Journal of Computational Mathematics, Vol.22, No.5, 2004, 651–660.

A QP FREE FEASIBLE METHOD *1)

Ding-guo Pu Yan Zhou Hai-yan Zhang
(*Department of Mathematics, Tongji University, Shanghai 200092, China*)

Abstract

In [12], a QP free feasible method was proposed for the minimization of a smooth function subject to smooth inequality constraints. This method is based on the solutions of linear systems of equations, the reformulation of the KKT optimality conditions by using the Fischer-Burmeister NCP function. This method ensures the feasibility of all iterations. In this paper, we modify the method in [12] slightly to obtain the local convergence under some weaker conditions. In particular, this method is implementable and globally convergent without assuming the linear independence of the gradients of active constrained functions and the uniformly positive definiteness of the submatrix obtained by the Newton or Quasi Newton methods. We also prove that the method has superlinear convergence rate under some mild conditions. Some preliminary numerical results indicate that this new QP free feasible method is quite promising.

Mathematics subject classification: 90C30, 65K10.
Key words: Constrained optimization, KKT point, Multiplier, Nonlinear complementarity, Convergence.

1. Introduction

Consider the constrained nonlinear optimization Problem (NLP):

$$\min f(x), \quad x \in R^n, \quad \text{s.t.G(x)} \le 0,$$

where $f : R^n \to R$ and $G(x) = (g_1(x), g_2(x), \cdots, g_m(x))^T : R^n \to R^m$ are Lipchitz continuously differentiable functions.

We denote by $D = \{x \in R^n | G(x) < 0\}$ and $\bar{D} = cl(D)$ the strictly feasible set and the feasible set of Problem (NLP), respectively.

The Lagrangian function associated with Problem (NLP) is the function

$$L(x, \lambda) = f(x) + \lambda^T G(x), \tag{1}$$

where $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_m)^T \in R^m$ is the multiplier vector. For simplicity, we use (x, λ) to denote the column vector $(x^T, \lambda^T)^T$.

A *Karush-Kuhn-Tucker* (KKT) point $(\bar{x}, \bar{\lambda}) \in R^n \times R^m$ is a point that satisfies the necessary optimality conditions for Problem (NLP):

$$\nabla_x L(\bar{x}, \bar{\lambda}) = 0, \ G(\bar{x}) \le 0, \ \bar{\lambda} \ge 0, \ \bar{\lambda}_i g_i(\bar{x}) = 0, \tag{2}$$

where $1 \le i \le m$. We also say $\bar{x}$ is a KKT point if there exists a $\bar{\lambda}$ such that $(\bar{x}, \bar{\lambda})$ satisfy (2).

Finding KKT points for Problem (NLP) can be equivalently reformulated as solving the mixed nonlinear complementarity problem (NCP) in (2), Problem (NCP) has attracted much

* Received April 4, 2002; final revised March 10, 2004.
1) This research is supported by National Science Foundation of China (No.10371089).

attention due to its various applications, see [4, 1]. One method to solve the nonlinear complementarity problem is to construct a Newton method for solving a system of nonlinear equations (see [11, 5]). Qi and Qi [12] proposed a new QP-free method which ensures the strict feasibility of all iterates. Their work is based on the Fischer-Burmeister NCP function. They proved the global convergence without isolatedness of the accumulation point and the strict complementarity condition. They also proved the superlinear convergence under mild conditions.

However, for the global convergence, [12] still used some stronger conditions. One is the linear independence of the gradients of active constrained functions at the solution; another is the uniformly positive definiteness of H^k which is obtained by the quasi Newton update. To overcome the shortcoming, in this paper, an algorithm is proposed for the minimization of a smooth function subject to smooth inequality constraints. This algorithm is based on the method in [12]. Our main work is to modify this method slightly for obtaining the global convergence under some weaker conditions. Comparing with the method in [12], our method is implementable and globally convergent without assuming the uniformly positive definiteness of H^k and the linear independence of the gradients of active constrained functions at the solution. In particular, for the superlinear convergence of the algorithm we used the same conditions as the method in [12].

In this paper, we use the Fischer-Burmeister function [2] as the following:

$$\psi(a,b)=\sqrt{a^2+b^2}-a-b.$$

Let $\phi_i(x,\lambda)=\psi((-g_i(x)),\lambda_i),\quad 1\le i\le m,\ \Phi_1(x,\lambda)=(\phi_1(x,\lambda)\cdots\phi_m(x,\lambda))^T$. We denote $\Phi(x,\lambda)=((\nabla_x L(x,\lambda))^T,(\Phi_1(x,\lambda))^T)^T$, Clearly, the KKT point conditions of (2) are equivalently reformulated as the condition $\Phi(x,\lambda)=0$.

