Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 4, No. 1, pp. 43 – 53, January 2008

FILTER QP-FREE METHOD WITH PIECEWISE LINEAR NCP FUNCTION *

Kangdi Li, Dingguo Pu and Weiwen Tian

Dedicated to Professor Liansheng Zhang on the occasion of his 70th birthday.

Abstract: In this paper, we introduce a piecewise linear NCP function and propose a filter QP-free infeasible method using this NCP function for constrained nonlinear optimization problems. This iterative method is based on the solution of nonsmooth equations which are obtained by the multipliers and the NCP function for the KKT first-order optimality conditions. Locally, each iteration of this method can be viewed as a perturbation of a mixed Newton and quasi-Newton iteration on both the primal and dual variables for the solution of the KKT optimality conditions. We also use the filter on line searches. This method is implementable and globally convergent. We also prove that the method has superlinear convergence rate under some mild conditions.

Key words: *filter, QP-free method, constraint, convergence, NCP function*

Mathematics Subject Classification: *90C30*

---

1 Introduction

We shall study the constrained nonlinear optimization problem (NLP):

$$\begin{aligned} \min \quad & f(x), \\ \text{s.t.} \quad & x \in D = \{x \in R^n \mid G(x) \le 0\}, \end{aligned} \tag{1.1}$$

where $G(x) = (g_1(x), g_2(x), \cdots, g_m(x))^T$.

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

$$\nabla_x L(\bar{x}, \bar{\mu}) = 0, \ G(\bar{x}) \le 0, \ \bar{\mu} \ge 0, \ \bar{\mu}_i g_i(\bar{x}) = 0, \quad 1 \le i \le m, \tag{1.2}$$

where $L(x, \mu) = f(x) + \mu^T G(x)$ is the Lagrangian function, $\mu = (\mu_1, \mu_2, \cdots, \mu_m)^T$ is the multiplier vector. For simplicity, we use (x, μ) to denote the column vector $(x^T, \mu^T)^T$.

*This research was supported by National Science Foundation of China (No. 10571137, 10771162), and Natural Science Foundation of Shanghai Educational Committee (No. 05LZ05).

Problem (1.2) is a mixed nonlinear complementarity problem (NCP). NCP has attracted much attention due to its various applications. One method to solve the nonlinear complementarity problem (1.2) is to construct a Newton method for solving a system of nonlinear equations:

$$\Phi(x, \mu) = 0,$$

which is a reformulation of (1.2).

Recently Pu, Li ad Xue [5] proposed a new QP-free infeasible method for minimizing a smooth function subject to smooth inequality constraints. This iterative method is based on the solution of nonsmooth equations which are obtained by the multipliers and the Fischer-Burmeister NCP function for the KKT first-order optimality conditions. They proved that the method has superlinear convergence rate under some mild conditions. For other QP-free methods, see [6, 7, 8].

Consider the constraint violation function defined by

$$p(G(x)) = \sum_{j=1}^{m} \max\{0, g_j(x)\}.$$

A nonlinear programming algorithm must deal with two conflicting criteria, f and p, which must be simultaneously minimized, with preference given to the infeasibility measure p, which must be driven to zero. Fletcher and Leyffer have proposed to solve problem (NLP) using filter method as an alternative to traditional merit functions approach. The underlying concept is fairly simple. Trial points generated from solving a sequence of trust region quadratic programming (QP) subproblems are accepted if there is a sufficient decrease in the objective function or the constraint violation function. In addition the computational results reported in Fletcher and Leyffer are also very encouraging (see [2, 3, 10]).

Definition 1.1. A pair $(p(G(x^k)), f(x^k))$, is said to dominate another pair $(p(G(x^l)), f(x^l))$ if and only if $p(G(x^k)) \leq p(G(x^l))$ and $f(x^k) \leq f(x^l)$.

Definition 1.2. A filter F is a list of pair $(p(G(x^k)),\ f(x^k))$ such that no pair dominates any other. A pair $(p(G(x^k)),\ f(x^k))$ is said to be accepted for inclusion in the filter if it is not dominated by another pair in the filter.

In this paper, we propose a filter QP-free infeasible method using a piecewise linear NCP function for constrained nonlinear optimization problems. This iterative method is based on the solution of nonsmooth equations which are obtained by the multipliers and the NCP function for the KKT first-order optimality conditions. Locally, each iteration of this method can be viewed as a perturbation of a mixed Newton and quasi-Newton iteration on both the primal and dual variables for the solution of the KKT optimality conditions. The filter is also used in the line searches of the algorithm. This method is implementable and globally convergent. We show that the method has superlinear convergence rate under some mild conditions. Some preliminary numerical results indicate that this new QP-free infeasible method is quite promising.

