xi_{k} \leq \theta \nabla f_{\bar{\rho}}(x^{k})^{T}\bar{d}^{k}$$

$$= -\theta \left(\bar{d}^{k}\right)^{T}Q_{k}\bar{d}^{k} \leq \theta a \|\bar{d}^{k}\|^{2}$$

$$= -\theta a \|d^{k}\|^{2} + o(\|d^{k}\|^{2}).$$
(51)

Therefore, for *k* large enough, relations (49)-(51) show that  $w_k \leq (\alpha - \frac{1}{2})\theta a ||d^k||^2 + o(||d^k||^2) \leq 0$ . Thus, inequality (19) holds for t = 1 and *k* large enough, and the entire proof of Theorem 3 is finished.

Finally, based on Theorem 3, by similar analysis in [1, 18, 22] (for two-step superlinear convergence) and [21], Appendix A (for one-step superlinear convergence), we can prove the following rate of superlinear convergence.

**Theorem 4** Suppose that  $\tilde{X} \neq \emptyset$  and the hypotheses H2-H5 hold. If the boundary parameters  $\underline{\varepsilon}$  and  $\overline{\varepsilon}$  satisfy (34), then the proposed Algorithm A is two-step superlinearly convergent, *i.e.*,  $\|x^{k+2} - x^*\| = o(\|x^k - x^*\|)$ . Moreover, if H5 is strengthened as H5<sup>+</sup>, then Algorithm A is one-step superlinearly convergent, *i.e.*,  $\|x^{k+1} - x^*\| = o(\|x^k - x^*\|)$ .

#### **5** Numerical experiments

In this section, to show the practical effectiveness of Algorithm A, we test 59 typical problems from [33]. The numerical experiments are implemented by using MATLAB R2013a, and on a PC with Inter(R) Core(TM) i5-4590 3.30 GHz CPU, 4.00 GB RAM. The details about the implementation are described as follows.

### 5.1 Computing matrix *H<sub>k</sub>*

During the process of iteration, to ensure the boundedness of  $\{H_k\}$ , by modifying the computing technique in [1] for the approximate Lagrangian Hessian, we introduce a slightly new computing method for the approximate Hessian matrix  $H_k$  in Step 2 as follows from second order derivative information. Denote vector  $\hat{z}^k$  and matrix  $\mathcal{M}_k$  by

$$\hat{z}^{k} = \left(z_{I_{k}^{k}}^{k}, z_{I_{k}^{k}}^{k}, \mathbf{0}_{I^{i} \setminus I_{k}^{i}}\right), \tag{52}$$

$$\mathcal{M}_{k} = \nabla_{xx}^{2} L_{\rho_{k}} \left( x^{k}, \hat{z}^{k} \right) - \sum_{j \in I} \frac{\hat{z}_{j}^{k}}{g_{j}^{k}} \nabla g_{j}^{k} \nabla g_{j}^{k^{T}}.$$
(53)

Then compute the smallest eigenvalue  $\vartheta_{\min}^k$  of matrix  $\mathcal{M}_k$ , and yield

$$\theta_{k} = \begin{cases}
0, & \text{if } \vartheta_{\min}^{k} > \underline{\varepsilon}; \\
-\vartheta_{\min}^{k} + \underline{\varepsilon}, & \text{if } |\vartheta_{\min}^{k}| \le \underline{\varepsilon}; \\
2|\vartheta_{\min}^{k}|, & \text{otherwise.} 
\end{cases}$$
(54)

Subsequently, compute matrix  $H_k$  in Step 2 by

$$H_{k} = \begin{cases} \nabla_{xx}^{2} L_{\rho_{k}}(x^{k}, \hat{z}^{k}) + \theta_{k} E_{n}, & \text{if } \rho_{k} \leq \overline{\varepsilon} \text{ and } \theta_{k} \leq \overline{\varepsilon}; \\ E_{n}, & \text{otherwise,} \end{cases}$$
(55)

where the positive parameters  $\underline{\varepsilon}$  and  $\overline{\varepsilon}$  same as the ones in Algorithm A are sufficiently small and sufficiently large, respectively.

The sequence  $\{H_k\}$  of matrices defined above possesses nice properties as follows.

