

A Nonmonotone Filter Method for Minimax Problems*

Qi Zhao, Nan Guo

Department of Basic, Jiangsu University of Science and Technology, Zhangjiagang, China

Department of Basic, Nanjing Institute of Technology, Nanjing, China

E-mail: johnzqzq@163.com

Received September 18, 2011; revised October 21, 2011; accepted October 28, 2011

Abstract

In this paper, we propose a modified trust-region filter method algorithm for Minimax problems, which based on the framework of SQP-filter method and associated with the technique of nonmonotone method. We use the SQP subproblem to acquire an attempt step, and use the filter to weigh the effect of the attempt step so as to avoid using penalty function. The algorithm uses the Lagrange function as a merit function and the nonmonotone filter to improve the effect of the algorithm. Under some mild conditions, we prove the global convergence.

Keywords: Minimax Problem, Nonmonotone, Global Convergence, Filter Methods

1. Introduction

Consider the following Minimax problem:

$$\min_{x \in R} \max_{1 \leq i \leq m} f_i(x) \quad (1)$$

where $f_i(x) : R^n \rightarrow R$ is a twice continuously differentiable function.

The problem (1) can be transformed into the following problem below:

$$\begin{aligned} \min t \\ s.t. f_i(x) - t \leq 0 \end{aligned} \quad (2)$$

The Minimax problem is one of the most important non-differentiable optimization problems. It does not only have broader applications in engineering design, electronic microcircuits programming, game theory and so on, but also has very close relationship with nonlinear equations, multi-object programming, nonlinear programming, etc. There are some methods e.g., line search method SQP method, trust region method and the active-set method, for solving Minimax problems. C. Charalambous and A.R. Conn [1] proposed the line search method. A. Vardi [2] presented the trust region method with the active-set methods. There are many other effective algorithms, see Z. B. Zhu [3], L. Gao [4], J. L. Zhou [5], Y. Xue [6].

Recently, the filter method for nonlinear programming has broader applications and good numerical effects, see [7-12]. The major filter methods are of two kinds: line search and trust-region methods.

R. Fletcher proposed the global convergent SQP-filter trust-region method [9], based on this idea, Huang [13] proposed a filter method for Minimax problems. In [14], Ulbrich S. used the Lagrange function to replace the function and gave the local superlinear convergence proof of the SQP-filter trust-region method.

The nonmonotone technique can improve the effect of the algorithm, relax the accept criteria of the attempt step. Recently, Su [15] and Shen [16] presented the idea of using nonmonotone filter methods for nonlinear programming. Motivated by their idea, we present a modified filter-method for Minimax problems. The algorithm uses the Lagrangian function instead of the function itself as a merit function, and combines it with a nonmonotone filter technique to improve the effect of the algorithm.

Consider the SQP subproblem of problem (2):

$$\begin{aligned} \min q_k(s) &= s^t + \frac{1}{2} s^t H_k s \\ s.t. f_i(x) + \begin{pmatrix} \nabla f_i(x) \\ -1 \end{pmatrix} s &\leq t \\ \|s\| &\leq \Delta \end{aligned} \quad (3)$$

We use the following notations:

$$s = (s^x, s^t) \in R^{n+1}, I = \{1, 2, \dots, m\}, e = (1, 1, \dots, 1)^T$$

$$h(x, t) = f(x) - te, f(x) = (f_1(x), f_2(x), \dots, f_m(x))^T$$

$H_k \in R^{(n+1) \times (n+1)}$ is a symmetric matrix, and it is the approximate Hessian matrix of the subproblem (3).

*This work was supported by Chinese NSF Grant 10201026.

Remark: H_k is updated by the Powell's safeguard BFGS update formula.

This paper is organized as follows. The new algorithm is described in Section 2. Basic assumptions and some important lemmas are given in Section 3. The analysis of the global convergence is given in Sections 4 and 5.

2. Algorithm

Now we introduce some definitions about the filter used in this paper.

Definition 1: [14]

Lagrange function:

$$l(x, t, \lambda) = t - \lambda^T h(x, t),$$

Constrain violation function:

$$\theta(x, \lambda, t) = \|h(x, t)^+\|_2^2 + (\lambda^T h(x, t))^2$$

For simplicity, we just use the following notations:

$$\theta(x, t, \lambda) = \theta, l(x, t, \lambda) = l,$$

$$\theta(x_j, t_j, \lambda_j) = \theta_j, l(x_j, t_j, \lambda_j) = l_j$$

Definition 2:

A pair (θ_k, l_k) obtained on iteration k is said to dominate another pair (θ_l, l_l) if and only if

$$\theta_k < \theta_l \text{ and } l_k < l_l$$

A filter set is a list of pairs (θ_j, l_j) such that no pair dominates any other. We denote the set by \mathcal{F}_k for each iteration k .