The paper is outlined as follows. In the next section, we give some preliminary results. In Section 3, we give the algorithm. In Section 4, we show that the algorithm is implementable. In Section 5, we discuss the convergence of the algorithm. Some numerical results are given in Section 6.

2 Preliminaries

Definition 2.1 (NCP pair and NCP function). We call a pair $(a, b) \in R^2$ an NCP pair if $a \geq 0, b \geq 0$ and $ab = 0$; a function $\psi : R^2 \to R$ is called an NCP function if $\psi(a, b) = 0$ if and only if (a, b) is an NCP pair.

Two most famous NCP functions are the min function and the Fischer-Burmeister NCP function. In this paper we define a 4-l piecewise linear NCP function ψ with a parameter $k > 0$ as follows.

$$\psi(a,b) = \begin{cases} k^2 a & \text{if } b \geq k|a|, \\ 2kb - b^2/a & \text{if } a > |b|/k, \\ 2k^2 a + 2kb + b^2/a & \text{if } a < -|b|/k, \\ k^2 a + 4kb & \text{if } b \leq -k|a| < 0. \end{cases} \tag{2.1}$$

We know that ψ is continuously differentiable everywhere except at the origin, but it is strongly semismooth at the origin. *i.e.*, if $a \neq 0$ or $b \neq 0$, then ψ is continuously differentiable at $(a, b) \in R^2$, and

$$\nabla\psi(a,b) = \begin{cases} \begin{pmatrix} k^2 \\ 0 \end{pmatrix} & \text{if } b \geq k|a|, \\ \begin{pmatrix} b^2/a^2 \\ 2k - 2b/a \end{pmatrix} & \text{if } a > |b|/k, \\ \begin{pmatrix} 2k^2 - b^2/a^2 \\ 2k + 2b/a \end{pmatrix} & \text{if } a < -|b|/k, \\ \begin{pmatrix} k^2 \\ 4k \end{pmatrix} & \text{if } \ b \leq -k|a| < 0, \end{cases} \tag{2.2}$$

and

$$A_\psi = \partial\psi(0,0) = \left\{ \begin{pmatrix} k^2 t^2 \\ 2k(1-t) \end{pmatrix} \cup \begin{pmatrix} 2k^2(1-t^2) \\ 2k(1-t) \end{pmatrix} \mid \ |t| \leq 1 \right\}. \tag{2.3}$$

Let

$$\phi_i(x,\mu) = \psi(-g_i(x), \mu_i), \quad 1 \leq i \leq m.$$

We denote $\Phi(x, \mu) = ((\nabla_x L(x, \mu))^T, (\Phi_1(x, \mu))^T)^T$, where $\Phi_1(x, \mu) = (\phi_1(x, \mu), \cdots \phi_m(x, \mu))^T$. Clearly, the KKT optimality conditions (1.2) can be equivalently reformulated as the nonsmooth equations $\Phi(x, \mu) = 0$.

If $(g_i(x), \mu_i) \neq (0, 0)$, then ϕ_i is continuously differentiable at $(x, \mu) \in R^{n+m}$. In this case, we have

$$\nabla\phi_i(x,\mu) = \begin{cases} \begin{pmatrix} -k^2\nabla g_i(x) \\ 0 \end{pmatrix} & \text{if } \ \mu_i \geq k|g_i(x)|, \\ \begin{pmatrix} -\mu_i^2 \nabla g_i(x)/g_i(x)^2 \\ (2k - 2\mu_i/g_i(x))e_i \end{pmatrix} & \text{if } \ -g_i(x) > |\mu_i|/k, \\ \begin{pmatrix} (-2k + \mu_i^2/g_i(x)^2)\nabla g_i(x) \\ (2k - 2\mu_i/g_i(x))e_i \end{pmatrix} & \text{if } -g_i(x) < -|\mu_i|/k, \\ \begin{pmatrix} -k^2\nabla g_i(x) \\ 4ke_i \end{pmatrix} & \text{if } \ \mu_i \leq -k|g_i(x)| < 0. \end{cases} \tag{2.4}$$

If $g_i(x) = 0$ and $\mu_i = 0$, $1 \le i \le m$, then $\phi_i(x, \mu)$ is strongly semismooth and directionally differentiable at (x, μ). We have

$$\partial\phi_i(x,\mu) = \left\{ \begin{pmatrix} -k^2t^2\nabla g_i(x) \\ 2k(1-t)e_i \end{pmatrix} \cup \begin{pmatrix} -2k^2(1-t^2)\nabla g_i(x) \\ (2k-2t)e_i \end{pmatrix} : \ |t| \le 1 \right\}, \qquad (2.5)$$

where $e_i = (0, \cdots, 0, 1, 0 \cdots, 0)^T \in R^m$ is the ith column of the $n \times n$ unit matrix. In the sequel, we set $k = 1$.

