A New Filter Method*

Pu Dingguo[1,2†] Shao Wenqiong[2] Liu Meiling[2] Liu Ciwen[2]

Abstract In this paper, we define a new filter and propose a filter QP-free infeasible method with some piecewise linear relational 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-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.

Keywords Operations research, convergence, filter method, QP-free method, NCP function

Subject Classification (GB/T13745-92) 110.74

新的滤子方法

濮定国 [1,2†] 邵雯琼 [2] 刘美玲 [2] 刘慈文 [2]

摘要 本文定义了一种新的滤子方法，并提出了求解光滑不等式约束最优化问题的滤子 QP-free 非可行域方法. 通过乘子和分片线性非线性互补函数，构造一个等价于原约束问题一阶 KKT 条件的非光滑方程组. 在此基础上，通过牛顿 - 拟牛顿迭代得到满足 KKT 最优条件的解，在迭代中采用了滤子线搜索方法，证明了该算法是可实现，并具有全局收敛性. 另外，在较弱条件下可以证明该方法具有超线性收敛性.

关键词 运筹学，收敛性，滤子方法，无二次子规划方法，非线性互补函数

学科分类号(GB/T13745-92) 110.74

收稿日期： 2010 年 7 月 3 日.

* This research is supported by National Science Foundation of China (No. 10771162).

1. School of Mathematics and Statistics, Henan University of Science and Technology, Henan Luoyang 471003, China; 河南科技大学数学与统计学院，河南洛阳， 471003

2. Department of Mathematics, Tongji University, Shanghai 200092, China; 同济大学数学系，上海，200092

† Corresponding author 通讯作者

1 Introduction

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

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

where $x \in R^n$ and $G = (g_1, g_2, \cdots, g_m)^{\mathrm{T}} : R^n \to R^m$ is the inequality constraint.

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):

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

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

Problem (**??**) 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 (**??**) is to construct a Newton method for solving a system of nonlinear equations:

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

which is a reformulation of (**??**).

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 [6,8-10].

On the other hand, we define the constraint violation 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. In addition the computational results reported in Fletcher and Leyffer are also very encouraging (see [2-3,7,12]).

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 $p(G(x^k)) \leqslant p(G(x^l))$ and $f(x^k) \leqslant f(x^l)$.

Definition 1.2 A filter F is a list of pairs $\{(p(G(x^k), f(x^k)))\}$ 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.

Call $\mathcal{F} = \{l : (p(G(x^l), f(x^l)) \in F\}$ is the index set accompanied with F.

For k-th iteration, We use F^k to denote the current filter and $\mathcal{F}^k$ to be the set of iteration indices j $(j \leqslant k)$ such that $(p(G(x^l)), f(x^l)) \in F^k$.

In this paper, we define a piecewise linear relational NCP function and propose a filter QP-free infeasible method with 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-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. Some preliminary numerical results indicate that this new QP-free infeasible method is quite promising.

2 Preliminaries

Definition 2.1 (NCP pair and NCP function) We call a pair $(a, b) \in R^2$ to be an NCP pair if $a \geqslant 0, b \geqslant 0$ 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 relational NCP function ψ with a parameter $k > 0$ as follows.

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

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

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

and

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

Let

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

We denote $\Phi(x,\mu) = ((\nabla_x L(x,\mu))^{\mathrm{T}}, (\Phi_1(x,\mu))^{\mathrm{T}})^{\mathrm{T}}$, where $\Phi_1(x,\mu) = (\phi_1(x,\mu), \cdots \phi_m(x,\mu))^{\mathrm{T}}$. Clearly, the KKT optimality conditions (**??**) 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

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

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

$$\partial\phi_i(x,\mu)=\left\{\left(\begin{array}{c}-k^2t^2\nabla g_i(x)\\ 2k(1-t)e_i\end{array}\right)\bigcup\left(\begin{array}{c}-2k^2(1-t^2)\nabla g_i(x)\\ (2k-2t)e_i\end{array}\right)\Bigg|\ |t|\leqslant 1\right\},\qquad(2.5)$$

where $e_i=(0,\cdots,0,1,0\cdots,0)^{\mathrm T}\in R^m$ is the ith column of the unit matrix, its ith element is 1, and other elements are 0. In this paper we take $k=1$

Another piecewise linear relational NCP function was proposed in [?]. For other properties of the NCP functions, see [1,8,10].

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 $(\Phi(x,\mu))^{\mathrm T}\Phi(x,\mu)=0$ is a descent direction of $\|\Phi\|$ or $\|\Phi\|^2$, respectively.

In this paper, we replace the constraint violation $p(G(x))$ in the filter F of Fletcher and Leyffer method by $p(G(x),\mu)=\|\Phi(x,\mu)\|$.

