SIAM J. OPTIM.
Vol. 13, No. 4, pp. 1222–1244

A FEASIBLE SEQUENTIAL LINEAR EQUATION METHOD FOR INEQUALITY CONSTRAINED OPTIMIZATION*

YU-FEI YANG†, DONG-HUI LI†, AND LIQUN QI‡

Abstract. In this paper, by means of the concept of the working set, which is an estimate of the active set, we propose a feasible sequential linear equation algorithm for solving inequality constrained optimization problems. At each iteration of the proposed algorithm, we first solve one system of linear equations with a coefficient matrix of size $m \times m$ (where m is the number of constraints) to compute the working set; we then solve a subproblem which consists of four reduced systems of linear equations with a common coefficient matrix. Unlike existing QP-free algorithms, the subproblem is concerned with only the constraints corresponding to the working set. The constraints not in the working set are neglected. Consequently, the dimension of each subproblem is not of full dimension. Without assuming the isolatedness of the stationary points, we prove that every accumulation point of the sequence generated by the proposed algorithm is a KKT point of the problem. Moreover, after finitely many iterations, the working set becomes independent of the iterates and is essentially the same as the active set of the KKT point. In other words, after finitely many steps, only those constraints which are active at the solution will be involved in the subproblem. Under some additional conditions, we show that the convergence rate is two-step superlinear or even Q-superlinear. We also report some preliminary numerical experiments to show that the proposed algorithm is practicable and effective for the test problems.

Key words. sequential linear equation algorithm, optimization, active set strategy, global convergence, superlinear convergence

AMS subject classifications. 90C30, 65K10

PII. S1052623401383881

1. Introduction.

We consider the nonlinear inequality constrained optimization problem

$$\text{(P)} \qquad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & g(x) \leq 0, \end{array}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are assumed to be twice continuously differentiable. We denote by

$$\mathcal{F} = \{x \in \mathbb{R}^n \mid g(x) \leq 0\}$$

the feasible set of problem (P).

The Lagrangian function associated with problem (P) is defined by

$$L(x, \lambda) = f(x) + \lambda^T g(x).$$

A pair $(x^*, \lambda^*) \in \mathbb{R}^{n \times m}$ is called a KKT point or a KKT pair of problem (P) if it satisfies the following KKT conditions:

$$\begin{array}{l} \nabla_x L(x^*, \lambda^*) = 0, \quad g(x^*) \leq 0, \quad \lambda^* \geq 0, \\ g_i(x^*)\lambda_i^* = 0 \ \ \forall\, i \in I, \end{array} \tag{1.1}$$

*Received by the editors January 22, 2001; accepted for publication (in revised form) November 6, 2002; published electronically May 2, 2003. This work was supported by the NSF (grant 10171030) of China, the RGC (grant PolyU/5314/01p) of Hong Kong, and the Australian Research Council.
http://www.siam.org/journals/siopt/13-4/38388.html

†Institute of Applied Mathematics, Hunan University, Changsha 410082, China. Current address: Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (mayfyang@polyu.edu.hk, madhli@polyu.edu.hk).

‡Corresponding author. Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (maqilq@polyu.edu.hk).

where $I := \{1, \dots, m\}$ and

$$\nabla_x L(x, \lambda) := \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x). \tag{1.2}$$

Sometimes, we also call the point x^* satisfying (1.1) a KKT point of problem (P). If (x^*, λ^*) satisfies all conditions in (1.1) except for the inequality $\lambda^* \geq 0$, we call the point x^* a stationary point of problem (P).

Throughout the paper, we assume that the following blanket hypotheses hold.

Assumption A1. The set $\mathcal{F}$ is bounded.

Assumption A2. At every $x \in \mathcal{F}$, the vectors $\nabla g_i(x)$, $i \in I_0(x)$, are linearly independent, where $I_0(x) := \{i \in I \mid g_i(x) = 0\}$.

Note that Assumption A1 is often substituted by the assumption that the level sets of the objective function of some unconstrained optimization problem are compact or the sequence of points generated by the algorithm is bounded, while Assumption A2 is a common assumption in dealing with the global convergence of most algorithms for solving problem (P).

The sequential quadratic programming (SQP) methods are a class of efficient methods for solving nonlinearly constrained optimization problems. They have received much attention in recent decades. We refer to a review paper [2] for a good survey on SQP methods.

The iterative process of a typical SQP method is as follows. Let the current iterate be x^k. Compute a search direction d^k by solving the following quadratic program (QP):

$$\begin{array}{ll} \min\limits_{d} & \frac{1}{2}\langle d, H_k d\rangle + \langle \nabla f(x^k), d\rangle, \\ \text{s.t.} & g_i(x^k) + \langle \nabla g_i(x^k), d\rangle \leq 0 \quad \forall\, i \in I, \end{array} \tag{1.3}$$

where $H_k \in \mathbb{R}^{n \times n}$ is symmetric positive definite. Perform a line search to determine a steplength t_k and let the next iterate be $x^{k+1} = x^k + t_k d^k$.

SQP methods possess global and superlinear convergence properties under certain conditions. However, in a traditional SQP method, the QP subproblem (1.3) may be inconsistent; that is, the feasible set of (1.3) may be empty. To overcome this shortcoming, various techniques have been proposed; see, e.g., [6, 18, 21, 24, 25, 29, 31]. In particular, Panier and Tits [21] presented a feasible SQP (FSQP) algorithm in which the generated iterates lie in the feasible region $\mathcal{F}$. Under certain conditions, this FSQP algorithm is globally convergent and locally two-step superlinearly convergent. Further study on FSQP algorithms can be found in [17, 22, 27, 28].

FSQP methods are particularly useful for solving those problems arising from engineering design where the objective function f might be undefined outside the feasible region $\mathcal{F}$. Another advantage of FSQP methods is that the objective function f can be used as a merit function to avoid the use of a penalty function. However, FSQP algorithms still require solving QP subproblems at each iteration, which is computationally expensive. In [23], Panier, Tits, and Herskovits proposed a feasible QP-free algorithm in which, at every iteration, only three systems of linear equations need to be solved. Specifically, the iterative process of the QP-free algorithm is as follows. Let (x^k, λ^k) be the current iterate. To guarantee the feasibility of the next iterate, they first solve two systems of linear equations of the form

$$\begin{pmatrix} H_k & \nabla g(x^k) \\ \operatorname{diag}(\mu^k)\nabla g(x^k)^T & \operatorname{diag}(g(x^k)) \end{pmatrix} \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) \\ c \end{pmatrix} \tag{1.4}$$

by choosing a different vector c, where $H_k \in \mathbb{R}^{n\times n}$ is positive definite, $\mu^k \in \mathbb{R}^m$, $c \in \mathbb{R}^m$, and $\text{diag}(\mu^k)$ denotes the $m \times m$ diagonal matrix whose ith diagonal element is μ_i^k. Then they further "bend" the primal search direction by solving a least squares subproblem to avoid the Maratos effect. It has been shown in [23] that under appropriate conditions, this QP-free method possesses global convergence as well as a locally two-step superlinear convergence rate. However, the QP-free algorithm proposed in [23] may have instability problems. The linear system (1.4) may become very ill-conditioned if some multiplier μ_i corresponding to a nearly active constraint g_i becomes very small. In addition, in the global convergence theorem, there is a restrictive condition which requires that the number of stationary points is finite. The idea of this QP-free algorithm has been further used by Urban, Tits, and Lawrence [34] to develop a primal-dual logarithmic barrier interior-point method; see also [1]. Under similar conditions, the method possesses global and fast local convergence properties.

Recently, by means of the Fischer–Burmeister function, Qi and Qi [26] presented a new feasible QP-free algorithm for solving problem (P). At each iteration, the subproblem of the new QP-free method consists of three systems of linear equations of the form

$$\begin{pmatrix} H_k & \nabla g(x^k) \\ \text{diag}(\eta^k)\nabla g(x^k)^T & -\sqrt{2}\text{diag}(\theta^k) \end{pmatrix} \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) \\ c, \end{pmatrix}, \tag{1.5}$$

where c is a suitable vector and for each $i \in I$

$$\eta_i^k := \frac{g_i(x^k)}{\sqrt{g_i^2(x^k) + (\mu_i^k)^2}} + 1 \quad \text{and} \quad \theta_i^k := \left(1 - \frac{\mu_i^k}{\sqrt{g_i^2(x^k) + (\mu_i^k)^2}}\right)^{1/2}.$$

To avoid the Maratos effect, they also solve a least squares subproblem. Their algorithm shares some advantages of the method in [23]. Moreover, the matrix in (1.5) is nonsingular even if the strict complementarity does not hold. The method achieves global convergence without requiring the isolatedness of the stationary points. The local one-step superlinear convergence rate of the method has also been established.

In this paper, we propose a feasible sequential linear equation (FSLE) algorithm for solving problem (P). At each step, we first solve three reduced systems of linear equations with the following form:

$$\begin{pmatrix} H_k & \nabla g_{A^k}(x^k) \\ \nabla g_{A^k}(x^k)^T & 0 \end{pmatrix} \begin{pmatrix} d \\ \lambda_{A^k} \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) \\ c_{A^k} \end{pmatrix}, \tag{1.6}$$

where $A^k \subset I$ is called a *working set* which is an estimate of the active set $I_0(x^k)$. The calculation of the working set depends on some multiplier function which is the solution of a system of linear equations. If x^k is sufficiently close to a KKT point x^*, then A^k is an identification of the active set $I_0(x^*)$. The working set and the identification of the active set have been studied by some authors [9, 10, 12, 13, 31, 32]. They are also very important in our algorithm. It is clear that the dimension of system (1.6) is no more than the dimension of system (1.5). Moreover, as we shall show in section 4 (see Lemma 4.1), under appropriate conditions we have $A^k = I_0(x^*)$ for all k sufficiently large. This means that after finitely many iterations, the inactive constraints at x^* will be neglected.

Like other existing feasible QP-free methods, the method proposed in this paper also generates a sequence of iterates that are interior points in the feasible region.

However, feasible QP-free methods are different from interior-point methods. An interior-point method follows a central path, while a feasible QP-free method does not.

In order to achieve a superlinear convergence rate, we solve another system of linear equations. This system is equivalent to a least squares problem. Unlike algorithms proposed in [23, 26], the coefficient matrix of the last linear system is the same as the previous reduced ones. Furthermore, our algorithm provides a special technique to update the working set and makes it possible to remove multiple inactive constraints in one iteration. This technique for updating the working set has also been used recently in [32].