Another piecewise linear NCP function was proposed in [6]. For other properties of the NCP functions, see [1, 6, 8].

If f and g_i are Lipschitz continuously differentiable, then $\psi(0, 0) = 0$ implies that $\psi^2(a, b)$ is continuously differentiable at $(0, 0)$ and $\|\Phi(x, \mu)\|^2$ is continuously differentiable. The Newton direction of $\Phi(x, \mu) = 0$ or $\|\Phi(x, \mu)\|^2 = 0$ is a descent direction of $\|\Phi(x, \mu)\|$ or $\|\Phi(x, \mu)\|^2$.

In this paper, in instead of using the constraint violation function $p(G(x))$ in the filter F of Fletcher and Leyffer method, we use the constraint violation function $p(G(x), \mu) = \|\Phi_1(x, \mu)\|$.

3 Algorithm

At the kth iteration of the algorithm, let F^k denote the current filter. If $(-g_j(x^k), \mu^k) = (0, 0)$, let $\xi_j^k = -2$, $\eta_j^k = 2$, otherwise, let

$$(-\xi_j^k, \eta_j^k) = \nabla\psi(-g_j(x^k), \mu_j^k).$$

We have

$$(\xi_j^k \nabla g_j(x^k), \eta_j^k e_j) = \nabla\phi_j(x^k, \mu^k).$$

Clearly $\xi_j^k \le 0$ and $\eta_j^k \ge 0$. Let

$$V^k = \begin{pmatrix} V_{11}^k & V_{12}^k \\ V_{21}^k & V_{22}^k \end{pmatrix} = \begin{pmatrix} H^k & \nabla G^k \\ \mathrm{diag}(\xi^k)(\nabla G^k)^T & \mathrm{diag}(\eta^k + c^k) \end{pmatrix}, \qquad (3.1)$$

where H^k is a symmetric positive definite matrix which may be modified by BFGS update and $\nabla G^k = \nabla G(x^k)$, $\mathrm{diag}(\xi^k)$ or $\mathrm{diag}(\eta^k + c^k)$ denotes the diagonal matrix whose jth diagonal element is ξ_j^k or $\eta_j^k + c_j^k$, respectively, and

$$c_j^k = c \min\{1, \|\Phi^k\|^\nu\},$$

where $\Phi^k = \Phi^k(x^k, \mu^k)$, $c > 0$ and $\nu > 1$ are given parameters.

Algorithm 3.1.

Step 0. Initialization.

Choose an initial guess $x^0 \in R^n$, $\tau \in (0, 1)$, $\bar{\mu} \ge \mu^0 > 0$, $1 > \theta_1 > \theta > 0$ $c > 0$, and $\nu > 1$, a symmetric positive definite matrix H^0. Let $F^0 = \{(f(x^0), \mu^0)\}$.

Step 1. Computation of the search direction.

If $\Phi^k \ne 0$ then compute d^{k0} and $\bar{\lambda}^{k0}$ by solving the following linear system in (d, λ):

$$V^k \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f^k \\ 0, \end{pmatrix}, \qquad (3.2)$$

where $\nabla f^k = \nabla f(x^k)$. If $\eta_j^k \neq 0$ then let $\lambda_j^{k0} = \eta_j^k \bar{\lambda}_j^{k0}/(-\eta_j^k + c_j^k)$, otherwise let $\lambda_j^{k0} = \bar{\lambda}_j^{k0}$. Compute d^{k1} and $\bar{\lambda}^{k1}$ by solving the following linear system in (d, λ):

$$V^k \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla L^k \\ -\Phi_1^k \end{pmatrix}, \tag{3.3}$$

where $\nabla L^k = \nabla L(x^k, \mu^k)$ and $\Phi_1^k = \Phi_1(x^k, \mu^k)$. If $\eta_j^k \neq 0$ then let $\lambda_j^{k1} = \eta_j^k \bar{\lambda}_j^{k1}/(-\eta_j^k + c_j^k)$, otherwise let $\lambda_j^{k1} = \bar{\lambda}_j^{k1}$.