3 Algorithm

If $(-g_j(x^k),\mu^k)=(0,0)$, let

$$\xi_j^k=-2,\qquad\eta_j^k=2,$$

otherwise, let

$$(-\xi_j^k,\eta_j^k)=\nabla\psi(a,b)|_{a=-g_j(x^k),b=\mu_j^k}.$$

We obtain

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

Clearly $\xi_j^k\leqslant 0$ and $\eta_j^k\geqslant 0$. Let

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

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

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

Step 1. Computation of the search direction

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

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

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

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

and (**??**) or (**??**), 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

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

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

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

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

for all $(f(x^l), \|\Phi^l\|) \in F^k$. 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 $\mu_i^{k+1} \leqslant \bar{\mu}$ then $\mu_i^{k+1} = \mu_i^{k+1}$; otherwise let $\mu_i^{k+1} = \bar{\mu}$, give H^{k+1} by BFGS update, $F^{k+1} = F^k \cup (f(x^{k+1}), \|\Phi^{k+1}\|)$ and delete all pairs $(f(x^l), \|\Phi^l\|)$ which are dominated by $(f(x^{k+1}), \Phi^{k+1})$ in F^{k+1}. Obtain $\mathcal{F} = \{l : (f(x^l), \Phi^l) \in F^{k+1}\}$, the index set corresponding filter F^{k+1}. Let $k = k+1$ and go to Step 1.

4 Implementation

We suppose that the following assumptions A1-A3 hold.

A1 The level set $\{x | f(x) \leqslant f(x^0)\}$ 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, $z \in R^{n+m}$,

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

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

$$m_1 \|d\|^2 \leqslant d^{\mathrm{T}} H^k d \leqslant m_2 \|d\|^2$$

for all $d \in R^n$ and all k.

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

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

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

and

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

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

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

Putting (**??**) into (**??**), we have

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

The fact that H^k is positive definite and $-\nabla G^k \mathrm{diag}(\xi^k)(\mathrm{diag}(\eta^k + c^k))^{-1}(\nabla G^k)^{\mathrm{T}}$ is positive semidefinite implies $u = 0$, and then $v = 0$ by (**??**). V^k is nonsingular. This lemma holds.

Clearly The following lemma holds (see [**?**, **?**]).

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

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

If $(d^{k1})^{\mathrm{T}}\nabla f^k \geqslant \theta(d^{k0})^{\mathrm{T}}\nabla f^k$, then (**??**) implies

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

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

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

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

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

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

$$\bar{\xi}_i^{k0}(\nabla g_i^k)^{\mathrm{T}}d^{k0} + \bar{\eta}_i^{k0}\lambda_i^{k0} = \phi_i'((x^k, \mu^k), (d^{k0}, \lambda^{k0})),$$

where $\phi_i'((x^k, \mu^k), (d^{k0}, \lambda^{k0}))$ is the direction derivative of $\phi_i(x, \mu)$ at (x^k, μ^k) in the direction (d^{k0}, λ^{k0}). Let $\mathrm{diag}(\bar{\xi}^{k0})$ or $\mathrm{diag}(\bar{\eta}^{k0})$ denote the diagonal matrix whose jth diagonal element is $\bar{\xi}_j^{k0}$ or $\bar{\eta}_j^{k0}$, respectively. Then $\phi_i(0,0) = 0$ implies

$$(\Phi_1^k)^{\mathrm{T}}(\mathrm{diag}(\bar{\xi}^{k0})(\nabla G^k)^{\mathrm{T}},\ \mathrm{diag}(\bar{\eta}^{k0})) = (\Phi_1^k)^{\mathrm{T}}(\mathrm{diag}(\xi^k)(\nabla G^k)^{\mathrm{T}},\ \mathrm{diag}(\eta^k)),$$

and

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

It is clear that

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

This lemma holds.

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

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

Proof If $\Phi_1^k\neq 0$, then (**??**) implies

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

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

$$\bar{\xi}_i^{k1}(\nabla g_i^k)^{\mathrm{T}}d^{k1}+\bar{\eta}_i^{k1}\lambda^{k1}=\phi_i'((x^k,\mu^k),(d^{k1},\lambda^{k1})),$$

where $\phi_i'((x^k,\mu^k),(d^{k1},\lambda^{k1}))$ is the direction derivative of $\phi_i(x,\mu)$ at (x^k,μ^k) in the direction (d^{k1},λ^{k1}). Let $\mathrm{diag}(\bar{\xi}^{k1})$ or $\mathrm{diag}(\bar{\eta}^{k1})$ denote the diagonal matrix whose ith diagonal element is $\bar{\xi}_i^{k1}$ or $\bar{\eta}_i^{k1}$, respectively.

Clearly, for all i,

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

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

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

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

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

Hence, this lemma holds.

From Lemmas 4.2-4.4 and (**??**), we know that if $\Phi_1^k\neq 0$, then (d^k,λ^k) is a descent direction of $\|\Phi^k\|$; 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.

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

Suppose that the assumptions A1-A4 hold in this section.

Lemma 5.1 *Consider sequences of* $\{\|\Phi(x^k)\|^2\}$ *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*

$$\text{either} \qquad \|\Phi(x^{k+1},\mu^{k+1})\| \leqslant \theta\|\Phi(x^l,\mu^l)\|, \tag{5.1}$$

$$\text{or} \qquad f(x^{k+1}) - f(x^l) \leqslant -\alpha_k\theta\|\Phi(x^{k+1},\mu^{k+1})\|, \tag{5.2}$$

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

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

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

Because $\{f_k\}$ is monotonically decreasing, (**??**) implies $f(x^k) \to -\infty$ as $k \to +\infty$ which contradict to the assumption. this lemma holds.