Similar to the definition in Fletcher. and Leyffer, S. [9], a point $((x, t), \lambda) + s$ can be accept to $\mathcal{F}_k \cup (\theta_k, l_k)$ if

$$\theta \leq \beta \theta_j, \text{ or } l + \gamma \theta \leq l_j, (\theta_j, l_j) \in \mathcal{F}_k \cup (\theta_k, l_k)$$

Here we use the nonmonotone filter idea in [14] a point $((x, t), \lambda) + s$ can be accept to $\mathcal{F}_k \cup (\theta_k, l_k)$ if

$$\theta \leq \beta \max_{0 \leq r \leq m(k)-1} \theta_{k-r} \text{ or}$$

$$l + \gamma \theta \leq \max \left[l_k, \sum_{r=0}^{m(k)-1} \xi_{k,r} l_{k-r} \right] \quad (4)$$

where $(\theta_{k-r}, l_{k-r}) \in \mathcal{F}_k \cup (\theta_k, l_k)$, $0 < \gamma < \beta < 1$

$m(0) = 0, 0 \leq m(k) \leq \min \{m(k-1) + 1, M_1\}$

$\sum_{r=0}^{m(k)-1} \xi_{k,r} = 1, \xi_{k,r} \in (0, 1), \xi_{k,r} \geq \bar{\xi} > 0, M_1 \geq 1$

Update of the filter set

If (θ_k, l_k) is added to the filter, then the new filter set is updated as follows:

$$\mathcal{F}_{k+1} := (\theta_k, l_k) \cup \left\{ (\theta_j, l_j) \in \mathcal{F}_k, \min \{ \theta_j - \theta_k, l_j - l_k \} < 0 \right\}$$

Definition 3: [14]

$$pred_k(s) = -q_k(s),$$

$$rared_k^l(s) = \max \left[l_k, \sum_{r=0}^{m(k)-1} \xi_{k,r} l_{k-r} \right] - l(x_k + s^x, t_k + s^t, \lambda_k(s))$$

$$ared_k^l(s) = l_k - l(x_k + s^x, t_k + s^t, \lambda_k(s)),$$

where $\lambda_k(s)$ is the Lagrange multiplier of the subproblem (3).

In order to improve both the feasibility and optimality if

$$\widehat{pred}_k(s) := pred_k(s) + \lambda_k^T h_k \geq \kappa_\theta \theta_k^{1/2}, \kappa_\theta > 0 \quad (5)$$

then we require

$$rared_k^l(s) \geq \widehat{pred}_k(s) \quad (6)$$

and call it is a f-type iteration.

If

$$\widehat{pred}_k(s_k) < \kappa_\theta \theta_k^{1/2} \quad (7)$$

then we add (θ_k, l_k) to the filter set and update the filter set, calling it is a h-type iteration ($k \in \mathcal{H}$).

If the subproblem (3) is not compatible or

$$\theta_k^{1/2} > \kappa_\Delta \Delta^{1+\zeta}, \kappa_\Delta > 0, \zeta \in (0, 1]$$

we also call it is a h-type iteration ($k \in \mathcal{H}$).

Now we describe the detailed algorithm below:

Step 0 (Initialization)

Give $\beta, \zeta, \sigma, \gamma, \varepsilon \in (0, 1), \kappa_\Delta, \kappa_\theta > 0,$

$\Delta \geq \Delta_{\min} \geq 0, k := 0,$ set the initial filter set \mathcal{F}_0 .

Step 1: Solve subproblem (3) to get an attempt step s .

Step 2: If the solution of (3) is not compatible or

$$\theta_k^{1/2} > \kappa_\Delta \Delta^{1+\zeta}, \zeta \in (0, 1]$$

Then add (θ_k, l_k) to the filter set and update the filter set, let: $t_k = \max_{1 \leq i \leq m} f_i(x_k), (x_{k+1}, t_{k+1}) = (x_k, t_k)$ go to Step 1;

If $\|s\| \leq \varepsilon$, stop; else go to step 3.

Step 3: If (4) fails, then $\Delta := 0.5\Delta$, go to Step 1; else go to Step 4.

Step 4: If (5) holds but (6) fails, then $\Delta := 0.5\Delta$, go to step 1, else go to Step 6.

Step 5: If (5) fails, add (θ_k, l_k) to the filter set and update the filter \mathcal{F}_k to \mathcal{F}_{k+1} , else go to Step 6.

