Journal of Computational Mathematics, Vol.24, No.5, 2006, 591–608.

A NEW CONSTRAINTS IDENTIFICATION TECHNIQUE-BASED QP-FREE ALGORITHM FOR THE SOLUTION OF INEQUALITY CONSTRAINED MINIMIZATION PROBLEMS *1)

Chang-yin Zhou
(*Department of Mathematics, Shanghai Jiao Tong University, Shanghai 200240, China; College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao 266510, China*)

Guo-ping He Yong-li Wang
(*College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao 266510, China*)

Abstract

In this paper, we propose a feasible QP-free method for solving nonlinear inequality constrained optimization problems. A new working set is proposed to estimate the active set. Specially, to determine the working set, the new method makes use of the multiplier information from the previous iteration, eliminating the need to compute a multiplier function. At each iteration, two or three reduced symmetric systems of linear equations with a common coefficient matrix involving only constraints in the working set are solved, and when the iterate is sufficiently close to a KKT point, only two of them are involved. Moreover, the new algorithm is proved to be globally convergent to a KKT point under mild conditions. Without assuming the strict complementarity, the convergence rate is superlinear under a condition weaker than the strong second-order sufficiency condition. Numerical experiments illustrate the efficiency of the algorithm.

Mathematics subject classification: 90C30, 65K10.
Key words: QP-free method, Optimization, Global convergence, Superlinear convergence, Constraints identification technique.

1. Introduction

Consider the following inequality constrained optimization problem.

$$min \ \ f(x) \qquad subject \ \ to \qquad c(x) \leq 0 \tag{1.1}$$

where $f(\cdot) : R^n \rightarrow R$ and $c(\cdot) : R^n \rightarrow R^m$ are twice continuously differentiable functions. Define $\mathcal{F} := \{x \in R^n \mid c(x) \leq 0\}$ and let $I = \{1, \cdots, m\}$. For any $x \in \mathcal{F}$, the active set is denoted by $I_0(x) = \{i \in I | c_i(x) = 0\}$.

It is well known that sequential quadratic programming(SQP) methods is one of the most efficient methods for solving nonlinear constrained optimization problems. Under certain conditions SQP methods also possess good global and superlinear convergence properties. Therefore, SQP methods have received much attention in recent decades. However, there are still some defaults existed in SQP methods. For example, the QP subproblem may be inconsistent, that is, the feasible set of the QP subproblem may be empty. Moreover, it is usually difficult to use some good spare and symmetric properties of the original problem, which may restrict the application of the SQP algorithm, especially for large-scale problems. We refer to the review paper [1] for an excellent survey about SQP methods.

* Received May 30, 2004; final revised December 9, 2005.
1) This work is supported by the National Natural Science Foundation of China (10571109).

Sequential systems of linear equations(SSLE in short) method, also called QP-free method, is proposed mainly to overcome the shortcomings of SQP method mentioned above. Panier, Tits and Herskovits [2] first proposed a feasible QP-free algorithm for solving problem (1.1). At each iteration, they first solve two systems of linear equations with the following form

$$\begin{pmatrix} H_k & A(x^k) \\ MA(x^k)^T & G(x^k) \end{pmatrix} \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) \\ \triangle_k \end{pmatrix} \tag{1.2}$$

where $A(x^k) = (\nabla c_1(x^k), \cdots, \nabla c_m(x^k))^T$, $G(x^k) = diag((c_i(x^k))$ and $M = diag(\mu_i)$. $diag((c_i(x^k))$ and $diag(\mu_i)$ denote the $m \times m$ diagonal matrix whose i-th diagonal element are $c_i(x^k)$ and $\mu_i(i = 1, \cdots, m)$, respectively. To avoid the Maratos effect, locally a second order correction is computed by solving a least squares subproblem. It is shown that any accumulation point of the iterates generated by the algorithm is a stationary point of problem (1.1). If further the stationary points are assumed to be finite, the point is proved to be a KKT point. It is also shown that the algorithm has a two-step superlinear convergence rate. Later, Z. Gao, G. He and F. Wu[4] improved the algorithm in the sense that any accumulation point is a KKT point without assuming the isolatedness of the accumulation point. To achieve this, they solve an extra linear system obtained from (1.2) by slightly perturbing the right-hand side of (1.2). Recently, their algorithm is further improved in [5] by themselves, which proved the one-step superlinear convergence under the strict complementarity and second order sufficient condition. The same idea is also used to develop a primal-dual logarithmic barrier interior-point method by Urban, Tits and Lawrence [3] and an infeasible SSLE algorithm for solving general constrained optimization problems in [6]. However, the coefficient matrix in linear system (1.2) may become ill-conditioned if the multiplier μ_i corresponding to a nearly active constraint $c_i(x)$ becomes very small. This may easily occur if the strict complementarity doesn't hold at the solution of problem (1.1). To avoid the ill-conditionedness, H. Qi and L. Qi [10] proposed a new QP-free algorithm for solving problem (1.1), that is based on a nonsmooth equation reformulation of the KKT system of problem (1.1) by using the Fisher-Burmeister function. It is shown that under the Linear Independence Constraint Qualification, the coefficient matrix in [10] is uniformly nonsingular and well-conditioned even if the strict complementarity does not hold at the accumulation point. But they still need strict complementarity to prove the superlinear convergence of their algorithm.

On the other hand, Z. Gao, G. He and F. Wu [7] proposed an infeasible SSLE method for general constrained optimization problems. At each iteration, they solve three symmetric systems of linear equations of the following form

$$\begin{pmatrix} H_k & \nabla c_{I_k}(x^k) \\ \nabla c_{I_k}(x^k)^T & 0 \end{pmatrix} \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) \\ \triangle_k \end{pmatrix} \tag{1.3}$$

where $I_k \subseteq I$ is a working set and an estimate of the active set $I_0(x^k)$. The same idea is also used in [8] by Y.F. Yang, D.H. Li and L. Qi to develop a feasible QP-free algorithm for solving problem (1.1), based on an active constraints identification technique proposed by Facchinei, Fischer and Kanzow [9]. Different from algorithms in [2-6,10], algorithms in [7,8] are not affected by the ill-conditionedness of coefficient matrix triggered by dual degeneracy. However, in order to prove superlinear convergence, they still need strict complementarity to hold. Other similar QP-free methods can also be seen in [11-13].

Ever since the first QP-free algorithm was proposed, it has always been an important research area to establish superlinear convergence without strict complementarity. Facchinei and Lucidi[14], Bonnans[15] have proposed several local QP-free algorithms whose rapid convergence does not need the strict complementarity to hold. Facchinei , Lucidi and Palagi[16] proposed a globally convergent truncated Newton method for solving box constrained optimization problem, whose superlinear convergence also does not rely on the strict complementarity. However, globally convergent QP-free algorithm for generally constrained optimization problems without strict complementarity still need to be further studied.

In this paper, we propose a new feasible SSLE algorithm for solving problem (1.1). This study is motivated by successful application of Facchinei-Fisher-Kanzow accurate identification technique for active set (see, [8] [9]). The new algorithm not only possesses almost all the favorable global convergence properties of existing QP-free methods, it also enjoys the local superlinear convergence without strict complementarity. Other interesting features of the algorithm includes:

(a) A different rule from that of [6], [10] and [11] is used to construct the term $\triangle_k$ in linear systems (1.2) and (1.3). Based on the new rule, only two or three symmetric systems of linear equations with a common coefficient matrix need to be solved at each iteration. Specially, when the iterate is sufficiently close to a KKT point, only two of them are involved.

(b) In order to determine the working set I_k, we use the multiplier λ^{k-1} from the last iteration instead of the following multiplier function used in [8,9].

$$\lambda(x) := (\nabla c(x)^T \nabla c(x) + diag(c_i^2(x)))^{-1} \nabla c(x)^T \nabla f(x) \tag{1.4}$$

(c) Another remarkable feature of the new algorithm is that under mild conditions, the working set I_k eventually identifies the strong active $I_0^+(x^*)$ as the iterate is sufficiently close to a KKT point x^*. It is also an accurate identification of $I_0(x^*)$ under additional assumptions. Moreover, we prove that $I_k = I_0(x^*)$ without using the corresponding results given in [9].

The remainder of the paper is organized as follows. In the following section, we give the new algorithm and prove that it is well defined. Section 3 is devoted to show the global convergence of the new algorithm. Its local superlinear convergence is discussed in section 4. We present in section 5 some preliminary numerical experiments. The last section is about some conclusions.

A few words for the notation. Throughout the paper, we use $\| \cdot \|$ to stand for the Euclidean vector norm. For $\lambda \in R^m$ and $J \subseteq I$, $\mid J \mid$ denotes the cardinal number of set J. We denote by $c_J(x)$ the subvector of $c(x)$ with components $c_i(x), i \in J$ and by $\nabla c_J(x)$ the transposed Jacobin matrix of $c_J(x)$. $e_J \in R^{|J|}$ and $0_J \in R^{|J|}$ are the vectors of all ones and zeros, respectively. Let K be an infinite subset of the integer set $\mathcal{N}$, the symbol $K-1$ denote the set $\{k-1 \mid k \in K\}$. We use the notation $u_k \mid_K = 0(or 1)$ if for all $k \in K$, $u_k = 0(or1)$. If the set $I_k = \emptyset$ for all $k \in K$, we denote $I_k \mid_K = \emptyset$ and $I_k \mid_K \neq \emptyset$ if $I_k \neq \emptyset$ for all $k \in K$. If the working set I_k keeps changeless for all $k \in K$, we write $I_k \mid_K = con$, and I_k is rewritten by I_K.