**Theorem 5** Suppose that  $\tilde{X} \neq \emptyset$  and assumptions H3 and H4(i) hold. Yield matrix  $H_k$  in Step 2 by (52)-(55). If the sequence  $\{x^k\}$  yielded by Algorithm A is bounded, then the following results hold.

- (i) The sequence {H<sub>k</sub>} is bounded and satisfies the positive definite restriction (27) with constant a = ε, so H2 holds.
- (ii) In addition, assume that H4(ii) and (34) are satisfied. Then, for k large enough, matrix  $\mathcal{M}_k$  is positive definite,  $\vartheta_{\min}^k > 0$  and  $\theta_k < \underline{\varepsilon}$ . Therefore,  $H_k$  is always yielded by the first case in (55), i.e.,

$$H_k = \nabla_{xx}^2 L_{\rho_k} \left( x^k, \hat{z}^k \right) + \theta_k E_n, \quad \text{when } k \text{ is large enough.}$$
(56)

Further, it follows that

$$\lim_{k\to\infty} \left\| \nabla_{xx}^2 L_{\bar{\rho}}(x^k,\lambda^k) - H_k \right\| = \lim_{k\to\infty} \theta_k \leq \underline{\varepsilon} \approx 0$$

when  $\varepsilon$  is sufficiently small. In this sense, we say assumption H5<sup>+</sup> is almost satisfied.

*Proof* (i) By the boundedness of  $\{(x^k, \hat{z}^k)\}$ , in view of H4(i), the boundedness of  $\{H_k\}$  follows immediately from (55). To show the second claim of part (i), it is sufficient to discuss the case  $H_k \neq E_n$ . For any  $d \in \mathbb{R}^n$ , from (22) and (52)-(55), one has

$$d^{T}Q_{k}d = d^{T}\left(H_{k} - \sum_{j \in I} \frac{\hat{z}_{j}^{k}}{g_{j}^{k}} \nabla g_{j}^{k} \nabla g_{j}^{k^{T}}\right) d = d^{T}(\mathcal{M}_{k} + \theta_{k}E_{n})d.$$

$$(57)$$

On the other hand, due to the symmetry of matrix  $\mathcal{M}_k$ , there exists a real orthogonal matrix  $U_k$  such that  $\mathcal{M}_k = U_k \operatorname{diag}(\vartheta^k) U_k^T$ , where  $\vartheta^k = (\vartheta_i^k)$  is the eigenvalue vector of  $\mathcal{M}_k$ . Therefore,

$$d^{T}\mathcal{M}_{k}d = d^{T}U_{k}\operatorname{diag}(\vartheta^{k})U_{k}^{T}d = (U_{k}^{T}d)^{T}\operatorname{diag}(\vartheta^{k})U_{k}^{T}d$$
$$= \sum_{i=1}^{n} (u_{i}^{k})^{2}\vartheta_{i}^{k} \ge \vartheta_{\min}^{k} \sum_{i=1}^{n} (u_{i}^{k})^{2} = \vartheta_{\min}^{k} (U_{k}^{T}d)^{T}U_{k}^{T}d = \vartheta_{\min}^{k} \|d\|^{2},$$

where  $u^k = (u_i^k, i = 1, ..., n) = U_k^T d$ . This, along with (57) and (54), shows that

$$d^{T}Q_{k}d = d^{T}\mathcal{M}_{k}d + \theta_{k}||d||^{2} \geq (\vartheta_{\min}^{k} + \theta_{k})||d||^{2} \geq \underline{\varepsilon}||d||^{2}.$$

So request (27) is satisfied with  $a = \underline{\varepsilon}$ .

(ii) First, under the given conditions, one knows that all the assumptions requested in Theorem 2 and Lemma 7 are satisfied. So, by Theorems 2 and Lemma 7, it follows that

$$I_{k}^{i} \equiv I^{i}(x^{*}), \qquad \rho_{k} \equiv \bar{\rho}, (x^{k}, \hat{z}^{k}) \to (x^{*}, \lambda^{*}),$$

$$\nabla_{xx}^{2} L_{\bar{\rho}}(x^{k}, \hat{z}^{k}) \to \nabla_{xx}^{2} L_{\bar{\rho}}(x^{*}, \lambda^{*}) = \nabla_{xx}^{2} L(x^{*}, \lambda').$$
(58)

Therefore, taking the above results and the SOSC in H4(ii) into account, it is not difficult to show that matrix  $\mathcal{M}_k$  is positive definite when k is large enough, and this together with (54)-(55) and (58) further implies that the remaining claims in part (ii) hold.