Step 6: $s_k := s, x_{k+1} := x_k + s^x, t_{k+1} := t_k + s^t, \lambda_{k+1} = \lambda_k(s)$ update H_k to $H_{k+1}, \Delta \geq \Delta_{\min}. k := k + 1,$ go to Step 1.

Remark. Step 2 to Step 5 are called inner circle iteration steps, while Step 1, Step 6 are called outer circle steps.

3. Basic Assumptions and Lemmas

First we make the following assumptions:

A1: All iterations (x_k, t_k) remain in a close and

bounded convex set Ω .

A2: The functions $f_i(x)$ are twice continuously differentiable.

A3: The matrix H_k is bounded, $\|H_k\| \leq M_H$.

Lemma 1: If the solution of the subproblem (3) is $s = 0$, then (x_k, t_k) is the KKT point of problem (2).

Proof: It is not difficult to get the result by using the first order necessary optimality conditions.

Lemma 2: All the elements (θ_j, l_j) in the filter set \mathcal{F}_k satisfy $\theta_j > 0$.

Proof. Suppose, by contradiction, that the result is not true, then there must exist a $(\theta_j, l_j) \in \mathcal{F}_k$ and $\theta_j = 0$, which means (x_j, t_j) is a feasible solution of the problem (2), and $\theta_j^{\frac{1}{2}} \leq \kappa_\Delta \Delta^{1+\zeta}, \lambda_j^T h_j = 0$; but $s_j \neq 0$ so $\widehat{pred}_j(s_j) = pred_j(s_j) > 0 = \kappa_\theta \theta_j^{1/2}$, based on the update mechanism of the filter set, (θ_j, l_j) can't be added to the filter set, which is a contradiction. Thus, the Lemma is proven.

Lemma 3: If $\{l_k\}$ is bounded below and one of the following holds when k is sufficiently large enough:

$$\theta_{k+1} \leq \beta \max_{0 \leq r \leq m(k)-1} \theta_{k-r}$$

$$l_{k+1} + \gamma \theta_{k+1} \leq \max \left[l_k, \sum_{r=0}^{m(k)-1} \xi_{k,r} l_{k-r} \right]$$

then $\lim_{k \rightarrow \infty} \theta_k = 0$.

Proof. We consider the following two cases:

Case 1: $\theta_{k+1} \leq \beta \max_{0 \leq r \leq m(k)-1} \theta_{k-r}$,

for k sufficiently large enough.

Let $\theta(l(k)) = \max_{0 \leq r \leq m(k)-1} \theta_{k-r}, (k - m(k) + 1 \leq l(k) \leq k)$

then

$$\begin{aligned} \theta(l(k+1)) &= \max_{0 \leq r \leq m(k+1)-1} \theta_{k+1-r} \leq \max_{0 \leq r \leq m(k)} \theta_{k+1-r} \\ &= \max(\theta(l(k)), \theta_{k+1}) = \theta(l(k)) \end{aligned}$$

which implies $\theta(l(k))$ converges.

By $\theta(l(k)) \leq \beta \theta(l(k-1)), \beta \in (0, 1)$ we get

$$\theta(l(k)) \rightarrow 0, k \rightarrow \infty$$

So $\theta_{k+1} \leq \beta \theta(l(k)) \rightarrow 0$.

Case 2: $l_{k+1} + \gamma \theta_{k+1} \leq \max \left[l_k, \sum_{r=0}^{m(k)-1} \xi_{k,r} l_{k-r} \right]$,

for k sufficiently large enough.

First we will show that, for all $k \geq 1$

$$l_k \leq l_0 - \xi \gamma \sum_{r=1}^{k-1} \theta_r - \gamma \theta_k \leq l_0 - \xi \gamma \sum_{r=1}^k \theta_r \quad (8)$$

We prove (8) by induction.

When $k = 1$, $l_1 \leq l_0 - \gamma \theta_1 \leq l_0 - \gamma \xi \theta_1$, assume that (8) holds for $1, 2, \dots, k$, we want to prove it also holds for $k + 1$.