Let $I_1(x,\lambda)=\{i|(g_i(x),\lambda_i)\neq(0,0)\}$ and $I_0(x,\lambda)=\{i|(g_i(x),\lambda_i)=(0,0)\}$. If $j\in I_1(x,\lambda)$, then we denote,

$$\xi_j(x,\lambda)=\frac{g_j}{\sqrt{(g_j)^2+(\lambda_j)^2}}+1;\qquad \gamma_j(x,\lambda)=\frac{\lambda_j}{\sqrt{(g_j)^2+(\lambda_j)^2}}-1.$$

We have $\nabla_x\phi_j=\xi_j\nabla g_j(x)$ and $\nabla_\lambda\phi_j=\gamma_j e_j$ where $e_j=(0,\cdots,0,1,0\cdots,0)^T\in R^m$ is the jth column of the unit matrix, its jth element is 1, and other elements are 0. If $j\in I_0(x,\lambda)$, then we denote

$$\xi_j(x,\lambda)=1-\sqrt{2}/2;\qquad \gamma_j=\gamma_j(x,\lambda)=-1+\sqrt{2}/2.$$

We have $\xi_j\nabla g_j(x)\in\partial_x\phi_j(x,\lambda)$ and $\gamma_j e_j\in\partial_\lambda\phi_j(x,\lambda)$. Clearly, $\xi_j^2+\gamma_j^2\ge 3-2\sqrt{2}>0$.

The paper is organized as follows. In Section 2, we propose a QP free feasible method. In Section 3, we show that the algorithm is well defined. In Section 4 and Section 5, we discuss the conditions of the global convergence and superlinear convergence of the algorithm, respectively. In Section 6, we give a brief discussion and some numerical tests.

2. Algorithm

In the following algorithm 2.1, let $\xi_j^k=\xi_j(x^k,\mu^k)$ and $\gamma_j^k=\gamma_j(x^k,\mu^k)$, $\eta_j^k=-\sqrt{-2\gamma_j^k}$,

$$V^k=\begin{pmatrix} V_{11}^k & V_{12}^k\\ V_{21}^k & V_{22}^k\end{pmatrix}=\begin{pmatrix} H^k+\bar{c}_1^k I_n & \nabla G^k\\ diag(\xi^k)(\nabla G^k)^T & diag(\eta^k-c^k)\end{pmatrix},$$

where I_n is the n order unit matrix, $\bar{c}_1^k=c_1 min\{1,\|\bar{\Phi}^k\|^\nu\}$, $\bar{\Phi}^k=\Phi(x^k,\bar{\lambda}^k)$, $\bar{\lambda}^k$ is obtained in Algorithm 2.1, $c_1\in(0,1)$, $diag(\xi^k)$ or $diag(\eta^k-c^k)$ denotes the diagonal matrix whose jth diagonal element is $\xi_j(x^k,\mu^k)$ or $\eta_j(x^k,\mu^k)-c_j^k$, respectively, and

$$c_j^k=\begin{cases} c_1\min\{1,\|\bar{\Phi}^k\|^\nu\}, & \eta_j^k=0 \text{ or } -\xi_j^k/\eta_j^k\ge 1, \eta_j^k\neq 0;\\ 0, & \text{otherwise.}\end{cases}$$

Algorithm 2.1.

Step 0. Initialization.

Given an initial guess (x^0,μ^0), $x^0\in D$, $\mu_0\ge 0$, $\bar{\lambda}^0=\mu^0$, $c_1\in(0,1)$, $\tau\in(0,1)$, $\nu>1$, $\bar{\mu}\ge\mu_0>0$ $\kappa\in(0,1)$, and $\theta\in(0,1)$. Given a symmetric positive definite matrix H^0. Denote $\nabla G^k=\nabla G(x^k)$, $V^k=V(x^k,\mu^k)$, $f^k=f(x^k)$ and so on.

Step 1. Compute.

Compute d^{k0} and λ^{k0} by solving the following linear system in (d,λ):

$$V^k\begin{pmatrix} d\\ \lambda\end{pmatrix}=\begin{pmatrix}-\nabla f^k\\ 0\end{pmatrix}. \tag{3}$$

If $d^{k0}=0$, then stop. Otherwise, compute d^{k1} and λ^{k1} by solving the following linear system in (d,λ):

$$V^k\begin{pmatrix} d\\ \lambda\end{pmatrix}=\begin{pmatrix}-\nabla f^k\\ diag(\xi^k)(\lambda_-^{k0})^3\end{pmatrix}. \tag{4}$$

where $\lambda_-^{k0}=\min\{\lambda^{k0},0\}$. Compute d^{k2} and λ^{k2} by solving the following linear system in (d,λ):

$$V^k\begin{pmatrix} d\\ \lambda\end{pmatrix}=\begin{pmatrix}-\nabla f^k\\ diag(\xi^k)(\lambda_-^{k0})^3-\|d^{k1}\|^\nu diag(\xi^k)e\end{pmatrix}, \tag{5}$$

where $e=(1,\cdots,1)^T\in R^m$. Let

$$\begin{pmatrix} d^k\\ \lambda^k\end{pmatrix}=b^k\begin{pmatrix} d^{k1}\\ \lambda^{k1}\end{pmatrix}+\rho^k\begin{pmatrix} d^{k2}\\ \lambda^{k2}\end{pmatrix}, \tag{6}$$

where $b^k=(1-\rho^k)$ and

$$\rho^k=(\theta-1)\frac{(d^{k1})^T\nabla f^k}{1+\left|\sum_{j=1}^m\lambda_j^{k0}\right|\|d^{k1}\|^\nu}. \tag{7}$$