Step 2. Line search with filter.

2.1. If

$$\|\Phi(x^k + d^{k1}, \mu^k + \lambda^{k1})\| \leq \theta_1 \|\Phi^k\| \tag{3.4}$$

and (3.6) or (3.7), at least one, holds, then let $x^{k+1} = x^k + d^{k1}$ and $\mu^{k+1} = \mu^k + \lambda^{k1}$. Go to Step 3.

2.2. If $\Phi_1^k = 0$ then let $b^k = 1$ and $\rho^k = 0$. Otherwise, if $d^{k0} = 0$ then let $b^k = 0$ and $\rho^k = 1$, else denote $b^k = 1 - \rho^k$ and

$$\rho^k = \begin{cases} 1, & \text{if } (d^{k1})^T \nabla f^k \leq \theta (d^{k0})^T \nabla f^k, \\ (1-\theta)\frac{(d^{k0})^T \nabla f^k}{(d^{k0} - d^{k1})^T \nabla f^k}, & \text{otherwise}, \end{cases} \tag{3.5}$$

and let

$$\begin{pmatrix} d^k \\ \lambda^k \end{pmatrix} = b^k \begin{pmatrix} d^{k0} \\ \lambda^{k0} \end{pmatrix} + \rho^k \begin{pmatrix} d^{k1} \\ \lambda^{k1} \end{pmatrix}.$$

Check whether (x^{k+1}, μ^{k+1}) is acceptable for the filter test: let $x^{k+1} = x^k + \alpha_k d^k$ and $\mu^{k+1} = \mu^k + \alpha_k \lambda^k$, where $\alpha^k = \tau^j$ and j is the smallest non-negative integer satisfying

$$\text{either} \qquad \|\Phi_1(x^{k+1}, \mu^{k+1})\| \leq \theta \|\Phi_1^l\|, \tag{3.6}$$

$$\text{or} \qquad f(x^{k+1}) - f(x^l) \leq -\alpha_k \theta \|\Phi_1^{k+1}\| \tag{3.7}$$

for all $(f(x^l), \|\Phi_1^l\|) \in F^k$. If there is no such $(x^{k+1}, \mu^{(k+1)})$ or α_k is too small, use the restoration phase to find $(x^{k+1}, \mu^{(k+1)})$ so that it is acceptable by the filter F^k. Go to Step 1.

Step 3 Update.

If x^{k+1} is a KKT point then stop. Otherwise, if $\mu_i^{k+1} \leq \bar{\mu}$ then $\mu_i^{k+1} = \mu_i^{k+1}$, otherwise let $\mu_i^{k+1} = \bar{\mu}$, give H^{k+1} by BFGS update, $F^{k+1} = F^k \cup (f(x^{k+1}), \|\Phi_1^k\|)$ and delete all pairs $(f(x^l), \|\Phi_1^l\|)$ which are dominated by $(f(x^{k+1}), \mu^{k+1})$ in F^{k+1}. Set $k := k+1$ and go to Step 1.

4 Implementation

We suppose that the following assumptions A1-A3 hold.

A1 The level set $\{x | f(x) \leq f(x^0)\}$ is bounded, and for sufficiently large k, $\|\mu^k + \lambda^{k0} + \lambda^{k1}\| < \bar{\mu}$.

A2 f and g_i are Lipschitz continuously differentiable, and for all y, $z \in R^{n+m}$,

$$\|\nabla L(y) - \nabla L(z)\| \leq m_0 \|y - z\|, \qquad \|\Phi(y) - \Phi(z)\| \leq m_0 \|y - z\|,$$

where $m_0 > 0$ is a Lipschitz constant.

A3 H^k is positive definite and there exist positive numbers m_1 and m_2 such that $m_1 \|d\|^2 \leq d^T H^k d \leq m_2 \|d\|^2$ for all $d \in R^n$ and all k.