We consider the following two cases:

$$(a) \max \left[l_k, \sum_{r=0}^{m(k)-1} \xi_{k,r} l_{k-r} \right] = l_k$$

$$l_{k+1} \leq l_k - \gamma \theta_{k+1} \leq l_0 - \xi \gamma \sum_{r=1}^k \theta_r - \gamma \theta_{k+1}$$

$$(b) \max \left[l_k, \sum_{r=0}^{m(k)-1} \xi_{k,r} l_{k-r} \right] = \sum_{r=0}^{m(k)-1} \xi_{k,r} l_{k-r}$$

Let $p = m(k) - 1$, then

$$\begin{aligned} l_{k+1} &\leq \sum_{t=0}^p \xi_{k,t} l_{k-t} - \gamma \theta_{k+1} \\ &\leq \sum_{t=0}^p \xi_{k,t} \left(l_0 - \xi \gamma \sum_{r=1}^{k-t-1} \theta_r - \gamma \theta_{k-t} \right) - \gamma \theta_{k+1} \\ &= \xi_{k,0} \left(l_0 - \xi \gamma \sum_{r=1}^{k-p-1} \theta_r - \xi \gamma \sum_{r=k-p}^{k-1} \theta_r - \gamma \theta_k \right) - \gamma \theta_{k+1} \\ &\quad + \xi_{k,1} \left(l_0 - \xi \gamma \sum_{r=1}^{k-p-1} \theta_r - \xi \gamma \sum_{r=k-p}^{k-2} \theta_r - \gamma \theta_{k-1} \right) \\ &\quad + \dots + \xi_{k,p} \left(l_0 - \xi \gamma \sum_{r=1}^{k-p-1} \theta_r - \gamma \theta_{k-p} \right) \\ &\leq \sum_{t=0}^p \xi_{k,t} l_0 - \xi \gamma \sum_{r=1}^{k-p-1} \left(\sum_{t=0}^p \xi_{k,t} \right) \theta_r \\ &\quad - \sum_{t=0}^p \gamma \xi_{k,t} \theta_{k-t} - \gamma \theta_{k+1} \end{aligned}$$

By the fact that $\sum_{t=0}^p \xi_{k,r} = 1, \xi_{k,r} \geq \xi, \theta_r \geq 0$

$$\begin{aligned} l_{k+1} &\leq l_0 - \xi \gamma \sum_{r=1}^{k-p-1} \theta_r - \xi \gamma \sum_{r=k-p}^k \theta_r - \gamma \theta_{k+1} \\ &= l_0 - \xi \gamma \sum_{r=1}^k \theta_r - \gamma \theta_{k+1} \\ &\leq l_0 - \xi \gamma \sum_{r=1}^{k+1} \theta_r \sum_{r=1}^{k+1} \theta_r \end{aligned}$$

So (8) is true. Since $\{l_k\}$ is bounded below, so

$$\xi \gamma \sum_{r=1}^{\infty} \theta_r < \infty, \theta_k \rightarrow 0, k \rightarrow \infty$$

Thus the Lemma is proven.

Some assumptions are needed for Lagrange multiplier estimates $\lambda_k(s)$.

A4: There exists constants $M_y, M_L > 0$, all of the Lagrange multiplier estimates $\lambda_k(s)$ satisfies:

$$0 \leq \lambda_k(s)_i \leq M_y, i = 1, 2, \dots, m$$

$$\left| (\lambda_k(s))_i [h(x_k, t_k) + s] \right| \leq M_L \Delta^2$$

Lemma 4: Under the assumptions A1-A4, we have

$$\left\| h((x_k, t_k) + s)^+ \right\|_2^2 \leq \frac{1}{4} m(n+1)^2 M_f^2 \Delta^4 \quad (9)$$

$$\begin{aligned} \theta((x_k, t_k) + s, \lambda_k(s)) &\leq \frac{1}{4} m(n+1)^2 M_f^2 \Delta^4 \\ &\quad + m^2 M_L^2 \Delta^4 := M_\theta \Delta^4 \end{aligned} \quad (10)$$

Proof. Use the Taylor Expansion:

$$\begin{aligned} f_i(x_k + s^x) - t_k - s^t &= -t_k + f_i(x_k) + \begin{pmatrix} \nabla f_i(x_k) \\ -1 \end{pmatrix}^T s \\ &\quad + \frac{1}{2} s^{xT} \nabla^2 f_i(y_k) s^x \leq \frac{1}{2} (n+1) M_f \Delta^2 \end{aligned}$$

which implies (9) holds.

$$\begin{aligned} \theta((x_k, t_k) + s, \lambda_k(s)) &= \left\| h((x_k, t_k) + s)^+ \right\|_2^2 \\ &+ (\lambda_k(s)^T (h((x_k, t_k) + s)))^2 \leq M_\theta \Delta^4 \end{aligned}$$

where $M_\theta = m^2 M_L^2 + \frac{1}{4} m(n+1)^2 M_f^2$ which implies (10) holds. Thus, the Lemma is proven.

4. The Well Definedness of the Algorithm

Lemma 5: Let $\tau_k = \min_{(\theta_j, l_j) \in \mathcal{F}_k} \theta_j$, when $\Delta^4 \leq \beta \tau_k / M_\theta$,

$((x_k, t_k) + s, \lambda_k(s))$ can be accept to \mathcal{F}_k .

