Mathematica Aeterna, Vol. 6, 2016, no. 3, 389 - 402

An improved ODE-type filter method with smooth approximation function

Mingxi Lu
Ke Su
College of Mathematics and Information Science, Hebei University, Baoding, China

Abstract

In this paper, we propose an improved ODE-type filter method. Compared to other methods, at each iteration, only one linear system needs to be solved in our algorithm . Also , this method is more flexible and less computational scale. By using the Fischer-Burmeister function, we reformulate complementarity problems to a system of nonlinear and nonsmooth equations. Furthermore, we use the Kanzow function to approximate the Fischer-Burmeister function so that smooth and nonlinear equations can be obtained. Under some reasonable conditions, the global convergence results of our algorithm is presented.

Mathematics Subject Classification: xxxxx

Keywords: smooth approximation function ; Trust region; ODE; Filter.

1 Introduction

The nonlinear complementarity problem (NCP) is equivalent to constraint optimization problem

$$x \geq 0, F(x) \geq 0, x^T F(x) = 0. \tag{1}$$

where $x \in R^n, F(x) : R^n \to R^n$ is second-order continuously differentiable. We use an NCP function to transform the nonlinear complementarity problem to optimization problem.

Definition 1.1 (**NCPfunction**) The function $\varphi : R^2 \to R$ is called NCP function, if $\varphi(a, b) = 0$, if and only if $a \geq 0, b \geq 0, ab = 0$.

We use Fischer - Burmeister function $\varphi(a,b)=\sqrt{a^2+b^2}-a-b$ to transform nonlinear complementarity problem (1) to nonlinear equations

$$\phi(x)=\begin{pmatrix} \sqrt{{x_1}^2+{F_1}^2(x)}-x_1-F_1(x) \\ \vdots \\ \sqrt{{x_n}^2+{F_n}^2(x)}-x_n-F_n(x) \end{pmatrix}. \tag{2}$$

Thus, problem (1) is equivalent to the optimization problem as follows

$$\min f(x)=\frac{1}{2}\phi(x)^T\phi(x) \tag{3}$$

problem (1) is equivalent to find a solution the following constrained optimization problem.

$$\min\ f(x) \qquad\qquad s.t.\ c_j(x)\geq 0,\ j=1,2,\cdots,2n. \tag{4}$$

where$f(x)=\|x^TF(x)\|_2^2$ and $c(x)=(x_1,x_2,\cdots,x_n,F_1(x),F_2(x),\cdots,F_n(x))^T: R^n\rightarrow R^{2n}$. let $c_1(x)=x_1, c_2(x)=x_2,\cdots,c_n(x)=x_n, c_{n+1}(x)=F_1(x), c_{n+2}(x)=F_2(x),\cdots,c_{2n}(x)=F_n(x).F(x)$ is second-order continuously differentiable.

There are many methods for inequality constrained nonlinear programming (NLP). Most of them are monotonic descent methods, that is only the trial point that makes merit function strictly descent can be accepted. There are two drawbacks, one is that the choice of penalty parameter is difficult, the other is that the monotonic descent methods can result on reduction of convergence rate when the iterate is trapped near a narrow curved valley. While, filter method, proposed by Fletcher and Leyffer [1], overcome the drawbacks above. In filter method, the use of a penalty function is replaced by the introduction of so-called filter. So, they have several advantages which is convenient to determine whether the value of constraint violation and objective function are descent and are accepted . Recently, this technique is applied to the many kinds of nonlinear problems [3-5,7].

ODE methods for minimizing a function f(x) which obtains a stepsize at each iteration by solving a system of linear equations and which proceed by the solution curve of a system of ordinary differentiable equations and is more reliability, accuracy and efficiency than conventional Newton and quasi-Newton algorithms [2].

Motivated by the referees [3,6,7], we transform the nonlinear inequality problems to a nonlinear equation, so that an improved ODE-type filter trust region method is proposed. the filter technique is employed to determine whether to accept the trial point or not. This paper is organized as follows. The next section introduces the symbols and notation we need. In section 3, an improved method is put forward. The convergent properties are analyzed in Section 4. Some numerical results are reported in the last section.

2 Preliminary Notes

Lagrangian function $L : R^{n^2+2n} \to R$ of the problem NLP is defined by

$$L(x,\lambda) = f(x) + \sum_{j=1}^{2n} \lambda_j c_j(x), \{j = 1,2,\cdots,2n.\} \tag{5}$$

It is easy to obtain the KKT conditions of problem (NLP) as following

$$\nabla_x L(x,\lambda) = 0 \tag{6}$$

$$\lambda_j \geq 0, c_j(x) \geq 0, \lambda_j c_j(x) = 0, j = \{1,2,\cdots,2n\}. \tag{7}$$

where $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_{2n})^T \in R^{2n}$ is multiplier vector.

Definition 2.1 Let $c : R^n \to R^n$ be locally Lipschitz continuous, then the B-differential of c at $x \in R^n$ is defined by

$$\partial_B(c(x)) = \{V \in R^{n\times n} | V = \lim_{x_k \to x} \nabla c(x_k), x_k \in D_c\} \tag{8}$$

The generalized Jacobian of c at x in the sense of Clarke is defined by

$$\partial(c(x)) = conv\partial_B(c(x)) \tag{9}$$

where $D_c = \{x \in R^n : c\text{at x is differential}\}$.Symbol conv(x) denotes the convex hull of set S

Definition 2.2 Let $G : R^n \to R^n$ be locally Lipschitz continuous, we call G at X semi-smooth if

$$\forall h \in R^n. \lim_{V\in\partial G(x+th').h'\to h.t\downarrow 0}(Vh') \tag{10}$$

exists for all h in R^n.