Lemma 4.1. *If $\Phi^k \neq 0$ then V^k is nonsingular.*

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

$$H^k u + \nabla G^k v = 0 \tag{4.1}$$

and

$$\mathrm{diag}(\xi^k)(\nabla G^k)^T u + \mathrm{diag}(\eta^k + c^k)v = 0. \tag{4.2}$$

From the definitions of ξ_j^k and η_j^k, we know that $\xi_j^k \leq 0$ and $\eta_j^k + c_j^k > 0$ for all j. So, $\mathrm{diag}(\eta^k + c^k)$ is nonsingular. We have

$$v = -(\mathrm{diag}(\eta^k + c_j^k))^{-1}\mathrm{diag}(\xi^k)(\nabla G^k)^T u. \tag{4.3}$$

Putting (4.3) into (4.1), we have

$$\begin{aligned} &u^T(H^k u + \nabla G^k v) \\ &= u^T H^k u - u^T \nabla G^k \mathrm{diag}(\xi^k)(\mathrm{diag}(\eta^k + c^k))^{-1}(\nabla G^k)^T u = 0. \end{aligned}$$

The fact that H^k is positive definite and $-\nabla G^k \mathrm{diag}(\xi^k)(\mathrm{diag}(\eta^k + c^k))^{-1}(\nabla G^k)^T$ is positive semidefinite imply $u = 0$, and then $v = 0$ by (4.3). Hence V^k is nonsingular. □

The following lemma holds (see [5, 7]).

Lemma 4.2. *If $d^{k0} \neq 0$, then*

$$(d^{k0})^T H^k d^{k0} \leq -(d^{k0})^T \nabla f^k.$$

We see that if $(d^{k1})^T \nabla f^k \geq \theta (d^{k0})^T \nabla f^k$, then (3.5) implies

$$\begin{aligned} &(d^k)^T \nabla f^k = (1-\rho^k)(d^{k0})^T \nabla f^k + \rho^k (d^{k1})^T \nabla f^k \\ = \;& (d^{k0})^T \nabla f^k \left[1 - (1-\theta)\frac{(d^{k0})^T \nabla f^k}{(d^{k0}-d^{k1})^T \nabla f^k} - (1-\theta)\frac{(d^{k1})^T \nabla f^k}{(d^{k0}-d^{k1})^T \nabla f^k}\right] \\ = \;& \theta (d^{k1})^T \nabla f^k \leq -\theta (d^{k0})^T H^k d^{k0}. \end{aligned} \tag{4.4}$$

Lemma 4.3. *There exists an $m_3 > 0$ such that, for any $0 < t \leq 1$,*

$$\|\Phi_1(x^k + td^{k0}, \mu^k + t\lambda^{k0})\|^2 - \|\Phi_1^k\|^2 \leq m_3 t^2.$$

Proof. If $\Phi_1^k = 0$, let $m_4 = m_0^2$. Then for any $0 < t \leq 1$, we have

$$\begin{aligned} &\|\Phi_1(x^k + td^{k0}, \mu^k + t\lambda^{k0})\|^2 = \|\Phi_1(x^k + td^{k0}, \mu^k + t\lambda^{k0}) - \Phi_1^k\|^2 \\ &\leq t^2 m_0^2 \|(d^{k0}, \lambda^{k0})\|^2 = t^2 m_4 \|(d^{k0}, \lambda^{k0})\|^2, \end{aligned}$$

So the lemma holds for $\Phi_1^k = 0$.

We define that if $(g_i^k, \mu_i^k) \neq (0,0)$ then $(\bar{\xi}_i^{k0}, \bar{\eta}_i^{k0}) = (\xi_i^k, \eta_i^k)$, otherwise $\bar{\xi}_i^{k0}(\nabla g_i^k)^T d^{k0} + \bar{\eta}_i^{k0}\lambda_i^{k0} = \phi_i'((x^k, \mu^k), (d^{k0}, \lambda^{k0}))$, where $\phi_i'((x^k, \mu^k), (d^{k0}, \lambda^{k0}))$ is the directional derivative of $\phi_i(x, \mu)$ at (x^k, μ^k) in the direction (d^{k0}, λ^{k0}). Then $\phi_i(0,0) = 0$ implies $(\Phi_1^k)^T(\mathrm{diag}(\bar{\xi}^{k0})(\nabla G^k)^T, \mathrm{diag}(\bar{\eta}^{k0})) = (\Phi_1^k)^T(\mathrm{diag}(\xi^k)(\nabla G^k)^T, \mathrm{diag}(\eta^k))$, and

$$\begin{aligned} &\|\Phi_1^k + t(\mathrm{diag}(\bar{\xi}^{k0})(\nabla G^k)^T d^{k0} + \mathrm{diag}(\bar{\eta}^{k0})\lambda^{k0})\|^2 \\ = \;& \|\Phi_1^k\|^2 + t^2 \|\mathrm{diag}(\bar{\xi}^{k0})(\nabla G^k)^T d^{k0} + \mathrm{diag}(\bar{\eta}^{k0})\lambda^{k0}\|^2. \end{aligned} \tag{4.5}$$