2. Algorithm

Define the Lagrangian function of problem (1.1) by

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

A point $x^* \in \mathcal{F}$ is called a stationary point of problem (1.1), if there exits $\lambda^* \in R^m$ such that

$$\nabla_x L(x^*, \lambda^*) = 0, \ \ and \ \ c_i(x^*)\lambda_i^* = 0, \ \forall i \in I \tag{2.1}$$

If further $\lambda^* \geq 0$, then x^* is called a KKT point. Sometimes, we also call (x^*, λ^*) satisfying (2.1) with $\lambda^* \geq 0$ a KKT point of problem (1.1). For a KKT point (x^*, λ^*), we denote the strong active set by $I_0^+(x^*) := \{i \in I_0(x^*) \mid \lambda_i^* > 0\}$. The working set is defined by

$$I(x, y, \lambda, \varepsilon) := \{i \in I \mid c_i(x) + \varepsilon \min\{\rho(y, \lambda), M\} > 0\} \tag{2.2}$$

where $x, y \in \mathcal{F}$, $\lambda \in R^m$, ε is a nonnegative parameter, M is a positive constant and

$$\rho(y, \lambda) := \sqrt{\| \Phi(y, \lambda) \|} \quad with \ \ \Phi(y, \lambda) := \begin{pmatrix} \nabla_y L(y, \lambda) \\ \min\{-c(y), \lambda\} \end{pmatrix}$$

It is easy to see that $\rho(\cdot)$ is continuous and nonnegative with $\rho(x^*, \lambda^*) = 0$ if and only if (x^*, λ^*) is a KKT point of problem (1.1). For a given x^k and H_k, we denote the matrix $M_J(x^k)$ by

$$M_J(x^k) := \begin{pmatrix} H_k & \nabla c_J(x^k) \\ \nabla c_J(x^k)^T & 0 \end{pmatrix}$$

Now we formally state our algorithm.

Algorithm 2.1.
(S.0)(Initialization)

Parameters: $\sigma \in (0,1), \sigma_1 \in (0,1), \alpha \in (0,1), \beta \in (0,1), \delta \in (0,1), \eta \in (2,3)$, $u \in (0, \frac{1}{2})$.

Data: $\varepsilon_0 > 0, w_0 > 0, M > 0$, $\lambda^0 = 0, u_0 = 0, x^0 = x^1 \in \mathcal{F}$ *and* $c_i(x^0) < 0, \forall i \in I$, $H_0 \in R^{n\times n}$ *is a symmetric positive definite matrix; set* $k := 1$

(S.1) (Compute working set)

(S1.1) Set $t = 0$.

(S1.2) Compute $J_t := I(x^k, x^{k-1}, \lambda^{k-1}, \varepsilon_t)$.

(S1.3) If $J_t = \emptyset$, *then set* $\varepsilon_k = \varepsilon_t$, $w_k := w_t$, $I_k = \emptyset$, $\bar{d}^k = d^k = -H_k^{-1}\nabla f(x^k)$, $\lambda^k = 0$, *goto (S.3).*

(S1.4) If $\det(\nabla c_{J_t}(x^k)^T \nabla c_{J_t}(x^k)) < w_t$, *then set* $t = t+1$, $\varepsilon_t = \sigma\varepsilon_{t-1}, w_t = \sigma_1 w_{t-1}$, *goto (S1.2).*

(S1.5) Set $\varepsilon_k = \varepsilon_t, w_k = w_t$, *and* $I_k = J_t$.

(S.2) (Computation of search direction)

(S2.1) Set $A_k := \| \nabla c_{I_k}(x^k)\hat{\lambda}^k_{I_k} + \nabla f(x^k) \|^3 + \| c_{I_k}(x^k) \|^3$, *where*

$$\hat{\lambda}^k_{I_k} = \left(\lambda^{k-1}_{I_k \cap I^+_{k-1}},\ 0_{I_k \setminus I^+_{k-1}}\right), \quad I^+_{k-1} = \{i \in I_{k-1} \mid \lambda^{k-1}_i > 0\}$$

Compute $(d^{k0}, \lambda^{k0}_{I_k})$ *by solving the following linear system in* (d, λ)

$$M_{I_k}(x^k) \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) \\ -c_{I_k}(x^k) - A_k \cdot e_{I_k} \end{pmatrix} \tag{2.3}$$

Set $\Gamma^-_{k0} := \{i \in I_k \mid \lambda^{k0}_i < 0\}$. *If*

$$\nabla f(x^k)^T d^{k0} \le -\delta (d^{k0})^T H_k d^{k0}, \| c_{I_k}(x^k) \| \le \sqrt{\| d^{k0} \|}, | \lambda^{k0}_i | \le \sqrt{\| d^{k0} \|}, i \in \Gamma^-_{k0} \tag{2.4}$$

then set $\lambda^k := \lambda^{k0} := \left(\lambda^{k0}_{I_k},\ 0_{I\setminus I_k}\right)$, *else set* $u_k = 1$ *and goto (S2.3).*

(S2.2) Compute $(d^{k1}, \lambda^{k1}_{I_k})$ *by solving the following linear system in* (d, λ).

$$M_{I_k}(x^k) \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) \\ \omega^k \end{pmatrix} \tag{2.5}$$

where $\omega^k = (\omega^k_i, i \in I_k)$ *and*

$$\omega^k_i = \nabla c_i(x^k)^T d^{k0} - c_i(x^k + d^{k0}) - \| d^{k0} \|^\eta, \quad i \in I_k$$

Set $d^k = d^{k0}$, $\bar{d}^k = d^{k1}$, $\lambda^{k1} := \left(\lambda^{k1}_{I_k},\ 0_{I\setminus I_k}\right)$; *goto (S.3).*

(S2.3) Compute $(d^{k2}, \lambda^{k2}_{I_k})$ *by solving the following linear system in* (d, λ)

$$M_{I_k}(x^k) \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) \\ 0 \end{pmatrix} \tag{2.6}$$

Set $B_k := \frac{-\alpha}{\left(1 + \sum_{i\in I_k} |\lambda^{k2}_i|\right)} \left(\nabla f(x^k)^T d^{k2} - \sum_{i\in I_k} \lambda^{k2}_i \min\{-c_i(x^k), \lambda^{k2}_i\}\right)$, $\lambda^{k2} := \left(\lambda^{k2}_{I_k},\ 0_{I\setminus I_k}\right)$.

(S2.4) Compute $(d^{k3}, \lambda^{k3}_{I_k})$ *by solving the following linear system in* (d, λ)

$$M_{I_k}(x^k) \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) \\ \min\{-c_{I_k}(x^k), \lambda^{k2}_{I_k}\} - B_k e_{I_k} \end{pmatrix} \tag{2.7}$$

Set $\lambda^k := \lambda^{k3} := \left(\lambda^{k3}_{I_k},\ 0_{I\setminus I_k}\right)$ *and* $\bar{d}^k = d^k = d^{k3}$

(S.3) If $\rho(x^k, \lambda^k) = 0$, *stop; else if* $u_k = 0$ *and* $\| \bar{d}^k - d^k \| > \| d^k \|$, *set* $\bar{d}^k = d^k$.

(S.4)(Line search) Choose t_k, *the first number* t *in the sequence* $\{1, \beta, \beta^2, \cdots\}$ *satisfying*

$$f(x^k + td^k + t^2(\bar{d}^k - d^k)) - f(x^k) \le ut\nabla f(x^k)^T d^k \tag{2.8}$$

$$c_i(x^k + td^k + t^2(\bar{d}^k - d^k)) < 0, i \in I \tag{2.9}$$

(S.5) Set $x^{k+1} := x^k + t_k d^k + t_k^2(\bar{d}^k - d^k)$, $u_{k+1} = 0$. *Generate a new symmetric positive definite matrix* H_{k+1}. *Set* $k := k+1$, *goto (S.1).*

Remarks. The main purpose of (S.1) is to generate a working set and ensure that the matrix $M_{I_k}(x^k)$ is nonsingular for every k. Thus, (d^{kj}, λ^{kj}) is well defined for all $j \in \{0, 1, 2, 3\}$. Besides, at each iteration, two or three reduced linear systems need to be solved, where the main purpose of solving (2.3) and (2.5) is to guarantee the one-step superlinear convergence of Algorithm 2.1 without strict complementarity. Moreover, a judging system of inequalities (2.4) is introduced, through which we can decide which equation need to be solved at each iteration.

The rest of the section is devoted to show that Algorithm 2.1 is well defined. To this end, we first assume that the following hypotheses hold throughout the paper.

Assumption A1. The set $\mathcal{F} \cap \{x | f(x) \leq f(x^0)\}$ is nonempty and compact.

Assumption A2. At every $x \in \mathcal{F}$, $\{\nabla c_i(x), i \in I_0(x)\}$ are linearly independent.

Assumption A3. There exist constants $C_1 > 0$ and $C_2 > 0$ such that for all k

$$C_1 \parallel d \parallel^2 \leq d^T H_k d \leq C_2 \parallel d \parallel^2, \quad \forall d \in R^n$$

Lemma 2.1. *The inner iteration (S1.2)-(S1.4) terminates finitely.*

Proof. Assume to contrary that for all $t \in \mathcal{N}$, $J_t \neq \emptyset$ and

$$\det(\nabla c_{J_t}(x^k)^T \nabla c_{J_t}(x^k)) < \omega_t \tag{2.10}$$

Besides, by Assumption A1 and the finiteness of set I, we can suppose without loss of generality that there exists an infinite set $\mathcal{N}_\infty \subseteq \mathcal{N}$ such that $J_t \mid_{\mathcal{N}_\infty} = con$. Letting $t \in \mathcal{N}_\infty \to \infty$ in (2.10) yields that

$$\det(\nabla c_{J_{\mathcal{N}_\infty}}(x^k)^T \nabla c_{J_{\mathcal{N}_\infty}}(x^k)) = 0 \tag{2.11}$$

On the other hand, since $\varepsilon_t \to 0$, we have that $J_{\mathcal{N}_\infty} \subseteq I_0(x^k)$. It follows from Assumption A2 that

$$\det(\nabla c_{J_{\mathcal{N}_\infty}}(x^k)^T \nabla c_{J_{\mathcal{N}_\infty}}(x^k)) \neq 0.$$

This is a contradiction. The proof is completed.

Similar to the proof of Lemma 2.1, we can easily obtain the following lemma:

Lemma 2.2. *There exists a* $\bar{\varepsilon} > 0$ *such that* $\varepsilon_k > \bar{\varepsilon}$, $\forall k \in \mathcal{N}$.

Assumption A1 and Lemma 2.2 then directly imply the following Lemmas 2.3-2.4.

Lemma 2.3. $\{M_{I_k}(x^k)\}$ *is nonsingular and uniformly bounded with respect to* k, *that is, there exist* $\bar{W} > 0$ *and* $\bar{M} > 0$ *such that, for all* k *and* $x^k \in \mathcal{F}$,

$$\bar{W} \leq \mid det(M_{I_k}(x^k)) \mid \leq \bar{M}$$

Lemma 2.4. *The sequences* $\{(d^{kj}, \lambda^{kj})\}, j = 0, 1, 2, 3$ *are all bounded.*

Lemma 2.5. *If* $u_k = 1$, *then* $\nabla f(x^k)^T d^k \leq -(1-\alpha)(d^{k2})^T H_k d^{k2}$.