Based on Theorem 5, comparing with [1], the following remark is given.

**Remark 11** The technique (52)-(55) yielding matrix  $H_k$  is a modification of the one in [1], Section 5.1, and they are unlike in two points. First, the former introduced in this work can ensure the boundedness of { $H_k$ } (see Theorem 5(i)), which plays a key role in the analysis of global and superlinear convergence; especially, in ensuring the penalty parameter  $\rho_k$ is increased at most finitely many times. However, the latter in [1], Section 5.1, cannot ensure the boundedness of the sequence { $W_k$ } yielded by [1], Section 5.1 (corresponds to { $H_k$ } in this paper) since this strict relies on the bounded property of { $(\rho_k, \theta_k)$ }, and one of the necessary conditions for the boundedness of { $(\rho_k, \theta_k)$ } is just the boundedness of  $\{W_k\}$  (see the proof of [1], Lemma 4.1). Second, by introducing  $\hat{z}^k$  in the computation technique (52)-(55) rather than  $z^k$  (corresponds the one denoted in [1], Section 5.1), the assumption H5<sup>+</sup> is almost satisfied (see Theorem 5(iii)). If one still uses  $z^k$  rather that  $\hat{z}^k$  in (52)-(55), then the second order approximate condition H5<sup>+</sup> even H5 would be difficult to be satisfied since  $z^k_{I^1 \setminus I^1(x^*)} \rightarrow \underline{\varepsilon} e_{I^1 \setminus I^1(x^*)} > 0 = \lambda^*_{I^1 \setminus I^1(x^*)}$  (by Theorem 2(iv)). Of course, in view of  $\lim_{k\to\infty} ||z^k - \hat{z}^k|| = \underline{\varepsilon}$  which is small enough, it can be thought that the numerical performances with  $z^k$  and  $\hat{z}^k$  should possess no distinct difference.

## 5.2 Choices of parameters

The parameters in our numerical testing are chosen as follows:

<i>r</i> = 0.5,	$\alpha$ = 0.45,	$\theta$ = 0.99,	$\beta = 0.5$ ,
σ = 0.8,	$\xi = 2.5,$	$\overline{\varepsilon} = 10^5$ ,	$\underline{\varepsilon} = 10^{-5}$ ,
v = 3,	$\rho_0 = p = 1$ ,	$\vartheta = 2,$	<i>M</i> = 100,
$\gamma_1 = 0.1$ ,	$\gamma_2 = \gamma_3 = 0$	$0.01,   z_0 =$	: (1, , 1).

**Remark 12** (Analysis for lower bound  $\underline{\varepsilon}$  and upper bound  $\overline{\varepsilon}$ ) First, by Theorems 1 and 2, it is known that, in terms of global and strong convergence of Algorithm A, there is no additional request on the lower bound  $\underline{\varepsilon}$  and upper bound  $\overline{\varepsilon}$ , i.e., any two positive constants should be suitable. Second, if one considers the rate of convergence of Algorithm A, by Theorem 4, parameters  $\underline{\varepsilon}$  and  $\overline{\varepsilon}$  should be sufficiently small and sufficiently large, respectively. However, if the initial values of  $\underline{\varepsilon}$  and  $\overline{\varepsilon}$  are chosen too small and/or too large, the numerical performances should be unstable. An ideal approach is to decrease  $\underline{\varepsilon}$  and increase  $\overline{\varepsilon}$  based on values min $\{z_i^k, i \in I\}$  and max $\{z_i^k, i \in I\}$ , respectively.

### 5.3 Termination rules

During the process of iteration, the implementation is terminated successfully if one of the following two conditions is satisfied:

(i)  $\|\Phi(x^k,\lambda^k)\| < 10^{-5}$ ; (ii)  $\|\bar{d}^k\| < 10^{-5}$  and  $\max\{-\bar{\lambda}_i^k, j \in I^i\} < 10^{-5}$ .