Definition 2.3 G is called P_0 function , if $\forall x, y \in R^n andx \neq y$, satisfied

$$\max_{i:x_i\neq y_i}(G_i(x) - G_i(y)) \geq 0 \tag{11}$$

and let F is P_0 function .

Definition 2.4 The function M holds

$$M_k = \max_{0\leq s\leq m(k)}\{M_{\varepsilon_k}(x^{k-s})\} \tag{12}$$

Definition 2.5 The function $\varphi : R^2 \to R$ is called NCP function,if $\varphi(a,b) = 0$ if and only if $a \geq 0, b \geq 0, ab = 0$

One of the most popular functions is Fischer-Burmeister NCP function [7]:

$$\varphi(a,b) = \sqrt{a^2+b^2} - a - b \tag{13}$$

One of the Kanzow [12] smooth approximation functions of F-B NCP function is

$$\varphi_\mu(a,b) = \sqrt{a^2+b^2+2\mu} - a - b, \mu > 0 \tag{14}$$

Lemma 2.1 Let the function ∇f be semi-smooth at X on R^{n^2},then ∇f is direction differentiable, and for $\forall V \in \partial(\bigtriangledown f(x+h))$,we have

$$(i) Vh - (\nabla f)'(x;h) = o(\|h\|); \tag{15}$$

$$(ii) \bigtriangledown f(x+h) = \nabla f(x) + (\nabla f)'(x;h) + o(\|h\|) \tag{16}$$

as h decreases infinitely, where

$$(\nabla f)'(x;h) = \lim_{t\downarrow 0}(\frac{\nabla f(x+th) - \nabla f(x)}{t}) \tag{17}$$

is called directional derivative that ∇f along the direction of h at x.

Lemma 2.2 Let V be a neighborhood of x, and $\varphi : R^n \to R^n$ is a LC^1 function, then for $x+d \in$ V , there exists $\beta > 0$, such that

$$|\varphi(x+d) - \varphi(x) - \nabla\varphi(x)^T d| \le \frac{\beta\|d\|^2}{2} \tag{18}$$

By the F-B NCP function φ , KKT conditions (6),(7) can be reformulated to the following form:

$$H_\mu(z) = \begin{pmatrix} \nabla_x(L(x,\lambda) \\ \phi_\mu(x,\lambda) \end{pmatrix} = 0 \tag{19}$$

where $Z = (x^T, \lambda^T) \in R^{n^2+2n}, \phi(x,\lambda) = (\varphi(c_1(x),\lambda_1), \varphi(c_2(x),\lambda_2), \dots, \varphi(c_{\{2n\}}(x), \lambda_{\{2n\}}))^T$..

In filter method, which is proposed by Fletcher and Leyffer [1], the acceptability of steps is determined by comparing the value of constraint violation and the objective function with previous iterates collected in a filter. Different from the traditional filter method, we define the objective function l(z) by $l(z) = \|\nabla_l(x,\lambda)\|$,and the violation function$\theta(z)$ by $\theta(z) = \|\phi_\mu(x,\lambda)\|$. So a trial point should either reduce the value of constraint violation $\theta(z)$ or that of the function l(z).

Definition 2.6 A pair $(l(z_k), \theta(z_k))$ is said to dominate another pair $(l(z_j), \theta(z_j))$ if both $l(z_k) \le l(z_j)$ and $\theta(z_k) \le \theta(z_j)$. We also call this a point z_k dominates another point z_j

Definition 2.7 A filter τ is a list of pairs $(l(z_j), \theta(z_j)))$ such that no pair dominates any other.

For convenience, we denote $z_j = (l_j, \theta_j)$. A new trial point z_k^+ is accepted if it is not dominated by any points in $\tau \bigcup z_k$. Consider the convergence property of the algorithm, we For convenience, we denote $z_j = (l_j, \theta_j)$. A new trial point z_k^+ is accepted if it is not dominated by any points in $\tau \bigcup z_k$.

Consider the convergence property of the algorithm, we call a new trial point is accepted, if and only if for $z_j \in \tau$ it holds

$$l_k^+ \le l_j - \gamma\delta(\|H_k^+\|, \|H_j\|) or \theta_k^+ \le \theta_j - \gamma\delta(\|H_k^+\|, \|H_j\|) \tag{20}$$

where $\delta(\|H_k^+\|, \|H_l\|)$=min $\{(\|H(z_k^+)\|, \|H(z_l)\|)\}$,and γ is a small positive number. Let $H_k^+ = H(z_k^+), \theta_k^+ = \theta(z_k^+)$, and so on.

If the trial point z_k is accepted in the sense of (20), then we add the pair (l_k^+, θ_k^+) into the filter, that is $\tau = \tau \bigcup (l_k^+, \theta_k^+)$. And removed those points which are dominated from the filter. For convenience, we call this the update of the filter.