The main advantage of the proposed algorithm lies in that it has the potential of saving computational cost. Moreover, it reserves all the advantages of algorithms proposed in [23, 26].

Interesting features of the proposed algorithm include the following:

- All iterates are feasible and the sequence of objective functions is decreasing.
- At each iteration, we need to solve only one $m \times m$ system of linear equations and four reduced systems of linear equations with a common coefficient matrix.
- Under appropriate conditions, the generated direction sequences are uniformly bounded.
- The iterative matrices are nonsingular without the requirement of strict complementarity.
- Every accumulation point of the sequence generated by the proposed algorithm is a KKT point of problem (P) without assuming that the stationary points are isolated.
- Locally two-step superlinear or Q-superlinear convergence rate is achieved.

Recently, Facchinei and Lazzari [11] presented a local feasible QP-free algorithm for solving problem (P) with an SC^1 objective function. Their algorithm possesses some favorable properties, such as fast local convergence and feasibility of all iterates. In addition, at each iteration, only systems of linear equations need to be solved. Their algorithm produces a sequence $\{x^k\}$ according to the following formula:

$$x^{k+1} = x^k + d^k + \hat{d}^k.$$

The local structure of our algorithm is similar to theirs. In some sense, our algorithm can be regarded as a globalization of their algorithm. However, compared with their algorithm, we used quasi-Newton algorithms. Moreover, the computation of the directions d^k and $\hat{d}^k$ is different from that in [11].

The paper is organized as follows. In the next section we introduce a multiplier function to define the working set. We then describe the algorithm and show that it is well defined. In section 3, we establish a global convergence theorem for the algorithm. In section 4, we prove that under appropriate conditions the sequence $\{x^k\}$ generated by the proposed algorithm is locally two-step superlinearly or Q-superlinearly convergent. We report some preliminary numerical results in section 5. In the last section, we give some remarks to conclude the paper.

A few words for the notation. The symbol $\|\cdot\|$ always stands for the Euclidean vector norm or its associated matrix norm. Given $h : \mathbb{R}^n \to \mathbb{R}^m$ and a subset A of I, we denote by $h_A(x)$ the subvector of $h(x)$ with components $h_i(x), i \in A$, and by $\nabla h_A(x)$ the transpose of the Jacobian of $h_A(x)$. We use $e \in \mathbb{R}^m$ to denote the vector of all ones, and $E \in \mathbb{R}^{m\times m}$ is the unit matrix.

2. Algorithm. In this section we first define the working set based on a multiplier function; then we present an FSLE algorithm for solving problem (P) and show that it is well defined.

The following proposition comes from [14] and [19].

PROPOSITION 2.1. *The following statements hold.*

(i) *For every $x \in \mathcal{F}$, there exists a unique minimizer $\lambda(x)$ of the quadratic function in λ,*

$$\|\nabla_x L(x,\lambda)\|^2 + \|G(x)\lambda\|^2$$

over $\mathbb{R}^m$, given by

$$\lambda(x) = -M^{-1}(x)\nabla g(x)^T \nabla f(x), \tag{2.1}$$

where

$$G(x) := \operatorname{diag}(g_i(x)) \quad \textit{and} \quad M(x) := \nabla g(x)^T \nabla g(x) + G^2(x).$$

(ii) *The multiplier function $\lambda(x)$ is continuously differentiable in $\mathcal{F}$.*

(iii) *If $(x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m$ is a KKT pair for problem (P), we have $\lambda(x^*) = \lambda^*$.*

For $x \in \mathcal{F}$, we now make the following "guess" for the active set $I_0(x)$:

$$A(x;\epsilon) := \{i \mid g_i(x) + \epsilon\rho(x, \lambda(x)) \geq 0\},$$

where ϵ is a nonnegative parameter and $\rho(x,\lambda) := \sqrt{\|\Phi(x,\lambda)\|}$ with

$$\Phi(x,\lambda) := \begin{pmatrix} \nabla_x L(x,\lambda) \\ \min\{-g(x), \lambda\} \end{pmatrix}.$$

It is obvious that (x^*, λ^*) is a KKT pair of problem (P) if and only if $\Phi(x^*, \lambda^*) = 0$ or $\rho(x^*, \lambda^*) = 0$. Facchinei, Fischer, and Kanzow [9] showed that if the second order sufficient condition and the Mangasarian–Fromovotz constraint qualification hold, then for any $\epsilon > 0$, when x is sufficiently close to x^*, the working set $A(x;\epsilon)$ is an exact identification of $I_0(x^*)$. It is not difficult to see from Assumption A1 and Proposition 2.1(ii) that $\rho(x, \lambda(x))$ is bounded on $\mathcal{F}$. This property will enable us to keep the parameter ϵ fixed after a finite number of iterations in our algorithm. Details will be given subsequently.

Let

$$V(x, H; A) = \begin{pmatrix} H & \nabla g_A(x) \\ \nabla g_A(x)^T & 0 \end{pmatrix},$$

where H is an $n \times n$ positive definite matrix and A is a subset of I. We now state the steps of our algorithm for solving problem (P).

ALGORITHM 2.1.

Parameters. $\beta \in (0,1)$, $\mu \in (0, 1/2)$, $\nu > 2$, $\tau \in (2,3)$, $\vartheta \in (0,1)$, *and* $\sigma \in (0,1)$.

Data. x^1, *a strictly feasible point in* $\mathcal{F}$; $H_1 \in \mathbb{R}^{n\times n}$, *a symmetric positive definite matrix; and* $\epsilon^0 > 0$, *an initial parameter.*

Set $k := 1$.

Step 1. *Set* $\epsilon := \epsilon^{k-1}$.

Step 2. *Set* $A^k(\epsilon) := A(x^k; \epsilon)$.

If $\nabla g_{A^k(\epsilon)}(x^k)$ *is not of full rank, then set* $\epsilon := \sigma\epsilon$ *and go to Step* 2.

Step 3. *Set* $\epsilon^k := \epsilon$, $A^k := A^k(\epsilon^k)$, *and* $V_k := V(x^k, H_k; A^k)$.

Step 4. *Computation of a search direction.*

(i) *Compute* $(d^{k0}, z^{k0}_{A^k})$ *by solving the system of linear equations in* (d, z_{A^k}),

$$V_k \begin{pmatrix} d \\ z_{A^k} \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) \\ 0 \end{pmatrix}. \tag{2.2}$$

(ii) *Compute* $(d^{k1}, z^{k1}_{A^k})$ *by solving the system of linear equations in* (d, z_{A^k}),

$$V_k \begin{pmatrix} d \\ z_{A^k} \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) \\ \varphi^k \end{pmatrix}, \tag{2.3}$$

where $\varphi^k \in \mathbb{R}^{|A^k|}$ *is defined by*

$$\varphi^k_i := \begin{cases} z^{k0}_i & \text{if } \; z^{k0}_i < 0, \\ -g_i(x^k) & \text{if } \; z^{k0}_i > 0, \\ 0 & \text{otherwise.} \end{cases}$$

If $d^{k1} = 0$, *stop.*

(iii) *Compute* $(d^{k2}, z^{k2}_{A^k})$ *by solving the system of linear equations in* (d, z_{A^k}),

$$V_k \begin{pmatrix} d \\ z_{A^k} \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) \\ \varphi^k - \|d^{k1}\|^\nu e_{A^k} \end{pmatrix}. \tag{2.4}$$

(iv) *Compute the search direction* d^k *and the approximate multiplier vector* $z^k_{A^k}$ *according to*

$$\begin{pmatrix} d^k \\ z^k_{A^k} \end{pmatrix} := (1 - \phi^k) \begin{pmatrix} d^{k1} \\ z^{k1}_{A^k} \end{pmatrix} + \phi^k \begin{pmatrix} d^{k2} \\ z^{k2}_{A^k} \end{pmatrix},$$

where

$$\phi^k := (\vartheta - 1) \frac{\langle \nabla f(x^k), d^{k1} \rangle}{1 + \|d^{k1}\|^\nu |\sum_{i \in A^k} z^{k0}_i|}.$$

Step 5. *Compute a correction* $\hat{d}^k$ *by solving the system of linear equations in* (d, z_{A^k}),

$$V_k \begin{pmatrix} d \\ z_{A^k} \end{pmatrix} = \begin{pmatrix} 0 \\ -\|d^k\|^\tau e_{A^k} - g_{A^k}(x^k + d^k) \end{pmatrix}. \tag{2.5}$$

If $\|\hat{d}^k\| > \|d^k\|$, *set* $\hat{d}^k := 0$.

Step 6. *Line search. Compute* t_k, *the first number* t *in the sequence* $\{1, \beta, \beta^2, \dots\}$ *satisfying*

$$f(x^k + td^k + t^2 \hat{d}^k) \le f(x^k) + \mu t \langle \nabla f(x^k), d^k \rangle \tag{2.6}$$

and

$$g_i(x^k + td^k + t^2 \hat{d}^k) < 0 \qquad \forall\, i \in I. \tag{2.7}$$

Step 7. *Set* $x^{k+1} := x^k + t_k d^k + t_k^2 \hat{d}^k$ *and generate a new symmetric definite positive matrix* H_{k+1}. *Set* $k := k + 1$ *and go to Step* 1.

Remarks.

(i) It follows from Assumption A2 that there exists some $\delta_0 > 0$ such that $\nabla g_{I(x;\delta)}(x)$ is of full rank, where $I(x;\delta) := \{i \in I \,:\, g_i(x) \geq -\delta\}$ and $0 \leq \delta \leq \delta_0$. By the continuity of $\lambda(x)$ and Assumption A1, there exists some $\bar{\epsilon}_0 > 0$ such that the inequality $\epsilon\rho(x, \lambda(x)) \leq \delta_0$ holds for all $\epsilon \leq \bar{\epsilon}_0$ and $x \in \mathcal{F}$, and hence $A(x;\epsilon) \subseteq I(x;\delta_0)$. This implies that $\nabla g_{A(x;\epsilon)}(x)$ is of full rank. Therefore, for symmetric positive definite matrix $H \in \mathbb{R}^{n\times n}$, the matrix $V(x, H; A(x;\epsilon))$ is nonsingular. Consequently, V_k is nonsingular for each k. This shows that $(d^{k0}, z^{k0}_{A^k})$, $(d^{k1}, z^{k1}_{A^k})$, $(d^{k2}, z^{k2}_{A^k})$, and $\hat{d}^k$ are well defined.