It is clear that

$$\|\Phi_1(x^k + td^{k0}, \mu^k + t\lambda^{k0})\|^2 = \|\Phi_1^k\|^2 + O(t^2).$$

This completes the proof of the lemma. □

Lemma 4.4. *If $\Phi_1^k \neq 0$ then for any given $\varepsilon > 0$ there is a $\bar{t} > 0$ such that, for any $0 < t \leq \bar{t}$,*

$$\|\Phi_1^k\|^2 - \|\Phi_1(x^k + td^{k1}, \mu^k + t\lambda^{k1})\|^2 \geq (2 - \varepsilon)t\|\Phi_1^k\|^2.$$

Proof. If $\Phi_1^k \neq 0$, then (3.3) implies

$$\mathrm{diag}(\xi^k)(\nabla G^k)^T d^{k1} + \mathrm{diag}(\eta^k + c^k)\lambda^{k1} = -\Phi_1^k. \tag{4.6}$$

We define that if $(g_i^k, \mu_i^k) \neq (0, 0)$, then $(\bar{\xi}_i^{k1}, \bar{\eta}_i^{k1}) = (\xi_i^k, \eta_i^k)$, otherwise $\bar{\xi}_i^{k1}(\nabla g_i^k)^T\ d^{k1} + \bar{\eta}_i^{k1}\lambda^{k1} = \phi_i'((x^k, \mu^k), (d^{k1}, \lambda^{k1}))$, where $\phi_i'((x^k, \mu^k), (d^{k1}, \lambda^{k1}))$ is the directional derivative of $\phi_i(x, \mu)$ at (x^k, μ^k) in the direction (d^{k1}, λ^{k1}).

Clearly, for all i,

$$\phi_i(x^k + td^{k1}, \mu^k + t\lambda^{k1}) - \phi_i^k - t(\bar{\xi}_i^{k1}(\nabla g_i^k)^T d^{k1} + (\bar{\eta}_i^{k1})\lambda^{k1}) = o(t). \tag{4.7}$$

Since $c_i^k \neq 0$, it follows by the definition of c_i^k, η_i^k and (4.6) that

$$\begin{aligned} & \|\Phi_1^k + t(\mathrm{diag}(\bar{\xi}^{k1})(\nabla G^k)^T d^{k1} + \mathrm{diag}(\bar{\eta}^{k1})\lambda^{k1})\|^2 \\ = \quad & (1 - 2t)\|\Phi_1^k\|^2 + t^2\|\mathrm{diag}(\bar{\xi}^{k1})(\nabla G^k)^T d^{k1} + \mathrm{diag}(\bar{\eta}^{k1})\lambda^{k1}\|^2. \end{aligned} \tag{4.8}$$

It follows from (4.7) and (4.8) that, given any $\varepsilon > 0$, there is a $\bar{t} > 0$ such that, for any $0 < t \leq \bar{t}$,

$$\|\Phi_1^k\|^2 - \|\Phi_1(x^k + td^{k1}, \mu^k + t\lambda^{k1})\|^2 \geq (2 - \varepsilon)t\|\Phi_1^k\|^2.$$

Hence, this lemma holds. □

From Lemmas 4.2-4.4 and (4.4), we know that if $\Phi_1^k \neq 0$, then (d^k, λ^k) is a descent direction of $\|\Phi^k\|^2$; if $d^{k0} \neq 0$, then d^k is a descent direction of f^k. If $\Phi_1^k = 0$ and $d^{k0} = 0$, then (x^k, μ^k) is a KKT point.

5 Convergence

In this section, we discuss the global and superlinear convergence of the method.

In addition to A1-A3, we need the following assumption:

A4 For all k and some $\alpha_{\min} > 0$, $\alpha_k > \alpha_{\min} > 0$.