3 An improved ODE-type filter algorithm

For non-smooth linear equations (19). At the k-th iteration, we use ODE trust region method [10] to obtain the search direction d_k. That is to solve the following system of linear equation

$$[h_k(\nabla H_\mu(z_k)\nabla H_\mu(z_k)^T + B_k) + \frac{I}{h_k}]d = -\nabla H_\mu(z_k)H_\mu(z_k) \tag{21}$$

Or the equivalent form

$$[(\nabla H_\mu(z_k)\nabla H_\mu(z_k)^T + B_k) + I]d = -h_k\nabla H_\mu(z_k)H_\mu(z_k) \tag{22}$$

where $h_k > 0$ is the integral step,$\nabla H_\mu(z_k)$ is the gradient of the smooth function p at the point z_k ,

the matrix B_k is the n^2+2n order symmetric matrix, which can be updated by SR1 correction [4], i.e.

$$B_{k+1} = B_k + \frac{(y_k - B_k d_k)(y_k - B_k d_k)^T}{(y_k - B_k d_k)^T d_k} \tag{23}$$

where $y_k = \nabla H_\mu(z_{k+1})H_\mu(z_{k+1}) - \nabla H_\mu(z_k)H_\mu(z_k) - \nabla H_\mu(z_{k+1})H_\mu(z_{k+1})^T d_k$

Remark 2 The computational scale of this method to obtain d_k is much less than that of solving quadratic subproblem, meanwhile, the adjustment of the step h_k is much easier in the parameter space.

Algorithm

Step 0 Initialization, choose $z_1, h_1 > 0, \overline{\epsilon} \ge 0, 0 < \eta < l, 0 < \gamma < 1$. Initial split control value $\epsilon_1 > 0$, let k=1.

Step 1 Compute $H_\mu(z)$and $\nabla H_\mu(z_k)$.

Step 2 If $\|\nabla p(z_k)H(z_k)\| \le \overline{\varepsilon}$, stop.

If $h_k(\nabla H_\mu(z_k)\nabla H_\mu(z_k)^T + B_k) + I$ is positive definite, go to Step 3. Otherwise,let m_k is the smallest integer with which the symmetric matrix $2^{-m_k}h_k(\nabla H_\mu(z_k)\nabla H_\mu(z_k)^T + B_k) + I$ is positive definite, let $h_k = 2^{-m_k}h_k$,

Step 3 Solve (22) to get d_k. Let $z_k^+ = z_k + d_k$,compute $r_k = \frac{M(z_k)-M(z_k^+)}{M(z_k)-q_k(d_k)}$. where $M(z) = \frac{1}{2}\|H_\mu(z)\|^2, q_(d) = \frac{1}{2}\|H_\mu(z) + \nabla H_\mu(z_k)^T d\|^2 + \frac{1}{2}d^T B_k d$

Step 4 Denote $l_k^+ = l(z_k^+), \theta_k^+ = \theta(z_k^+)$.If (l_k^+, θ_k^+) is not accepted by the filter, go to Step 5. Otherwise $z_{k+1} = z_k^+$, go to Step 6.

Step 5 If $r_k \geq \eta, z_{k+1} = z_k^+, h_{k+1} = 2h_k$.go to Step 7. Otherwise,$h_k = \frac{1}{2}h_k, \varepsilon_k = \|d_k\|^2$go to Step 3. (Inner loop)

Step 6 If $r_k \geq \eta, h_{\{k+1\}} = 2h_k$go to Step 7. Otherwise,add the pair (l_k^+, θ_k^+) to the filter τ and update,$h_k = \frac{1}{2}h_k$, go to Step 7

Step 7 Let $\varepsilon_{\{k+1\}} = \|d_k\|^2$,update B_k to B_{k+1}, k=k+1, go to Step 1.(Outside loop)

4 Main Results

In this section, in order to present a proof of global convergence of algorithm, we always assume that following conditions hold.

Assumptions

A1 The iterate x^k remains in a closed, bounded convex subset $S \in R^n$.

A2 There exist two constants $0 < a \leq b$ such that $a\|d\|^2 \leq d^T B_k d \leq b\|d\|^2$, for all k and for all $d \in R^n$.

Suppose that $Ared_k = M(z) - M(z_k^+), Pred_k = M(z_k) - q_k(d_k)$.then $r_k = \frac{Ared_k}{Pred_k}$

By Lemma 2.2, we obtain some results as following.

Lemma 4.1 $|Ared_k - Pred_k| = O(\|d_k\|^2)$

Lemma 4.2 $Pred_k \geq \frac{1}{2h_k}\|d_k\|^2$.

Proof By (22),we have

$$Pred_k = M(z_k) - q_k(d_k) \tag{24}$$

$$= \frac{1}{2}\|H_\mu(z)\|^2 - \frac{1}{2}d^T B_k d_k - \frac{1}{2}\|h(z_k) + \nabla H_\mu(z_k)^T d_k\|^2 \tag{25}$$

$$= -(\nabla H_\mu(z_k)^T H_\mu(z_k) d_k - \frac{1}{2}d_k^T[\nabla H_\mu(z_k)\nabla H_\mu(z_k)^T + B_k]d_k) \tag{26}$$

$$= d_k^T[\nabla H_\mu(z_k)\nabla H_\mu(z_k)^T + B_k + \frac{1}{h_k}I]d_k - \frac{1}{2}d_k^T[\nabla H_\mu(z_k)\nabla H_\mu(z_k)^T + B_k]d_k^T \tag{27}$$

$$= \frac{1}{2}d_k^T[\nabla H_\mu(z_k)\nabla H_\mu(z_k)^T + B_k + \frac{1}{h_k}I]d_k + \frac{1}{2h_k}d_k^T d_k \tag{28}$$

$$\geq \frac{1}{2h_k}\|d_k\|^2 \tag{29}$$

Lemma 4.3 (i) If $k_{s+1} - k_s \leq N$, for $\forall t = 1, 2, \ldots, k_{s+1} - k_s$ it holds

$$M_{k_s+t-1} = \max_{1 \leq i \leq k_{s+1}-k_s} M_{\varepsilon_{k_t}}(x^{k_s+t-1}) = M_{\varepsilon_{k_s}}(x^{k_s}) \tag{30}$$