Proof. From (2.6) and (2.7), we can obtain that

$$\nabla f(x^k)^T d^{k2} = -(d^{k2})^T H_k d^{k2}, \qquad \nabla f(x^k)^T d^{k2} = -(d^{k2})^T H_k d^{k3}$$
$$\nabla f(x^k)^T d^{k3} = -(d^{k3})^T H_k d^{k2} - (\lambda_{I_k}^{k2})^T \left(\min\{-c_{I_k}(x^k), \lambda_{I_k}^{k2}\} - B_k e_{I_k}\right)$$

Since

$$\sum_{i \in I_k} \lambda_i^{k2} \min\{-c_i(x^k), \lambda_i^{k2}\} \geq 0 \;\; and \;\; \left(1 - \frac{\alpha \sum_{i \in I_k} \lambda_i^{k2}}{(1 + \sum_{i \in I_k} \mid \lambda_i^{k2} \mid)}\right) \geq (1-\alpha) > 0,$$

it follows that

$$\begin{aligned}
&\nabla f(x^k)^T d^k = \nabla f(x^k)^T d^{k3} = \nabla f(x^k)^T d^{k2} - (\lambda_{I_k}^{k2})^T \left(\min\{-c_{I_k}(x^k), \lambda_{I_k}^{k2}\} - B_k e_{I_k}\right)\\
&= \nabla f(x^k)^T d^{k2} - \textstyle\sum_{i\in I_k} \lambda_i^{k2} \min\{-c_i(x^k), \lambda_i^{k2}\} + B_k \sum_{i\in I_k} \lambda_i^{k2}\\
&= \nabla f(x^k)^T d^{k2} - \textstyle\sum_{i\in I_k} \lambda_i^{k2} \min\{-c_i(x^k), \lambda_i^{k2}\}\\
&+ \left(\frac{-\alpha}{\left(1+\sum_{i\in I_k}|\lambda_i^{k2}|\right)} \left(\nabla f(x^k)^T d^{k2} - \textstyle\sum_{i\in I_k} \lambda_i^{k2} \min\{-c_i(x^k), \lambda_i^{k2}\}\right)\right) \textstyle\sum_{i\in I_k} \lambda_i^{k2}\\
&= \left(1 - \frac{\alpha \sum_{i\in I_k} \lambda_i^{k2}}{\left(1+\sum_{i\in I_k}|\lambda_i^{k2}|\right)}\right) \left(\nabla f(x^k)^T d^{k2} - \textstyle\sum_{i\in I_k} \lambda_i^{k2} \min\{-c_i(x^k), \lambda_i^{k2}\}\right)\\
&\le -(1-\alpha)(d^{k2})^T H_k d^{k2} - (1-\alpha) \textstyle\sum_{i\in I_k} \lambda_i^{k2} \min\{-c_i(x^k), \lambda_i^{k2}\} \le -(1-\alpha)(d^{k2})^T H_k d^{k2}
\end{aligned} \tag{2.12}$$

This completes the proof.

From Lemma 2.5, inequalities (2.4) and (S1.3) of Algorithm 2.1, we can find that d^k is actually a descent direction of $f(\cdot)$. Hence, it follows from a similar proof of Proposition 3.3 in [2] that there exists a t_k, the first number of the sequence $\{1, \beta, \beta^2, \cdots\}$, that satisfies the inequalities (2.8) and (2.9). That is, we can always obtain a new iterate x^{k+1} from the current iterate x^k. Therefore, we can claim as follows.

Proposition 2.6. *Algorithm 2.1 is well-defined.*

3. Global Convergence

In this section, we will show that Algorithm 2.1 is globally convergent.

Lemma 3.1. $\rho(x^k, \lambda^k) = 0$ *if and only if* $\nabla f(x^k)^T d^k = 0$.

Proof. From the construction of Algorithm 2.1, the conclusion is obvious if $I_k = \emptyset$. If $I_k \neq \emptyset$, it is not difficult to see that $u_k = 0$ will not occur. Therefore, we need only consider the case: $I_k \neq \emptyset$ and $u_k = 1$.

If $\rho(x^k, \lambda^k) = 0$, it is easy to see from the definition of $\rho(x^k, \lambda^k)$ that $H_k d^{k3} = 0$, and therefore $d^k = d^{k3} = 0$. On the contrary, if $\nabla f(x^k)^T d^k = 0$, we have from (2.12) that

$$0 = \nabla f(x^k)^T d^k \le -(1-\alpha)(d^{k2})^T H_k d^{k2} - (1-\alpha) \sum_{i\in I_k} \lambda_i^{k2} \min\{-c_i(x^k), \lambda_i^{k2}\} \le 0 \tag{3.1}$$

Hence,

$$d^{k2} = 0 \quad and \quad (\lambda_{I_k}^{k2})^T \min\{-c_{I_k}(x^k), \lambda_{I_k}^{k2}\} = 0 \tag{3.2}$$

This and the fact that $c_i(x^k) < 0$, $\forall i \in I$ further implies that $\lambda^{k2} = 0$. Therefore, by (2.6), $\nabla f(x^k) = 0$. Moreover, from (3.2) and the definition of B_k, we have that

$$\min\{-c_{I_k}(x^k), \lambda_{I_k}^{k2}\} - B_k e_{I_k} = 0 \tag{3.3}$$

Since $M_{I_k}(x^k)$ is nonsingular, it follows from (2.7) and (3.3) that $d^{k3} = 0$ and $\lambda^k = 0$. This completes the proof.

From Lemma 3.1 we know that if Algorithm 2.1 finitely terminate at iteration k, i.e. $\rho(x^k, \lambda^k) = 0$, then x^k is an unconstrained stationary point and also a KKT point of problem (1.1). Therefore, in the following discussion we assume, without loss of generality, that Algorithm 2.1 generates an infinite sequence $\{x^k\}$, that is, $\rho(x^k, \lambda^k) \neq 0$ for all $k \in \mathcal{N}$.

Lemma 3.2. *Suppose the following conditions hold.*

(i) $\{(x^k, \lambda^k)\}_K \to (x^*, \lambda^*)$, *(ii)* $I_k \,|_K = \emptyset$

(iii) There exists $K_0 \subseteq K$ *such that* $\{(x^{k-1}, \lambda^{k-1})\}_{K_0} \to (\bar{x}^*, \bar{\lambda}^*)$ *and* $\rho(\bar{x}^*, \bar{\lambda}^*) \neq 0$ *then* $\lambda^* = 0$ *and* $\nabla f(x^*) = 0$.

Proof. It is obvious from $\lambda^k = 0$ for all $k \in K$ that $\lambda^* = 0$. Hence, we need only show that $\nabla f(x^*) = 0$. Assume to the contrary that $\nabla f(x^*) \neq 0$. First, by condition (ii), $d^k = -H_k^{-1}\nabla f(x^k)$. This combined with Assumption A3 implies that there exists a $\gamma_1 > 0$ such that

$$\nabla f(x^k)^T d^k = -\nabla f(x^k)^T H_k^{-1} \nabla f(x^k) < -\gamma_1, \quad \forall k \in K \tag{3.4}$$

Besides, from condition (ii) , (iii) and the definition of I_k, we have that a $\gamma_2 > 0$ exists such that

$$c_i(x^k) \leq -\bar{\varepsilon}\rho(x^{k-1}, \lambda^{k-1}) \leq -\gamma_2, \quad \forall k \in K_0, \quad i \in I \tag{3.5}$$

Let $\gamma = min\{\gamma_1, \gamma_2\}$. Since the sequences $\{x^k + \delta d^k \mid \delta \in [0,1]\}$ and $\{d^k\}$ are bounded and $f(x)$ is twice continuously differentiable, we have from (3.4) that there is a $\xi \in [0,1]$ such that for all $k \in K$

$$\begin{aligned} f(x^k + td^k) - f(x^k) &= t\nabla f(x^k)^T d^k + \frac{1}{2}t^2 (d^k)^T \nabla^2 f(x^k + \xi t d^k) d^k \\ &= t\nabla f(x^k)^T d^k + o(t) \leq ut\nabla f(x^k)^T d^k - (1-u)t\gamma + o(t) \end{aligned} \tag{3.6}$$

Similar to the proof of (3.6), we have from (3.5) that for all $i \in I$ and $k \in K_0$

$$c_i(x^k + td^k) = c_i(x^k) + O(t) \leq -\gamma + O(t) \tag{3.7}$$

Hence, we have from (3.6) and (3.7) that there exists a $\bar{t} > 0$ independent of k such that for any $t \in (0, \bar{t}\,]$, both (2.8) and (2.9) hold. Moreover, (3.4) and (3.6) implies that there exists a k_0 such that for all $k \geq k_0$ and $k \in K_0$, $t_k \in [\beta\bar{t}, \bar{t}]$ and

$$f(x^k + t_k d^k) - f(x^k) \leq -u\bar{t}\beta\gamma \tag{3.8}$$

On the other hand, it is easy to see that, for all k, $\nabla f(x^k)^T d^k \leq 0$. Thus, together with (2.8), we get

$$\sum_{k=k_0}^{\infty} (f(x^{k+1}) - f(x^k)) \leq \sum_{k \geq k_0, k \in K_0}^{\infty} (-u\bar{t}\beta\gamma) \to -\infty$$

This implies that $f(x^k) \to -\infty$, a contradiction with Assumption A1. The proof is completed.

Lemma 3.3. *Suppose the following conditions hold.*
(i) $\{(x^k, \lambda^k)\}_K \to (x^*, \lambda^*)$, *(ii)* $I_k \mid_K = \emptyset$,
(iii) There is a $K_0 \subseteq K$ *such that* $\{(x^{k-1}, \lambda^{k-1})\}_{K_0} \to (\bar{x}^*, \bar{\lambda}^*)$ *and* $\rho(\bar{x}^*, \bar{\lambda}^*) = 0$.
If $I_{k-1} \mid_{K_0-1} = \emptyset$, *then* $\lambda^* = 0$ *and* $\nabla f(x^*) = 0$.

