

## A TRUST REGION FILTER METHOD FOR MINIMAX PROBLEMS

**XIAOCHUAN LI, RUYUE HOU and KE SU**

College of Mathematics and Information Science  
Hebei University  
Baoding, 071002  
P. R. China  
e-mail: [lxc491524706@163.com](mailto:lxc491524706@163.com)

### Abstract

In this paper, we propose a trust region filter method for minimax problems. Based on the filter technique, the minimax problem is transformed to a constrained optimization problem and solved by the traditional filter idea. In the presented algorithm, the acceptable criterion of the trial points is relaxed, so compared to the existed SQP and Newton-type methods for minimax method, this method is more flexible. Under some suitable conditions, we establish the global convergence of the algorithm.

### 1. Introduction

Consider the following minimax optimization problems:

$$\begin{aligned} \min \quad & f(x) = \max_{1 \leq j \leq m} f_j(x), \\ \text{s.t.} \quad & x \in R^n, \end{aligned} \tag{1.1}$$

where  $f_j(j = 1, \dots, m)$  are continuously differentiable real-valued functions defined on  $R^n$ .

2010 Mathematics Subject Classification: 90C30.

Keywords and phrases: filter method, minimax problem, trust region filter method.

Received February 27, 2017; Revised March 8, 2017

Due to the non-differentiability of the object function  $f(x)$ , the authors use the classical gradient methods directly to solve such optimization problems [1-3, 6]. Many skills have been proposed for solving the problem (1.1), by shifting it to a smooth constrained optimization problem in  $R^{n+1}$  as below:

$$\begin{aligned} \min \quad & z \\ \text{s.t.} \quad & f_j(x) - z \leq 0, \quad j \in I = \{1, \dots, m\}. \end{aligned} \quad (1.2)$$

Obviously, from the problem (1.2), the K-T condition of (1.1) is defined as follows:

$$\begin{aligned} \sum_{j=1}^m \lambda_j \nabla f_j(x) = 0, \quad \sum_{j=1}^m \lambda_j = 1, \\ \lambda_j (f_j(x) - f(x)) = 0, \quad \lambda_j \geq 0, \quad j = 1, \dots, m. \end{aligned} \quad (1.3)$$

There are various methods for solving the minimax problems. For example, Zhu et al. [4] propose an improved SQP method with nonmonotone line search, but its also suffers from the Maratos effects. Wang and Wang [5] present an algorithm with nonmonotone strategy and second-order correction technique for minimax optimization problems, using this algorithm can overcome the difficulties of the Maratos effect occurred in the nonsmooth optimization. As we all know, it is difficult to generate a descent direction for nonsmooth optimization. To deal with the above drawbacks, researcher always choose line search rather than trust region based. Zhou and Tits [7] proposed a nonmonotone SQP line search method with second-order correction for finite minimax problems developed a class of line search algorithms based on a new approximation model to problem (1.1). Typically, under mild assumptions, these algorithms exhibit a locally superlinear rate of convergence.

But the traditional SQP problems is how to select the value function. While the majority of the merit function contains a penalty parameter. As we know, it is difficult to choose an appropriate parameter. Too small a choice of the penalty parameter may result in an infeasible point, or even an unbounded increase in the penalty. On the other hand, too large a choice may result in any monotonic method to be forced to follow the nonlinear constraint manifold very closely. This may lead to much shortened Newton steps and slow convergence. In order to avoid these drawbacks, recently, a filter method was proposed by Fletcher and Leyffer in 2002 [8], which does not require any penalty parameter and has promising numerical results. Filter methods have several advantages over penalty function methods. Firstly, no penalty parameter estimates, which could be difficult to obtain, are required. Secondly, practical experience shows that they exhibit a certain degree of non-monotonicity which may be beneficial. Finally, filter approaches play an important role to balance the objective function and constraints.

Motivated by the above ideas, we propose a trust region filter method for minimax problems, in which the acceptance criterion of trial points is relaxed. More trial points are accepted by the algorithm so that the Maratos effects can be avoid to a certain degree. Under reasonable assumptions, the proposed algorithm is globally convergent.

This paper is organized as follows: in Section 2, we introduce a trust region filter method for minimax problems. The convergent properties are analyzed in Section 3.