(ii) If $k_{s+1} - k_s > N$, setting $Z_l = \min\{k_{s+1} - k_s - Nl, N\}$, l is positive integer and $k_{s+1} - k_s - Nl > 0$, it holds

$$\max_{1\leq i\leq Z_l} M_{\varepsilon_{k_t}}(x^{k_s+Nl+t-1}) \leq \max_{1\leq t\leq N} M_{\varepsilon_{k_t}}(x^{k_s+N(l-1)+t-1}) \leq M_{\varepsilon_{k_t}}(x^{k_s}) \tag{31}$$

Proof (i) first , if t = 1 ,as $m(k_s) = 0$, it holds $M_{k_s} = M_{\varepsilon_{k_s}}(x^{k_s})$. If t = 2 , so $M_{k_s} = M_{\varepsilon_{k_s}}(x^{k_s})$. as $x^{k_s+1} = x^{k_s} + d^{k_s}$, by $M_{\varepsilon_{k_s}}(x^{k_s+1}) \leq M_{k_s} = M_{\varepsilon_{k_s}}(x^{k_s})$,

$$M_{k_s+1} = \max\{M_{\varepsilon_{k_s}}(x^{k_s}), M_{\varepsilon_{k_s}}(x^{k_s+1})\} = M_{\varepsilon_{k_s}}(x^{k_s}), \tag{32}$$

then t = 2 is true .

The following method of induction , if t = 1 , . . . , i-1 is true , we proof t = i is true too.

On the basis of induction assumption to get $M_{k_s+i-2} = M_{\varepsilon_{k_s}}(x^{k_s})$ as $x^{k_s+i} = x^{k_s+i-1} + d^{k_s+i-1}$, then ,

$$M_{\varepsilon_{k_s}}(x^{k_s+i}) \leq M_{k_s+i-1} = M_{\varepsilon_{k_s}}(x^{k_s}) \tag{33}$$

hence

$$M_{k_s+i-1} = \max\{M_{\varepsilon_{k_s}}(x^{k_s+i-1}), M_{k_s+i-2}\} = M_{\varepsilon_{k_s}}(x^{k_s}) \tag{34}$$

(ii) To prove the first section of inequation , we need prove $\forall t$ and $1 \leq i \leq Z_l$,

$$M_{\varepsilon_{k_s}}(x^{k_s+Nl+t-1}) \leq \max_{1\leq i\leq N} M_{\varepsilon_{k_t}}(x^{k_s+N(l-1)+i-1}) \tag{35}$$

first we pay attention to the definition of function M_k

$$M_{k_s+Nl-1} = \max_{1\leq b\leq N} M_{\varepsilon_{k_t}}(x^{k_s+N(l-1)+b-1}) \tag{36}$$

when t=1 $x^{k_s+Nl} = x^{k_s+Nl-1} + d^{k_s+Nl-1}$, by r_k , then

$$M_{\varepsilon_{k_s}}(x^{k_s+Nl}) \leq M_{k_s+Nl-1} = \max_{1\leq i\leq N} M_{\varepsilon_{k_t}}(x^{k_s+N(l-1)+t-1}). \tag{37}$$

Hence , t = 1 the inequation is true .

The following we use induction method . Assuming when t = 1, . . . ,i-1 , the inequation is true .

Then , t =i on the basis of induction Assumption $M_{\varepsilon_{k_s}}(x^{k_s+Nl+i-1}) \leq M_{k_s+Nl+i-2} \leq \max\{\max_{t=1,...,i-t}\{M_{\varepsilon_{k_s}}(x^{k_s+Nl+t-1})\}, M_{k_s+Nl-1}\}$

$$\leq M_{k_s+Nl-1} = \max_{1\leq t\leq N} M_{\varepsilon_{k_s}}(x^{k_s+N(l-1)+t-1}) \tag{38}$$

hence the first section of statement (ii) comes into existence . and the like ,and by (i)

$$\max_{1\leq t\leq Z_l} M_{\varepsilon_{k_t}}(x^{k_s+Nl+t-1}) \leq \max_{1\leq t\leq N} M_{\varepsilon_{k_t}}(x^{k_s+N(l-1)+t-1}) \tag{39}$$

$\max_{1\leq t\leq N} M_{\varepsilon_{k_t}}(x^{k_s+N(l-2)+t-1}) \leq \cdots \leq \max_{1\leq t\leq N} M_{\varepsilon_{k_t}}(x^{k_s+t-1})$

$$= M_{\varepsilon_{k_t}}(x^{k_s}) \tag{40}$$

so the second section of statement (ii) comes into existence .