On the other hand, the above analysis also indicates that at Step 2 of Algorithm 2.1 the parameter ϵ is reduced only finitely many times. In other words, ϵ_k will remain fixed after finitely many iterations. Without loss of generality, we assume that $\epsilon^k = \tilde{\epsilon}$ for all k.

(ii) In order to guarantee the feasibility of all iterates and the decrease of the objective function at each iteration, we solve three linear systems with the same coefficient matrix but different right vectors. This technique is similar to that in [26]. Notice that the choice of φ^k at Step 4(ii) ensures that x^k is a trivial KKT point of problem (P) whenever $d^{k1} = 0$ (see Lemma 2.2).

(iii) The role of Step 5 is to avoid the Maratos effect. It is not difficult to see that $\hat{d}^k$ is also the unique solution of the least squares problem in d,

$$\begin{aligned} &\min \quad \frac{1}{2}\|d\|^2_{H_k} \\ &\text{s.t.} \quad g_i(x^k + d^k) + \langle \nabla g_i(x^k), d\rangle = -\|d^k\|^\tau \qquad \forall\, i \in A^k. \end{aligned} \tag{2.8}$$

An important difference between our algorithm and those in [23, 26] lies in the fact that the coefficient matrix in (2.5) is the same as that in Step 4. Hence, our algorithm needs fewer computational efforts. If H_k is taken to be the unit matrix for every k, $A^k = I_0(x^*)$, and $\tau = 2$, then problem (2.8) reduces to the subproblem of computing the correction direction $\hat{d}^k$ in [11].

(iv) It is not difficult to deduce that the direction $(d^k, z^k_{A^k})$ is the unique solution of the following system of linear equations:

$$V_k \begin{pmatrix} d \\ z_{A^k} \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) \\ \varphi^k - \phi^k \|d^{k1}\|^\nu e_{A^k} \end{pmatrix}. \tag{2.9}$$

We now analyze the updating technique for the working set. For $i \in A^k$, we obtain from (2.5) and (2.9)

$$\begin{aligned} &g_i(x^k + td^k + t^2\hat{d}^k) \\ &= g_i(x^k) + t\nabla g_i(x^k)^T d^k + t^2 \nabla g_i(x^k)^T \hat{d}^k + O((t\|d^k\|)^2) \\ &= g_i(x^k) + t\nabla g_i(x^k)^T d^k - t^2 g_i(x^k + d^k) + O((t\|d^k\|)^2) \\ &= (1-t^2) g_i(x^k) + (t - t^2)\nabla g_i(x^k)^T d^k + O((t\|d^k\|)^2) \\ &= O((t\|d^k\|)^2) - (t - t^2)\phi^k \|d^{k1}\|^\nu \\ &\quad + \begin{cases} (1-t^2) g_i(x^k) + (t-t^2) z^{k0}_i & \text{if } z^{k0}_i < 0, \\ (1-t) g_i(x^k) & \text{if } z^{k0}_i > 0, \\ (1-t^2) g_i(x^k) & \text{otherwise.} \end{cases} \end{aligned}$$

Hence, if $z^{k0}_i < 0$ is not small and $t_k - t_k^2$ is not very small, it is likely that $i \notin A^{k+1}$ because g_i becomes strongly negative now. Thus, it is reasonable to exclude these i from A^{k+1}. This technique was also used by Spellucci [32].

For the sake of convenience, we let for each k

$$z_i^{k0} = z_i^{k1} = z_i^{k2} = z_i^k = 0 \qquad \forall\, i \notin A^k.$$

To analyze the well-definedness and convergence of the above algorithm, we make the following hypothesis on the choice of matrix H_k.

Assumption A3. There exist positive constants C_1 and C_2 such that, for all k and $d \in \mathbb{R}^n$,

$$C_1\|d\|^2 \le d^T H_k d \le C_2\|d\|^2.$$

It is not difficult to see from the discussion of Remark (i) that every limit of the sequence $\{\nabla g_{A^k}(x^k)\}$ is also of full rank. Therefore, Assumption A3 shows that every limit of the sequence $\{V_k\}$ is nonsingular, which implies that $\{\|V_k^{-1}\|\}$ is bounded. We assume that $\|V_k^{-1}\| \le \tilde{M}$ for all k.

Let $N_{A^k} := \nabla g_{A^k}(x^k)$. Then, by Step 2, N_{A^k} is of full rank. Since V_k is nonsingular, it is clear that matrix $D_k := N_{A^k}^T H_k^{-1} N_{A^k}$ is also nonsingular. Let

$$B_k := H_k^{-1} N_{A^k} D_k^{-1} \quad \text{and} \quad Q_k := H_k^{-1}(E - N_{A^k} B_k^T).$$

By Step 4 of the algorithm, it is not difficult to deduce the following relations:

$$\text{(2.10)} \qquad \begin{cases} d^{k0} = -Q_k \nabla f(x^k), & z_{A^k}^{k0} = -B_k^T \nabla f(x^k), \\ d^{k1} = d^{k0} + B_k \varphi^k, & z_{A^k}^{k1} = z_{A^k}^{k0} - D_k^{-1}\varphi^k, \\ d^{k2} = d^{k1} - \|d^{k1}\|^\nu B_k e_{A^k}, & z_{A^k}^{k2} = z_{A^k}^{k1} + \|d^{k1}\|^\nu D_k^{-1} e_{A^k}, \\ d^k = d^{k1} - \phi^k \|d^{k1}\|^\nu B_k e_{A^k}, & z_{A^k}^k = z_{A^k}^{k1} + \phi^k \|d^{k1}\|^\nu D_k^{-1} e_{A^k}. \end{cases}$$

LEMMA 2.2. *If the algorithm stops at Step* 4(ii)*, i.e.,* $d^{k1} = 0$*, then* $\nabla f(x^k) = 0$.

Proof. If $d^{k1} = 0$, then it follows from (2.3) that

$$\text{(2.11)} \qquad \begin{cases} \nabla f(x^k) + \nabla g_{A^k}(x^k) z_{A^k}^{k1} = 0, \\ \varphi^k = 0. \end{cases}$$

By the construction of φ^k, we have $z_{A^k}^{k0} = 0$, and hence by (2.10) $z_{A^k}^{k1} = 0$. The assertion then follows from the first equation of (2.11). ∎

The above lemma shows that if the algorithm stops at Step 4(ii), then x^k is an unconstrained stationary point of f. Since we always have $x^k \in \mathcal{F}$, this means that x^k is actually a KKT point of problem (P). In what follows, we assume that the algorithm never stops at Step 4(ii). Therefore, the algorithm generates an infinite sequence $\{x^k\}$.

LEMMA 2.3. (i) $\langle \nabla f(x^k), d^{k0}\rangle = -\langle d^{k0}, H_k d^{k0}\rangle$.

(ii) $\langle \nabla f(x^k), d^{k1}\rangle = \langle \nabla f(x^k), d^{k0}\rangle - \langle \varphi^k, z_{A^k}^{k0}\rangle \le \langle \nabla f(x^k), d^{k0}\rangle$.

(iii) $\langle \nabla f(x^k), d^k\rangle \le \vartheta \langle \nabla f(x^k), d^{k1}\rangle$.

Proof. By (2.2), we deduce

$$\begin{aligned} \langle \nabla f(x^k), d^{k0}\rangle &= -\langle d^{k0}, H_k d^{k0}\rangle - \langle d^{k0}, \nabla g_{A^k}(x^k) z_{A^k}^{k0}\rangle \\ &= -\langle d^{k0}, H_k d^{k0}\rangle - \langle \nabla g_{A^k}(x^k)^T d^{k0}, z_{A^k}^{k0}\rangle \\ &= -\langle d^{k0}, H_k d^{k0}\rangle. \end{aligned}$$

This establishes (i). From (2.10), we have

$$\begin{aligned}\langle \nabla f(x^k), d^{k1}\rangle &= \langle \nabla f(x^k), d^{k0}\rangle + \langle \nabla f(x^k), B_k \varphi^k\rangle\\ &= \langle \nabla f(x^k), d^{k0}\rangle + \langle B_k^T \nabla f(x^k), \varphi^k\rangle\\ &= \langle \nabla f(x^k), d^{k0}\rangle - \langle z_{A^k}^{k0}, \varphi^k\rangle\\ &\le \langle \nabla f(x^k), d^{k0}\rangle,\end{aligned}$$

where the last inequality holds because by the definition of φ^k we have $\langle z_{A^k}^{k0}, \varphi^k\rangle \ge 0$. This establishes (ii). We now turn to verify (iii).

It follows from (2.10) and the definition of ϕ^k that

$$\begin{aligned}\langle \nabla f(x^k), d^k\rangle &= \langle \nabla f(x^k), d^{k1}\rangle - \phi^k \|d^{k1}\|^\nu \langle \nabla f(x^k), B_k e_{A^k}\rangle\\ &= \langle \nabla f(x^k), d^{k1}\rangle - \phi^k \|d^{k1}\|^\nu \langle B_k^T \nabla f(x^k), e_{A^k}\rangle\\ &= \langle \nabla f(x^k), d^{k1}\rangle + \phi^k \|d^{k1}\|^\nu \sum_{i\in A^k} z_i^{k0}\\ &\le \vartheta \langle \nabla f(x^k), d^{k1}\rangle.\end{aligned}$$

This establishes (iii). ∎

Lemma 2.3 shows that the direction d^k is a descent direction of the merit function f. Similar to the proof of Proposition 3.3 in [23], we can deduce that for each k there is a nonnegative integer $j(k)$ such that inequalities (2.6) and (2.7) are satisfied with $t_k = \beta^{j(k)}$.

The above discussion has shown that Algorithm 2.1 is well defined.

3. Global convergence. In this section we will show that Algorithm 2.1 is globally convergent. First, we see from Step 2 of Algorithm 2.1 and the discussion after Assumption A3 that $\|V_k^{-1}\| \le \tilde{M}$ for all k. The following lemma is then obvious.