## 2. A Trust Region Filter Method for Minimax Problems

In a trust region method, a trial step is obtained by solving some trust region model. Then some merit function is utilized to evaluate the new iterate to decide whether to accept the new point or not. If the new point is accepted, increase the trust region radius and a new quadratic model is formed by a certain method. If the new point is rejected, decrease the trust region radius and recompute. Certainly, there are

many approaches to get a quadratic model at the current point  $x_k$ , as following:

$$\bar{\Psi}_k(d) = g_k^T d + \frac{1}{2} d^T H_k d.$$

The quadratic model to trust region method is given as follows:

$$\begin{aligned} \min \quad & \bar{\Psi}_k(d), \\ \text{s.t.} \quad & \|d\|_\infty \leq \Delta_k. \end{aligned}$$

Moreover, the subproblem is

$$\begin{aligned} \min \quad & \Psi_k(d) = g_k^T d + \frac{1}{2} d^T H_k d, \\ \text{s.t.} \quad & \left\| (C(x_k^i) + A(x_k^i)^T d)^+ \right\|_\infty = 0, \\ & \|d\|_\infty \leq \Delta_k. \end{aligned} \tag{1.4}$$

So it is very difficult to tackle (1.4), we need the approximate without very strong conditions. Hence, compute (1.2) by solving the following quadratic problem:

$$\begin{aligned} \min \quad & z + g(x)^T d + \frac{1}{2} d^T H_k d, \\ \text{QP}(z_k, x_k, H_k) \quad \text{s.t.} \quad & f_j(x_k) - f(x_k) + \nabla f_j(x_k)^T d \leq z, \quad j \in I, \\ & \|d\| \leq \Delta_k, \end{aligned} \tag{1.5}$$

where  $H \in R^{n \times n}$  is an approximate Hessian of the Lagrangian function of the objective function, which is usually obtained by quasi-Newton techniques. We denote

$$g(x) = \nabla f(x), \quad C(x) = (f_1(x) - z, f_2(x) - z, \dots, f_m(x) - z)^T,$$

and

$$A(x) = (\nabla(f_1(x) - z), \nabla(f_2(x) - z), \dots, \nabla(f_m(x) - z)).$$

Clearly, the convex programming  $(z_k, x_k, H_k)$  is feasible. Since  $\bar{d}_k$  is an obvious feasible solution. Furthermore, if  $H_k$  is positive definite, the solution is unique. Let  $d_k$  be the solution of  $QP(z_k, x_k, H_k)$ . The  $d_k$  can be used as the search direction at the current point  $x_k$ .

**Lemma 1** ([9]). *If  $d_k = 0, z_k = 0$  is the solution to  $QP(z_k, x_k, H_k)$ , then  $x_k$  is a KKT point of the problem (1.3).*

The proof is similar to that of Lemma 4.1 in [10].

Now, we turn to solve  $QP(z_k, x_k, H_k)$ , in traditional filter method, originally proposed by Fletcher and Leyffer [8], solving problem (1.2) equivalent to minimize the objective function  $f(x)$  and to satisfy the constraints. To test whether constraints are satisfied or not, we denote the violation function  $h$  as follows:

$$h(x) = \|C(x)^+\|,$$

where  $C_i(x)^+ = \max\{C_i(x), 0\}$ ,  $i \in I$ , the notion  $\|\cdot\|$  denotes the Euclidean norm on  $R^m$ .

It is easy to see that  $h(x) = 0$  if and only if  $x$  is a feasible point ( $h(x) > 0$  if and only if  $x$  is infeasible). The filter will be employed as criterion to accept or to reject a trial step generated by a subproblem.

We aim to reduce both the value of  $P(x)$  and the constraint violation function  $h(x)$ , where

$$P(x_k) = \max f_j(x_k),$$

$$h(x) = \|C(x)^+\| = (\max\{f_j(x_k) - z\})^2.$$

Definition of filter set is based on the definition of dominance as following:

**Definition 1.** A pair  $(h_k, P_k)$  is dominated by  $(h_j, P_j)$  if and only if  $h_k \leq h_j$  and  $P_k \leq P_j$  for each  $j \neq k$ .

**Definition 2.** A filter set  $F$  is a set of pairs  $(h, P)$  such that no pair dominates any other. To ensure the convergence, some additional condition are required to decide whether to accept a trial point to the filter or not. The traditional acceptable criterion is as following.