Lemma 4.4 If K is an finite set , then there exists infinity index set K^* includes in D to let that formula be true $0 \leq \sum_{k\in K^*}[M_k - M_{\varepsilon_k}(x^{k+1})] < +\infty$

Proof First $\forall k \in D$, on the basis of end conditions of algorithm to get $M_k - M_{\varepsilon_k}(x^{k+1}) \geq 0$. It is easy to get $\forall K^* includesinD$,

$$0 \leq \sum_{k\in K^*}[M_k - M_{\varepsilon_k}(x^{k+1})]. \tag{41}$$

As K is an finite , we assume $\widehat{k}$ is biggest index of K . Then , $\forall k \geq \widehat{k}, \varepsilon_k = \varepsilon_{\widehat{k}}, \xi_k = \xi_{\widehat{k}}$ means $\widetilde{\varepsilon} = \varepsilon_{\widehat{k}}, \widetilde{\xi} = \xi_{\widehat{k}}$

Let

$$K^* = \{k|M_{\widetilde{\varepsilon}}(x^{k+1})\} = \max_{1\leq t\leq N} M_{\widetilde{\varepsilon}}(x^{k_s+Nl+t-1}), l = 1, 2, \ldots\} \tag{42}$$

for $k \in K^*$, suppose $k_j + Nl \leq k < k_s + N(l+1)$. by (ii) of lemma 4.3 ,we can get

$$M_{\widetilde{\varepsilon}}(x^k) = \max_{1\leq t\leq N} M_{\widetilde{\varepsilon}}(x^{k_s+N(l_1)+t-1}) \tag{43}$$

by (42) , (43)

$$\sum_{k\in K^*}[M_k - M_{\varepsilon_k}(x^{k+1})] = \sum[M_k - \max_{1\leq t\leq N} M_{\widetilde{\varepsilon}}(x^{k_s+Nl+t-1})] \tag{44}$$

$$\leq \sum_{l=1}^{\infty}\{\max_{1\leq t\leq N} M_{\widetilde{\varepsilon}}(x^{k_s+N(l-1)+t-1}) - \max_{1\leq t\leq N} M_{\widetilde{\varepsilon}}(x^{k_s+Nl+t-1})\} \tag{45}$$

$$\leq \max_{1\leq t\leq N} M_{\widetilde{\varepsilon}}(x^{k_s+t-1}) < +\infty \tag{46}$$

Lemma 4.5 let F is P_0 function . If the point of $\{x^k\}$ has an accumulation point at least , then index set K is finite and

$$\lim_{k\to\infty} \varepsilon_k = 0, \lim_{k\to\infty} H(x^k) = 0and \lim_{k\to\infty} H_{\varepsilon_k}(x^k) = 0 \tag{47}$$

Proof first we prove K is a endless set. Otherwise , we assume the K is an finite set On the basis of Lemma 4.4 , for the max value $\widehat{k}$ of K , let $\widetilde{\varepsilon} = \varepsilon_{\widehat{k}}, \widetilde{\xi} = \xi_{\widehat{k}}$. By means of definition of K , for $\forall k > \widehat{k}$

$$\|H(x^k)\| > \max\{\eta\widehat{\xi}, \mu^{-1}\|p(x^k)\|\}, H(x^k) = H_{\widetilde{\varepsilon}}(x^k) + p(x^k). \tag{48}$$

The following ,we prove the point of $\{x^k\}$ has an accumulation at least $\overline{x} \in L_0$ to satisfy this formula $\nabla M_{\widetilde{\varepsilon}}(\overline{x}) = 0$

Assume these conclusions are not true ,there are

$$\liminf_{k\to\infty} \|\nabla M_{\widetilde{\varepsilon}}(x^k)\| > 0 and \lim_{k\to\infty} \frac{\|\nabla M_{\widetilde{\varepsilon}}(x^k)\|}{\|\nabla M_{\widetilde{\varepsilon}}(x^k)\|^2} > 0 \tag{49}$$

necessarily .

According to $r_k = \frac{M(x_k)-M(x_k^+)}{M(x_k)-q_k(d_k)}$ and step 3 , $M_{\varepsilon_k}(x^k) - Q_k(d^k) \geq \frac{1}{2}\|\nabla M_{\varepsilon_k}(x^k)\| \min\{h_k, \frac{\|\nabla M_{\varepsilon_k}(x^k)}{\|\nabla M_{\varepsilon_k}(x^k)}$ and lemma 4.4 to obtain

$$\sum_{k\in K^*} (M_{\varepsilon_k}(x^k) - Q_k(d^k)) < \infty, \tag{50}$$

the following

$$\sum_{k\in K^*} \|\nabla M_{\overline{\varepsilon}}(x^k)\| \min\{h_k, \frac{\|\nabla M_{\overline{\varepsilon}}(x^k)\|}{\|\nabla H_{\overline{\varepsilon}}(x^k)\|^2}\} < \infty \tag{51}$$

where K^* is the set of

$$K^* = \{k | M_{\widetilde{\varepsilon}}(x^{k+1}) = \max_{1\leq i\leq N} M_{\widetilde{\varepsilon}}(x^{k_j+Nl+i-1}), l = 1, 2, \cdots\}. \tag{52}$$

This shows

$$\sum_{k\in K^*} h_k < \infty, \tag{53}$$

hence, there exists one accumulation point $\overline{x}$

$$\lim_{k\in K^*, k\to\infty} x^k = \overline{x}, \lim_{k\in K^*, k\to\infty} h_k = 0. \tag{54}$$

It is contradicted with the choice of h_k, hence, $\nabla M_{\overline{\varepsilon}}(\overline{x}) = 0$ is true .