Proof. From the definition of τ_k , Use the result of (10), when $\Delta^4 \leq \beta \tau_k / M_\theta$, we have

$$\theta((x_k, t_k) + s, \lambda_k(s)) \leq \beta \max_{0 \leq r \leq m(k)-1} \theta_{k-r}, (\theta_{k-r}, l_{k-r}) \in \mathcal{F}_k,$$

so $((x_k, t_k) + s, \lambda_k(s))$ can be accepted to \mathcal{F}_k .

Thus, the Lemma is proven.

Lemma 6: Under the assumptions A1-A4, if $(\tilde{x}, \tilde{t}) \in \Omega$ is a feasible point, but not a KKT point of the problem (2) then there must exit a neighborhood \tilde{N} of (x_*, t_*) , and constants $\varepsilon, \hat{\mu}, \hat{\kappa} > 0$ for all $(x_k, t_k) \in \tilde{N} \cap \Omega$, and all the Δ satisfying (11), the subproblem must have a solution, and

$$\frac{1}{\theta_k^2} \leq \kappa_\Delta \Delta^{1+\zeta},$$

$$\hat{\mu}(\theta_k^{1/2+2\zeta} + \theta_k^{1/2}) \leq \Delta \leq \hat{\kappa} \tag{11}$$

$$\widehat{pred}_k(s) \geq \frac{1}{3} \varepsilon \Delta + \kappa_\theta \theta_k^{1/2} \tag{12}$$

$$rared_k^l(s) \geq \sigma \widehat{pred}_k(s) \tag{13}$$

$$rared_k^l(s) \geq \gamma \theta((x_k, t_k) + s, \lambda(s)) \tag{14}$$

Proof. From the results of Lemma7 in [13], there must exit a neighborhood \tilde{N} of (x_*, t_*) , and some constants $\varepsilon, \mu, \kappa \geq 0$, for all $(x_k, t_k) \in \tilde{N} \cap \Omega$ and

$$\mu \left\| h(x_k, t_k)^+ \right\|_2 \leq \Delta \leq \kappa,$$

the subproblem (3) has a solution, and

$$pred_k(s) \geq \frac{1}{2} \Delta \varepsilon, ared_k(s) \geq \sigma pred_k(s), \tag{15}$$

$$ared_k(s) \geq \gamma \left\| h((x_k, t_k) + s)^+ \right\|_2$$

next we will prove, there must exit $\hat{\kappa}, \hat{\mu}$, when (11)

holds, then $\frac{1}{\theta_k^2} \leq \kappa_\Delta \Delta^{1+\zeta}$ and (12)-(14) holds.

If $\frac{1}{\theta_k^2} \leq \kappa_\Delta \Delta^{1+\zeta}$, then $\Delta \geq \kappa_\Delta^{-\frac{1}{1+\zeta}} \theta_k^{\frac{1}{2+2\zeta}}$, from

$\left| \lambda_k^T h(x_k, t_k) \right| \leq \theta_k^{1/2}$ and the results of (15) we can deduce

$$\begin{aligned} pred_k(s) + \lambda_k^T h(x_k, t_k) &\geq pred_k(s) - \theta_k^{1/2} \\ &\geq \frac{1}{3} \Delta \varepsilon + \frac{1}{6} \Delta \varepsilon - \theta_k^{1/2} \end{aligned}$$

If $\Delta \geq 6 \frac{(\kappa_\theta + 1) \theta_k^{1/2}}{\varepsilon}$ then

$$pred_k(s) + \lambda_k^T h(x_k, t_k) \geq \frac{1}{3} \Delta \varepsilon + \kappa_\theta \theta_k^{1/2}$$

So if we take

$$\hat{\mu} = \max \left\{ \mu, \kappa_\Delta^{-\frac{1}{1+\zeta}}, 6 \frac{\kappa_\theta + 1}{\varepsilon} \right\}$$

then $\widehat{pred}_k(s) \geq \frac{1}{3} \varepsilon \Delta + \kappa_\theta \theta_k^{1/2}$, (12) holds.