**Definition 3.** A trial point  $x$  is called acceptable to the filter if and only if

$$\text{either } h(x) \leq \beta h_j \text{ or } P(x) \leq P_j - \gamma h(x) \text{ for } \forall (h_j, P_j) \in F, \quad (1.6)$$

where  $0 < \gamma < \beta < 1$  are constants. In practice,  $\beta$  is close to 1 and  $\gamma$  is close to 0.

**Definition 4.** For any  $x^*, \exists w \in R^n$ , that set up

$$\begin{aligned} \nabla C_E(x^*)^T d &= 0, \\ \nabla C_i(x^*)^T d &> 0, \quad i \in I(x^*). \end{aligned}$$

We defined

$$P_k^+ = \min_{x \in F} \{P(x) : h(x) = 0, i \in F_k\}, \quad (1.7)$$

$$h_k^+ = \min_{x \in F} \{h(x) : h(x) > 0, i \in F_k\}, \quad (1.8)$$

and  $P_k^+$  is the corresponding value to  $h_k^+$  for  $k = 0, 1, 2, \dots$ . If  $x_k + d_k$  is accepted by the filter, then  $x_k^+ = x_k + d_k$  and  $D_{k+1} = \{i | x_i \geq x_{k+1}, i \in F_k\}$ . Filter set is updated as the following rule  $F_{k+1} = F_k \cup \{k+1\} \setminus D_{k+1}$ . The purpose of optimization methods is to reduce both  $h(x)$  and  $P(x)$ . When  $h(x)$  is reduced,  $P(x)$  may be increased. Meanwhile, we hope to avoid  $h(x)$  being very large. When  $h(x)$  is very large, we take steps to control it.

The constraint violation function is defined as follows:

$$h_{k+1} = C(x_k + d_k)^T C(x_k + d_k). \quad (1.9)$$

In fact, in traditional filter method, some good point such as superlinear convergent step may be rejected due to the increase of both objective function value and constraint violation value compared to other entries in filter. That is the reason why the Maratos effect occurs.

We can be more detailed description of the algorithm:

**Algorithm A**

**Step 0.** Choose  $x_0 \in R^n$ ,  $H_0$  is a symmetric and positive definite matrix and an initial region radius  $\Delta_0 \geq \Delta_{\min} > 0$ ,  $\beta \in \left(\frac{1}{2}, 1\right)$ .

**Step 1.**  $QP(z_k, x_k, H_k)$  is it solvable? If the unsolvable, it entered a recovery phase.

**Step 2.** Solve  $QP(z_k, x_k, H_k)$  to get  $d_k, z_k$ . If  $(z_k, d_k) = 0$ , stop, otherwise go to Step 3.

**Step 3.** Let  $x_k^+ = x_k + d_k$ , compute  $h_k^+, P_k^+$ .

**Step 4.** If  $x_k^+$  is acceptable to the filter  $F_k$ , go to Step 5, otherwise go to Step 6.

**Step 5.** Compute  $r_k = \frac{M_k(x_k) - M_k(x_k + d_k)}{\Psi_k(0) - \Psi_k(d_k)} = \frac{ared_k}{pred_k}$ .

If  $r_k = \frac{ared_k}{pred_k} > \eta$ , then update  $\Delta_k^i$ , let  $\Delta_k^{i+1} = 2\Delta_k^i$ ,  $k := k + 1$ , go to Step 1, otherwise go to Step 7.

**Step 6.** Update  $\Delta_k^i$ , let  $\Delta_k^{i+1} = \frac{1}{2}\Delta_k^i$ , go to Step 1.

**Step 7.** Update the filter set. Let  $x_{k+1} = x_k^+$ , update  $H_k$  to  $H_{k+1}$ , go to Step 6.

In a restoration algorithm, it is therefore described to decrease the value of  $h(x)$ . The direct way is utilized Newton method or the similar ways to attack  $C(x_k^+)^+ = 0$ . Let

$$M_k^i(d) = h(x_k^i) - h(x_k^i + d),$$

$$\Psi_k^i(d) = h(x_k^i) - \left\| \left( c(x_k^i) + A(x_k^i)^T d \right)^+ \right\|,$$

and  $r_k^i = \frac{M_k^i(d)}{\Psi_k^i(d)}$ .