Proof. First, conditions (i)-(iii) imply that $\forall k \in K$, $\lambda^k = 0$ and $\nabla f(\bar{x}^*) = 0$. Hence, by Assumption A3, we have $\| d^{k-1} \| = \| {H_{k-1}}^{-1} \nabla f(x^{k-1}) \| \to 0$ as $k \in K_0 \to \infty$. Therefore, $x^k = x^{k-1} + t_{k-1}d^{k-1} \to \bar{x}^* = x^*$ as $k \in K_0 \to \infty$. This completes the proof.

Lemma 3.4. *Suppose the conditions (i)-(iii) in Lemma 3.3 hold. If* $I_{k-1} \mid_{K_0-1} \neq \emptyset$, *then* $\lambda^* = 0$ *and* $\nabla f(x^*) = 0$.

Proof. Without loss of generality, suppose that $I_{k-1} \mid_{K_0-1} = con$. Then, from condition (iii) in Lemma 3.3 we have

$$\{\rho(x^{k-1}, \lambda^{k-1})\}_{K_0} \to 0 \tag{3.9}$$

If $u_{k-1} \mid_{K_0-1} = 1$, (3.9) implies that $\bar{d}^{k-1} = d^{k-1} \to 0$ as $k \in K_0 \to \infty$. If $u_{k-1} \mid_{K_0-1} = 0$, again by (3.9), $\{d^{k_0-1}\}_{K_0} \to 0$. Hence, by (2.4),

$$\{c_{I_{K_0-1}}(x^{k-1})\}_{K_0} \to c_{I_{K_0-1}}(\bar{x}^*) = 0 \tag{3.10}$$

This implies that $\{\omega^{k-1}\}_{K_0} \to 0$. Meanwhile, (3.10) and Assumption A2 also imply that $M_{I_{K_0-1}}(\bar{x}^*)$ is nonsingular. Thus, by (2.3) and (2.5), $\{d^{k_1-1}\}_{K_0} \to 0$. Letting $k \in K_0 \to \infty$, we get

$$x^k = x^{k-1} + t_{k-1}d^{k-1} + t_{k-1}^2(\bar{d}^{k-1} - d^{k-1}) \to \bar{x}^* \tag{3.11}$$

Denote $I_0^+(\bar{x}^*) = \{i \in I \mid \bar{\lambda}_i^* > 0\}$. Since $(\bar{x}^*, \bar{\lambda}^*)$ is a KKT point of problem (1.1), it follows from (3.9) and the definition of I_{K_0-1} that $I_0^+(\bar{x}^*) \subseteq I_{K_0-1}$. Hence, by (3.11),

$$x^* = \bar{x}^* \ \ \textit{and} \ \ c_i(x^*) = c_i(\bar{x}^*) = 0, i \in I_0^+(\bar{x}^*) \tag{3.12}$$

Moreover, for $\forall i \in I_0^+(\bar{x}^*)$ and sufficiently large $k \in K_0$, we have

$$\begin{aligned}
&c_i(x^k) + \bar{\varepsilon}\rho(x^{k-1}, \lambda^{k-1}) \\
&= c_i(x^k) + \bar{\varepsilon}[\| \nabla_x L(x^{k-1}, \lambda^{k-1}) \|^2 + \sum_{i=1}^m | \min\{-c_i(x^{k-1}), \lambda_i^{k-1}\} |^2]^{\frac{1}{4}} \\
&= c_i(x^k) + \bar{\varepsilon}[\| H_{k-1}d^{k-1} \|^2 + \sum_{i=1}^m | \min\{-c_i(x^{k-1}), \lambda_i^{k-1}\} |^2]^{\frac{1}{4}} \\
&\geq c_i(x^k) + 2^{-\frac{3}{4}}\bar{\varepsilon}[\| H_{k-1}d^{k-1} \|^{\frac{1}{2}} + (\sum_{i=1}^m | \min\{-c_i(x^{k-1}), \lambda_i^{k-1}\} |^2)^{\frac{1}{4}}] \\
&\quad = c_i(x^{k-1}) + O(\| d^{k-1} \|) \\
&\quad\quad + 2^{-\frac{3}{4}}\bar{\varepsilon}[\| H_{k-1}d^{k-1} \|^{\frac{1}{2}} + (\sum_{i=1}^m | \min\{-c_i(x^{k-1}), \lambda_i^{k-1}\} |^2)^{\frac{1}{4}}] > 0
\end{aligned}$$

This implies that

$$I_0^+(\bar{x}^*) \subseteq I_K \subseteq I_0(x^*) \tag{3.13}$$

Since $I_K = \emptyset$ from (ii), it follows that $I_0^+(\bar{x}^*) = \emptyset$. Hence, $\bar{\lambda}^* = \lambda^* = 0$ and $\nabla f(x^*) = 0$. This completes the proof.

Lemma 3.5. *Suppose that $\{(x^k, \lambda^k)\}_K \to (x^*, \lambda^*)$. If $\{\nabla f(x^k)^T d^k\}_K \to 0$, then (x^*, λ^*) is a KKT point of problem (1.1).*

Proof. If $I_k \mid_K = \emptyset$, the result can be directly obtained from $d^k = -H_k^{-1}\nabla f(x^k)$ and $\lambda^k = 0$. Hence, without loss of generality, we suppose that $I_k \mid_K \neq \emptyset$ and denote $I_k \mid_K := con$. If $u_k \mid_K = 0$, by (2.3) and (2.4), the conclusion is obvious. If, instead, $u_k \mid_K = 1$, we have from (3.1) that

$$\nabla f(x^k)^T d^k \leq -(1-\alpha)(d^{k2})^T H_k d^{k2} - (1-\alpha) \sum_{i \in I_K} \lambda_i^{k2} \min\{-c_i(x^k), \lambda_i^{k2}\} \to 0$$

This implies that

$$\{d^{k2}\}_K \to 0 \ \ \textit{and} \ \ \{(\lambda_{I_K}^{k2})^T \min\{-c_{I_K}(x^k), \lambda_{I_K}^{k2}\}\}_K \to 0 \tag{3.14}$$

Let $\eta^* = (\lambda_i^* \mid i \in I_K)$. Since $\{\lambda^{k2}\}$ and $\{d^{k3}\}$ are bounded, we can suppose that $\{\lambda^{k2}\}_{\bar{K}} \to \bar{\lambda}^*$, $\{d^{k3}\}_{\bar{K}} \to \bar{d}$, where $\bar{K} \subseteq K$ is an infinite index set. Therefore, it follows from (2.6) and (3.14) that $(x^*, \bar{\lambda}^*)$ is a KKT point of problem (1.1). Moreover, it is easy to see from (3.14) that

$$\{\min\{-c_{I_K}(x^k), \lambda_{I_K}^{k2}\} - B_k e_{I_K}\}_K \to 0$$

Thus, by (2.6) and (2.7), both $(\bar{d}, \eta^*)$ and $(0, \bar{\lambda}_{I_K}^*)$ are solutions of the following linear system

$$M_{I_K}(x^*) \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f(x^*) \\ 0 \end{pmatrix} \tag{3.15}$$

Since $M_{I_K}(x^*)$ is nonsingular by Lemma 2.3, it follows that $\eta^* = \bar{\lambda}_{I_K}^*$. Hence, $\lambda^* = \bar{\lambda}^*$. This completes the proof.

Lemma 3.6. *Assume that* $\{(x^k, \lambda^k)\}_K \to (x^*, \lambda^*)$ *and* $I_k \mid_K \neq \emptyset$. *If either* $\{B_k\}_K \to 0$, $u_k \mid_K = 1$ *or* $\{A_k\}_K \to 0$, $u_k \mid_K = 0$, *then* (x^*, λ^*) *is a KKT point of problem(1.1).*

Proof. If $u_k \mid_K = 1$, by the definition of B_k and in a way similar to the proof of Lemma 3.5, we can easily obtain the conclusion. If $u_k \mid_K = 0$, without loss of generality, we can assume that $I_k \mid_K = con$. Since $\{\hat{\lambda}^k_{I_K}\}$ is bounded by Lemma 2.4, we can suppose that $\{\hat{\lambda}^k_{I_K}\}_{K_0} \to \hat{\lambda}^*_{I_K}$, where $K_0 \subseteq K$ is an infinite index set. Thus, by $\{A_k\}_K \to 0$ and (2.3), we get

$$\nabla c_{I_K}(x^*)\hat{\lambda}^*_{I_K} + \nabla f(x^*) = 0, \quad c_{I_K}(x^*) = 0 \tag{3.16}$$

Let $\hat{\lambda}^* = \left(\hat{\lambda}^*_{I_K}, \ 0_{I \setminus I_K}\right)$. Then, (3.16) and the definition of $\hat{\lambda}^k_{I_K}$ imply that $(x^*, \hat{\lambda}^*)$ is a KKT point of problem (1.1). Moreover, from (3.16) and linear system (2.3), we have

$$\begin{aligned} H_k d^{k0} + \nabla c_{I_K}(x^k)(\lambda^{k0}_{I_K} - \hat{\lambda}^*_{I_K}) &= -(\nabla f(x^k) - \nabla f(x^*)) - (\nabla c_{I_K}(x^k) - \nabla c_{I_K}(x^*))\hat{\lambda}^*_{I_K} \\ \nabla c_{I_K}(x^k)^T d^{k0} &= -c_{I_K}(x^k) - A_k e_{I_K} \end{aligned} \tag{3.17}$$

Since $c_{I_K}(x^*) = 0$, it follows from Assumption A2 that $M_{I_K}(x^*)$ is nonsingular. Hence,

$$\{d^{k0}\}_{K_0} \to 0 \quad and \quad \{\lambda^{k0}_{I_K}\}_{K_0} \to \hat{\lambda}^*_{I_K}$$

This implies that $\lambda^* = \hat{\lambda}^*$. This proof is completed.

By the definition of B_k, the following Lemma 3.7 is obvious.

Lemma 3.7. $\rho(x^k, \lambda^k) = 0$ *if and only if* $B_k = 0$ *and* $u_k = 1$.