On the other hand , assuming there is an point set ${x^k}_{k\in K_3}$ which is convergency to $\overline{x}$. On the basis of F is P_0 function , $\nabla H_\varepsilon(x)$ is nonsingular and $\nabla M_{\overline{\varepsilon}}(\overline{x}) = 0$ to obtain $\{H_{\widetilde{\varepsilon}}(x^k)\}_{k\in K_0} \to H_{\widetilde{\varepsilon}}(\overline{x}) = 0$. Thus , there exists $\widetilde{k} \geq \overline{k}$ so that $\forall k \in K_0$ and $k \geq \widetilde{k}$

$$\|H_{\widetilde{\varepsilon}}(x^k)\| \leq (1-\mu)\eta\widetilde{\xi}. \tag{55}$$

Relying on that formula and (4.8) , $\forall k \in K_0$, and $\widetilde{k} \geq \overline{k}$ $\|H_{\widetilde{\varepsilon}}(x^k)\| \leq (1-\mu)\|H(x^k)\|$

$$\leq (1-\mu)(\|p(x^k)\| + \|H_{\widetilde{\varepsilon}}(x^k)\|) \tag{56}$$

That is

$$\|H_{\widetilde{\varepsilon}}(x^k)\| < (\mu^{-1} - 1)\|p(x^k)\|. \tag{57}$$

So that formula $\|H(x^k)\| \leq \|H_{\widetilde{\varepsilon}}(x^k)\| + \|p(x^k)\|$

$$< \mu^{-1}\|p(x^k)\|, \forall k \in K_0, k \geq \widetilde{k} \tag{58}$$

is true . But , that is contradicted with (48) .

Hence, the set of K is infinite

On the basis of iteration condition $\beta_{k+1} = H(x^{k+1}), 0 < \varepsilon_{k+1} \leq \min\{(\frac{\mu}{2\kappa}\beta_{k+1}{}^2)^2, \frac{\varepsilon_k}{4}, \overline{\varepsilon}(x^{k+1}, \nu\beta_{k+}$ immediately , $\{\varepsilon^k\} \to 0$. further more , on the basis of F which is P_0 function ,so iteration point is subset of one set L_0 and let the set be $L_0 = \{x \in R^n | M(x) \leq (1+\mu)^2 M(x^0)\}$. as K is a infinite set , hence by level set L_0 and q_k ,

$$\lim_{k\to\infty} H(x^k) = 0 and \lim_{k\to\infty} H_{\varepsilon_k}(x^k) = 0 \tag{59}$$

Theorem 4.6 the inner loop of the Algorithm terminates in finite number of times.

Proof Assume the conclusion is not true, we have $r_k < \eta$ by the algorithm as k increases infinitely, and $h_k \to 0$.By (22),it holds $d_k \to 0$,so

$$|r_k - 1| = |\frac{Ared_k - Pred_k}{Pred_k}| \leq \frac{O(\|d_k\|^2)}{\frac{1}{2h_k}\|d_k\|^2} \to 0 \tag{60}$$

It shows that $r_k \geq \eta$ as k increases infinitely and the desired conclusion holds.

In order to analysis the convergence properties of the algorithm, we define some sets as following:

Let $A = \{k | (l_k^+, \theta_k^+)\}$ is accepted by $tau_k\}$ represent the index set which is accepted by the filter.

$E = \{k | z_{\{k+1\}} = z_k^+\}$ represent the index set which successful iteration.

$I = \{k | (l_k^+, \theta_k^+)$ add to the filter$\}$,represent the index set which filter update.

Based on Algorithm, we have A includes in N , and there are several cases as following:

1) $|A| < \infty, |E| < \infty$ (It must hold $|I| < \infty$)

2) $|A| < \infty, |E| = \infty$ (It must hold $|I| < \infty$)

3) $|A| = \infty, |E| < \infty$ (It must hold $|I| = \infty$)

4) $|I| < \infty$ (It must hold $|A| = \infty, |E| = \infty$)

Then we discuss the convergence properties of the algorithm according to these four cases respectively.

Lemma 4.7 Suppose there are finitely many points added to the filter ($|A| < \infty$) , and there are finitely times successful iteration ($|E| < \infty$), then the algorithm can terminate in finite times. It holds $\nabla H_\mu(z_k) H_\mu(z_k) = 0$.

Proof Suppose by contradiction that there exists a constant $\varepsilon > 0$, such that $\|\nabla H_\mu(z_k) H_\mu(z_k)\| \geq \varepsilon$. Suppose k_0 be the last successful iteration, then for $\forall j$.we have $z_{k_0+1} = z_{k_0+j}$ and $z_{k_0+j} < \eta$.Based on the algorithm, it holds $h_k \to 0 (k \to \infty))$.Similar to Theorem 4.6, we obtain the contradiction and the result follows.

Lemma 4.8 Suppose there are finitely many points added into the filter ($|A| < \infty$),and there are infinitely times successful iteration ($|E| = \infty$), then

$$\liminf_{k\to\infty} \|\nabla H_\mu(z_k) H_\mu(z_k)\| = 0 \tag{61}$$

Proof By algorithm, if $|A| < \infty, |E| = \infty$, that means (l_k^+, θ_k^+) is not accepted by filter as k increases infinitely. By Theorem 4.6, it holds $r_k \geq \eta$, and $\exists h > 0$ such that $h_k \geq h$ as k increases infinitely .