LEMMA 3.1. *The sequences* $\{(d^{k0}, z^{k0})\}$, $\{(d^{k1}, z^{k1})\}$, *and* $\{(d^{k2}, z^{k2})\}$ *are all bounded.*

Proof. By (2.2), Assumption A1, and the boundedness of $\{\|V_k^{-1}\|\}$, we deduce that $\{(d^{k0}, z^{k0})\}$ is bounded, which implies that $\{(d^{k1}, z^{k1})\}$ is also bounded by (2.3). The boundedness of $\{(d^{k2}, z^{k2})\}$ directly follows from (2.4) and the boundedness of $\{d^{k1}\}$ and $\{z^{k0}\}$. ∎

LEMMA 3.2. *There exists a constant* $\kappa > 0$ *such that, for all* $k = 1, 2, \ldots$,

$$\|d^k - d^{k1}\| \le \kappa \|d^{k1}\|^\nu.$$

Proof. Assumption A1 and Lemma 3.1 imply that $\{\phi^k\}$ is bounded. It follows from Step 4 of Algorithm 2.1 that

$$\begin{pmatrix} d^k - d^{k1} \\ z_{A^k}^k - z_{A^k}^{k1} \end{pmatrix} = V_k^{-1} \begin{pmatrix} 0 \\ -\phi^k \|d^{k1}\|^\nu e_{A^k} \end{pmatrix},$$

which shows that the assertion holds with $\kappa := \tilde{M} \sup\{\phi^k\}$. ∎

The following proposition gives a sufficient condition for the global convergence of Algorithm 2.1.

PROPOSITION 3.3. *Let* x^* *be an accumulation point of the sequence* $\{x^k\}$ *generated by Algorithm* 2.1 *and suppose that* $\{x^k\}_{K_0} \to x^*$. *If*

$$\{\langle \nabla f(x^k), d^{k1}\rangle\}_{K_0} \to 0, \tag{3.1}$$

then x^* is a KKT point of problem (P) *and* $\{z^{k0}\}_{K_0}$ *converges to the unique multiplier vector λ^* associated with x^*.*

Proof. It follows from Assumption A3, Lemma 2.3, and (3.1) that

$$\{d^{k0}\}_{K_0} \to 0 \quad \text{and} \quad \{\langle \varphi^k, z^{k0}_{A^k} \rangle\}_{K_0} \to 0. \tag{3.2}$$

Let z^* be an arbitrary accumulation point of $\{z^{k0}\}_{K_0}$, and let $\{z^{k0}\}_{K_1}$ be a subsequence of $\{z^{k0}\}_{K_0}$ such that $\{z^{k0}\}_{K_1} \to z^*$. The boundedness of $\{z^{k0}\}$ implies that z^* exists. From (2.2), (3.2), and the definition of φ^k, we deduce

$$\begin{cases} \nabla f(x^*) + \nabla g(x^*) z^* = 0, \\ z_i^* \geq 0, \quad z_i^* g_i(x^*) = 0 \ \ \forall\, i \in I. \end{cases}$$

It is also obvious that $g(x^*) \leq 0$. Thus, x^* is a KKT point of problem (P) and z^* is its associated multiplier vector (i.e., $z^* = \lambda^*$). The uniqueness of the multiplier vector implies that $\{z^{k0}\}_{K_0} \to \lambda^*$. □

By Proposition 3.3, we establish a global convergence theorem for Algorithm 2.1.

THEOREM 3.4. *If (x^*, λ^*) is an accumulation point of the sequence $\{(x^k, z^{k0})\}$ generated by Algorithm* 2.1, *then (x^*, λ^*) is a KKT pair of problem* (P).

Proof. We prove the theorem by contradiction. Suppose that there is a subsequence $\{(x^k, z^{k0})\}_K$ converging to (x^*, λ^*), but (x^*, λ^*) is not a KKT pair of problem (P). We first prove that there must be a subset K_0 of K such that (3.1) holds. Otherwise, there exist $\gamma > 0$ and $\underline{d} > 0$ such that

$$\langle \nabla f(x^k), d^{k1} \rangle \leq -\gamma \ \ \forall\, k \in K \qquad \text{and} \qquad \liminf_{k \in K} \|d^{k1}\| \geq \underline{d}. \tag{3.3}$$

By the definition of ϕ^k, Lemma 3.1, and (3.3), it follows that there exists $\tilde{\phi} > 0$ such that

$$\phi^k \geq \tilde{\phi} \qquad \forall\ \ k \in K.$$

In a way similar to the proof of Lemma 3.9 in [23], we deduce

$$\begin{aligned} &f(x^k + td^k + t^2\hat{d}^k) - f(x^k) - \mu t \langle \nabla f(x^k), d^k \rangle \\ &\leq t \Bigg\{ \sup_{\xi \in [0,1]} \|\nabla f(x^k + t\xi d^k + t^2 \xi \hat{d}^k) - \nabla f(x^k)\| \|d^k\| \\ &\quad + 2t \sup_{\xi \in [0,1]} \|\nabla f(x^k + t\xi d^k + t^2 \xi \hat{d}^k)\| \|\hat{d}^k\| - (1-\mu)\vartheta C_1 \underline{d}^2 \Bigg\}, \end{aligned} \tag{3.4}$$

where C_1 and $\underline{d}$ are specified by Assumption A3 and (3.3), respectively. We also have for each $i \in I$

$$g_i(x^k + td^k + t^2\hat{d}^k) \leq g_i(x^k) + t\{u_i^k(t) + \langle \nabla g_i(x^k), d^k \rangle\} \tag{3.5}$$

with

$$\begin{aligned} u_i^k(t) := \ &\sup_{\xi \in [0,1]} \|\nabla g_i(x^k + t\xi d^k + t^2\xi \hat{d}^k) - \nabla g_i(x^k)\| \|d^k\| \\ &+ 2t \sup_{\xi \in [0,1]} \|\nabla g_i(x^k + t\xi d^k + t^2 \xi \hat{d}^k)\| \|\hat{d}^k\|. \end{aligned}$$

Hence, by (2.9), (3.3), (3.5), and the definition of φ^k, we have, for $i \in A^k$,

$$
\begin{aligned}
& g_i(x^k + td^k + t^2\hat{d}^k) \\
& \le g_i(x^k) + t\{u_i^k(t) + \varphi_i^k - \phi^k \|d^{k1}\|^\nu\} \\
& \le g_i(x^k) + t\{u_i^k(t) + \varphi_i^k - \tilde{\phi}\underline{d}^\nu\} \\
& = \begin{cases} g_i(x^k) + tz_i^{k0} + t\{u_i^k(t) - \tilde{\phi}\underline{d}^\nu\} & \text{if } z_i^{k0} < 0, \\ (1-t)g_i(x^k) + t\{u_i^k(t) - \tilde{\phi}\underline{d}^\nu\} & \text{if } z_i^{k0} > 0, \\ g_i(x^k) + t\{u_i^k(t) - \tilde{\phi}\underline{d}^\nu\} & \text{otherwise} \end{cases} \\
& \le t\{u_i^k(t) - \tilde{\phi}\underline{d}^\nu\}.
\end{aligned} \tag{3.6}
$$

On the other hand, for $i \notin A^k$, $g_i(x^k) < -\tilde{\epsilon}\rho(x^k, \lambda(x^k))$, and hence by (3.5), we get

$$
g_i(x^k + td^k + t^2\hat{d}^k) < -\tilde{\epsilon}\rho(x^k, \lambda(x^k)) + t\{u_i^k(t) + \langle \nabla g_i(x^k), d^k \rangle\}. \tag{3.7}
$$

Since $\rho(x^*, \lambda(x^*)) > 0$, $\{x^k\}_K \to x^*$, $\|\hat{d}^k\| \le \|d^k\|$, and $\{d^k\}$ is bounded, it follows from (3.6) and (3.7) that for all $i \in I$ there exists $\bar{t}_i > 0$, independent of k, such that, for all $t \in [0, \bar{t}_i]$ and $k \in K$ sufficiently large,

$$
g_i(x^k + td^k + t^2\hat{d}^k) < 0.
$$

Moreover, (3.4) implies that there exists $\bar{t}_f > 0$, independent of k, such that, for all $t \in [0, \bar{t}_f]$ and $k \in K$ sufficiently large,

$$
f(x^k + td^k + t^2\hat{d}^k) - f(x^k) - \mu t \langle \nabla f(x^k), d^k \rangle \le 0. \tag{3.8}
$$

Let

$$
\bar{t} := \min\{\bar{t}_f, \bar{t}_1, \ldots, \bar{t}_m\}.
$$

The line search rules (2.6) and (2.7) show that $t_k \ge \beta\bar{t}$ for all $k \in K$ sufficiently large, and hence by Lemma 2.3, (3.3), and (3.8) we deduce

$$
f(x^k + t_k d^k + t_k^2 \hat{d}^k) - f(x^k) \le -\mu\beta\bar{t}\vartheta\gamma, \quad k \in K. \tag{3.9}
$$

Since $\{f(x^k)\}$ is monotonically decreasing and bounded below, it converges. Taking limits in (3.9) as $k \to \infty$ with $k \in K$ yields a contradiction. The contradiction shows that (3.1) holds for some $K_0 \subseteq K$. It then follows from Proposition 3.3 that (x^*, λ^*) is a KKT pair of problem (P). The proof is complete. ∎

4. Superlinear convergence. In this section we analyze the rate of convergence of Algorithm 2.1. Let (x^*, λ^*) be an accumulation point of the sequence $\{(x^k, z^{k0})\}$. Then it follows from Theorem 3.4 that (x^*, λ^*) is a KKT pair of problem (P). For simplicity, we let $I_0 = I_0(x^*)$.

Assumption A4. The strict complementarity condition holds at (x^*, λ^*), i.e., $\lambda^* - g(x^*) > 0$.

Assumption A5. The second order sufficiency condition holds at (x^*, λ^*); i.e., the Hessian $\nabla_{xx}^2 L(x^*, \lambda^*)$ is positive definite on the space $\{u | \langle \nabla g_i(x^*), u \rangle = 0$ for all $i \in I_0\}$.

We first show that under the conditions of Assumptions A1–A3 and A5, the whole sequence $\{x^k\}$ converges to x^* and the sequence $\{z^k\}$ converges to λ^*. Then we prove that under Assumptions A1–A5, together with Assumption A6′, which will be introduced later in this section, the unit steplength is accepted for all k sufficiently large, and hence the Maratos effect does not occur. Finally, we show that the convergence rate is two-step superlinear or even Q-superlinear.

The following lemma follows from Theorems 2.3 and 3.7 in [9] directly.

LEMMA 4.1. *Let x^* be a KKT point of problem* (P) *and assume that Assumption* A5 *holds. Then there exists a neighborhood of x^* such that, for each x in this neighborhood,*

$$A(x; \tilde{\epsilon}) = I_0.$$

The above lemma indicates that the active constraints can be accurately identified close to a KKT point even if the strict complementarity condition does not hold at that point. To prove that the whole sequence $\{x^k\}$ converges to x^*, we cite another useful result from Proposition 7 in [16]. The original version of this result is due to Moré and Sorensen [20], which is slightly different from this version.