Lemma 5.1. *Consider sequences of $\{\|\Phi_1(x^k)\|\}$ and $\{f^k\}$ such that $\{f^k\}$ is monotonically decreasing and bounded below. Let a positive constant θ satisfy, for all k and $l \in \mathcal{F}^{\ k}$, that*

$$\textit{either} \qquad \|\Phi_1(x^{k+1}, \mu^{k+1})\| \leq \theta\|\Phi_1(x^l, \mu^l)\|, \tag{5.1}$$

$$\textit{or} \qquad f(x^{k+1}) - f(x^l) \leq -\alpha_k\theta\|\Phi_1(x^{k+1}, \mu^{k+1})\|, \tag{5.2}$$

where $\alpha_k \geq \alpha_{\min} > 0$ is the step length. Then $\Phi_1(x^k, \mu^k) \to 0$.

Proof. Suppose the theorem is not true. Then there exists an $\varepsilon > 0$ and an infinite index set K such that $\|\Phi_1(x^k, \mu^k)\| \geq \varepsilon > 0$ and $\|\Phi_1(x^{k+1}, \mu^{k+1})\| \geq \theta\|\Phi_1(x^k, \mu^k)\|$ for any $k \in K$. We have

$$f(x^k) - f(x^{k+1}) \geq \alpha_k\theta\|\Phi_1(x^k, \mu^k)\| > \alpha_{\min}\theta\varepsilon. \tag{5.3}$$

Because $\{f^k\}$ is monotonically decreasing, (5.3) implies $f(x^k) \to -\infty$ as $k \to +\infty$ which contradicts to the assumption. This lemma holds. □

Lemma 5.2. *Suppose the assumptions in Lemma 5.1 hold. Consider an infinite sequence of iterations on which* $\{f^k, \|\Phi_1(x^k, \mu^k)\|\}$ *entered the filter, where* $\|\Phi_1(x^k, \mu^k)\| > 0$ *and* $\{f^k\}$ *is bounded below. Then* $\Phi_1(x^k, \mu^k) \to 0$.

Proof. Suppose the theorem is not true. Then there exist an $\varepsilon > 0$ and an infinite index set K such that either $\|\Phi_1(x^k, \mu^k)\| \geq \varepsilon > 0$ and $\|\Phi_1(x^k, \mu^k)\| \leq \theta\|\Phi_1(x^l, \mu^l)\|$ for any $k \in K$ and $k > l$, implying $\{\|\Phi_1(x^k, \mu^k)\|\}_{k\in K} \to 0$, or $\{f^k\}$ is monotonically decreasing, which, by lemma 5.1, implies $\|\Phi_1(x^k, \mu^k)\| \to 0$. So, this lemma holds. □

The following Lemmas hold (see [5]).

Lemma 5.3. $d^{k0} \to 0$.

Lemma 5.4. $d^{k0} = 0$ *if and only if* $\nabla f^k = 0$*, and* $d^{k0} = 0$ *implies* $\bar{\lambda}^{k0} = 0$ *and* $\lambda^{k0} = 0$*. If* (x^*, μ^*) *is an accumulation point of* $\{(x^k, \mu^k)\}$ *then* $d^{*0} = 0$*, and* $(d^{*0}, \bar{\lambda}^{*0})$ *is the solution of the following equations*

$$V^* \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f^* \\ 0 \end{pmatrix}, \tag{5.4}$$

where $\nabla f^* = \nabla f(x^*)$ *and* $\nabla L(x^*, \mu^*) = 0$.

Lemmas 5.2-5.4 imply the following theorem.

Theorem 5.5. *If* (x^*, μ^*) *is an accumulation point of* $\{(x^k, \mu^k)\}$ *then* x^* *is a KKT point of problem (NLP).*

Now we consider the superlinear convergence of the method. We need the following assumptions.

A5 $\{\nabla g_i(x^*)\}_{i\in I(x^*)}$ are linearly independent, where $I(x^*) = \{i : g_i(x^*) = 0\}$ and x^* is an accumulation point of $\{x^k\}$ and a KKT point of problem (NLP).

A6 The sequence $\{H^k\}$ satisfies

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

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

Assumption A7 implies that Φ is continuously differentiable at each KKT point (x^*, μ^*). Similar to Lemma 4.1 we have (see [5, 7]):

Lemma 5.6. $V(x^*, \mu^*)$ *is nonsingular.*

Lemma 5.7. *For sufficiently large* k, $x^{k+1} = x^k + d^{k1}$ *and* $\mu^{k+1} = \mu^k + \lambda^{k1}$.