Lemma 3.8. *Assume that the following conditions hold.*

(i) $\{(x^k, \lambda^k)\}_K \to (x^*, \lambda^*)$, *(ii)* $I_k \mid_K \neq \emptyset$,

(iii) There exists $K_0 \subseteq K$ *such that* $\{(x^{k-1}, \lambda^{k-1})\}_{K_0} \to (\bar{x}^*, \bar{\lambda}^*)$ *and* $\rho(\bar{x}^*, \bar{\lambda}^*) \neq 0$, *then* (x^*, λ^*) *is a KKT point of problem (1.1).*

Proof. Assume to the contrary that (x^*, λ^*) is not a KKT pair of problem (1.1). Then, it follows from Lemma 3.5 that there exists a $\gamma_1 > 0$ such that

$$\nabla f(x^k)^T d^k < -\gamma_1, \forall k \in K \tag{3.18}$$

Without loss of generality, suppose that $I_k \mid_K = con$. In a way similar to the proof of Lemma 3.2, we have, for all $k \in K$,

$$f(x^k + td^k + t^2(\bar{d}^k - d^k)) - f(x^k) \le ut\nabla f(x^k)^T d^k - (1-u)t\gamma_1 + o(t)$$

If $u_k \mid_K = 0$, by Lemma 3.6, there is a $\delta_1 > 0$ such that $A_k > \delta_1$, $\forall k \in K$. If $u_k \mid_K = 1$, then Lemmas 3.6 and 3.7 imply that there is a $\delta_2 > 0$ such that $B_k > \delta_2$, $\forall k \in K$. Since $\rho(\bar{x}^*, \bar{\lambda}^*) \neq 0$, there must exist a $\rho_0 > 0$ such that

$$\rho(x^{k-1}, \lambda^{k-1}) > \rho_0, \forall k \in K_0$$

Hence, $I_K \supseteq I_0(x^*)$. Moreover, for every $i \in I_K$ and all $k \in K_0$

$$\begin{aligned} c_i(x^k + td^k + t^2(\bar{d}^k - d^k)) &= c_i(x^k) + t\nabla c_i(x^k)^T d^k + o(t) \\ &= \begin{cases} c_i(x^k) - tc_i(x^k) - tA_k + o(t) \le -t\delta_1 + o(t), & u_k = 0 \\ c_i(x^k) + t \cdot \min\{-c_i(x^k), \lambda^{k0}_i\} - tB_k + o(t) \le -t\delta_2 + o(t), & u_k = 1 \end{cases} \end{aligned}$$

If $i \notin I_K$, then

$$c_i(x^k + td^k + t^2(\bar{d}^k - d^k)) = c_i(x^k) + O(t) \le -\bar{\varepsilon}\rho_0 + O(t)$$

In a way similar to the proof of Lemma 3.2, we get that $\{f(x^k)\}_{K_0} \to -\infty$, which contradicts with the fact that $\{f(x^k)\}$ has a lower bound. This completes the proof.

Lemma 3.9. *Assume that the following conditions hold.*

(i) $\{(x^k,\lambda^k)\}_K \to (x^*,\lambda^*)$, *(ii)* $I_k\mid_K \neq \emptyset$,

(iii) There exists $K_0 \subseteq K$ *such that* $\{(x^{k-1},\lambda^{k-1})\}_{K_0} \to (\bar{x}^*,\bar{\lambda}^*)$ *and* $\rho(\bar{x}^*,\bar{\lambda}^*)=0$

If $I_{k-1}\mid_{K_0-1}=\emptyset$, *then* $\nabla f(x^*)=0$ *and* $\lambda^*=0$.

Proof. Without loss of generality, we suppose that $I_k\mid_K = con$. Since $I_{k-1}\mid_{K_0-1}=\emptyset$, condition (iii) implies that

$$\nabla f(\bar{x}^*)=0 \quad and \quad \{d^{k-1}\}_{K_0}\to 0 \tag{3.19}$$

where $d^{k-1}=-H_{k-1}^{-1}\nabla f(x^{k-1})$. Let $k\in K_0\to\infty$, then $x^k=x^{k-1}+t_{k-1}d^{k-1}\to\bar{x}^*=x^*$. Again, from condition (iii) and the definition of A_k, we get

$$I_K\subseteq I_0(x^*), \quad \forall k\in K_0 \tag{3.20}$$

Hence, by (3.19) and (3.20), $\{c_{I_K}(x^k)+A_k\cdot e_{I_K}\}_{K_0}\to 0$. Meanwhile, Assumption A2 and (3.20) imply that $M_{I_K}(x^*)$ is nonsingular. Therefore, by linear systems (2.3) and (2.6), we obtain that $\{d^{kj}\}_{K_0}\to 0$ and $\{\lambda_{I_K}^{k_j}\}_{K_0}\to 0$, $j=0,2$. Consequently, if $u_k\mid_K=0$, then $\lambda^*=0$ while if $u_k\mid_K=1$, then $\{\min\{-c_{I_K}(x^k),\lambda_{I_K}^{k2}\}-B_k\cdot e_{I_K}\}_{K_0}\to 0$, which also implies that $\lambda^*=0$. This completes the proof.

Lemma 3.10. *Suppose the conditions (i)-(iii) in Lemma 3.9 hold. If* $I_{k-1}\mid_{K_0-1}\neq\emptyset$, *then* (x^*,λ^*) *is a KKT point of problem (1.1).*

Proof. Without loss of generality, we suppose that $I_k\mid_K = con$ and $I_{k-1}\mid_{K_0-1}:= con$, In a way similar to the proof of Lemma 3.4, we get

$$x^*=\bar{x}^* \quad and \quad I_0^+(\bar{x}^*)\subseteq I_K\subseteq I_0(x^*) \tag{3.21}$$

Let $\bar{\eta}^*=(\bar{\eta}_i^*, i\in I_K)$, where

$$\bar{\eta}_i^*=\begin{cases}\bar{\lambda}_i^* & i\in I_0^+(x^*)\\ 0 & i\in I_K\setminus I_0^+(x^*)\end{cases}$$

Since $(x^*,\bar{\lambda}^*)$ is a KKT point of (1.1), we have from (3.21) that (0, $\bar{\eta}^*$) is the solution of the following linear system.

$$M_{I_K}(x^*)\begin{pmatrix} d\\ \lambda\end{pmatrix}=\begin{pmatrix}-\nabla f(x^*)\\ 0\end{pmatrix} \tag{3.22}$$

Therefore, the definition of A_k implies that $\{A_k\}_{K_0}\to 0$. On the other hand, since $M_{I_K}(x^*)$ is nonsingular, it follows from condition (i) in Lemma 3.9 that there exist $\bar{d}$ and $\bar{\lambda}_{I_K}$ such that $\{d^{k0}\}_{K_0}\to\bar{d}$ and $\{\lambda_{I_K}^{k0}\}_{K_0}\to\bar{\lambda}_{I_K}$ with $(\bar{d},\bar{\lambda}_{I_K})$ being the unique solution of (3.22). Hence, $\bar{d}=0$, $\bar{\lambda}_{I_K}=\bar{\eta}^*$. Thus, if $u_k\mid_K=0$, then $\lambda^*=\bar{\lambda}^*$. Analogously, if $u_k\mid_K=1$, we can also get that $\lambda^*=\bar{\lambda}^*$. This completes the proof.

From Lemmas 3.2-3.10, we have

Theorem 3.11. *Algorithm 2.1 either terminates at a KKT point of (1.1) in finite number of steps or generates an infinite sequence* $\{(x^k,\lambda^k)\}$, *any accumulation point of which is a KKT point of (1.1).*

4. Rate of Convergence

In this section, we will establish the superlinear convergence of Algorithm 2.1. Let (x^*,λ^*) be an accumulation point of the sequence $\{(x^k,\lambda^k)\}$ generated by Algorithm 2.1, then from Theorem 3.11, (x^*,λ^*) is a KKT point of (1.1). In the following discussion, we further assume that $\nabla^2 f(x),\nabla^2 c_i(x), i\in I$ are locally Lipschitz continuous on a neighborhood of x^*. Moreover, to get the superlinear convergence, we need the following assumption.

Assumption A4. (x^*, λ^*) is a quasi-regular point of (1.1), i.e., the matrix $M(J)$ is nonsingular for every $J \subseteq I_0(x^*) \setminus I_0^+(x^*)$ (empty set included), where $M(J)$ is defined by

$$M(J) := \begin{bmatrix} \nabla^2_{xx} L(x^*, \lambda^*) & -\nabla c_{I_0^+(x^*)}(x^*) & -\nabla c_J(x^*) \\ \nabla c_{I_0^+(x^*)}(x^*)^T & 0 & 0 \\ \nabla c_J(x^*)^T & 0 & 0 \end{bmatrix}$$

Note: The definition of quasi-regular point is first introduced by F. Facchinei, A. Fischer and C. Kanzow in [9]. Assumption A4 ensures that (x^*, λ^*) is an isolated KKT point(see Theorem 3.15 in [9]). This is essential for our superlinear convergence analysis. We first introduce a useful proposition as follows.

Proposition 4.1 ([10], Proposition 4.1). *Assume that $\omega^* \in R^t$ is an isolated accumulation point of a sequence $\{\omega^k\} \subset R^t$ such that for every subsequence $\{\omega^k\}_K$ converges to ω^*, there is an infinite subset $\bar{K} \subseteq K$ such that $\{\| \omega^{k+1} - \omega^k \|\}_{\bar{K}} \to 0$, then the whole sequence $\{\omega^k\}$ converges to ω^*.*

Lemma 4.2. *If $\{x^k\}_K \to x^*$, then $\{\lambda^k\}_K \to \lambda^*$ and $\{d^k\}_K \to 0$.*

Proof. Assume that λ^* is an arbitrary accumulation of $\{\lambda^k\}_K$. Then from Theorem 3.11, we know that (x^*, λ^*) is a KKT point of problem (1.1). Since Lagrangian multiplier with respect to x^* is unique due to Assumption A2, it follows that $\{\lambda^k\}_K \to \lambda^*$ by the boundedness of $\{\lambda^k\}$. Next we show that $\{d^k\}_K \to 0$ by contradiction. Suppose that there is a $K_0 \subseteq K$ such that $\{d^k\}_{K_0} \to \bar{d} \neq 0$ and $\{H_k\}_{K_0} \to H_*$. By Algorithm 2.1 and the definition of $\rho(\cdot)$, we have $\rho(x^k, \lambda^k) \geq \sqrt{\| H_k d^k \|}$. Therefore, by Assumption A3,

$$\rho(x^*, \lambda^*) \geq \sqrt{\| H_* \bar{d} \|} > 0 \tag{4.1}$$

which contradicts with $\rho(x^*, \lambda^*) = 0$. This completes the proof.