From the definition of $rared_k^l(s)$

$$\begin{aligned} rared_k^l(s) &\geq ared_k(s) + \lambda_k^T h_k - \lambda_k(s)^T h((x_k, t_k) + s) \\ &\geq ared_k(s) + \lambda_k^T h_k - \lambda_k(s)^T h((x_k, t_k) + s)^+ \\ &\geq ared_k(s) + \lambda_k^T h_k - M_y \frac{1}{2} m(n+1) M_f \Delta^2 \end{aligned}$$

and by Taylor Expansion we know

$$\left| ared_k(s) - pred_k(s) \right| \leq \frac{1}{2} (n+1) M_H \Delta^2.$$

So we can deduce

$$\begin{aligned} rared_k^l(s) &\geq ared_k(s) + \lambda_k^T h_k \\ &- \left(\frac{1}{2} (n+1) M_H + M_y \frac{1}{2} m(n+1) M_f \right) \Delta^2 \\ &\geq \sigma \widehat{pred}_k(s) + (1-\sigma) \widehat{pred}_k(s) \\ &- \left(\frac{1}{2} (n+1) M_H + M_y \frac{1}{2} m(n+1) M_f \right) \Delta \\ &\geq \sigma \widehat{pred}_k(s) + (1-\sigma) \frac{1}{3} \Delta \varepsilon \\ &- \left(\frac{1}{2} (n+1) M_H + M_y \frac{1}{2} m(n+1) M_f \right) \Delta^2 \end{aligned}$$

By simple calculation we know that if

$$\Delta \leq \frac{(1-\sigma) \frac{1}{3} \Delta \varepsilon}{\frac{1}{2} (n+1) M_H + M_y \frac{1}{2} m(n+1) M_f} := \kappa_1$$

then $rared_k^l(s) \geq \sigma \widehat{pred}_k(s)$

$$\begin{aligned} rared_k^l(s) - \gamma \theta((x_k, t_k) + s, \lambda_k(s)) &\geq \sigma \frac{1}{3} \Delta \varepsilon - \gamma M_\theta \Delta^4 \end{aligned}$$

So if we take

$$\Delta \leq \left(\frac{\sigma \frac{1}{3} \epsilon}{\gamma M_\theta} \right)^{\frac{1}{3}} := \kappa_2$$

then $rared_k^l(s) \geq \gamma \theta((x_k, t_k) + s, \lambda_k(s))$.

From the discuss above we know if we take $\hat{\kappa} = \max\{\kappa_1, \kappa_2\}$ then (13)-(14) holds. Thus, the Lemma is proven.

Lemma 7: Under the assumptions A1-A4, the inner iteration terminates finitely.

Proof. If $s = 0$, the algorithm terminates at a KKT point, otherwise, if the inner iteration does not terminate finitely, then the rule for decreasing Δ ensures $\Delta \rightarrow 0$.

We consider the following two cases:

Case 1: $\theta_k > 0$

When $\Delta \rightarrow 0$, $\theta_k^{1/2} \leq \kappa_\Delta \Delta^{1+\zeta}$ cannot hold, so the algorithm will enter a restoration phrase at step 2 and terminates finitely.

Case 2: $\theta_k = 0$ This means (x_k, t_k) is a feasible point.

When $\Delta \rightarrow 0$, if Δ satisfies

$$0 = \hat{\mu}(\theta_k^{1/2+2\zeta} + \theta_k^{1/2}) \leq \Delta \leq \min\left\{\hat{\kappa}, (\beta\tau_k/M_\theta)^{\frac{1}{4}}\right\}$$

From the result of Lemma 6, the subproblem (3) must be compatible and (12)-(14) holds. From Lemma 5 and (14) we know that $(x_k, t_k) + s$ can be accept to

$\mathcal{F}_k \cup (\theta_k, l_k)$. So all the conditions for f-type step are satisfied and the inner iteration terminates successfully.

5. Global Convergence

Lemma 8: Under the assumptions A1-A4, if the algorithm does not terminate finitely, and there are infinite points added to the filter set, then $\lim_{k \in \mathcal{H}} \theta_k \rightarrow 0$

Proof. Let

$$\max_{0 \leq r \leq m(k)-1} \theta_{k-r} = \tilde{\theta}_k, \max\left[l_k, \sum_{t=0}^{m(k)-1} \xi_{k,r} l_{k-r}\right] = \tilde{l}_k$$

We consider the following two cases:

Case 1: $\theta_{k+1} \leq \beta \tilde{\theta}_k, k \in \mathcal{H}$ for k sufficiently large, then from the results of Lemma 3, we know $\lim_{k \in \mathcal{H}} \theta_k \rightarrow 0$

Case 2: if Case 1 is not true, define a subset \mathcal{H}_1 as follows: the first element k of \mathcal{H}_1 is the first index which satisfies $\theta_j > \beta \tilde{\theta}_k$, and the next element k^+ is the index which satisfies $\theta_j > \beta \tilde{\theta}_k, j > k$. From the definition of the filter set, we can deduce that:

$$\tilde{l}_k - l_{k^+} \geq \gamma \theta_{k^+}, k \in \mathcal{H}_1$$

From the proof of the Lemma 3, Case 2, we know that

$$\lim \theta_k \rightarrow 0, k \in \mathcal{H}_1.$$

For $k < j < k^+$, it is obviously $\theta_j \leq \beta \theta_k \rightarrow 0$.