Theorem 5.8. *Assume A1-A7 hold. Let Algorithm 3.1 be implemented to generate a sequence* $\{(x^k, \mu^k)\}$ *and* (x^*, μ^*) *be an accumulation point of* $\{(x^k, \mu^k)\}$*. Then* (x^*, μ^*) *is an KKT point of problem (NLP), and* (x^k, μ^k) *converges to* (x^*, μ^*) *superlinearly.*

6 Numerical Tests

In this section, we report some preliminary numerical results of Algorithm 3.1 for some constrained optimization problems from [9].

In the implementation of the algorithm, the termination criterion is $\|\phi\| \leq 10^{-5}$. The parameters of the algorithm are chosen as: $c = 0.1$, $\nu = 2$, $\tau = 0.7$, $\theta_1 = 0.8$, $\theta = 0.6$,

Table 1: Numerical results for Algorithm 3.1

Problem No.	Initial point	NIT	NF	NG	Initial points	NIT	NF	NG
227	(0.5, 0.5)	11	25	31	(1, 1)	12	26	32
227	(10, 10)	15	27	37	(-10, -10)	13	18	27
215	(0.5, 0.5)	10	13	24	(1.5, 1.5)	13	35	91
215	(1, 1)	7	17	28	(2, 2)	6	15	35
232	(2, 0.5)	5	7	9	(4, 1)	5	7	13
232	(4,2)	5	9	12	(6,2)	8	10	13
250	(10, 10, 10)	10	15	27	(-10, -10, -10)	10	16	28
250	(15, 15, 15)	8	13	18	(5, 5, 5)	9	17	19

$\bar{\mu} = 10000$, $\mu^0 = 1$. The initial $H^0 = I$, where I is the unit matrix. The matrices H^ks are updated by BFGS method (see [8]).

Numerical results are summarized in Table 1, where

- Problem No=the same number of the problem in [9];
- NIT =the number of iterations;
- NF= the number of evaluations of the objective and constraint functions;
- NG=the number of evaluations of Φ.

It has been found in our numerical experiments that if $\|\phi\| \leq 10^{-6}$, then the algorithm converges very quickly. This is due to the fact that each iteration of Algorithm 3.1 can be viewed as a perturbation of a mixed Newton and quasi-Newton iteration locally. We also found that the parameter c can not be chosen too small. This is because a small c may influence the convergence rate when the strict complementarity conditions are not satisfied at some iteration points. So, we may consider some modification to the algorithm when the strict complementarity conditions are not satisfied near an iteration point. For example, instead of using a constant c, we may use $c^k \in [0.001, 0.5]$, whose value depends on $\|\Phi^k\|$, the strict complementarity and the termination criterion.

References

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

[2] R. Fletcher, S. Leyffer and P. Toint, On the global convergence of a filter-SQP algorithm, *SIAM J. Optim.* 12 (2002) 44–59.

[3] N.I.M. Gould, C. Sainvitu and P.L. Toint, A filter-trust-region method for unconstrained optizization, *SIAM J. Optim.* 16 (2006) 310–340.

[4] 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.

[5] D. Pu, K. Li and W. Xue, Convergence of QP-free infeasible methods for nonlinear inequality constrained optimization problems, *J. Tongji University* 33 (2005) 525–529.

[6] D. Pu and Y. Zhou, Piecewise linear NCP function for QP-free feasible method, *Appl. Math. A Journal of Chinses Universities* 21 (2006), 289–301.

[7] D. Pu, Y. Zhou and Z. Zhang, A QP free feasible method, *J. Comput. Math.* 22 (2004) 651-660.

[8] H. Qi and L. Qi, A new QP-free, globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization, *SIAM J. Optim.* 11 (2000) 113–132.

[9] K. Schittkowski, More Test Examples for Nonlinear Programming Codes, Springer-Verlag, New York, 1988.

[10] M. Ulbrich and S. Ulbrich, Non-monotone trust region methods for nonlinear equality constrained optizization without a penalty function, *Math. Program.* 95 (2003) 103–135.

Manuscript received 3 October 2006
revised 27 September 2007
accepted for publication 28 September 2007

Kangdi Li
School of Science, Shanghai University of Electric Power
Pingliang Lu No. 2103, Shanghai 200090, P.R. China
E-mail address: `lkd123@citiz.net`

Dingguo Pu
Department of Mathematics, Tongji University
Siping Lu, No.1239, Shanghai 200092, P.R. China
E-mail address: `madpu@tongji.edu.cn`

Weiwen Tian
Department of mathematics, Shanghai University
Shangda Lu No. 99, Shanghai 200436, P.R. China
E-mail address: `wwtian@mail.shu.edu.cn`