Compute a correction $\hat{d}$, the solution of the least square problem in d.

$$\min d^THd,\quad \text{s.t. } g_i(x^k+d^k)+d^T\nabla g_i^k=-\psi^k \text{ for any } i\in I_k, \tag{8}$$

where $I_k=\{i|g_i^k\ge-\lambda_i^k\}$ and

$$\psi^k=\max\left\{\|d^k\|^\nu,\max_{i\in I_k}\left\{\left|\frac{\xi_i^k}{-\eta_i^k\lambda_i^k}-1\right|^\kappa\right\}\|d^k\|^2\right\}. \tag{9}$$

If (8) has no solution or if $\|\hat{d}^k\|\ge\|d^k\|$, set $\|\hat{d}^k\|=0$.

Step 2. Line search.

Let $t^k=\tau^j$, where j is the smallest non-negative integer such that

$$f(x^k+t^kd^k+(t^k)^2\hat{d}^k)\le f^k+\theta t^k(d^k)^T\nabla f^k \tag{10}$$

and

$$g_i(x^k+t^kd^k+(t^k)^2\hat{d}^k)<0,\ 1\le i\le m. \tag{11}$$

Step 3. Updates.

Set $x^{k+1} = x^k + t^k d^k + (t^k)^2 \hat{d}^k$. Set $\bar{\lambda}^{k+1} = \min\{\lambda^{k0}, \bar{\mu}e\}$, $\bar{\Phi}^{k+1} = \Phi(x^{k+1}, \bar{\lambda}^{k+1})$ and $\mu^{k+1} = \min\{\max\{\lambda^{k0}, \|d^k\|e\}, \bar{\mu}e\}$. If $\bar{\Phi}^{k+1} = 0$ or $\Phi(x^{k+1}, \mu^{k+1}) = 0$ then stop; otherwise update H^k and obtain a symmetric positive definite matrix H^{k+1}. Set $k = k+1$. Go to Step 1.

3. Implement of Algorithm

In this section, we assume that the following assumptions A1-A3 hold.

A1 The strictly feasible set D is nonempty. The level set $S = \{x | f(x) \le f(x_0) \text{ and } x \in \bar{D}\}$ is bounded.

A2 f and g_i are Lipschitz continuously differentiable, and for all y, $z \in R^{n+m}$, $\|L(y) - L(z)\| \le c_2 \|y - z\|$.

A3 H^k is positive definite and there exists a positive number m_1 such that $0 < d^T H^k d \le m_1 \|d\|^2$ for all $d \in R^n$, $d \ne 0$ and all k.

Lemma 3.1. *Assume that Φ^* is an accumulation point of $\{\bar{\Phi}^k\}$, $\bar{\Phi}^{k(i)} \to \Phi^*$, $\xi^{k(i)} \to \xi^*$ and $\eta^{k(i)} \to \eta^*$. If $\Phi^* \ne 0$, then $\eta_j^* - c_j^* < 0$ for all j, $j = 1, 2, \cdots, m$.*

Proof. Without loss of generality, we may assume that $V^{k(i)} \to V^*$ for any j. If $\eta_j^* = 0$, then $\gamma_j^* = 0$. $(\gamma_j^{k(i)})^2 + (\xi_j^{k(i)})^2 \ge 3 - 2\sqrt{2} > 0$ implies $(\gamma_j^*)^2 + (\xi_j^*)^2 \ge 3 - 2\sqrt{2} > 0$ and $\xi_j^* > 0$. It is easy to see either $\eta_j^{k(i)} = 0$ or $-\xi_j^{k(i)} / \eta_j^{k(i)} > 1$ for sufficiently large $k(i)$. From the definition of $c_j^{k(i)}$ we have $c_j^* > 0$ and $\eta_j^* - c_j^* < 0$. This lemma holds.

Lemma 3.2. *If $\bar{\Phi}^k \ne 0$, then V^k is nonsingular. Furthermore, assume that (x^*, μ^*) is an accumulation point of $\{(x^k, \mu^k)\}$, $(x^{k(i)}, \mu^{k(i)}) \to (x^*, \mu^*)$, $\bar{\Phi}^{k(i)} \to \Phi^*$ and $V^{k(i)} \to V^*$. If $\Phi^* \ne 0$, then $\|(V^{k(i)})^{-1}\|$ is bounded and V^* is nonsingular.*