Lemma 4.3. *If $\{(x^k, \lambda^k)\}_K \to (x^*, \lambda^*)$, then $\{(x^{k-1}, \lambda^{k-1})\}_K \to (x^*, \lambda^*)$*

Proof. Suppose that $(\bar{x}^*, \bar{\lambda}^*)$ is an arbitrary accumulation point of $\{(x^{k-1}, \lambda^{k-1})\}_K$. Then, from Theorem 3.11, there is a $K_0 \subseteq K$ such that

$$\{\rho(x^{k-1}, \lambda^{k-1})\}_{K_0} \to \rho(\bar{x}^*, \bar{\lambda}^*) = 0,$$

In a way similar to the proof of Lemmas 3.3, 3.4, 3.9 and 3.10, we get that $x^* = \bar{x}^*, \lambda^* = \bar{\lambda}^*$. Therefore, the accumulation point of $\{(x^{k-1}, \lambda^{k-1})\}_K$ is unique, and the result follows.

Lemma 4.4. *Assume that $\{x^k\}_K \to x^*$, $I_k \mid_K \neq \emptyset$, then $\{d^{kj}\}_K \to 0$, $\{\lambda^{kj}\}_K \to \lambda^*$, $j = 0, 1, 2, 3$.*

Proof. If $u_k \mid_K = 0$, by Lemma 4.2, the result is obvious for $j = 0$. Suppose that $(\bar{d}, \bar{\lambda})$ is an arbitrary accumulation point of $\{(d^{k1}, \lambda^{k1})\}_K$. That is, there exists a $K_0 \subseteq K$ such that $\{(d^{k1}, \lambda^{k1})\}_{K_0} \to (\bar{d}, \bar{\lambda})$. Since the set I is finite, we can suppose without loss of generality that $I_k \mid_{K_0} = con$. Let $\bar{\eta} = (\bar{\lambda}_i \mid i \in I_{K_0})$, $\eta^* = (\lambda^*_i \mid i \in I_{K_0})$. It follows from Lemma 4.3 that $I_k \subseteq I_0(x^*)$ for all $k \in K$ large enough. Hence, $\{\omega^k\}_{K_0} \to 0$ due to $\{d^{k0}\}_K \to 0$. This implies that both $(0, \eta^*)$ and $(\bar{d}, \bar{\eta})$ are solutions of the following linear system.

$$M_{I_{K_0}}(x^*) \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f(x^*) \\ 0 \end{pmatrix} \tag{4.2}$$

Since $M_{I_{K_0}}(x^*)$ is nonsingular, it follows that $\bar{d} = 0$ and $\bar{\lambda} = \lambda^*$. Thus, $\{d^{k1}\}_K \to 0$ and $\{\lambda^{k1}\}_K \to \lambda^*$ by the boundedness of $\{(d^{k1}, \lambda^{k1})\}$. By (4.2) and linear system (2.6), the conclusion is obvious for $j = 2$. Therefore, from the fact that $I_k \subseteq I_0(x^*)$ for all $k \in K$ large enough, we get

$$\{\| \min\{-c_{I_K}(x^k), \lambda^{k2}_{I_K}\} - B_k e_{I_K} \|\}_K \to 0 \tag{4.3}$$

This further implies that the result holds for $j = 3$. In a similar way to the above proof, the conclusion can be proved easily for $u_k\mid_K = 1$.

Lemma 4.5. *Under Assumptions A1-A4, the whole sequence $\{(x^k, \lambda^k)\}$ converges to (x^*, λ^*).*

Proof. Suppose that $\{x^k\}_K \to x^*$. If $I_k\mid_K = \emptyset$, by Lemmas 3.2 and 3.3, we have $\bar{d}^k = d^k \to 0$. If, instead, $I_k\mid_K \neq \emptyset$, then (S.3) in Algorithm 2.1 and Lemma 4.2 imply that $\| \bar{d}^k - d^k \| \leq \| d^k \| \to 0$. Let $k \in K \to \infty$, we have

$$\| x^{k+1} - x^k \| \leq \| d^k \| + \| \bar{d}^k - d^k \| \leq 2 \| d^k \| \to 0$$

Since x^* is an isolated accumulation point of $\{x^k\}$ by Assumption A4, it follows from Proposition 4.1 that the whole sequence $\{x^k\}$ converges to x^*. Moreover, we have shown in Lemma 4.2 that the whole sequence $\{\lambda^k\}$ converges to λ^*. Hence, the result follows directly.

Assumption A5. The sequence of matrices $\{H_k\}$ satisfies

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

Lemma 4.6. *If $I_k = \emptyset$ for all k large enough, then the step $t_k = 1$ in Algorithm 2.1 is accepted for sufficiently large k.*

Proof. Since $I_k = \emptyset$, it follows from Algorithm 2.1 and Lemmas 3.2-3.4 that the whole sequence $\{\nabla f(x^k)\}$ converges to 0 and $\lambda^* = \lambda^k = 0$. Hence,

$$d^k = -H_k^{-1} \nabla f(x^k) \to 0 \quad \textit{and} \quad c_i(x^k) \leq -\bar{\varepsilon} \sqrt{\| \nabla f(x^{k-1}) \|}$$

Moreover, by Assumption A5, we have, for sufficiently large k,

$$\begin{aligned} f(x^k + d^k) - f(x^k) &= \nabla f(x^k)^T d^k + \frac{1}{2}(d^k)^T \nabla^2 f(x^k) d^k + o(\| d^k \|^2) \\ &= -\frac{1}{2}(d^k)^T H_k d^k + \frac{1}{2}(d^k)^T (\nabla^2 f(x^k) - H_k) d^k + o(\| d^k \|^2) \\ &= -\frac{1}{2}(d^k)^T H_k d^k + o(\| d^k \|^2) \leq -\mu (d^k)^T H_k d^k = \mu \nabla f(x^k)^T d^k \end{aligned}$$

and

$$\begin{aligned} c_i(x^k - H_k^{-1} \nabla f(x^k)) &= c_i(x^k - H_k^{-1} \nabla f(x^{k-1} - t_{k-1} H_{k-1}^{-1} \nabla f(x^{k-1}))) \\ &= c_i(x^k - H_k^{-1} \nabla f(x^{k-1}) + O(\| \nabla f(x^{k-1} \|)) \\ &\leq -\bar{\varepsilon} \sqrt{\| \nabla f(x^{k-1}) \|} + O(\| \nabla f(x^{k-1} \|)) < 0 \end{aligned}$$

This completes the proof.

Lemma 4.6 shows that if $I_k = \emptyset$ for sufficiently large k, then Algorithm 2.1 reduces to a quasi-Newton method for unconstrained optimization problems. It is well known that a sequence $\{x^k\}$ generated by quasi-Newton methods converges to an unconstrained stationary point x^* superlinearly if the sequence $\{H_k\}$ satisfies

$$\lim_{k\to\infty} \frac{\| (H_k - \nabla^2 f(x^*)) d^k \|}{\| d^k \|} = 0$$

Lemma 4.7. *If $I_k = \emptyset$ when k is sufficiently large, then $I_0(x^*) = \emptyset$.*

Proof. Since $\{x^k\}$ converges to x^* superlinearly by Lemma 4.6, it follows that

$$\| x^k + d^k - x^* \| = o(\| x^k - x^* \|)$$

Therefore, $\| d^k \| = O(\| x^k - x^* \|)$. Moreover, for every $i \in I_0(x^*)$ and sufficiently large k,

$$\begin{aligned} & c_i(x^{k+1}) + \bar{\varepsilon}\rho(x^k, \lambda^k) = c_i(x^k + d^k) + \bar{\varepsilon} \| H_k d^k \|^{\frac{1}{2}} \\ & = c_i(x^k) - c_i(x^*) + O(\| d^k \|) + \bar{\varepsilon} \| H_k d^k \|^{\frac{1}{2}} \\ & = O(\| x^k - x^* \|) + O(\| d^k \|) + \bar{\varepsilon} \| H_k d^k \|^{\frac{1}{2}} \\ & = O(\| d^k \|) + \bar{\varepsilon} \| H_k d^k \|^{\frac{1}{2}} > 0 \end{aligned} \tag{4.4}$$

Hence, $I_{k+1} = I_0(x^*)$, which implies that $I_k = I_0(x^*)$ for all sufficiently large k. This completes the proof.

From the above discussion, we know that if $I_k = \emptyset$ for all k large enough, then x^* is a strictly feasible KKT point and the superlinear convergence holds. Therefore, next we only need to consider the case that the active set $I_0(x^*)$ is nonempty.

Lemma 4.8. *If there is an infinite subset K such that $I_k \neq \emptyset$, $\forall k \in K$, then for sufficiently large $k \in K$, $u_k = 0$ and*

$$\| x^k + d^{k0} - x^* \| = o(\| x^k - x^* \|), \quad \| \lambda^{k0} - \lambda^* \| = o(\| x^k - x^* \|)$$

Proof. First, it follows from Lemma 4.5 that the whole sequence $\{(x^k, \lambda^k)\}$ converges to a KKT point (x^*, λ^*) and $\rho(x^{k-1}, \lambda^{k-1}) \to 0$. Lemmas 3.9 and 3.10 further imply that

$$I_0^+(x^*) \subseteq I_k \subseteq I_0(x^*) \tag{4.5}$$

Hence, by linear system (2.6), we can obtain that

$$A_k = o(\| x^k - x^* \|^2)$$

Moreover, by linear system (2.3), we have

$$\begin{aligned} & \nabla^2_{xx} L(x^*, \lambda^*)(x^k + d^{k0} - x^*) + \nabla c_{I_k}(x^k)(\lambda^{k0}_{I_k} - \lambda^*_{I_k}) \\ & \qquad = -(H_k - \nabla^2_{xx} L(x^*, \lambda^*)) d^{k0} + o(\| x^k - x^* \|) \end{aligned} \tag{4.6}$$

$$-\nabla c_{I_k}(x^k)^T (x^k + d^{k0} - x^*) = o(\| x^k - x^* \|) \tag{4.7}$$

Let

$$G_k := \begin{pmatrix} \nabla^2_{xx} L(x^*, \lambda^*) & \nabla c_{I_k}(x^k) \\ -\nabla c_{I_k}(x^k)^T & 0 \end{pmatrix}$$