Thus, the Lemma is proven.

Theorem 1: Under the assumptions A1-A4, if the algorithm-m doesn't terminate finitely, then there must exit an accumulation point which is a KKT point.

Proof. We discuss it in two cases:

1) Case 1: there are infinite h-type iterations.

From Lemma 8 we know $\lim_{k \in \mathcal{H}} \theta_k \rightarrow 0$, from the update mechanism of the filter set, there must exit a subset $S \subseteq \mathcal{H}$, $\tau_{k+1} = \theta_k < \tau_k, k \in S$.

Without loss of generality, we can assume that

$$\lim_{k \in S} (x_k, t_k) \rightarrow (x_*, t_*)$$

from which we know (x_*, t_*) is a feasible point.

If, suppose, by contradiction, (x_*, t_*) is not a KKT point, then from Lemmas 6 and 7, we know that if:

$$\hat{\mu}(\theta_k^{1/2+2\zeta} + \theta_k^{1/2}) \leq \Delta \leq \min\left\{\hat{\kappa}, (\beta\tau_k/M_\theta)^{\frac{1}{4}}\right\} := \hat{\Delta}. \quad (15)$$

Then all the conditions for f-type step are satisfied and the inner iteration terminates successfully.

Because $\theta_k < \tau_k \rightarrow 0$, the upper bound $\hat{\Delta}$ will be greater than twice the lower bound $\hat{\mu}(\theta_k^{1/2+2\zeta} + \theta_k^{1/2})$.

Initially a value $\Delta \geq \Delta_{min} > \hat{\Delta}$ is chosen, successively halving Δ will eventually (a) locate a value in the interval (15), or (b) locate a value to the right of this interval. It is obviously that $\widehat{pred}_k(s_k) > 0$ under case (a). From the optimality of s , if (b) is true, note that $\widehat{pred}_k(s)$ is nondecreasing if Δ increases, so we know $\widehat{pred}_k(s) > 0$, which means a f-type iteration will occur, and it is a contradiction.

2) Case 2: there are only finite h-type iterations.

That means for $k \geq K$ sufficiently large, no filter entries are made and

$$rared_k^l(s) \geq \sigma \left(pred_k(s) + \lambda_k^T h_k \right)$$

$$rared_k^l(s) \geq \sigma \kappa_\theta \theta_k^{1/2}.$$

Similar to the proof of Lemma 3 we can deduce:

$$l_{k+1} \leq \max_{K-m(K) \leq r \leq K} l_r - \sigma \sum_{r=K}^k \xi^{\min(k-r, M_1)} \kappa_\theta \theta_k^{1/2}$$

$$l_{k+1} \leq \max_{K-m(K) \leq r \leq K} l_r - \gamma \sum_{r=K}^k \xi^{\min(k-r, M)} (pred_r + y_r^T h_r)$$

Because $\{l_k\}$ is bounded below, we know that

$$\widehat{pred}_k(s) = pred_k(s) + \lambda_k^T h_k \rightarrow 0,$$

$$\theta_k \rightarrow 0, k \rightarrow \infty.$$

Without loss of generality, we can assume that

$\lim(x_k, t_k) \rightarrow (x_*, t_*)$ from which we know (x_*, t_*) is a feasible point. If, suppose, on the contrary, (x_*, t_*) is not a KKT point. Note that for $k \geq K$, $\tau_k = \tau_K$ from Lemma 6 and Lemma 7 we know that when

$$\hat{\mu}(\theta_k^{1/2+2\zeta} + \theta_k^{1/2}) \leq \Delta \leq \min \left\{ \hat{\kappa}, (\beta\tau_K/M_\theta)^{\frac{1}{4}} \right\} := \bar{\kappa}$$

then the iteration is f-type and terminates successfully.

Because the right side is a constant, and the left side tends to zero, the upper bound $\hat{\Delta}$ will be greater than twice the lower bound $\hat{\mu}(\theta_k^{1/2+2\zeta} + \theta_k^{1/2})$, and when inner iteration terminates we must have

$$\Delta \geq \min \left(\Delta_{\min}, \frac{1}{2} \hat{\kappa} \right)$$

so $\widehat{pred}_k(s) \geq \frac{1}{3} \min \left(\Delta_{\min}, \frac{1}{2} \hat{\kappa} \right)$, this is a contradiction

to $\widehat{pred}_k(s) = pred_k(s) + \lambda_k^T h_k \rightarrow 0$. Thus, the theorem is proven.