Proof. If $V^k(u, v) = 0$ for some $(u, v) \in R^{n+m}$, where $u = (u_1 \cdots, u_n)^T$, $v = (v_1 \cdots, v_m)^T$. Then we have

$$(H^k + \bar{c}_1^k I_n) u + \nabla G^k v = 0, \tag{12}$$

and

$$diag(\xi^k)(\nabla G^k)^T u + (diag(\eta^k - c^k)) v = 0. \tag{13}$$

Assume $\bar{\Phi}^k \ne 0$. Obviously, $\bar{c}_1^k \ne 0$ and, by the definitions of ξ_j^k and η_j^k, $\xi_j^k \ge 0$ and $\eta_j^k - c_j^k < 0$, $j = 1, 2, \cdots, m$. Thus, $diag(\eta^k - c^k)$ is nonsingular. We have

$$v = -(diag(\eta^k - c^k))^{-1} diag(\xi^k)(\nabla G^k)^T u. \tag{14}$$

Putting (14) into (12), we have

$$u^T (H^k + \bar{c}_1^k I_n) u - u^T \nabla G^k diag(\xi^k)(diag(\eta^k - c^k))^{-1} (\nabla G^k)^T u = 0. \tag{15}$$

$u^T (H^k + \bar{c}_1^k I_n) u = 0$ and $u = 0$ are implied by the fact that $H^k + \bar{c}_1^k I_n$ is positive definite and $-\nabla G^k diag(\xi^k)(diag(\eta^k - c^k))^{-1} (\nabla G^k)^T$ is positive semi-definite, then $v = 0$ by (13). The first part of this lemma holds.

On the other hand, without loss of generality we may assume that $\bar{c}_1^{k(i)} \to \bar{c}_1^* \ne 0$, $diag(\xi^{k(i)}) \to diag(\xi^*)$, $diag(\eta^{k(i)}) \to diag(\eta^*)$, $\eta_j^{k(i)} - c_j^{k(i)} \to \eta_j^* - c_j^*$, $H^{k(i)} \to H^*$ and $(x^{k(i)}, \mu^{k(i)}) \to (x^*, \mu^*)$. By the Lemma 3.1, we know that $\eta_j^* - c_j^* < 0$ for all $j = 1, 2, \cdots, m$. $H^{k(i)} \to H^*$ implies that H^* is positive semi-definite and $H^* + \bar{c}_j^* I_n$ is positive definite. By replacing index k by $*$ in the above proof, It is easy to check that V^* is nonsingular. Assumption $V^{k(i)} \to V^*$ implies that $\|(V^{k(i)})^{-1}\|$ is bounded. This lemma holds.

If $\bar{\Phi}^k = 0$ or $\Phi(x^k, \mu^k) = 0$, then $(x^k, \bar{\lambda}^k)$ or (x^k, μ^k) is a KKT point of Problem (NLP). Without loss of generality, in the sequel, we may assume that $\bar{\Phi}^k \ne 0$ and $\Phi(x^k, \mu^k) \ne 0$ for all k.

Because V^k is nonsingular, (3) or (4) always has a unique solution.

V^k is nonsingular and $A^k = (V^k)^{-1}$ exists. Let

$$A^k = \begin{pmatrix} H^k + \bar{c}_1^k I_n & \nabla G^k \\ diag(\xi^k)(\nabla G^k)^T & diag(\eta^k - c^k) \end{pmatrix}^{-1} = \begin{pmatrix} A_{11}^k & A_{12}^k \\ A_{21}^k & A_{22}^k \end{pmatrix}. \quad (16)$$

Let $Q^k = diag(\eta^k - c^k) - diag(\xi^k)(\nabla G^k)^T(H^k + \bar{c}_1^k I_n)^{-1}(\nabla G^k)$.

Lemma 3.3. *If $\bar{\Phi}^k \neq 0$, then $d^{k0} = 0$ if and only if $\nabla f(x^k) = 0$, and $d^{k0} = 0$ implies $\lambda^{k0} = 0$ and (x^k, λ^{k0}) is a KKT point of Problem (NLP).*

Proof. If $\nabla f(x^k) = 0$, then $d^{k0} = 0$ and $\lambda^{k0} = 0$ by (3). If $d^{k0} = 0$, then (3) implies $\nabla G^k \lambda^{k0} = -\nabla f(x^k)$ and $diag(\eta^k - c^k)\lambda^{k0} = 0$. So, $\lambda^{k0} = 0$ and $\nabla f(x^k) = 0$.

Without loss of generality, we assume that the algorithm never terminates at any k, *i.e.*, $d^{k0} \neq 0$ for all k in the remainder part of this paper.

Lemma 3.4. *If $d^{k0} \neq 0$, then*

1. $\bar{c}_1^k \|d^{k0}\|^2 \leq (d^{k0})^T(H^k + \bar{c}_1^k I_n)d^{k0} \leq -(d^{k0})^T \nabla f^k$.
2. $(d^{k1})^T \nabla f^k = (d^{k0})^T \nabla f^k - \sum_{i:\lambda_i^{k0}<0} (\lambda_i^{k0})^4$.
3. $(d^k)^T \nabla f^k \leq \theta (d^{k1})^T \nabla f^k$.