LEMMA 4.2. *Assume that $\omega^* \in \mathbb{R}^t$ is an isolated accumulation point of a sequence $\{\omega^k\} \subset \mathbb{R}^t$ such that, for every subsequence $\{\omega^k\}_K$ converging to ω^*, there is an infinite subset $K' \subseteq K$ such that $\{\|\omega^{k+1} - \omega^k\|\}_{K'} \to 0$. Then the whole sequence $\{\omega^k\}$ converges to ω^*.*

The next proposition claims the convergence of the whole sequence $\{x^k\}$.

PROPOSITION 4.3. *Under Assumptions* A1–A3 *and* A5*, the whole sequence $\{x^k\}$ converges to x^* and the sequence $\{z^{k0}\}$ converges to λ^*.*

Proof. Assumptions A2 and A5 imply that x^* is an isolated accumulation point of $\{x^k\}$ (see [30]). Let $\{x^k\}_K$ be a subsequence converging to x^*. It is clear from Lemma 4.1 that $A^k = I_0$ holds for $k \in K$ sufficiently large. We first prove that there must exist an infinite subset $K' \subseteq K$ such that

$$\{\|d^k\|\}_{K'} \to 0. \tag{4.1}$$

Suppose by contradiction that (4.1) does not hold for any infinite subset of K. Then

$$\liminf_{k\in K} \|d^k\| > 0,$$

which implies by Lemma 3.2 that

$$\liminf_{k\in K} \|d^{k1}\| > 0. \tag{4.2}$$

Without loss of generality, by Lemma 3.1 we assume that

$$\{(d^{k0}, z^{k0})\}_K \to (d^{*0}, z^{*0}) \quad \text{and} \quad \{(d^{k1}, z^{k1})\}_K \to (d^{*1}, z^{*1}).$$

Furthermore, we assume that $\{H_k\}_K \to H_*$. Taking limit in both sides of (2.2) as $k \to \infty$ with $k \in K$, we deduce that $(d^{*0}, z^{*0}_{I_0})$ solves the following system of linear equations:

$$V_* \begin{pmatrix} d \\ z_{I_0} \end{pmatrix} = \begin{pmatrix} -\nabla f(x^*) \\ 0 \end{pmatrix}, \tag{4.3}$$

where $V_* := V(x^*, H_*; I_0)$ is nonsingular. On the other hand, it is easy to see from the KKT system (1.1) that $(0, \lambda^*_{I_0})$ is the solution of system (4.3). So, we have $z^{*0}_{I_0} = \lambda^*_{I_0}$.

It then follows from the definition of φ^k that $\{\varphi^k\}_K \to 0$. Taking limit in (2.3) as $k \to \infty$ with $k \in K$, we see that $(d^{*1}, z_{I_0}^{*1})$ also satisfies (4.3), and hence $d^{*1} = 0$, which contradicts (4.2). This contradiction shows that (4.1) holds for some infinite subset $K' \subseteq K$. Therefore, we get from (4.1)

$$\|x^{k+1} - x^k\| \le \|d^k\| + \|\hat{d}^k\| \le 2\|d^k\| \to 0, \quad \text{as } k \to \infty \text{ with } k \in K'.$$

By means of Lemma 4.2, we claim that the whole sequence $\{x^k\}$ converges to x^*. Moreover, the uniqueness of the multiplier vector λ^* implies that $\{z^{k0}\}$ converges to λ^*. ∎

The following results are a direct corollary of Proposition 4.3 and will play an important role in the analysis of the convergence rate.

COROLLARY 4.4. *Let Assumptions* A1–A3 *and* A5 *hold. Then the equality* $A^k = I_0$ *holds for all* k *sufficiently large. Furthermore, we have*

(i) $d^k \to 0,\ d^{k0} \to 0,\ d^{k1} \to 0$*, as* $k \to \infty$.

(ii) $z^k \to \lambda^*,\ z^{k1} \to \lambda^*$*, as* $k \to \infty$.

(iii) *If, in addition, Assumption* A4 *holds, then for* k *sufficiently large it holds that* $\varphi^k = -g_{I_0}(x^k)$.

Proof. We have by Lemma 4.1 and Proposition 4.3 that $\{z^{k0}\} \to \lambda^*$ and that $A^k = I_0$ holds for all k sufficiently large. It is also not difficult to see from Proposition 4.3 and the uniqueness of the solution of system (4.3) that the sequences $\{(d^{k0}, z_{I_0}^{k0})\}$, $\{(d^{k1}, z_{I_0}^{k1})\}$, and $\{(d^{k2}, z_{I_0}^{k2})$ converge to the unique solution of system (4.3). This shows (i) and (ii).

If Assumption A4 holds, then we have $z_{I_0}^{k0} > 0$ for all k sufficiently large. This implies (iii). ∎

System (2.3) and Corollary 4.4(iii) show that for k sufficiently large $(d^{k1}, z_{I_0}^{k1})$ is the unique solution of the following system of linear equations:

$$\text{(4.4)} \qquad \begin{cases} H_k d + \nabla g_{I_0}(x^k) z_{I_0} = -\nabla f(x^k), \\ \nabla g_{I_0}(x^k)^T d = -g_{I_0}(x^k). \end{cases}$$

This means that d^{k1} produced by (2.3) can be regarded as a quasi-Newton direction for the equality constrained optimization problem

$$\text{(4.5)} \qquad \begin{array}{ll} \min & f(x), \\ \text{s.t.} & g_{I_0}(x) = 0. \end{array}$$

It is interesting to note that the local algorithm proposed by Facchinei and Lazzari [11] generates a direction d^k which is a Newton direction of (4.5). In other words, d^k generated by the algorithm in [11] is the solution of (4.4) with H_k taken from the generalized Hessian $\partial^2_{xx} L(x^k, \lambda(x^k))$. Our method is slightly different from the method in [11] in that d^k in our method is only an approximate solution of (4.4) because we have $d^k = d^{k1} + O(\|d^{k1}\|^\nu)$ with $\nu > 2$ by Lemma 3.2.

We are going to prove the superlinear convergence of the proposed method. It is well known that the Dennis and Moré condition [7] is necessary and sufficient for superlinear convergence of a quasi-Newton method for solving nonlinear equations or unconstrained optimization problems. Boggs, Tolle, and Wang [3] extended this result to the quasi-Newton method for solving equality constrained optimization problems (see also [33]). We will extend this result to our algorithm.

Assumption A6′. The sequence of matrices $\{H_k\}$ satisfies

$$\frac{\|P_k(H_k - \nabla^2_{xx}L(x^*, \lambda^*))P_k d^k\|}{\|d^k\|} \to 0,$$

where

$$P_k := E - N_k(N_k^T N_k)^{-1} N_k^T \quad \text{and} \quad N_k := \nabla g_{I_0}(x^k).$$

We will show that Assumption A6′ is a sufficient condition for our algorithm to be two-step superlinearly convergent. To this end, we first prove two lemmas.

LEMMA 4.5. *When k is sufficiently large, the direction d^k can be decomposed into*

$$d^k = P_k d^k + \tilde{d}^k$$

with

$$\|\tilde{d}^k\| = O(\|g_{I_0}(x^k)\|) + o(\|d^{k1}\|^2).$$

Proof. It follows from (2.9) and Corollary 4.4(iii) that for k sufficiently large

$$\begin{aligned} \langle \nabla g_i(x^k), d^k \rangle &= \varphi_i^k - \phi^k \|d^{k1}\|^\nu \\ &= -g_i(x^k) - \phi^k \|d^{k1}\|^\nu \qquad \forall\, i \in I_0. \end{aligned}$$

This implies that

$$N_k^T d^k = h^k,$$

where

$$h^k := -g_{I_0}(x^k) - \phi^k \|d^{k1}\|^\nu e_{I_0}.$$

Thus, we have

$$\begin{aligned} d^k &= P_k d^k + N_k(N_k^T N_k)^{-1} N_k^T d^k \\ &= P_k d^k + N_k(N_k^T N_k)^{-1} h^k \\ &= P_k d^k + \tilde{d}^k, \end{aligned}$$

where

$$\tilde{d}^k := N_k(N_k^T N_k)^{-1} h^k$$

satisfies

$$\|\tilde{d}^k\| = O(\|h^k\|) = O(\|g_{I_0}(x^k)\|) + o(\|d^{k1}\|^2). \qquad \square$$

LEMMA 4.6. *When k is sufficiently large, the direction $\hat{d}^k$ is determined by solving system* (2.5)*, and it satisfies*

$$\|\hat{d}^k\| = O(\|d^k\|^2).$$

Proof. It follows from (2.5) and Corollary 4.4 that when k is sufficiently large the direction $\hat{d}^k$ is first computed by solving the following system of linear equations:

$$V_k \begin{pmatrix} d \\ z_{I_0} \end{pmatrix} = \begin{pmatrix} 0 \\ -\|d^k\|^\tau e_{I_0} - g_{I_0}(x^k + d^k) \end{pmatrix} \tag{4.6}$$

with $V_k = V(x^k, H_k; I_0)$.

By Taylor's expansion, we get for each $i \in I_0$

$$\begin{aligned} &-\|d^k\|^\tau - g_i(x^k + d^k) \\ &= -\|d^k\|^\tau - [g_i(x^k) + \langle \nabla g_i(x^k), d^k \rangle + O(\|d^k\|^2)] \\ &= -\|d^k\|^\tau + \phi^k \|d^{k1}\|^\nu + O(\|d^k\|^2) \\ &= O(\|d^k\|^2), \end{aligned}$$

where the second equality follows from (2.9) and Corollary 4.4(iii), and the last equality follows from Lemma 3.2, respectively. The assertion then follows from (4.6) and the fact that $\|V_k^{-1}\| \le \tilde{M}$ for all k. ∎

We are now in a position to prove that a unit step is eventually accepted by Algorithm 2.1.

PROPOSITION 4.7. *Let Assumptions* A1–A5 *and* A6′ *hold. Then when k is sufficiently large the step $t_k = 1$ is accepted.*

Proof. By the line search rules (2.6) and (2.7), we need only to show that for k sufficiently large the following two conditions hold:

(i) The sufficient decrease condition (2.6) on f holds for $t = 1$.

(ii) The strict feasibility condition (2.7) on g holds for $t = 1$.