#### 5.4 Numerical reports

For the sake of comparing equally, the same initial points as in [33] should be selected. However, Algorithm A starts with a feasible interior point, namely,  $x^0 \in \tilde{X}_0$ , and some initial points given in [33] do not satisfy this request. So, other initial points for these problems are selected and listed in Table 1.

 Table 1 Feasible initial interior points for testing problems

Prob.	<i>x</i> <sup>0</sup>	Prob.	x <sup>0</sup>	Prob.	x <sup>0</sup>
HS6	(2, 2)	HS32	(0.1, 0.7, 0.1)	HS63	(1, 1, 1)
HS7	(0, 1)	HS39	(-1,-1,0,0)	HS73	(1, 1, 1, 1)
HS8	(4, 2)	HS40	(2,-1,0,1)	HS78	(-2, 1.5, 1, -1, -1)
HS25	(50, 25, 1.5)	HS42	(1, 1, 1, 1)	HS79	(0, 0, 0, 0, 0)
HS26	(0.2, 0.2, 0.2)	HS52	(1,-0.5,-1,0,1)	HS80	(-2, -1.5, 1, 1, 1)
HS27	(0,0,0)	HS53	(-6 2 2 2 2)	HS81	(-1.7, 1, 1.5, -0.8, -0.8)
HS28	(0,0,0)	HS60	(0,0,0)	HS107	(0.8 0.8 10 10 1 1 1 1 1)
HS111	(-1, -1, -1, -1,	, -1, -1, -1	,-1,-1,-1)	HS114	(1,745 12,000 110 3,048 1,974 89.2 92.8 8 3.6 145)

The numerical results are reported and compared with the ones from [1] in Table 2, where the columns have the following meanings:

- Prob.: the problem number given in [33];
- Itr: the number of iterations;
- Nf: the number of function evaluations for *f*;
- N: the total number of function evaluations for  $g_i$ ;
- $\bar{\rho}$ : the final value of  $\rho_k$ ;
- Tcpu: the CPU time (seconds);
- $f_{\text{final}}$ : the objective function value at the final iterate.

Same as the way of counting the number of iterations in [1], due to only a little change at the right side vector of SLE (10) in the loop between Step 3(i) and Step 3(ii), which leads to low computational cost, the number of this loop is not counted in the total number of iterations Itr.

From Table 2 it is clear that, for almost all test problems, the two algorithms (Algorithms A and the one in [1]) have the same optimal objective value. Relatively speaking, it also shows that Algorithm A is a promising one in terms of the CPU time, the number of function evaluations Nf and the total number of function evaluations N.

In particular, the following four performances are worth to be mentioned. First, for HS66, HS107 and HS111, the algorithm [1] yields the associated  $f_{\text{final}}$  after 1000 iterations for each problem, while Algorithm A needs only 15, 41 and 26 iterations, respectively. Second, for HS107, the two algorithms yield two large different final objective function values  $f_{\text{final}}$ , namely, 3,748e–08 and 5.0545e+38. Third, for HS118 with the same dimension as HS117, Algorithm A has a good numerical performance, while it is not reported in [1]. Fourth, for HS54, HS75, HS85 and HS117, Algorithm A fails to produce an invertible coefficient matrix after some iterations, then it cannot obtain the optimal objective value, so they are not listed in Table 2.

For more clarity, we also give the output of Algorithm A for problem HS8 in Table 3. It is found from  $\rho_k$ -column of Table 3 that the penalty parameter needs to be increased one, two, four and six times at 2nd, 3rd, 4th and 5th iterations, respectively; and it can be fixed in the subsequent iterations.