Then, Assumption A4 and (4.5) imply that when k is sufficiently large, G_k is nonsingular. Since I is finite, it follows that there exist $M > 0$ and $\bar{M} > 0$ such that $\bar{M} \leq \| G_k^{-1} \| \leq M$, $\forall I_k \subseteq I_0(x^*)$. So we have from Assumption A5 that

$$\| x^k + d^{k0} - x^* \| = o(\| x^k - x^* \|) + o(\| d^{k0} \|)$$

$$\| \lambda^{k0} - \lambda^* \| = o(\| x^k - x^* \|) + o(\| d^{k0} \|)$$

which implies that $\| x^k + d^{k0} - x^* \| = o(\| x^k - x^* \|)$ and $\| \lambda^{k0} - \lambda^* \| = o(\| x^k - x^* \|)$. Hence,

$$\| d^{k0} \| = O(\| x^k - x^* \|), \quad \| \lambda^{k0} - \lambda^* \| = o(\| d^{k0} \|) \tag{4.8}$$

Furthermore, we have from linear system (2.3) that for sufficiently large k,

$$\nabla f(x^k)^T d^{k0} = -(d^{k0})^T H_k d^{k0} + c_{I_k}(x^k)^T \lambda^{k0} + o(\| x^k - x^* \|^2)$$

$$\leq -(d^{k0})^T H_k d^{k0} + \sum_{\lambda_i^*=0} \lambda_i^{k0} \nabla c_i(x^*)(x^k - x^*) + o(\| x^k - x^* \|^2)$$

$$= -(d^{k0})^T H_k d^{k0} + o(\| d^{k0} \|^2) \leq -\delta (d^{k0})^T H_k d^{k0}$$

It is clear from (4.8) that for sufficiently large k, $| \lambda_i^{k0} | \leq \sqrt{\| d^{k0} \|}, i \in \Gamma_{k0}^-$ and

$$\| c_{I_k}(x^k) \| \leq \| \nabla c_{I_k}(x^*)(x^k - x^*) \| + o(\| x^k - x^* \|) \leq \sqrt{\| d^{k0} \|}$$

Therefore, $u_k = 0$. This completes the proof.

Lemma 4.9. *When k is sufficiently large, $I_k = I_0(x^*)$.*

Proof. Since $I_0(x^*) \neq \emptyset$, it follows from Lemma 4.7 that there exits an infinite subset K such that $I_k \neq \emptyset$, $\forall k \in K$. Lemma 4.8 shows that for sufficiently $k \in K$, $u_k = 0$ and $\| d^{k0} \| = O(\| x^k - x^* \|)$. Hence, in a similar way to the proof of Lemma 4.7, we get that, for sufficiently large $k \in K$, $I_{k+1} = I_0(x^*)$, which implies that $I_k = I_0(x^*)$ for sufficiently large k. This completes the proof.

By Lemmas 4.2-4.5, the following lemma is obvious.

Lemma 4.10. *Under Assumptions A1-A4, when $k \to \infty$, we have*

$$d^k \to 0, \quad \bar{d}^k \to 0, \quad \lambda^k \to \lambda^*, \quad d^{kj} \to 0, \quad \lambda^{kj} \to \lambda^*, \quad j = 0, 1, 2, 3.$$

Lemma 4.11. *When k is sufficiently large,*

$$\| d^{k1} - d^{k0} \| = O(\| d^{k0} \|^2), \quad \| \lambda^{k1} - \lambda^{k0} \| = O(\| d^{k0} \|^2) \tag{4.9}$$

Proof. Since $u_k = 0$ and $\| d^{k0} \| = O(\| x^k - x^* \|)$ from Lemma 4.8, it follows directly from linear systems (2.3) and (2.5) that the result holds.

Lemma 4.12. *When k is sufficiently large, $\| x^k + d^{k1} - x^* \| = o(\| x^k - x^* \|)$.*

Proof. By Lemmas 4.8 and 4.11, we have

$$\| x^k + d^{k1} - x^* \| \leq \| x^k + d^{k0} - x^* \| + \| d^{k1} - d^{k0} \|$$

$$= o(\| x^k - x^* \|) + O(\| d^{k0} \|^2) = o(\| x^k - x^* \|)$$

This completes the proof.

Lemma 4.13. *When k is sufficiently large , the stepsize $t_k = 1$ is accepted.*

Proof. For $i \notin I_0(x^*)$, since $c_i(x^*) < 0$, it is not difficult to see that when k is sufficiently large, $c_i(x^k + d^{k1}) < 0$. For $i \in I_0(x^*)$, by the definition of A_k, there is a $M > 0$ such that $\| A_k \| \leq M \| d^{k0} \|^3$. Hence, we have from linear system (2.5) and Lemma 4.11 that

$$\begin{aligned} c_i(x^k + d^{k1}) &= c_i(x^k) + \nabla c_i(x^k)^T d^{k1} + \tfrac{1}{2}(d^{k1})^T \nabla^2 c_i(x^k) d^{k1} + O(\| d^{k1} \|^3) \\ &= \tfrac{1}{2}(d^{k1} - d^{k0})^T \nabla^2 c_i(x^k) d^{k0} + \tfrac{1}{2}(d^{k1})^T \nabla^2 c_i(x^k)(d^{k1} - d^{k0}) - \| d^{k0} \|^\eta + O(\| d^{k0} \|^3) \\ &= - \| d^{k0} \|^\eta + O(\| d^{k0} \|^3) \end{aligned} \tag{4.10}$$

Since $\eta \in (2, 3)$, it follows that for sufficiently large k, $c_i(x^k + d^{k1}) < 0$. So, when k is sufficiently large , $x^k + d^{k1}$ is a strictly feasible point of problem (1.1). Moreover, by linear systems (2.3), (2.5) and Lemma 4.11, we have that, for sufficiently large k,

$$\begin{aligned} &\nabla f(x^k)^T (d^{k1} - d^{k0}) \\ &= -(d^{k0})^T H_k (d^{k1} - d^{k0}) - \textstyle\sum_{i \in I_0(x^*)} \lambda_i^{k0} \nabla c_i(x^k)^T (d^{k1} - d^{k0}) \\ &= \tfrac{1}{2} \textstyle\sum_{i \in I_0(x^*)} \lambda_i^{k0} (d^{k0})^T \nabla^2 c_i(x^k) d^{k0} + o(\| d^{k0} \|^2) \end{aligned} \tag{4.11}$$

Linear system (2.3), Lemmas 4.8 and 4.11 also imply that

$$\begin{aligned}&\nabla c_{I_0(x^*)}(x^k)(\lambda^k-\lambda^*)+(H_k-\nabla^2_{xx}L(x^*,\lambda^*))d^{k0}\\&=-\nabla^2_{xx}L(x^*,\lambda^*)(x^k+d^{k0}-x^*)+o(\| x^k-x^* \|)=o(\| d^{k0} \|)\end{aligned} \tag{4.12}$$

Since $A_k=o(\| x^k-x^* \|^2)$, it follows from linear system (2.3) that

$$\begin{aligned}&(d^{k0})^T\nabla c_{I_0(x^*)}(x^k)\lambda^*+o(\| d^{k0} \|^2)\\&=(d^{k0})^T\nabla c_{I_0(x^*)}(x^k)\lambda^*+A_k\textstyle\sum_{i\in I_0(x^*)}\lambda_i^*=-c_{I_0(x^*)}(x^k)^T\lambda^*\geq 0\end{aligned} \tag{4.13}$$

Hence, when k is sufficiently large, we get from (4.11)-(4.13) that

$$\begin{aligned}f(x^k+d^{k1})-f(x^k)&=\nabla f(x^k)^Td^{k1}+\frac{1}{2}(d^{k0})^T\nabla^2f(x^k)d^{k0}\\&\quad+\frac{1}{2}[(d^{k1}-d^{k0})^T\nabla^2f(x^k)d^{k0}+(d^{k1}-d^{k0})^T\nabla^2f(x^k)d^{k1}]+o(\| d^{k0} \|^2)\\&=\nabla f(x^k)^Td^{k0}+\nabla f(x^k)^T(d^{k1}-d^{k0})+\frac{1}{2}(d^{k0})^T\nabla^2f(x^k)d^{k0}+o(\| d^{k0} \|^2)\\&=\nabla f(x^k)^Td^{k0}+\frac{1}{2}\sum_{i\in I_0(x^*)}\lambda_i^{k0}(d^{k0})^T\nabla^2c_i(x^k)d^{k0}+\frac{1}{2}(d^{k0})^T\nabla^2f(x^k)d^{k0}+o(\| d^{k0} \|^2)\\&\leq\frac{1}{2}\nabla f(x^k)^Td^{k0}-\frac{1}{2}(d^{k0})^T(\nabla c_{I_0(x^*)}(x^k)(\lambda^{k0}-\lambda^*)+(H_k-\nabla^2_{xx}L(x^*,\lambda^*))d^{k0})+o(\| d^{k0} \|^2)\\&=\frac{1}{2}\nabla f(x^k)^Td^{k0}+o(\| d^{k0} \|^2)\\&=u\nabla f(x^k)^Td^{k0}+(u-\frac{1}{2})[(d^{k0})^TH_kd^{k0}+(d^{k0})^T\nabla c_{I_0(x^*)}(x^k)\lambda^{k0}]+o(\| d^{k0} \|^2)\\&\leq u\nabla f(x^k)^Td^{k0}+(u-\frac{1}{2})[(d^{k0})^TH_kd^{k0}+\sum_{\lambda_i^*=0}-c_i(x^k)\lambda_i^{k0}]+o(\| d^{k0} \|^2)\\&=u\nabla f(x^k)^Td^{k0}+(u-\frac{1}{2})(d^{k0})^TH_kd^{k0}+o(\| d^{k0} \|^2)\leq u\nabla f(x^k)^Td^{k0}\end{aligned}$$

This completes the proof.

The superlinear convergence rate of Algorithm 2.1 is a direct consequence of Lemmas 4.12 and 4.13.

Theorem 4.14. *Under the stated assumptions, Algorithm 2.1 converges superlinearly, i.e.* $\| x^{k+1}-x^* \|=o(\| x^k-x^* \|)$.