**Algorithm B** (Restoration Algorithm)

**Step 0.** Let  $x_k^0 := x_k$ ,  $\Delta_k^0 := \Delta_k$ ,  $i = 0$ ,  $\alpha_1, \alpha_2, \eta, \eta_2 \in (0, 1)$ .

**Step 1.** If  $h(x_k^i) \leq \eta_2 \min\{h_k^+, \alpha_1 \Delta_k^2\}$ , then  $x_k^+ := x_k^i$ , stop.

**Step 2.** Compute

$$\begin{aligned} \min \quad & \Psi_k^i(d), \\ \text{s.t.} \quad & \|d\| \leq \Delta_k^i \end{aligned} \tag{1.10}$$

to get  $d_k^i$ . Calculate  $r_k^i$ .

**Step 3.** If  $r_k^i > \eta$ , then let  $x_k^{i+1} = x_k^i$ ,  $\Delta_k^{i+1} = \Delta_k^i / 2$ ,  $i = i + 1$  and go to Step 2.

**Step 4.** If  $r_k^i > \eta$ , then let  $x_k^{i+1} = x_k^i + d_k^i$ ,  $\Delta_k^{i+1} = 2\Delta_k^i$ ,  $i = i + 1$  and go to Step 1.

The above restoration algorithm is a Newton method for  $C(x_k^+)^+ = 0$ . This method is utilized frequently. Of course, there are other restoration algorithms, the interior point restoration algorithm in [11].

### 3. The Convergence Properties

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

#### Assumption 1

A1: The objective function  $f$  and the constraint functions  $C_i (j \in I)$  are twice continuously differentiable.

A2: The iterate  $\{x_k\}$  remains in a closed, bounded subset  $S \subset R^n$ .

A3: The matrix sequence  $\{H_k\}$  is uniformly bounded.

A4: When solving (1.10), we have

$$\Psi_k^i(d) = h(x_k^{l(i)}) - \left\| \left( C(x_k^i) + A(x_k^i)^T d \right)^+ \right\| \geq \beta_2 \min \{ h(x_k^i), \Delta_k^i \},$$

where  $\beta_2 > 0$  is a constant.

A5: The functions  $A = \nabla C$  and  $\nabla f$  are uniformly bounded on  $S$ .

By the above assumptions, we can suppose that there exist constants  $M_1, M_2$  such that

$$\|P(x)\| \leq M_1, \|\nabla P(x)\| \leq M_1, \|\nabla^2 P(x)\| \leq M_1, \|C(x)\| \leq M_2,$$

$$\|\nabla C(x)\| \leq M_2, \|\nabla^2 C(x)\| \leq M_2.$$

A1 and A2 are the standard assumptions. A4 is the sufficient reduction condition. A4 is very moderate because Cauchy step satisfies A4. In a trust region method, A4 guarantees the convergence. To simplify the problem, we regard A4 as a condition. A3 and A5 plays an important role to obtain the convergence result. But it has minor effects to the local convergence rate. The following result are based on Assumption 1. In terms of the restoration algorithm we obtain the following result.

**Lemma 1.** *The restoration Algorithm B terminates in a finite number of iteration.*

**Proof.** Assume the restoration algorithm does not terminate finitely. Certainly, the termination criterion will be satisfied if  $\lim_{i \rightarrow \infty} h(x_k^i) = 0$ . We hence assume that there exists a constant  $\varepsilon > 0$  with  $h(x_k^i) > \varepsilon$  for all  $i$ . We show that it will lead to a contradiction.

Let

$$K = \{i : r_k^i \geq \eta\}.$$

According to Assumption 1 and  $r_k^i = \frac{M_k^i(d)}{\Psi_k^i(d)}$ , we have

$$\infty > \sum_{i=1}^{\infty} (h(x_k^{i-1}) - h(x_k^i)) \geq \eta \sum_{i=1}^{\infty} \Psi_k^{i-1}(d_k^{i-1}) \geq \eta \beta_2 \sum_{k \in K} \min\{\varepsilon, \Delta_k^i\}.$$

Therefore,  $\lim_{i \rightarrow \infty} \Delta_k^i \rightarrow 0$ . Thus,  $\Delta_k^{i-1} \geq \Delta_k^i$  by virtue of Step 4 of Restoration Algorithm. Namely,  $\{\Delta_k^i\}$  is increased when  $\Delta_k^i$  is very small, which contradicts  $\lim_{i \rightarrow \infty} \Delta_k^i \rightarrow 0$ . The result therefore holds and the proof is complete.