It follows from Lemma 4.6 that

$$\begin{aligned} f(x^k + d^k + \hat{d}^k) = f(x^k) + \langle \nabla f(x^k), d^k + \hat{d}^k \rangle& \\ + \frac{1}{2}\langle d^k, \nabla^2_{xx} f(x^k) d^k \rangle + o(\|d^k\|^2).& \end{aligned} \tag{4.7}$$

In view of (2.5), (2.9), and Corollary 4.4, for k sufficiently large

$$H_k d^k + \nabla f(x^k) + \sum_{i \in I_0} z_i^k \nabla g_i(x^k) = 0, \tag{4.8}$$

$$\langle \nabla g_i(x^k), d^k \rangle = -g_i(x^k) - \phi^k \|d^{k1}\|^\nu \quad \forall\, i \in I_0, \tag{4.9}$$

and

$$g_i(x^k + d^k) + \langle \nabla g_i(x^k), \hat{d}^k \rangle = -\|d^k\|^\tau \quad \forall\, i \in I_0. \tag{4.10}$$

By (4.8) and Lemma 4.6, we have

$$\langle \nabla f(x^k), d^k \rangle = -\langle d^k, H_k d^k \rangle - \sum_{i \in I_0} z_i^k \langle \nabla g_i(x^k), d^k \rangle \tag{4.11}$$

and

$$\langle \nabla f(x^k), \hat{d}^k \rangle = -\sum_{i \in I_0} z_i^k \langle \nabla g_i(x^k), \hat{d}^k \rangle + o(\|d^k\|^2). \tag{4.12}$$

Thus, from (4.7), (4.11), and (4.12) we deduce

$$\begin{aligned} &f(x^k + d^k + \hat{d}^k) \\ &= f(x^k) + \frac{1}{2}\langle \nabla f(x^k), d^k \rangle + \frac{1}{2}\langle d^k, (\nabla^2_{xx} f(x^k) - H_k) d^k \rangle \\ &\quad - \frac{1}{2}\sum_{i \in I_0} z_i^k \langle \nabla g_i(x^k), d^k \rangle - \sum_{i \in I_0} z_i^k \langle \nabla g_i(x^k), \hat{d}^k \rangle + o(\|d^k\|^2). \end{aligned} \tag{4.13}$$

Furthermore, for all $i \in I_0$ it follows from (4.10) that

$$g_i(x^k) + \langle \nabla g_i(x^k), d^k + \hat{d}^k \rangle + \frac{1}{2}\langle d^k, \nabla^2_{xx} g_i(x^k) d^k \rangle = o(\|d^k\|^2). \tag{4.14}$$

Using (4.9), (4.14), and Lemma 3.2, we obtain

$$\begin{aligned}
&-\frac{1}{2}\sum_{i\in I_0} z_i^k \langle \nabla g_i(x^k), d^k \rangle - \sum_{i\in I_0} z_i^k \langle \nabla g_i(x^k), \hat{d}^k \rangle \\
&= \frac{1}{2}\sum_{i\in I_0} z_i^k \langle \nabla g_i(x^k), d^k \rangle - \sum_{i\in I_0} z_i^k \langle \nabla g_i(x^k), d^k + \hat{d}^k \rangle \\
&= \frac{1}{2}\sum_{i\in I_0} z_i^k g_i(x^k) - \frac{1}{2}\sum_{i\in I_0} \phi^k \|d^{k1}\|^\nu z_i^k \\
&\quad + \frac{1}{2}\sum_{i\in I_0} z_i^k \langle d^k, \nabla^2_{xx} g_i(x^k) d^k \rangle + o(\|d^k\|^2) \\
&= \frac{1}{2}\sum_{i\in I_0} z_i^k g_i(x^k) + \frac{1}{2}\sum_{i\in I_0} z_i^k \langle d^k, \nabla^2_{xx} g_i(x^k) d^k \rangle + o(\|d^k\|^2).
\end{aligned} \tag{4.15}$$

Clearly, Assumption A4 and Corollary 4.4 imply that, for each $i \in I_0$ and any k sufficiently large, $z_i^k \geq 0.5\lambda_i^* > 0$; hence we get for k sufficiently large

$$\frac{1}{2}\sum_{i\in I_0} z_i^k g_i(x^k) + o(\|g_{I_0}(x^k)\|) < 0. \tag{4.16}$$

In view of (4.15)–(4.16) and Assumption A6′, we obtain from (4.13)

$$\begin{aligned}
&f(x^k + d^k + \hat{d}^k) \\
&= f(x^k) + \frac{1}{2}\langle \nabla f(x^k), d^k \rangle + \frac{1}{2}\sum_{i\in I_0} z_i^k g_i(x^k) \\
&\quad + \frac{1}{2}\langle d^k, (\nabla^2_{xx} f(x^k) + \sum_{i\in I_0} z_i^k \nabla^2_{xx} g_i(x^k) - H_k) d^k \rangle + o(\|d^k\|^2) \\
&= f(x^k) + \frac{1}{2}\langle \nabla f(x^k), d^k \rangle + \frac{1}{2}\sum_{i\in I_0} z_i^k g_i(x^k) + o(\|g_{I_0}(x^k)\|) \\
&\quad + \frac{1}{2}\langle d^k, P_k(\nabla^2_{xx} f(x^k) + \sum_{i\in I_0} z_i^k \nabla^2_{xx} g_i(x^k) - H_k) P_k d^k \rangle + o(\|d^k\|^2) \\
&\leq f(x^k) + \frac{1}{2}\langle \nabla f(x^k), d^k \rangle \\
&\quad + \frac{1}{2}\|d^k\| \left\| P_k \left(\nabla^2_{xx} f(x^k) + \sum_{i\in I_0} z_i^k \nabla^2_{xx} g_i(x^k) - H_k \right) P_k d^k \right\| + o(\|d^k\|^2) \\
&= f(x^k) + \frac{1}{2}\langle \nabla f(x^k), d^k \rangle + o(\|d^k\|^2),
\end{aligned} \tag{4.17}$$

where the second equality follows from Lemmas 4.5 and 3.2. We also have from (4.4)

$$\begin{aligned}
\langle \nabla f(x^k), d^{k1} \rangle &= -\langle d^{k1}, H_k d^{k1} \rangle - \langle d^{k1}, \nabla g_{I_0}(x^k) z_{I_0}^{k1} \rangle \\
&= -\langle d^{k1}, H_k d^{k1} \rangle + \langle g_{I_0}(x^k), z_{I_0}^{k1} \rangle \\
&< -\langle d^{k1}, H_k d^{k1} \rangle,
\end{aligned} \tag{4.18}$$

where the last inequality is due to $g_{I_0}(x^k) < 0$ and for k sufficiently large $z_{I_0}^{k1} > 0$. This, together with Lemma 2.3(iii), Assumption A3, and Lemma 3.2, implies that

$$\begin{aligned} \langle \nabla f(x^k), d^k \rangle &\leq \vartheta \langle \nabla f(x^k), d^{k1} \rangle \\ &\leq -\vartheta \langle d^{k1}, H_k d^{k1} \rangle \\ &\leq -\vartheta C_1 \|d^{k1}\|^2 \\ &= -\vartheta C_1 \|d^k\|^2 + o(\|d^k\|^2). \end{aligned} \tag{4.19}$$

Due to $\mu < \frac{1}{2}$, inequalities (4.17) and (4.19) show that for k sufficiently large $t = 1$ satisfies inequality (2.6), i.e.,

$$f(x^k + d^k + \hat{d}^k) \leq f(x^k) + \mu \langle \nabla f(x^k), d^k \rangle.$$

This proves (i). We now turn to prove (ii).

It is clear from Corollary 4.4 and Lemma 4.6 that $d^k \to 0$ and $\hat{d}^k \to 0$. For $i \notin I_0$, $g_i(x^*) < 0$ implies that for k sufficiently large

$$g_i(x^k + d^k + \hat{d}^k) < 0. \tag{4.20}$$

For $i \in I_0$, we have from (2.5) and Lemma 4.6 that for k sufficiently large

$$\begin{aligned} g_i(x^k + d^k + \hat{d}^k) &= g_i(x^k + d^k) + \langle \nabla g_i(x^k + d^k), \hat{d}^k \rangle + O(\|\hat{d}^k\|^2) \\ &= g_i(x^k + d^k) + \langle \nabla g_i(x^k), \hat{d}^k \rangle + O(\|d^k\| \|\hat{d}^k\|) \\ &= -\|d^k\|^\tau + O(\|d^k\|^3) \\ &= -\|d^k\|^\tau + o(\|d^k\|^\tau) \\ &\leq -\frac{1}{2}\|d^k\|^\tau < 0. \end{aligned}$$

This, together with (4.20), shows (ii). This completes the proof. □

Proposition 4.7 shows that the use of $\hat{d}^k$ on the search direction makes the unit step accepted for all k sufficiently large. Consequently, the Maratos effect does not appear. The next theorem indicates that Algorithm 2.1 is two-step superlinearly convergent.

THEOREM 4.8. *Let Assumptions* A1–A5 *and* A6′ *hold. Then the sequence* $\{x^k\}$ *generated by Algorithm* 2.1 *converges two-step superlinearly, i.e.,*

$$\lim_{k \to \infty} \frac{\|x^{k+2} - x^*\|}{\|x^k - x^*\|} = 0.$$

The proof of the above theorem follows step by step, with minor modifications, that of Theorem 4.6 in [23]. The details are omitted.

Furthermore, in the following, we show the Q-superlinear convergence of Algorithm 2.1 if Assumption A6′ is replaced by a stronger assumption.