#### 6 Conclusions

In this paper, based on a simple and effective penalty parameter update rule and using the idea of primal-point interior method, a primal-dual interior point QP-free algorithm for nonlinear constrained optimization is proposed and analyzed. A 'working set' technique for estimating the active set is used in this work, then we need to solve only two or three reduced systems of linear equations with the same coefficient matrix at each iteration. Under suitable CQ and assumptions including a relaxed positive definite restriction on the Lagrangian Hessian estimate  $H_k$ , but without the isolatedness of the stationary points, the proposed algorithm is globally and superlinearly convergent. Moreover, a slightly new computation technique for  $H_k$  based on second order derivative information is introduced such that the associated assumptions, i.e., the boundedness of  $\{H_k\}$ , the relaxed positive definiteness and the 1-sided projection second order approximation H5<sup>+</sup>, are all (almost) satisfied. The numerical experiments based on the proposed computation technique for  $H_k$  show that the proposed algorithm is promising.

Prob.	n	me	mi	Algorithm A in this paper					Algorithm from [1]			
		-	-	ltr	Nf	Ν		<b>f</b> <sub>final</sub>	Тсри	ltr	Ā	<b>f</b> <sub>final</sub>
HS1	2	0	1	28	73	70	1	1.7825e-18	0.02	24	1	6.5782e-27
HS3	2	0	1	6	7	8	1	2.3501e-06	0.01	4	1	8.5023e-09
HS4	2	0	2	7	13	29	1	2.6667e+00	0.01	4	1	2.6667e+00
HS5	2	0	4	5	13	47	1	-1.9132e+00	0.01	6	1	-1.9132e+00
HS6	2	1	0	9	364	718	1	2.4199e-07	0.03	7	2	0.0000e+00
HS7	2	1	0	8	15	28	32	-1.7320e+00	0.01	9	2	-1.7321e+00
HS8	2	2	0	9	16	59	8,192	-1.0000e+00	0.01	14	1	-1.0000e+00
HS9	2	1	0	18	34	66	8,192	-4.9985e-01	0.02	10	1	-5.0000e+01
HS12	2	0	1	9	19	39	1	-3.0000e+01	0.01	5	1	-3.0000e+01
HS24	2	0	5	16	29	179	1	-1.0000e+00	0.02	14	1	-1.0000e+00
HS25	3	0	6	1	1	6	1	9.4934e-31	0.01	62	1	1.8185e-16
HS26	3	1	0	16	76	142	2	1.6085e-04	0.02	19	2	2.8430e-12
HS27	3	1	0	28	484	939	4	3.9958e-02	0.05	14	32	4.0000e-02
HS28	3	1	0	11	38	71	1,024	7.5674e-08	0.01	6	1	0.0000e+00
HS29	3	0	1	11	24	53	1	-2.2627e+01	0.01	8	1	-2.2627e+01
HS30	3	0	7	7	10	63	1	1.0000e+00	0.02	7	1	1.0000e+00
HS32	3	1	4	19	33	166	128	9.8818e-01	0.02	24	4	1.0000e+00
HS33	3	0	6	15	20	189	1	-4.5178e+00	0.02	29	1	-4.5858e+00
HS34	3	0	8	10	15	104	1	-8.3403e-01	0.02	30	1	-0.8340e+00
HS36	3	0	7	10	15	144	1	-3.3000e+03	0.02	10	1	-3.3000e+03