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

Proof Suppose the theorem is not true. Then there exist an $\varepsilon > 0$ and an infinite index set K such that either

$$\|\Phi(x^k,\mu^k)\| \geqslant \varepsilon > 0 \quad \text{and} \quad \|\Phi(x^k,\mu^k)\| \leqslant \theta\|\Phi(x^l,\mu^l)\|$$

for any $k \in K$. and $l < k \in K$ then we obtain that

$$\{\|\Phi(x^k,\mu^k)\|\}_{k\in K} \to 0,$$

or $\{f^k\}$ is monotonically decreasing, then lemma 5.1 implies $\|\Phi(x^k,\mu^k)\| \to 0$. So, this lemma holds.

The following Lemmas 5.3-5.4 hold (see [**?**]).

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 5.1 *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 a accumulation point of $\{x^k\}$ and a KKT point of problem (NLP).

A6 The sequence of $\{H^k\}$ 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^*).

It follows from that ϕ^k is differentiable 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 [**?**, **?**]):

Lemma 5.5 *$V(x^*,\mu^*)$ is nonsingular.*

Assumption A6 shows that (x^k,μ^k) is a Newton direction of Φ^k with a high order perturbation. We obtain the following Lemma 5.6 and Theorem 5.2 (see [**?**]).

Lemma 5.6 *For sufficiently large k, $x^{k+1}=x^k+d^{k1}$ and $\mu^{k+1}=\mu^k+\lambda^{k1}$.*

Lemma 5.6 implies the following theorem.

Theorem 5.2 *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

We carry out some numerical experiments on the Algorithm 3.1 in the table 1. All of test examples are the constrained optimization problems in [**?**]. The problem No.

in the table 1 is the number of this problem in [?]. These preliminary numerical results indicate that this new QP-free infeasible method may be promising.

In the implements, the termination criterion is $\|\phi\| \leqslant 10^{-5}$. The parameters are chosen as:

$c = 0.1$, $\nu = 2$, $\tau = 0.7$, $\theta_1 = 0.8$, $\theta = 0.6$, $\bar{\mu} = 10000$, $\mu^0 = 1$.

$H^0 = I$ is the unit matrix. The H^k is updated by BFGS method (see [?]).

In the "NIT, NF and NG" entries of the table below is as follows.

NIT=the number of iterations.

NF=the number of objective function and constraints are evaluations. The number of NF increases one only if all functions are evaluated once.

NG=the number of Φ (or gradient) evaluations.

Table 1

problem No.	Initial point	NIT	NF	NG	Initial points	NIT	NF	NG
227	0.5, 0.5	9	18	31	1, 1	13	23	31
227	10, 10	14	28	37	-10, -10	11	28	37
215	0.5, 0.5	7	16	25	1.5, 1.5	13	25	41
215	1, 1	7	17	28	2, 2	5	11	25
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	9	15	29	-10, -10, -10	11	16	28
250	15, 15, 15	7	14	28	5, 5, 5	9	19	29

Because each iteration of Algorithm 3.1 can be viewed as a perturbation of a mixed Newton-quasi Newton iteration locally. During numerical experiments, we find that if $\|\phi\| \leqslant 10^{-6}$ then iteration points converge very quickly. We may also use the termination criterion $\|\phi\| \leqslant 10^{-5}$.

On the other hand, we can not choose the parameter c too small. When the strict complementarity conditions are not satisfied on some iteration point, small c may influence the convergence rate. So, we may consider to make some small modification in the algorithm when the strict complementarity conditions are not satisfied near a solution. For example, instead of constant c, we may use the various $c^k \in [0.001, 0.5]$, which may be depend on $\|\Phi^k\|$, strict complementarity and the termination criterion.

References

[1] Fischer A. A special Newton-type optimization method[J]. *Optimization*, 1992, 24: 269-284.

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

[3] Gould N.I.M., Sainvitu C., Toint P.L. A filter-trust-regionmethod for unconstrained optizization[J]. *SIAM J. Optim.*, 2006, 16: 310-340.

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

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

[6] Li K., Pu D., Tian W. Filter QP-free Method with piecewise linear NCP function[J]. *Pacific J. of Optimization*, 2008, **4**(1): 43-53.

[7] Pu D., Jin Z. Nonmonotone line search technique for QP-free infeasible method[J]. *J. of Tongji University*, 2010, 38: 311-316.

[8] Pu D., Zhou Y. Piecewise linear NCP function for QP-free feasible method[J]. *Applied Mathematics A J. of Chinses Universities*, 2006, 21: 289-301.

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

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

[11] Schittkowski K. More test examples for nonlinear programming codes[M]. Berlin: Springer-Verlag, 1988.

[12] Ulbrich M., Ulbrich S. Non-monotone trust region methods for nonlinear equality constrained optizization without a penalty function[J]. *Math. Program.*, 2003, 95: 103-135.