Proof. (3) implies

$$(H^k + \bar{c}_1^k I_n)d^{k0} + \nabla G^k \lambda^{k0} = -\nabla f^k \quad (17)$$

and

$$diag(\xi^k)(\nabla G^k)^T d^{k0} + (diag(\eta^k - c^k))\lambda^{k0} = 0. \quad (18)$$

We have

$$\lambda^{k0} = -(diag(\eta^k - c^k))^{-1} diag(\xi^k)(\nabla G^k)^T d^{k0}. \quad (19)$$

Putting (19) into (17), we have

$$\begin{aligned} & -(d^{k0})^T \nabla f^k = (d^{k0})^T((H^k d^{k0} + \bar{c}_1^k I_n) + \nabla G^k \lambda^{k0}) \\ = \; & (d^{k0})^T(H^k + \bar{c}_1^k I_n)d^{k0} \\ & -(d^{k0})^T \nabla G^k (diag(\eta^k - c^k))^{-1} diag(\xi^k)(\nabla G^k)^T d^{k0}. \end{aligned} \quad (20)$$

$(d^{k0})^T \nabla G^k (diag(\eta^k - c^k))^{-1} diag(\xi^k)(\nabla G^k)^T d^{k0} \leq 0$ implies

$$\bar{c}_1^k \|d^{k0}\|^2 \leq (d^{k0})^T(H^k d^{k0} + \bar{c}_1^k I_n)d^{k0} \leq -(d^{k0})^T \nabla f^k. \quad (21)$$

The first part of the lemma holds. (3) and (16) imply

$$(d^{k0})^T = -A_{11}^k \nabla f^k, \quad \lambda^{k0} = -A_{21}^k \nabla f^k. \quad (22)$$

The property of the matrix implies

$$(Q^k)^{-1} diag(\xi^k) = ((diag(\xi^k))^{-1} Q^k)^{-1} = diag(\xi^k)((Q^k)^T)^{-1}. \quad (23)$$

(4), (16) and (23) imply $A_{12}^k diag(\xi^k) = (A_{21}^k)^T$ and

$$\begin{aligned} & (d^{k1})^T \nabla f^k = -(A_{11}^k \nabla f^k)^T \nabla f^k - [(A_{12}^k)^T diag(\xi^k)]^T (\nabla f^k)(\lambda_-^{k0})^3 \\ = \; & (d^{k0})^T \nabla f^k - \sum_{i:\lambda_i^{k0}<0} (\lambda_i^{k0})^4. \end{aligned} \quad (24)$$

The second part of this lemma holds. Finally, (5)-(7) and (24) imply

$$(d^{k2} - d^{k1})^T \nabla f^k = \|d^{k1}\|^\nu [A_{12}^k diag(\xi^k) e]^T \nabla f^k = \|d^{k1}\|^\nu \sum_{j=1}^m \lambda^{k0}$$

and

$$(d^k)^T\nabla f^k = (1-\rho^k)(d^{k1})^T\nabla f^k + \rho^k (d^{k2})^T\nabla f^k \le \theta (d^{k1})^T\nabla f^k. \tag{25}$$

This lemma holds.

It is easy to prove that (see [12]):

Lemma 3.5. *If $d^{k0}\neq 0$, then there is a $\bar t$ such that, for all $t\in(0,\bar t)$, (10) and (11) are satisfied.*

Proof. If $d^{k0}\neq 0$, then by the continuous differentiability of f, we have

$$f^k - f(x^k + td^k + t^2\hat d^k) \ge -t(d^k)^T\nabla f^k + O(t^2). \tag{26}$$

(26), $g_i^k<0$ and the continuous differentiability of g_i imply that there is a $\bar t>0$ such that, for any $0<t\le\bar t$, (10) and (11) are satisfied. This lemma holds.

Lemmas 3.1-3.5 show that Algorithm 2.1 can be implemented.

4. Convergence

In this section, we assume that assumptions A1-A3 hold in this section.

Lemma 4.1. *Assume $x^{k(i)}\to x^*$ and $\bar\Phi^{k(i)}>\varepsilon>0$ for some ε, then the sequence of $\{(d^{k(i)0},\lambda^{k(i)0})\}$, $\{(d^{k(i)1},\lambda^{k(i)1})\}$ and $\{(d^{k(i)2},\lambda^{k(i)2})\}$ are all bounded on $k=0,1,\cdots$.*

Proof. If $x^{k(i)}\to x^*$ and $\bar\Phi^{k(i)}>\varepsilon>0$ for some ε, then the matrix sequence $\{(V^{k(i)})^{-1}\}$ is proved to be uniformly bounded in the Lemma 3.2. $\{x^{k(i)}\}$ is bounded due to the assumption A3. The solubility of system (3) implies that $\{(d^{k(i)0},\lambda^{k(i)0})\}$ is bounded, which implies the boundedness of $\{d^{k(i)1}\}$ of the right-hand side of (4). Hence $\{(d^{k(i)1},\lambda^{k(i)1})\}$ is also bounded. Finally, the boundedness of $\{d^{k(i)1},\lambda^{k(i)1})\}$ implies the boundedness of the right-hand side of (5). Hence $\{(d^{k(i)2},\lambda^{k(i)2})\}$ is also bounded.