HS37	3	0	8	12	19	200	1	-3.4560e+03	0.02	7	1	-3.4560e+03
HS38	4	0	8	73	153	1.218	1	1.9761e-11	0.06	37	1	3.1594e-24
HS39	4	2	0	11	19	63	1	2.5328e-04	0.02	19	4	-1.0000e+00
HS40	4	3	0	49	108	726	2	-2.5000e-01	0.05	4	2	-2.500e+00
HS42	4	2	0	36	70	290	1 024	1 3883e+01	0.03	6	4	1 3858e+01
HS43	4	0	3	12	29	73	1	-4 4000e+01	0.02	9	1	-4 4000e+01
HS46	5	2	0	101	234	735	1	1 3088e-04	0.05	25	2	6.6616e-12
HS47	5	3	0	21	54	276	1	2.0468e-04	0.04	25	16	8.0322e-14
HS48	5	2	0	21	55	202	2 048	3.1361e-09	0.02	6	4	0.0000e+00
HS49	5	2	0	51	87	276	64	1.1761e-02	0.03	69	64	3.5161e-12
H\$50	5	3	0	50	200	1 0 6 5	128	93190e-05	0.04	11	512	4 0725e-17
HS51	5	3	0	29	132	722	256	2 2808e-05	0.03	8	4	0.0000e+00
HS52	5	3	0	31	45	225	256	5 2930e+00	0.03	4	8	5 3266e+00
H\$53	5	3	10	36	69	1 694	256	4.0734e+00	0.06	5	8	4.0930e+00
HS56	7	4	0	21	43	2 482	4	-2.6183e+00	0.06	12	4	-34560e+00
H\$57	2	0	3 3	34	53	141	1	2.8461e-02	0.03	15	18	2 8460e-02
HS60	3	1	6	18	43	574	1	3.2650e-02	0.04	7	1	3 2568e-02
HS61	3	2	0	16	255	986	256	-17195e+02	0.03	44	128	-1 4365e+02
HS62	3	1	6	8	19	153	1	-2.6273e+04	0.02	5	120	-2.6273e+04
HS63	3	2	3	15	27	200	1	9.6232e+02	0.02	5	2	9.6172e+02
HS66	3	0	8	15	42	200	1	5.1816e-01	0.02	1 000+	1	5.1817e-01
H\$70	4	0	q	16	22	215	1	1.0085e-02	0.02	22	1	1 7981e-01
H\$73	4	1	6	17	35	211	1	2.9896e+01	0.03	16	1	2 9894e+01
H\$77	5	2	0	21	141	587	1	4 5981e-01	0.05	13	1	2.909 1c 101
H\$78	5	3	0	23	66	329	1	-2.9197e+00	0.03	4	4	-2.9197e+00
н570 Н570	5	2	0	16	26	122	128	7.86816-02	0.03	7	י כ	787770-02
HS80	5	3	10	66	196	3 975	120	6.01/9e_02	0.02	6	2	5 30500-02
HS00	5	3	10	10	37	708	т Q	6.01490 02	0.14	0	2	5 30500 02
HC81	5	0	16	30	57	1 25 2	1	5 28030+06	0.05	30	1	5 28030+06
H203	6	0	10 Q	21	/3	1,202	1	136200+02	0.00	12	1	1 35080+02
HS00	7	2	11	2 I 1 Q	21	57	1	831080+08	0.04	12 Q	1	0.00000+02
ЦС100	7	~	14	0	27	96	1	6 9062 0 0 0 0	0.02	0		6.0000000000000000000000000000000000000
113100	/ 0	6	4	41	67	1.006	1	1.27490.09	0.02	1 000+	0 100	0.0003e+02
HSIU/	10	0	20	41	0/ F10	1,080	1	1.3748e-08	0.00	1,000	8,192	5.05450+38
	10	2	20	11 24	21U 264	6 5 4 2	1 0 2 4	-4.31340+UI	0.15	1 000+	1	-4.J//0E+UI
	10	3 2	20	20 2	204 11	0,542	024 I د	-3.03310+UI	0.14	1,000 °	1	-4.//0UE+UI
	10	5	10	0	11	199	2	-5.519/0+01	0.02	10	1	-4.//010+01
	10	U c	ठ २०	2 I 1 1	48 124	519	17	2.4300e+01	0.03	10	ן אבר	2.43066+01
	10	5	28	11	130	4,4/4	10	-1.340/0+03	0.13	39	250	-1./0886+03
H2118	15	U	59	34	51	2,554	1	0.0482e+02	0.12	-	-	-