Assumption A6. The sequence of matrices $\{H_k\}$ satisfies

$$\frac{\|P_k(H_k - \nabla^2_{xx} L(x^*, \lambda^*)) d^k\|}{\|d^k\|} \to 0.$$

THEOREM 4.9. *Let Assumptions* A1–A6 *hold. Then the sequence* $\{x^k\}$ *generated by Algorithm* 2.1 *converges Q-superlinearly, i.e.,*

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

If, in addition, supposing that $\nabla^2 f$ *and* $\nabla^2 g_i$*, for all* $i \in I$*, are Lipschitz continuous and* $H_k = \nabla^2_{xx} L(x^k, \lambda(x^k))$*, then the convergence rate is Q-quadratic, i.e.,*

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

Proof. In view of Proposition 4.7 and Lemmas 3.2 and 4.6, we have for k sufficiently large

$$x^{k+1} - x^k = d^k + \hat{d}^k = d^{k1} + O(\|d^{k1}\|^2) = d^{k1} + o(\|d^{k1}\|). \tag{4.21}$$

For k sufficiently large, d^{k1} can be viewed as a quasi-Newton direction for the equality constrained optimization problem (4.5). It follows from Lemma 3.2 that $d^k = d^{k1} + o(\|d^{k1}\|)$. By the boundedness of P_k, H_k, and $\nabla^2_{xx} L(x^*, \lambda^*)$, Assumption A6 is equivalent to

$$\frac{\|P_k(H_k - \nabla^2_{xx} L(x^*, \lambda^*))d^{k1}\|}{\|d^{k1}\|}.$$

Combining this expression and the results in [33], we have

$$\|x^k + d^{k1} - x^*\| = o(\|x^k - x^*\|). \tag{4.22}$$

Furthermore, by the use of Lemma 3.1 in [8], we get

$$\lim_{k \to \infty} \frac{\|d^{k1}\|}{\|x^k - x^*\|} = 1. \tag{4.23}$$

So, by (4.21)–(4.23), it holds that

$$\|x^{k+1} - x^*\| = \|x^k + d^{k1} + o(\|d^{k1}\|) - x^*\| = o(\|x^k - x^*\|),$$

which shows that $\{x^k\}$ converges to x^* Q-superlinearly.

If $H_k = \nabla^2_{xx} L(x^k, \lambda(x^k))$ for k sufficiently large, we get from Theorem 3.1 in [12] that

$$\|x^k + d^{k1} - x^*\| = O(\|x^k - x^*\|^2).$$

This, together with (4.21) and (4.23), yields

$$\|x^{k+1} - x^*\| = \|x^k + d^{k1} + O(\|d^{k1}\|^2) - x^*\| = O(\|x^k - x^*\|^2),$$

which shows that the convergence rate is Q-quadratic. ☐

5. Numerical experiments. In this section we report the numerical results on a test set that includes some of Hock and Schittkowski's problems [15] as well as several other large-scale real-world problems from the CUTE [4] and the COPS [5] collections. The algorithm was implemented by a Matlab code. For each test problem, we chose $H_1 = E$ as the initial guess of the Lagrangian Hessian. At each step, the matrix H_k was updated by the damped BFGS formula from Powell [24] as in [17, 26]. Specifically, we set

$$H_{k+1} = H_k - \frac{H_k s_k s_k^T H_k}{s_k^T H_k s_k} + \frac{y_k y_k^T}{s_k^T y_k},$$

where

$$y_k = \begin{cases} \hat{y}_k & \text{if } \hat{y}_k^T s_k \geq 0.2 s_k^T H_k s_k, \\ \theta_k \hat{y}_k + (1-\theta_k) H_k s_k & \text{otherwise} \end{cases}$$

and

$$\begin{cases} s_k = x^{k+1} - x^k, \\ \hat{y}_k = \nabla f(x^{k+1}) - \nabla f(x^k) + (\nabla g(x^{k+1}) - \nabla g(x^k))\lambda^{k0}, \\ \theta_k = 0.8 s_k^T H_k s_k / (s_k^T H_k s_k - s_k^T \hat{y}_k). \end{cases}$$

We set the parameters as follows:

$$\beta = 0.5, \quad \mu = 0.1, \quad \nu = 3.0, \quad \tau = 2.5, \quad \vartheta = 0.5, \quad \sigma = 0.1, \text{ and } \epsilon^0 = 3.0.$$

The algorithm stops if one of the following termination criteria is satisfied:

(a) $\|\Phi(x^k, \lambda(x^k))\| \leq 10^{-5}$.
(b) $\|\Phi(x^k, z^{k0})\| \leq 10^{-5}$.
(c) $\|d^{k1}\|/(1 + \|x^k\|) \leq 10^{-5}$.

The first and second criteria state the KKT conditions for problem (P). At Step 2 of the algorithm $\Phi(x^k, \lambda(x^k))$ has to be computed so as to estimate the working set and to update the parameter ϵ. Hence the first criterion is used here. Moreover, Lemma 2.2 implies that x^k is only a trivial KKT point of problem (P) if $d^{k1} = 0$. Hence in our implementation the second or third criterion is used at Step 4(ii) as the termination criterion.

The check of full rankness in Step 2 is done by using the Matlab command "rank."

We first tested some problems taken from [15]. For these test problems, we used the initial point given in [15] if it was strictly feasible. For some problems whose initial points given in [15] were not strictly feasible, we chose other initial points which were strictly feasible. These initial points are listed in Table 1.

Table 1
Starting points for some HS problems.

Problem	Starting point
HS25	(25, 5, 1)
HS30	(3, 2, 1)
HS31	(4, 3, -2)
HS33	(1, 3, 4)
HS34	(0.1, 1.15, 3.2)
HS65	(0, 0, 0)
HS66	(0.5, 2, 8)

The computational results are shown in Table 2, where the columns have the following meanings:

TABLE 2
Numerical results on the HS problems.

Problem	n	m	Iter	Nf	Ng	Fv	Term	Prec	Final-ϵ	Aset
HS1	2	1	36	57	59	6.662078e-14	(a)	3.862116e-06	0.3000	0
HS3	2	1	10	17	26	2.293930e-08	(a)	2.388519e-08	0.3000	1
HS4	2	2	3	4	7	2.666667	(a)	4.049708e-11	0.3000	2
HS5	2	4	5	8	8	-1.913223	(c)	6.007380e-06	0.3000	0
HS12	2	1	7	24	28	-30.000000	(c)	2.902889e-06	3.0000	1
HS24	2	5	9	13	22	-1.000000	(a)	7.192276e-08	0.3000	2
HS25	3	6	14	55	62	3.318784e-06	(c)	1.268899e-06	0.3000	0
HS29	3	1	9	28	34	-22.627417	(a)	6.168718e-06	3.0000	1
HS30	3	7	10	27	34	1.000000	(a)	6.986504e-07	0.3000	2
HS31	3	7	11	32	40	6.000000	(c)	1.925959e-06	0.3000	1
HS33	3	6	15	74	87	-4.585782	(a)	7.512854e-07	0.3000	3
HS34	3	8	17	76	92	-0.834024	(c)	1.611863e-06	0.3000	3
HS35	3	4	7	13	19	0.111111	(c)	8.006110e-06	0.3000	1
HS36	3	7	11	33	44	-3.300000e+03	(a)	6.296636e-07	0.3000	3
HS37	3	8	15	45	58	-3.456000e+03	(c)	6.947600e-06	0.3000	1
HS38	4	8	49	91	91	5.128073e-11	(c)	1.890126e-06	0.0300	0
HS43	4	3	12	36	45	-44.000000	(c)	6.631011e-06	0.3000	2
HS44	4	10	17	60	73	-14.999860	(c)	3.117109e-06	0.3000	4
HS65	3	7	8	19	22	0.953529	(a)	6.193657e-08	0.3000	1
HS66	3	8	11	24	35	0.518164	(a)	7.452867e-06	3.0000	2
HS76	4	7	9	15	23	-4.681818	(a)	8.369451e-10	0.3000	2
HS93	6	8	18	51	69	135.075964	(c)	2.708803e-06	0.3000	2
HS100	7	4	14	44	58	680.630057	(c)	6.183027e-06	3.0000	2
HS113	10	8	21	58	79	24.306209	(c)	1.424558e-06	3.0000	6

Problem: the problem number given in [15],
n: the number of variables,
m: the number of constraints (including bound constraints),
Iter: the number of iterations,
Nf: the number of function evaluations for f,
Ng: the number of function evaluations for g,
Fv: the objective function value at the final iterate,
Term: the label of the termination criterion,
Prec: the final value of the norm function used in the termination criteria,
Final-ϵ: the value of the parameter ϵ at the final iterate,
Aset: the number of indices in the final working set.

We succeeded in solving all test problems chosen in Table 2, and for most of these problems the number of iterations was small. The computational results illustrate that our algorithm is competitive with those in [26, 34].

All of the problems in the Hock and Schittkowski set [15] are very small. To see more clearly the effectiveness of our algorithm, we tested several problems from the CUTE collection [4] and two problems from the COPS collection [5] that contained no equality constraints. Some of these problems are larger and therefore more interesting. Table 3 lists starting points of these problems, except for the last problem, whose initial points vary with its dimension. We also succeeded in solving all these test problems. The computational results are listed in Tables 4 and 5, where the termination criterion (c) is changed to $\|d^{k1}\| \leq 10^{-5}$.

The results reported in Tables 4 and 5 are encouraging. First, we note that here the number of iterations and hence the number of objective function evaluations are

TABLE 3
Starting points for problems in Tables 4 and 5.

Problem	Starting point
Expfit	$x = (6, 1, 6, 0, 0)^T$
Ngone-k	$x_i = 0.8 * i/k, i = 1, \cdots, k; y_i = 0.6, i = 1, \cdots, k-1$
Obstclae-k	$x_{i,j} = 1, i, j = 2, \cdots, k-1$
Svanberg-k	$x_i = 0, i = 1, \cdots, k$
Polygon-k	$r_i = 0.5, \theta_i = \pi * i/k, i = 1, \cdots, k-1$

TABLE 4
Numerical results on the CUTE problems.

Problem	n	m	Iter	Nf	Ng	Fv	Term	Prec	Final-ϵ	Aset
Expfita	5	22	228	1758	1897	0.00113661	(a)	1.866081e-08	0.0300	4
Expfitb	5	102	157	1107	1204	0.00501937	(a)	1.159079e-08	0.0300	4
Expfitc	5	502	273	2824	2964	0.02330257	(a)	2.493369e-06	0.0003	3
Ngone-3	8	9	5	11	15	-0.500000	(c)	8.261035e-06	0.3000	2
Ngone-5	12	20	14	26	39	-0.620366	(a)	9.202883e-07	0.3000	7
Ngone-24	50	324	241	95	1199	-0.643097	(b)	8.010034e-06	0.0300	26
Ngone-49	100	1274	1414	10876	12290	-0.643421	(c)	8.811006e-06	0.0300	51
Obstclae-4	16	32	4	5	9	0.753660	(a)	5.825214e-07	3.0000	4
Obstclae-10	100	200	165	979	1144	1.397898	(a)	8.353902e-06	3.0000	29
Obstclae-23	529	1058	908	7179	8087	1.678027	(a)	7.187001e-06	3.0000	221
Obstclae-32	1024	2048	2438	22024	24462	1.748270	(a)	9.658341e-06	3.0000	472
Svanberg-10	10	30	36	227	258	15.731517	(c)	5.365582e-06	0.0300	6
Svanberg-30	30	90	101	777	864	49.142526	(c)	9.130506e-06	0.0300	22
Svanberg-50	50	150	108	881	968	82.581912	(c)	9.472167e-06	0.0300	38
Svanberg-80	80	240	190	1666	1835	132.749819	(c)	4.663239e-06	0.0300	61
Svanberg-100	100	300	178	1628	1782	166.197171	(c)	7.281111e-06	0.0300	77
Svanberg-500	500	1500	402	4020	4407	835.186918	(c)	5.299494e-06	0.0300	398

TABLE 5
Numerical results on the COPS problems.