Lemma 4.2. *Assume $x^{k(i)}\to x^*$ and $\bar\Phi^{k(i)}>\varepsilon>0$ for some ε. There is a $c_3>0$ such that, for all $k=1,2,\cdots$,*

$$\|d^{k(i)} - d^{k(i)1}\| \le c_3\|d^{k(i)0}\|.$$

Proof. It is from the Lemma 3.2 that there exists a $c_3>0$ such that, for all $k=0,1,\cdots$, $c_3\ge 2m\rho^k\|(V^{k(i)})^{-1}\|$. Let

$$\Delta d^{k(i)} = d^{k(i)} - d^{k(i)1} \text{ and } \Delta\lambda^{k(i)} = \lambda^{k(i)} - \lambda^{k(i)1}.$$

Then by (4)-(6), $(\Delta x^{k(i)},\Delta\lambda^{k(i)})$ is the solution of

$$V^{k(i)}\begin{pmatrix}\Delta d^{k(i)}\\ \Delta\lambda^{k(i)}\end{pmatrix} = \begin{pmatrix}0\\ -\rho^k\|d^{k(i)0}\|^\nu diag(\xi^k)e\end{pmatrix}. \tag{27}$$

It is easy to see that

$$\|(\Delta x^{k(i)},\Delta\lambda^{k(i)})\| \le c_3\|d^{k(i)0}\|^\nu$$

The lemma holds.

Lemma 4.3. *Assume $x^{k(i)}\to x^*$ and $\lambda^{k(i)}\to\lambda^*$.*

1. *If $d^{k(i)}\to 0$, then $\lambda_j^*\ge 0$ for any $1\le j\le m$.*
2. *If $d^{k(i)0}\to 0$, then x^* is a KKT point of Problem (NLP).*
3. *If $d^{k(i)}\to 0$ and $\bar\Phi^{k(i)}>\varepsilon>0$ for some ε, then x^* is a KKT point of Problem (NLP).*

Proof. It follows from the Lemma 3.4 that

$$(d^{k(i)})^T\nabla f^{k(i)} \le -\bar c_1^{k(i)}\theta\|d^{k(i)0}\|^2 - \theta\sum_{i:\lambda_i^{k(i)0}<0}(\lambda_i^{k(i)0})^4. \tag{28}$$

Hence $\{d^{k(i)}\} \to 0$ implies that

$$\sum_{i:\lambda_i^{k(i)0}<0} (\lambda_i^{k(i)0})^4 \to 0 \text{ and } \lambda_j^* \geq 0,\ 1 \leq j \leq m. \tag{29}$$

The first part of this lemma holds.

Because $\{\bar{\lambda}^{k(i)}\} \leq \bar{\mu}$ and $\mu^{k(i)}$ are bounded, there is an accumulation point $\bar{\lambda}^*$ of $\{\bar{\lambda}^{k(i)}\}$. Without loss of generality we may assume that $c^{k(i)} \to c^*$, $\mu^{k(i)} \to \mu^*$ and $\bar{\lambda}^{k(i)} \to \bar{\lambda}^*$. (28) implies that, for any accumulation point λ^* of $\{\lambda^{k(i)}\}$, $\lambda_i^* \geq 0$, $1 \leq i \leq m$. Taking the limitations in both sides of (3), by noting $d^{k(i)0} \to 0$,we obtain $(\lambda^*)^T \nabla G^* = -\nabla f^*$ and $diag(\eta^* - c^*)\lambda^* = 0$. If $-g_i^* > 0$, for some $1 \leq i \leq n$, then $-\eta_i^* + c_i^* \geq \delta > 0$ *and* $\lambda_i^* = 0$, that is, for any $1 \leq i \leq m$, $g_i^* \lambda_i^* = 0$. The second part of this lemma holds. If $d^{k(i)} \to 0$ and $\bar{\Phi}^{k(i)} > \varepsilon > 0$ for some ε, then (28) implies $d^{k(i)0} \to 0$. So, x^* is a KKT point of Problem (NLP). This lemma holds.

By using the Lemma 4.3, the proof of the following lemma is the same as the proof of the Lemma 3.8 in [5].

Lemma 4.4. *Assume $x^{k(i)} \to x^*$ and $\bar{\Phi}^{k(i)} > \varepsilon > 0$ for some ε. If $d^{k(i)-1} \to 0$, then (x^*, λ^*) is a KKT point of Problem (NLP), where λ^* is an accumulation point of $\{\lambda^{k(i)}\}$.*

The following result is the same as the Lemma 3.6 of [12].

Lemma 4.5. *Assume $x^{k(i)} \to x^*$ and $\bar{\Phi}^{k(i)} > \varepsilon > 0$ for some ε. If*

$$\liminf\{\|d^{k(i)-1\|}\|\} > 0,$$

then (x^*, λ^*) is a KKT point of Problem (NLP), where λ^* is an accumulation point of $\{\lambda^{k(i)}\}$.

The following global convergence theorem holds.

Theorem 4.1. *If x^* is a limit point of $\{x^k\}$, then x^* is a KKT point of Problem (NLP).*

Proof. Assume $(x^{k(i)}, \bar{\lambda}^{k(i)}) \to (x^*, \bar{\lambda}^*)$. If $\bar{\Phi}^* = 0$, then $(x^*, \bar{\lambda}^*)$ is a KKT point of Problem (NLP).