Table 2 Numerical experiment compared reports

k	ρκ	x <sup>k</sup>	f(x <sup>k</sup> )	∥ <i>ā</i> <sup>k</sup> ∥	∥ <i>g</i> ℓ ∥
0	1	(4, 2)	-1.00000e+00		5.09902e+00
1	1	(4.56235, 1.91528)	-1.00000e+00	5.70399e-01	5.09902e+00
2	2	(4.60502, 1.91721)	-1.00000e+00	1.02139e-01	5.79231e-01
3	8	(4.60222, 1.94248)	-1.00000e+00	1.07010e-01	2.08008e-01
4	128	(4.60158, 1.95367)	-1.00000e+00	1.83034e-01	7.60755e-02
5	8,192	(4.60159, 1.95516)	-1.00000e+00	1.93245e-01	1.32456e-02
6	8,192	(4.60159, 1.95575)	-1.00000e+00	5.86992e-04	4.14795e-03
7	8,192	(4.60159, 1.95581)	-1.00000e+00	1.24840e-04	5.82513e-04
8	8,192	(4.60159, 1.95583)	-1.00000e+00	2.60839e-05	2.36077e-04
9	8,192	(4.60159, 1.95584)	-1.00000e+00	1.21999e-05	7.55776e-05

Table 3 Output of Algorithm A for problem HS8

## 7 Results and discussion

In this work, a new primal-dual interior point QP-free algorithm for nonlinear optimization with equality and inequality constraints is proposed. The global and superlinear convergence are analyzed. Some effective numerical results are reported. As further work, there are several interesting problems worthy of discussing. First, refer to [21], improve the algorithm such that it can start from an arbitrary initial point. Second, try to get rid of the strict complementarity condition. Third, apply the ideas in the paper to minimax optimization problems, engineering problems and so on.

#### Acknowledgements

Project supported by the Natural Science Foundation of Guangxi Province (Nos. 2016GXNSFDA380019 and 2014GXNSFFA118001) and the Natural Science Foundation of China (Nos. 11771383 and 11561005).

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

JBJ carried out the idea of this paper, conceived of the description of Algorithm A and drafted the manuscript. HJZ carried out the convergence analysis of Algorithm A. GDM participated in the numerical experiments and helped to draft the manuscript. ZBZ participated in the convergence analysis. All authors read and approved the final manuscript.

#### Author details

<sup>1</sup>School of Mathematics and Statistics, Guangxi Colleges and Universities Key Laboratory of Complex System Optimization and Big Data Processing, Yulin Normal University, Yulin, China. <sup>2</sup>College of Mathematics and Information Science, Guangxi University, Nanning, China. <sup>3</sup>College of Mathematics and Computing Science, Guilin University of Electronic Technology, Guilin, China.

#### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Received: 10 March 2017 Accepted: 6 September 2017 Published online: 29 September 2017

#### References

- Tits, AL, Wächter, A, Bakhtiari, S, Urban, TJ, Lawrence, CT: A primal-dual interior-point method for nonlinear programming with strong global and local convergence properties. SIAM J. Optim. 14(1), 173-199 (2003). doi:10.1137/S1052623401392123
- Mayne, DQ, Polak, E: Feasible direction algorithms for optimization problems with equality and inequality constraints. Math. Program. 11, 67-80 (1976)
- Gao, ZY, He, GP, Wu, F: Sequential systems of linear equations algorithm for nonlinear optimization problems with general constraints. J. Optim. Theory Appl. 95, 371-397 (1997)
- 4. Lian, SJ, Duan, YQ: Smoothing of the lower-order exact penalty function for inequality constrained optimization. J. Inequal. Appl. **2016**, 185 (2016)
- Jian, JB, Xu, QJ, Han, DL: A strongly convergent norm-relaxed method of strongly sub-feasible direction for optimization with nonlinear equality and inequality constraints. Appl. Math. Comput. 182, 854-870 (2006)
- Jian, JB, Tang, CM, Hu, QJ, Zheng, HY: A feasible descent SQP algorithm for general constrained optimization without strict complementarity. J. Comput. Appl. Math. 180(2), 391-412 (2006)
- Herskovits, J: Feasible directions interior-point technique for nonlinear optimization. J. Optim. Theory Appl. 99, 121-146 (1998)