Problem	n	m	Iter	Nf	Ng	Fv	Term	Prec	Final-ϵ	Aset
Polygon-4	6	17	6	17	21	-0.500000	(a)	9.911252e-09	0.3000	2
Polygon-6	10	34	13	26	38	-0.674981	(a)	1.813151e-07	0.3000	6
Polygon-10	18	80	18	31	49	-0.749137	(a)	3.669190e-06	0.3000	10
Polygon-15	28	160	43	159	199	-0.768622	(a)	7.178912e-06	0.0300	15
Polygon-20	38	265	78	348	422	-0.776859	(a)	9.648048e-06	0.0300	20
Polygon-25	48	395	96	403	494	-0.780232	(a)	9.167416e-06	0.0300	25
Polygon-30	58	550	137	739	872	-0.781674	(a)	9.128916e-06	0.0300	30
Polygon-40	78	935	416	3113	3509	-0.783069	(a)	7.798918e-06	0.0030	40
Polygon-50	98	1420	1416	14855	16192	-0.783799	(b)	8.082852e-06	0.0003	50
Cam-10	10	43	15	155	170	-43.85994	(a)	1.336130e-07	3.0e-4	10
Cam-20	20	83	14	110	124	-86.55864	(a)	3.498434e-06	3.0e-4	20
Cam-50	50	203	14	166	180	-214.6961	(b)	6.369979e-06	3.0e-6	50
Cam-100	100	403	17	244	261	-427.8899	(b)	9.198429e-06	3.0e-6	100
Cam-200	200	803	55	552	607	-855.7000	(c)	7.568989e-08	3.0e-7	200
Cam-400	400	1603	98	1207	1305	-1710.275	(c)	9.040695e-08	3.0e-8	400

generally larger than those reported in [17] for a feasible SQP method. This is understandable because the subproblems of Algorithm 2.1 are low dimensional, which use only partial information of the problems. The number of constraint function evaluations here is competitive with that of a feasible SQP method. On the other

TABLE 6
Number of indices in the working set on the problem "Obstclae-10."

Iteration	1	120	123	131	134	135	140	143	144	147
Working set	64	62	64	62	60	58	56	55	56	55
Iteration	148	149	150	154	156	158	161	162	163	
Working set	53	39	36	35	34	33	32	31	29	

hand, Tables 4 and 5 also show that the cardinality of the final working set "Aset" is generally much smaller than the number of constraints. This means that the subproblems of Algorithm 2.1 are generally much smaller than that of the full dimensional feasible SQP methods. Moreover, as the number of constraints in problem (P) increases, this benefit becomes extremely apparent. This shows the potential advantage of our algorithm when applied to solving problems with large numbers of constraints. Table 6 positively supports this possibility. Table 6 lists the numbers of indices in the working set corresponding to iterations when Algorithm 2.1 is applied to solving problem "Obstclae-10." The results show that as iteration increases, the number of corresponding indices in the working set exhibits the decreasing tendency.

6. Conclusion. In this paper an FSLE algorithm for inequality constrained optimization is proposed. The proposed algorithm is based on an efficient identification technique of the active constraints and has some nice properties. We have proved that every accumulation point of the sequence generated by the proposed algorithm is a KKT point of problem (P) without requiring the isolatedness of the stationary points. We have also established locally two-step superlinear or Q-superlinear or Q-quadratic convergence for the proposed algorithm under mild assumptions. The preliminary numerical experiments show that the proposed method is effective for the test problems. However, to achieve superlinear convergence of the algorithm we still need the strict complementarity condition. Recently, Facchinei, Lucidi, and Palagi [13] proposed a globally and superlinearly convergent truncated Newton method for solving the box constrained optimization. In particular, they established superlinear convergence without requiring the strict complementarity condition. How to remove this condition for the general constrained optimization is an important topic for further research.

REFERENCES

[1] S. Bakhtiari and A. Tits, *A simple primal-dual feasible interior-point method for nonlinear programming with monotone descent*, Comput. Optim. Appl., 25 (2003), pp. 17–38.

[2] P. T. Boggs and J. W. Tolle, *Sequential Quadratic Programming*, Acta Numer. 4, Cambridge University Press, Cambridge, UK, 1995, pp. 1–51.

[3] 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 (1982), pp. 161–171.

[4] I. Bongartz, A. R. Conn, N. I. M. Gould, and Ph. L. Toint, *CUTE: Constrained and unconstrained testing environment*, ACM Trans. Math. Software, 21 (1995), pp. 123–160.

[5] A. S. Bondarenko, D. M. Bortz, and J. J. Moré, *COPS: Large-Scale Nonlinearly Constrained Optimization Problems*, Technical report ANL/MCS-TM-237, Argonne National Laboratory, Argonne, IL, 1998.

[6] J. V. Burke and S. P. Han, *A robust sequential quadratic programming method*, Math. Programming, 43 (1989), pp. 277–303.

[7] J. E. Dennis and J. J. Moré, *A characterization of superlinear convergence and its application to quasi-Newton methods*, Math. Comp., 28 (1974), pp. 549–560.

[8] F. Facchinei, *Minimization of SC^1-functions and the Maratos effect*, Oper. Res. Lett., 17 (1995), pp. 131–137.

[9] F. Facchinei, A. Fischer, and C. Kanzow, *On the accurate identification of active constraints*, SIAM J. Optim., 9 (1998), pp. 14–32.
[10] F. Facchinei, J. Júdice, and J. Soares, *An active set Newton algorithm for large-scale nonlinear programs with box constraints*, SIAM J. Optim., 8 (1998), pp. 158–186.
[11] F. Facchinei and C. Lazzari, *Local feasible QP-free algorithm for the constrained minimization of SC^1 functions*, J. Optim. Theory Appl., to appear.
[12] F. Facchinei and S. Lucidi, *Quadratically and superlinearly convergent algorithms for the solution of inequality constrained minimization problems*, J. Optim. Theory Appl., 85 (1995), pp. 265–289.
[13] F. Facchinei, S. Lucidi, and L. Palagi, *A truncated Newton algorithm for large scale box constrained optimization*, SIAM J. Optim., 12 (2002), pp. 1100–1125.
[14] T. Glad and E. Polak, *A multiplier method with automatic limitation of penalty growth*, Math. Programming, 17 (1979), pp. 140–155.
[15] W. Hock and K. Schittkowski, *Test Examples for Nonlinear Programming Codes*, Lecture Notes in Econom. and Math. Systems 187, Springer-Verlag, Berlin, New York, 1981.
[16] C. Kanzow and H. D. Qi, *A QP-free constrained Newton-type method for variational inequality problems*, Math. Program., 85 (1999), pp. 81–106.
[17] C. T. Lawrence and A. L. Tits, *A computationally efficient feasible sequential quadratic programming algorithm*, SIAM J. Optim., 11 (2001), pp. 1092–1118.
[18] X.-W. Liu and Y.-X. Yuan, *A robust algorithm for optimization with general equality and inequality constraints*, SIAM J. Sci. Comput., 22 (2000), pp. 517–534.
[19] S. Lucidi, *New results on a continuously differentiable exact penalty function*, SIAM J. Optim., 2 (1992), pp. 558–574.
[20] J. J. Moré and D. C. Sorensen, *Computing a trust region step*, SIAM J. Sci. Statist. Comput., 4 (1983), pp. 553–572.
[21] E. R. Panier and A. L. Tits, *A superlinearly convergent feasible method for the solution of inequality constrained optimization problems*, SIAM J. Control Optim., 25 (1987), pp. 934–950.
[22] E. Panier and A. L. Tits, *On combining feasibility, descent and superlinear convergence in inequality constrained optimization*, Math. Programming, 59 (1993), pp. 261–276.
[23] E. R. Panier, A. L. Tits, and J. N. Herskovits, *A QP-free globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization*, SIAM J. Control Optim., 26 (1988), pp. 788–811.
[24] M. J. D. Powell, *A fast algorithm for nonlinearly constrained optimization calculations*, in Numerical Analysis, Lecture Notes in Math. 630, Springer, Berlin, 1978, pp. 144–157.
[25] M. J. D. Powell, *The convergence of variable metric methods for nonlinearly constrained optimization calculations*, in Nonlinear Programming 3, O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, eds., Academic Press, New York, 1978, pp. 27–63.
[26] H.-D. Qi and L. Qi, *A new QP-free globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization*, SIAM J. Optim., 11 (2000), pp. 113–132.
[27] L. Qi and Z. Wei, *On the constant positive linear dependence condition and its application to SQP methods*, SIAM J. Optim., 10 (2000), pp. 963–981.
[28] L. Qi and Z. Wei, *Corrigendum: On the constant positive linear dependence condition and its application to SQP methods*, SIAM J. Optim., 11 (2001), pp. 1145–1146.
[29] L. Qi and Y. F. Yang, *A globally and superlinearly convergent SQP algorithm for nonlinear constrained optimization*, J. Global Optim., 21 (2001), pp. 157–184.
[30] S. M. Robinson, *Strongly regular generalized equations*, Math. Oper. Res., 5 (1980), pp. 43–62.
[31] P. Spellucci, *A new technique for inconsistent QP problems in the SQP methods*, Math. Methods Oper. Res., 47 (1998), pp. 355–400.
[32] P. Spellucci, *An SQP method for general nonlinear programs using only equality constrained subproblems*, Math. Programming, 82 (1998), pp. 413–448.
[33] J. Stoer and A. Tapia, *On the characterization of q-superlinear convergence of quasi-Newton methods for constrained optimization*, Math. Comp., 49 (1987), pp. 581–584.
[34] T. Urban, A. L. Tits, and C. T. Lawrence, *A Primal-Dual Interior-Point Method for Nonconvex Optimization with Multiple Logarithmic Barrier Parameters and with Strong Convergence Properties*, Institute for Systems Research Technical report TR 98-27, University of Maryland, College Park, MD, 1998.