6. References

- [1] C. Charalambous and A. R. Conn, "An Efficient Method to Solve the Minimax Problem Directly," *SIAM Journal on Numerical Analysis*, Vol. 15, No. 1, 1978, pp. 162-187. [doi:10.1137/0715011](https://doi.org/10.1137/0715011)
- [2] A. Vardi, "New Minmax Algorithm," *Journal of Optimization Theory and Applications*, Vol. 75, No. 3, 1992, pp. 613-634. [doi:10.1007/BF00940496](https://doi.org/10.1007/BF00940496)
- [3] Z. B. Zhu, "An Improved SQP Algorithm for Solving Minimax problems," *Applied Mathematics Letters*, Vol. 22, No. 1, 2009, pp. 464-469. [doi:10.1016/j.aml.2008.06.017](https://doi.org/10.1016/j.aml.2008.06.017)
- [4] Y. H. Yu and L. Gao, "Nonmonotone Line Search Algorithm for Constrained Minimax Problems," *Journal of Optimization Theory and Applications*, Vol. 115, No. 2, 2002, pp. 419-446. [doi:10.1023/A:1020896407415](https://doi.org/10.1023/A:1020896407415)
- [5] J. L. Zhou and A. L. Tits, "Nonmonotone Line Search Method for Minimax Problems," *Journal of Optimization Theory and Applications*, Vol. 76, No. 3, 1993, pp. 455-476. [doi:10.1007/BF00939377](https://doi.org/10.1007/BF00939377)
- [6] Y. Xue, "A SQP Method for Minimax Problems," in Chinese, *Journal of System Science and Math Science*, Vol. 22, No. 3, 2002, pp. 355-364.
- [7] R. Fletcher and S. Leyffer, "Nonlinear Programming without a Penalty Function," *Mathematical Programming*, Vol. 91, No. 2, 2002, pp. 239-269. [doi:10.1007/s101070100244](https://doi.org/10.1007/s101070100244)
- [8] R. Fletcher, N. I. M. Gould, S. Leyffer, P. L. Toint and A. Wächter, "Global Convergence of a Trust-Region SQP Filter Algorithm for General Nonlinear Programming," *SIAM Journal on Optimization*, Vol. 13, No. 3, 2002, pp. 635-659. [doi:10.1137/S1052623499357258](https://doi.org/10.1137/S1052623499357258)
- [9] R. Fletcher, S. Leyffer and P. L. Toint, "On the Global Convergence of a Filter-SQP Algorithm," *SIAM Journal on Optimization*, Vol. 13, No. 1, 2002, pp. 44-59. [doi:10.1137/S105262340038081X](https://doi.org/10.1137/S105262340038081X)
- [10] A. Wächter and L. T. Biegler, "Line Search Filter Methods for Nonlinear Programming: Motivation and Global Convergence," *SIAM Journal on Optimization*, Vol. 16, No. 1, 2005, pp. 1-31. [doi:10.1137/S1052623403426556](https://doi.org/10.1137/S1052623403426556)
- [11] A. Wächter and L. T. Biegler, "Line Search Filter Methods for Nonlinear Programming: Local Convergence," *SIAM Journal on Optimization*, Vol. 16, No. 1, 2005, pp. 32-48. [doi:10.1137/S1052623403426544](https://doi.org/10.1137/S1052623403426544)
- [12] R. Fletcher, S. Leyffer and P. L. Toint, "A Brief History of Filter Methods," *SIAG/OPT Views and News*, Vol. 18, No. 1, 2006, pp. 2-12.
- [13] L. P. Huang, "A Filter Method for Minimax Problems," in Chinese, Post Graduate Thesis, Suzhou University, Suzhou, 2009.
- [14] S. Ulbrich, "On the Superlinear Convergence of Trust Region SQP-Filter Algorithm," *Mathematical Programming*, Vol. 100, Series B, 2004, pp. 217-245.
- [15] K. Su and D. G. Pu, "A Nonmonotone Filter Trust Region Method for Nonlinear Constrained Optimization," *Journal of Computational and Applied Mathematics*, Vol. 22, No. 1, 2009, pp. 230-239. [doi:10.1016/j.cam.2008.01.013](https://doi.org/10.1016/j.cam.2008.01.013)
- [16] R. Fletcher, S. Leyffer and C. G. Shen, "Nonmonotone Filter Method for Nonlinear Optimization," Argonne National Laboratory, Lemont, 2009.