Suppose by contradiction that there exists a constant $\varepsilon > 0$, and a positive index set k_0 such that $\|\nabla H_\mu(z_k)H_\mu(z_k)\| \geq \varepsilon$ as $k \geq k_0$.By the algorithm and Lemma 4.7, we have

$$Ared_k \geq \eta Pred_k \geq \frac{\eta}{2} d_k^T[\nabla H_\mu(z_k)\nabla H_\mu(z_k)^T + B_k + \frac{1}{h_k} I]d_k = -\frac{\eta}{2}(\nabla H_\mu(z_k)H_\mu(z_k))^T d_k \geq 0. \tag{62}$$

Thus ,the sequence $\{h_k\}$ is monotonically decreasing, and by A_1 , it is lower bounded. So $h_k - h_{k+1} \to 0 (k \to \infty)$,hence

$$(\nabla H_\mu(z_k)H_\mu(z_k))^T d_k \to 0(k \to \infty). \tag{63}$$

On the other hand

$$|(\nabla H_\mu(z_k)H_\mu(z_k))^T d_k| = (\nabla H_\mu(z_k)H_\mu(z_k))(\nabla H_\mu(z_k)\nabla H_\mu(z_k)^T + B_k + \frac{1}{h_k} I)^{\{-1\}}(\nabla H_\mu(z_k)H_\mu(z_k)) \tag{64}$$

$$\geq \frac{\|\nabla H_\mu(z_k)H_\mu(z_k)\|^2}{\|\nabla H_\mu(z_k)\nabla H_\mu(z_k)^T + B_k + \frac{1}{h_k} I\|} \tag{65}$$

$$\geq \frac{\varepsilon^2}{C^2 + M + \frac{1}{h}} \tag{66}$$

$$> 0 \tag{67}$$

which contradicts to the (63)),Hence the desired result follows.

Lemma 4.9 If there exists an infinite sequence of points is accepted by the filter ($|A| = \infty$), and finitely many points added to the filter ($|I| < \infty$) ,it holds

$$\liminf_{k \to \infty} \|\nabla H_\mu(z_k)H_\mu(z_k)\| = 0 \tag{68}$$

By Algorithm ,it is similar to Lemma 4.8.

Lemma 4.10 Suppose there exist infinitely many points added to the filter ($|I| = \infty$). Then

$$\liminf_{k \to \infty} \|\nabla H_\mu(z_k)H_\mu(z_k)\| = 0. \tag{69}$$

Proof Suppose by contradiction that there exists a constant $\epsilon > 0$ such that $\|\nabla H_\mu(z_k)H_\mu(z_k)\| \geq \epsilon$. Consider the iteration in I. Suppose there exists a subsequence $\{k_i\}$ such that $I = \{k_i\}$, then $z_{k_i} = z_{k_i-1}^+$. It follows that there exists a subsequence $\{k_l\}$ includes in $\{k_i\}$ such that

$$\liminf_{l \to \infty}(\nabla H_{\mu(z_{k_l})} H_{\mu(z_{k_l})}) = \nabla H_\mu(z_\infty)H_\mu(z_\infty) and \|\nabla H_\mu(z_\infty)H_\mu(z_\infty)\| \geq \varepsilon \tag{70}$$

By the definition of $\{k_i\}$, z_{k_l} is accepted by the filter for $\forall l$. Then as l increases infinitely, it holds

$$h_{k_l} \leq h_{k_l-1} - \gamma \min\{\|H_{\mu k_l}\|, H_{\mu(k_l-1)}\} or \theta_{k_l} \leq \theta_{k_l-1} - \gamma \min\{\|H_{\mu k_l}\|, H_{\mu(k_l-1)}\}. \tag{71}$$

By the assumptions, there exists a number $\overline{\delta} > 0$ such that $\min\{\|H_{\mu k_l}\|, H_{\mu(k_l-1)}\} \geq \overline{\delta}$. Then

$$h_{k_l} - h_{k_l-1} \leq -\gamma\overline{\delta} < 0 or \theta_{k_l} - \theta_{k_l-1} \leq -\gamma\overline{\delta} < 0. \tag{72}$$

By (70), the left of inequality tends to 0, it is a contradiction. So we have

$$\liminf_{i\to\infty} \|\nabla H_{\mu(z_{k_i})} H_{\mu(z_{k_i})}\| = 0. \tag{73}$$

Now consider l does not include in $\{k_i\}$, let $\{k_{i(l)}\}$) be the last iteration before l make $z_{k_{i(l)}}$ add to the filter.By algorithm, $r_k \geq \eta$ for l does not include in $\{k_i\}$. Then it holds

$$\|\nabla H_{\mu(z_{k_l})} H_{\mu(z_{k_l})}\| \leq \|\nabla H_{\mu(z_{k_{i(l)}})} H_{\mu(z_{k_{i(l)}})}\|. \tag{74}$$

together with (73) and $k_{i(l)} \in \{k_i\}$, the desired conclusion follows.

By lemma 4.7 and lemma 4.10, we obtain the convergence conclusion.

Theorem 4.11 Suppose that Assumptions hold, and $\nabla H_\mu(z)$ is nonsingular for all $z \in S$. Then the sequence $\{z_k\}$ generated by the algorithm satisfies two cases as following,

1) Iteration terminated at the KKT point of the original problem (NLP).