If $\bar{\Phi}^* \neq 0$, then $\lambda^{k(i)}$ is bounded by the Lemma 4.1. It follows from the Lemmas 4.2-4.5 that (x^*, λ^*) is a KKT point, where λ^* is an accumulation point of $\{\lambda^{k(i)}\}$. The proof of this theorem completes.

5. Superlinear Convergence

Let $I_2(x, \lambda) = \{i | g_i(x) = 0, \lambda_i > 0\}$ and $X(x, \lambda) = \{d | d^T \nabla g_i(x) \geq 0, i \in I_2(x, \lambda); \}$.

We need the following conception and assumptions for the superlinear convergence.

Definition 5.1. *A point (x, λ) is said to satisfy the strong second-order sufficiency condition for Problem(NLP) if it satisfies the first-order KKT conditions (2) and if $d^T V d > 0$ for all $d \in X(x, \lambda)$, $d \neq 0$ and any $V \in \partial_B \Phi(x, \lambda)$.*

Notice that the strong second-sufficiency condition implies that x is a strict local minimum of Problem (NLP) (see [10]).

A4 $\{\nabla g_i(x^*)\}$ are linear independent, where $i \in I(x^*) = \{i | g_i(x^*) = 0\}$, x^* is a accumulation point of $\{x^k\}$ and a KKT point of Problem (NLP).

A5 H^k is uniformly positive definite and there exist two positive numbers m_1 and m_2 such that $0 < m_2 \|d\|^2 \leq d^T H^k d \leq m_1 \|d\|^2$ for all $d \in R^n$ and all k.

A6 The strong second-order sufficiency condition for Problem(NLP) holds at each KKT point (x^*, λ^*).

A7 The strict complementarity condition holds at each KKT point (x^*, μ^*).

A8 The sequence of $\{H^k\}$ satisfies

$$\frac{\|P^k (H^k - \nabla_x^2 L(x^k, \mu^k)) d^{k1}\|}{\|d^{k1}\|} \to 0, \tag{30}$$

where $P^k = I - N^k((N^k)^T N^k)^{-1} N^k)^T$ and $N^k = (\nabla g_i^k),\ i \in I^k = \{i | g_i^k = 0\}$.

Assumption A7 implies that Φ is continuously differentiable at each KKT point (x^*, μ^*).

Same as the proof of the Lemma 3.2, we get

Lemma 5.1. *Assume A4 and A5 hold. Then* $\{\|(V^k)^{-1}\|\}$ *and* $\{\|(\hat{V}^k)^{-1}\|\}$ *are bounded.*

Furthermore if V^* is an accumulation matrix of $\{V^k\}$, then V^* is nonsingular.

The Lemma 5.1 and the proof of the Lemma 4.1 imply

Lemma 5.2. *Assume A4 and A5 hold. The sequence of* $\{(d^{k(i)0}, \lambda^{k(i)0})\}$, $\{(d^{k(i)1}, \lambda^{k(i)1})\}$ *and* $\{(d^{k(i)2}, \lambda^{k(i)2})\}$ *are all bounded on* $k = 0, 1, \cdots$.

A9 $\lambda^{k0} < \bar{\mu}$ and $\bar{\lambda}^k = \lambda^{k0}$ for sufficiently large k.

It is easy to check that $\lim_{k\to\infty} \Phi(x^k, \bar{\lambda}^k) = \lim_{k\to\infty} \Phi(x^k, \lambda^{k0}) = 0$. (30) is equivalent to the following

$$\frac{\|(P^k(H^k + \bar{c}_1^k) - \nabla_x^2 L(x^k, \mu^k))d^{k1}\|}{\|d^{k1}\|} \to 0. \tag{31}$$

We have $\|c_j^k\|/\|\bar{\Phi}^k\| \to 0$ and $\bar{c}_1^k/\|\bar{\Phi}^k\| \to 0$ as $k \to \infty$, respectively.

The proof of following Theorem 5.1 is the same as Theorem 3.7, Lemma 3.2 and Corollary 3.3 in [12].

Theorem 5.1. *Assume A1-A3 and A6 hold. If* (x^*, λ^*) *is a accumulation point of* $\{(x^k, \lambda^{k0})\}$, *then*

1. (x^*, λ^*) *is a KKT point of Problem (NLP),*

2.$(x^k, \lambda^{k0}) \to (x^*, \lambda^*)$,

3. $d^{k0} \to 0$, $d^{k1} \to 0$ *and* $d^{k2} \to 0$.

Assume A1 and A3-A9 hold. The proofs of following the Lemma 5.2 and Theorem 5.1 are the same as the Lemma 4.6 and Theorem 4.9 in [12].