- Boggs, PT, Tolle, JW, Wang, P: On the local convergence of quasi-Newton methods for constrained optimization. SIAM J. Control Optim. 20, 161-171 (1982)
- Boggs, PT, Tolle, JW: Sequential Quadratic Programming, vol. 4, pp. 1-51. Cambridge University Press, Cambridge (1995)
- Gill, PE, Murray, W, Saunders, MA: SNOPT: an SQP algorithm for large-scale constrained optimization. SIAM Rev. 47, 99-131 (2005)
- 11. Jian, JB, Tang, CM: An SQP feasible descent algorithm for nonlinear inequality constrained optimization without strict complementarity. Comput. Math. Appl. 49, 223-238 (2005)
- 12. Jian, JB, Zheng, HY, Hu, QJ, Tang, CM: A new norm-relaxed method of strongly sub-feasible direction for inequality constrained optimization. Appl. Math. Comput. **168**, 1-28 (2005)
- Jian, JB, Zheng, HY, Tang, CM, Hu, QJ: A new superlinearly convergent norm-relaxed method of strong sub-feasible direction for inequality constrained optimization. Appl. Math. Comput. 182, 955-976 (2006)
- Lawrence, CT, Tits, AL: A computationally efficient feasible sequential quadratic programming algorithm. SIAM J. Optim. 11, 1092-1118 (2001)
- Panier, ER, Tits, AL: A superlinearly convergent feasible method for the solution of inequality constrained optimization problems. SIAM J. Control Optim. 25, 934-950 (1987)
- Panier, ER, Tits, AL: On combining feasibility, descent and superlinear convergence in inequality constrained optimization. Math. Program. 59, 261-276 (1993)
- Spellucci, P: An SQP method for general nonlinear programs using only equality constrained subproblem. Math. Program. 82, 413-448 (1998)
- Bakhtiari, S, Tits, AL: A simple primal-dual feasible interior-point method for nonlinear programming with monotone descent. Comput. Optim. Appl. 25, 17-38 (2003)
- 19. El-bakry, AS, Tapia, RA, Tsuchiya, T, Zhang, Y: On the formulation and theory of the Newton interior-point method for nonlinear programming. J. Optim. Theory Appl. 89, 507-541 (1996)
- 20. Forsgren, A, Gill, PE, Wright, MH: Interior methods for nonlinear optimization. SIAM Rev. 44, 525-597 (2002)
- Jian, JB, Pan, HQ, Tang, CM, Li, JL: A strongly sub-feasible primal-dual quasi interior-point algorithm for nonlinear inequality constrained optimization. Appl. Math. Comput. 266, 560-578 (2015)
- Panier, ER, Tits, AL, Herskovits, JN: A QP-free, globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization. SIAM J. Control Optim. 26, 788-811 (1988)
- Qi, HD, Qi, LQ: A new QP-free, globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization. SIAM J. Optim. 11, 113-132 (2000)
- Wang, YL, Chen, L, He, GP: Sequential systems of linear equations method for general constrained optimization problems without strict complementarity. J. Comput. Appl. Math. 182, 447-471 (2005)
- Yang, YF, Li, DH, Qi, LQ: A feasible sequential linear equation method for inequality constrained optimization. SIAM J. Optim. 13, 1222-1244 (2003)
- Zhu, ZB: An interior point type QP-free algorithm with superlinear convergence for inequality constrained optimization. Appl. Math. Model. 31, 1201-1212 (2007)
- 27. Maratos, N: Exact penalty function algorithm for finite dimensional and control optimization problems. Dissertation, Imperial College Science, Technology, University of London (1978)
- 28. Cai, XZ, Wu, L, Yue, YJ, Li, MM, Wang, GQ: Kernel-function-based primal-dual interior-point methods for convex quadratic optimization over symmetric cone. J. Inequal. Appl. **2014**, 308 (2014)
- 29. Facchinei, F, Fischer, A, Kanzow, C: On the accurate identification of active constraints. SIAM J. Optim. 9, 14-32 (1998)
- Chen, L, Wang, YL, He, GP: A feasible active set QP-free method for nonlinear programming. SIAM J. Optim. 17, 401-429 (2006)
- Liu, Y, Jian, JB, Zhu, ZB: New active set identification for general constrained optimization and minimax problems. J. Math. Anal. Appl. 421, 1405-1416 (2015)
- 32. Wäachter, A, Biegler, LT: Failure of global convergence for a class of interior point methods for nonlinear programming. Math. Program. 88, 565-574 (2000)
- Hock, W, Schittkowski, K: Test Examples for Nonlinear Programming Codes. Lecture Notes in Economics and Mathematical Systems., vol. 187. Springer, Heidelberg (1981)

## Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- ► Rigorous peer review
- Open access: articles freely available online
- ► High visibility within the field
- Retaining the copyright to your article

#### Submit your next manuscript at > springeropen.com