2) Every accumulation point is a KKT point of the original problem (NLP).

Theorem 4.12 Suppose $z_k \to z^*, h_k \to \infty$, and $\nabla H_\mu(z_k) H_\mu(z_k)$ is semi-smooth in ∇z^*.there exist constants μ_1, μ_2 for $\nabla\forall k$ and $V_k \in \partial(\nabla H_\mu(z_k) H_\mu(z_k))$ it holds

$$d^T(\theta_k + \frac{1}{h_k})d \geq \mu_1 \|d\|^2 and \|V_k - Q_k\| \leq \frac{\mu_2}{h_k} \tag{75}$$

where $Q_k = \nabla H_\mu(z_k) \nabla H_\mu(z_k)^T + B_k$. If $\forall z \in S$ and $\nabla H_\mu(z)$ is nonsingular, then

1) z^* is the KKT point of the original problem.

2) $\{z_k\}$ converges to z^* superlinearly.

Proof The former part of the theorem follows the Theorem 4.11. Now we turn to prove the second part of the theorem.$\nabla H_\mu(z_k) H_\mu(z_k)$ is semi-smooth in z^*,by Lemma 2.1, as h decreases infinitely, it holds

$$\nabla H_\mu(z_k + h) H_\mu(z_k + h) = \nabla H_\mu(z_k) H_\mu(z_k) + V_k + o(\|z^* - z_k\|) \tag{76}$$

Based on that $z_k \to z^*$ as k increase infinitely, it holds

$$\nabla H_\mu(z^*) H_\mu(z^*) = \nabla H_\mu(z_k) H_\mu(z_k) + V_k(z_k - z^*) + o(\|z^* - z_k\|). \tag{77}$$

And z^* is KKT point of the original problem, that is $H_\mu(z^*) = 0$, then we have

$$\nabla H_\mu(z_k)H_\mu(z_k) = V_k(z_k - z^*) + o(\|z_k - z^*\|) \tag{78}$$

By (4.21)-(4.23) and $h_k \to \infty$, we obtain $\|z_{k+1} - z^*\| = \|z_k + d_k - z^*\|$

$$= \|z_k - z^* - (Q_k + \frac{1}{h_k}I)^{-1}\nabla H_\mu(z_k)H_\mu(z_k)\| \tag{79}$$

$$\geq \|(Q_k + \frac{1}{h_k}I)^{-1}\|\|(Q_k + \frac{1}{h_k}I)(z_k - z^*) - \nabla H_\mu(z_k)H_\mu(z_k)\| \tag{80}$$

$$\geq \|(Q_k + \frac{1}{h_k}I)^{-1}\|(\|Q_k(z_k - z^*) - \nabla H_\mu(z_k)H_\mu(z_k)\| + \frac{\|z_k - z^*\|}{h_k}) \tag{81}$$

$$\geq \|(Q_k+\frac{1}{h_k}I)^{-1}\|(\|V_k(z_k-z^*)-\nabla H_\mu(z_k)H_\mu(z_k)\|+\|V_k-Q_k\|\|z_k-z^*\|+\frac{\|z_k - z^*\|}{h_k}) \tag{82}$$

$$= o(\|z_k - z^*\|) \tag{83}$$

which yields the desired conclusion.

5 Conflict of interest

There has not any conflict of interest in this paper.

References

[1] Fletcher R, Leyffer S. Nonlinear programming without a penalty function. Mathematical Programming ,2002, 91(2): 239-269.

[2] A A Brown, M C Biggs. Some effective methods for unconstrained optimization based on the solution of system of ordinary differentiable equations[J] . JOTA, 1989, 62:211- 224

[3] Gould N I M, Leyffer S, Toint Ph L. A multidimensional filter algorithm for nonlinear equation and nonlinear least-aquares. SIAM Journal on Control and Optimization, 2004, 15(1): 17-38.

[4] Y X Yuan, W Y Sun. Optimization Theory and Methods. Science Press, Beijing, China, 1997.

[5] Sun W, Sampaio R J B, Yuan J. QuasiNewton trust region algorithm for nonlinear least aquares problems. Applied Mathematics and Computation, 1999, 105(2): 183-194.

[6] Y G Ou. A trust region method for constrained optimization problems with nonlinear inequalities. Applied Mathematics, 2006, 19(1): 80-85.

[7] Gould N I M, Sainvitu, Toint Ph L. A filter trust-region method for unconstrained optimization. SIAM Journal on Optimization, 2006, 16(2): 341-357.

[8] B Chen, X Chen, C A Kancow. Fischer-Bufmeister NCP-function. Mathematical Programming, 2000, 88(1): 211-216.

[9] To ng Xiaojiao , Zhou Shuzi. A trust-region algorithm for inequality constrained optimization, J.Compu.Math., 2003, 21(2):207-220.

[10] Y G Ou, D P Hou. A supearlinearly convergent ODE-type trust region algorithm for LC^1optimization n problems. China Quart Journal Mathematics, 2003, 18(2): 140-145.

[11] Hock W, Schittknowskik, Test examples for nonlinear programming codes, Beakma M, Kunzi H P. Lecture Notes in Economics and Mathematical Systems 187 Berlin: SpringerVerlag, 1981.

[12] Kanzow C , Some noninterior continuation methods for linear complementarity problems . SIAM J Matrix Anal Appl , 1996 ,17: 851-868 .

Received: June, 2016