Lemma 5.3. *For k large enough the step* $t_k = 1$ *is accepted by the line search.*

Theorem 5.2. *Let Algorithm 2.1 be implemented, to generate a sequence* $\{(x^k, \lambda^k)\}$ *and* (x^*, λ^*) *be an accumulation point of* $\{(x^k, \lambda^k)\}$, *then* (x^*, λ^*) *is an KKT point of Problem (NLP), and* (x^k, λ^k) *converges to* (x^*, λ^*) *superlinearly.*

6. Discussion and Numerical Tests

In Algorithm 2.1 the matrix H^k is replaced by $H^k + \bar{c}_1^k I_n$. This idea is stimulated by [6, 9, 7, 8]. [6, 8] proposed a class of revised Broyden Algorithm in which the search direction $-H^k \nabla f(x^k)$ is replaced by $-(H^k \nabla f(x^k) + c_1^k \nabla f(x^k))$, where the matrix H^k is obtained by Broyden updates, c_1^k is a positive number and f is the objective function. [9, 7] proposed a class of inexact Newton methods in which the search direction $-(\nabla^2 f(x^k))^{-1} \nabla f(x^k)$ is replaced by $-((\nabla^2 f(x^k))^{-1} \nabla f(x^k) + c_1^k \nabla f(x^k))$, where c_1^k is a positive number and f is the objective function. In both above two classes of methods, the algorithms are globally convergent without convexity assumption and superlinearly convergent under some mild conditions.

In Algorithm 2.1, the H^k is updated by the BFGS method. In particular, we set

$$H^{k+1} = H^k - \frac{H^k s^k (s^k)^T H^k}{(s^k)^T H^k s^k} + \frac{y^k (y^k)^T}{(s^k)^T y^k}, \tag{32}$$

where

$$y^k = \begin{cases} \hat{y}^k, & (s^k)^T \hat{y}^k \geq 0.2 (s^k)^T H^k s^k, \\ \theta^k \hat{y}^k + (1 - \theta^k) H^k s^k, & \text{otherwise}, \end{cases} \tag{33}$$

and

$$\begin{cases} s^k = x^{k+1} - x^k, \\ \hat{y}^k = \nabla f(x^k) - \nabla f(x^{k+1}) + (\nabla G(x^{k+1}) - \nabla G(x^k))\lambda^{k0}, \\ \theta^k = 0.8 (s^k)^T H^k s^k / ((s^k)^T H^k s^k - (s^k)^T \hat{y}^k). \end{cases} \tag{34}$$

In table 1 which presents some results of the numerical experiments, we use the following notation:

F1 method: the QP-free feasible method in [12]; F2 method: the method QP-free feasible in this paper. Problem: number of problem in [3]; NIT=the number of iterations; NF=the number of function $f(x)$ evaluations, NG=the number of functions $G(x)$ evaluations.

	F1 method			F2 method		
Problem	NIT	NF	NG	NIT	NF	NG
1	40	66	66	17	31	49
3	12	17	23	11	17	19
4	4	9	11	6	11	13
5	6	11	11	5	10	13
12	7	15	17	5	10	18
24	11	19	24	12	16	18
29	8	15	18	9	12	13
30	7	10	14	10	13	14
31	10	37	41	9	21	23
33	10	28	34	11	15	19
34	23	68	78	18	39	44
35	7	12	15	8	11	13
36	13	72	74	14	35	49
37	17	79	85	16	41	47
43	12	25	30	11	25	29
44	17	39	42	14	21	29
76	10	39	42	11	29	35
100	15	39	45	13	27	37
113	22	50	58	16	24	31

References

[1] M.C. Ferris and J.S. Pang, Engineering and economic applications of complementarity problems, *SIAM Review*, **39** (1997), 669-713.

[2] A. Fischer, A special Newton-type optimization method, *Optimization,* **24** (1992), 269-284.

[3] W. Hock and K. Schittkowski, Test Example for Nonlinear Programming Codes, Lecture Notes in Econom. and Math. Systems 187, Springer-Verlag, Berlin, 1981.

[4] P.T. Harker and J.S. Pang, Finite-dimensional variational and nonlinear complementarity problems: a survey of theory, algorithm and applications, *Mathematical Programming*, **48** (1990), 161-220.

[5] E.R. Panier A.L. Tits and J.N. Herskovits, A QP-free, globally, locally superlinear convergent method for the inequality constrained optimization problems, *SIAM J. Control Optim.*, **36** (1988), 788-811.

[6] D. Pu and W. Tian, A class of Broyden Algorithms with revised search direction, *Asia-Pacific J. of Operation Research*, **14** (1997), 93-109.

[7] D. Pu and W. Tian, Gallobally inexact generalized Newton methods for nonsmooth equation, *J. of Computational and Applied Mathematics*, 2001.

[8] D. Pu, S. Gui and W. Tian, A class of revised Broyden algorithms without exactline search, *J, computational Mathematics*, **22** (2004), 11-20.

[9] D. Pu and J. Zhang, An inexact generalized Newton method for second order C-Differentiable optimization, *J. of Computational and Applied Mathematics,* **93** (1998), 107-122.

[10] L. Qi, Convergence analysis of some method for solving nonsmooth equations, *Mathematics of Operations Research,* **18** (1993) 227-243.

[11] D. Sun, A regularization Newton method for solving nonlinear complementarity problems, *Applied Mathematics and Optimization*, **40** (1999), 315-339.

[12] H. Qi and L. Qi, A New QP-free, globally V, locally superlinear convergent feasible method for the solution of inequality constrained optimization problems, *SIAM J. Optim.*, **11** (2000), 113-132.