5. Numerical Results

In order to evaluate the performance of the proposed algorithm, we made some preliminary implementation using Mathlab 6.5. The test is done on a Pentium III PC with 128 MB of RAM. Experiments were conducted on all test problems from Hock and Schittkowski[17], where a feasible initial point is provided for each problem except for problem 17 and no equality constraints are present. The initial point in Ref.[17] is not feasible for problem 17. Moreover, the solutions of problems 25 are different from that given in Ref.[17]. Therefore, we choose other feasible starting points for problems 17 and 25 as follows:

HS17': $x_0=(0.3,-3)$,

HS25': $x_0 = (3, 10, 1)$,

The initial Lagrangian Hessian estimate is $H_0 = E$ and at each iteration H_k is updated by the following BFGS formula described in [18]:

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

where

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

and

$$\begin{cases} s_k = x^{k+1} - x^k \\ \hat{y}_k = \nabla f(x^{k+1}) - \nabla f(x^k) + (\nabla c_{I_k}(x^{k+1}) - \nabla c_{I_k}(x^k))\lambda_{I_k} \\ \hat{v}_k = 0.8 s_k^T H_k s_k / (s_k^T H_k s_k - s_k^T \hat{y}_k) \end{cases}$$

The parameters used in the implementation of Algorithm 2.1 are as follows:

$$\varepsilon_0 = w_0 = \sigma = \sigma_1 = \beta = 0.5, \alpha = 0.2, \delta = 0.8, \eta = 2.5, u = 0.1, M = 10.$$

Algorithm 2.1 stops at iteration k if any of the following termination criteria is satisfied with $\varepsilon_{stop} = 10^{-7}$:

$$(i) \parallel d^k \parallel / (1 + \parallel x^k \parallel) < \varepsilon_{stop}; \quad (ii) \parallel \Phi(x^k, \lambda^k) \parallel < \varepsilon_{stop}$$

The numerical results are summarized in Table 1. In the table, we denote by HS-No. the label of test problems in Ref.[17], by n the number of variables, by m the number of inequality constraints, by I the number of iterations, by $\triangle f$ the difference between the final value of f and the optimal value $f(x^*)$ reported in [17], namely $\triangle f = f - f(x^*)$, by T the label of termination criteria, by V-stop the final value of the norm function used in the stopping criterion, by W the number of indices in the final working set, by N the number of linear equations solved at the last iteration, and by S the step size of the last iteration.

From the results reported in Table 1, we can see that the new method is able to find a very good approximation of optimal solution within an acceptable number of function evaluations. In general, the number of iterations are very small. For almost all problems except problem HS25 , our algorithm can find the solution. For problem HS25, the starting point already satisfies one of the termination criteria. However, the solution of problems HS25' can be found with the given initial points above. For problem HS33, the final iterates are approximate KKT points of the original problems in the numerical results of Ref.[5] and [10], while our algorithm can find the approximate optimal solution of problem HS33. Besides, problems HS17' and HS30 show that Algorithm 2.1 is especially applied to solving problems that do not satisfy the strict complementarity condition. Moreover, from the last three columns, we can see that, for almost all problems except problem HS30, the final iteration have satisfied the following conclusions which are obtained in our superlinear convergence analysis: (1)the working set $I_k = I(x^*)$, (2) the scalar variable $u_k = 0$, (3)the step size $t_k = 1$. For problem HS30, the final working set I_k is not equal to $I(x^*)$. This is understandable because the linear independence condition does not hold at the optimal point for problem HS30. Finally, we note that the choice of parameters does not influence the behavior of the proposed algorithm significantly .

Table 1: Numerical results

HS-No	n	m	I	$\triangle f$	T	V-stop	W	N	S
HS1	2	1	24	1.9864e-16	(i)	1.1060e-08	0	0	1
HS3	2	1	10	3.2258e-10	(ii)	3.8955e-10	1	2	1
HS4	2	2	4	4.8850e-14	(i)	7.0351e-15	2	2	1
HS5	2	2	9	6.6613e-15	(i)	2.2514e-008	0	0	1
HS12	2	1	8	7.1054e-15	(i)	6.2603e-14	1	2	1
HS17'	2	5	8	5.7082e-11	(i)	4.0362e-11	2	2	1
HS24	2	5	9	1.8993e-10	(i)	6.3767e-11	2	2	1
HS25	3	6	1	3.2835+01	(i)	1.9663e-10	1	0	0
HS25'	3	6	46	1.3916e-18	(i)	4.7007e-09	0	0	1
HS29	3	1	13	1.7145e-11	(i)	7.7715e-08	1	2	1
HS30	3	7	5	1.4513e-12	(i)	2.4949e-08	1	2	1
HS31	3	7	12	3.8192e-14	(i)	3.5221e-08	1	2	1
HS33	3	6	15	4.3965e-11	(i)	9.3125e-12	3	2	1
HS34	3	8	42	4.0350e-09	(i)	7.8347e-10	3	2	1
HS35	3	4	12	1.0270e-14	(i)	7.2593e-09	1	2	1
HS36	3	7	8	2.5784e-10	(i)	6.4332e-14	3	2	1
HS37	3	8	14	1.3642e-12	(i)	2.1037e-10	1	2	1
HS38	4	8	51	3.0230e-17	(i)	1.0585e-09	0	0	1
HS43	4	3	11	1.8545e-12	(i)	5.6873e-08	2	2	1
HS44	4	10	10	2.4432e-06	(i)	5.4293e-08	4	2	1
HS57	2	3	23	2.9910e-12	(i)	7.2473e-10	1	2	1
HS66	3	8	8	8.1551e-11	(i)	9.5866e-08	2	2	1
HS76	4	7	11	-8.1634e-10	(i)	7.5056e-08	2	2	1
HS84	5	16	20	1.5076e-02	(i)	6.8875e-10	5	2	1
HS93	6	8	20	1.8293e-06	(i)	9.5824e-08	2	2	1
HS100	7	4	18	7.4404e-08	(i)	7.1672e-08	2	2	1
HS110	10	20	6	2.5538e-09	(i)	5.7202e-09	0	0	1
HS113	10	8	35	-3.1819e-08	(i)	1.6528e-09	6	2	1
HS117	15	20	70	-3.4273e-08	(i)	1.7556e-08	11	2	1
HS118	15	59	38	1.9054e-10	(i)	8.6767e-13	15	2	1

6. Conclusions

In this paper, we have proposed a feasible QP-free algorithm for nonlinear optimization problems with inequality constraints. A new technique is suggested to determine the working set. An important feature of the new technique is that to update the working set I_k, we directly make use of the multiplier λ^{k-1} generated by Algorithm 2.1 at the k-$1th$ iteration instead of using a multiplier funtiuon. This greatly reduces the computational cost. Another remarkable feature of this technique is that when the sequence $\{x^k\}$ is sufficiently close to a KKT point x^*, the working set I_k identifies the strong active set $I_0^+(x^*)$ of problem (1.1) under mild conditions. Specially, the working set is also an accurate identification of active set $I_0(x^*)$ under additional conditions. We have shown that arbitrary accumulation point of the sequence generated by the new algorithm is a KKT point of the problem (1.1). Without strict complementarity, the convergence rate is proved to be locally superlinear under Assumption A4, an assumption weaker than the strong second order sufficient condition. However, the new QP-free method is only applied to solving inequality constrained problems. How to extend the QP-free method to general constrained optimization problems is an important issue for further research. In addition, it is also an important topic to use QP-free method in the solution of degenerate optimization problems.

References

[1] P.T. Boggs and J.W. Tolle, Sequential quadratic programming, Acta Numerica, Cambridge University, Press, Cambridge, (1995), 1-51.

[2] E.R. Panier, A.L. Tits and J. N. Herskovits, A QP-free globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization, *SIAM Journal on Control and Optimization*, **26** (1988), 788-811.

[3] 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, 1998.

[4] Ziyou Gao, Guoping He and Fang Wu, Sequential systems of linear equations algorithm for nonlinear optimization problems with genernal contraints, *Journal of Optimization Theory and Applications*, **95** (1997), 371-397.

[5] Ziyou Gao, Tiande Guo, Guoping He and Fang Wu, Sequential systems of linear equations algorithm for nonlinear optimization problems-inequality constrained problems, *Journal of Computational Mathematics*, **20**:3 (2002), 301-312.

[6] Ziyou Gao, Guoping He and Fang Wu, Sequential systems of linear equations algorithm for nonlinear optimization problems-general constrained problems, *Applied Mathematics and Computation*, **147** (2004), 211-226.

[7] Ziyou Gao, Guoping He and Fang Wu, Sequential systems of linear equations algorithm with arbitrary initial point, *Science in China (Series A)*, **27** (1997), 24-33.

[8] Yufei Yang, Donghui Li and Liqun Qi, A feasible sequential system of linear equations method for inequality constrained optimization, Applied mathematics report, University of New South Wales, Sydney, 2000.

[9] F. Facchinei, A. Fischer and C. Kanzow, On the accurate identification of active constraints, *SIAM Journal on Optimization*, **9** (1999), 14-32.

[10] Houduo Qi and Liqun Qi, A new QP-free globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization, *SIAM Journal on Optimization*, **11** (2000), 113-132.

[11] Liqun Qi , Yufei Yang, Globally and superlinearly convergent QP-free algorithm for nonlinear constrained optimization, *Journal of Optimization Theory and Applications*, **113**:2 (2002), 297-323.

[12] A.L. Tits, A. Wachter, S. Bakhtiari, T.J. Urban and C.T. Lawrence, A Primal-Dual Interior-Point Method for Nonlinear Programming with Strong Global and Local Convergence Properties, *SIAM Journal on Optimization*, **14**:1 (2003), 173–199.

[13] S. Bakhtiari and A.L. Tits, A Simple Primal-Dual Feasible Interior-Point Method for Nonlinear Programming with Monotone Descent, *Computational Optimization and Applications*, **25** (2003), 17–38.

[14] F. Facchinei, S.Lucidi, Quadratically and superlinearly convergent algorithm for the solution of inequality constrainted minimization problems. *Joural of Optimization Theory and Applications*, **85** (1995), 265-289.

[15] J.F. Bonnans, Local analysis of Newton type methods for variational inequality and nonlinear programming. *Applied mathmetics and Optimization*, **29** (1994), 161-186.

[16] F. Facchinei, S.Lucidi, L.Palagl, A truncated Newton algorithm for large scale box constrained optimization, *SIAM Journal on Optimization*, **12**:4 (2002), 1100-1125.

[17] W.Hock, K.Schittkowski, Test problems for nonlinear programming codes. Lecture Notes in Economics and Mathematical Systems, 187, Springer-Verlag, 1981.

[18] M.J.D.Powell, A fast algorithm for nonlinearly constrained optimization calculations, in Numerical Analysis, Proceedings. Biennial Conference, Dundee 1977, Lecture Notes in Math. 630, G.A. Watson, ed., Springer-Verlag, Berlin, New York, (1978), 144-157.