**Lemma 2.** *Suppose there are infinitely many points added to the filter. Then  $\lim_{k \rightarrow \infty} h(x_k) = 0$ .*

**Proof.** If theorem were not true, there would have infinite components in  $K_1$ , which is defined as follows:

$$K_1 = \{k | h_k > \varepsilon\}.$$

Because of the Assumption 1 we assume that  $|P_k| < M$  for all  $k$  without loss of generality, where  $M$  is a positive constant. Then, we analyze with two cases.

(1) If  $\min_{i \in K_1} \{P_i\}$  exists. Let  $\bar{P} = \min_{i \in K_1} \{P_i\}$ . And  $\bar{h}$  is the corresponding value relate to  $\bar{P}$ . Let  $P(\bar{x}) = \bar{P}$ ,  $h(\bar{x}) = \bar{h}$ . Then, according to the definition of the filter,  $x_k$  which lie behind  $\bar{x}$  in the filter, satisfy:

$$h_k \leq \bar{h} \text{ and } P_k \geq \bar{P}.$$

Consider the area of the south-west part I of the graph  $area_0$  will be decreased. When a new point  $x_k$  enters into the filter, the area of square II is satisfied:

$$area_{II} = (h_2 - h_k)(P_1 - P_k).$$

According to the definition of filter,

$$h_k \leq \beta h_j \quad \text{or} \quad P_k \leq P_j - \gamma h_k.$$

We have

$$area_{II} \geq (h_2 - \beta h_2)\gamma h_k \geq (1 - \beta)\gamma h_2 h_k \geq (1 - \beta)\gamma \varepsilon^2.$$

After updates the filter set, we have the area of the new square, denoted by

$$\begin{aligned} area_1 &= area_0 - area_{II} \leq area_0 - (1 - \beta)\gamma \varepsilon^2 \\ 0 \leq area_{j+1} &\leq area_j - (1 - \beta)\gamma \varepsilon^2 \\ &\leq area_{j-1} - 2(1 - \beta)\gamma \varepsilon^2 \\ &\leq \dots \\ &\leq area_0 - (j + 1)(1 - \beta)\gamma \varepsilon^2, \end{aligned}$$

$j \rightarrow 0$ ,  $area_{j+1} \rightarrow 0$ , thus, the area will be reduced to 0 after finite times. When the area is zero, it means that a point cannot enter  $K_1$ , which is contradicted with the infiniteness of  $K_1$ .

(2) If  $\min_{i \in K_1} \{P_i\}$  doesn't exist. From the conditions in this lemma, let  $P_c = \inf_{i \in K_1} \{P_i\}$ . From the definition of *inf* there exists  $\bar{P} \geq P_c$  and  $\bar{P} \leq (P_c + \gamma\varepsilon)$ . Then, according to the definition of the filter, the other components, which lie behind  $x_{kc}$  in filter, satisfy:

$$h_k \leq \bar{h} \text{ and } P_k \geq (\bar{P} - \gamma\varepsilon).$$

Using the same techniques as that in (1), the result is gotten.

Thus, the conclusion is obtained.

As for the case of finitely many points added to the filter, it is apparent that the following result be true.

**Lemma 3.** *Suppose there are finitely many points added to the filter. Then  $h(x_k) = 0$ , for certain  $k_1$  and  $k > k_1$ .*

**Proof.** This proof is similar to that of Theorem 2 in [12].

**Theorem 1.** *Let Assumption A hold and let  $x^*$  be a feasible point of problem (1.2) at which MFCQ holds, but which is not a KKT point. Then there exists a neighbourhood  $N^0$  of  $x^*$  and positive constants  $\xi_1, \xi_2, \xi_3$  such that for all  $x_k \in N^0 \cap S$  and all  $\Delta_k$  for which*

$$\xi_2 h(x_k) \leq \Delta_k \leq \xi_3,$$

*it follows that SQP subproblem has a feasible solution  $d_k$ , and the predicted reduction satisfies  $-\Psi_k^*(d_k) \geq \frac{1}{3} \xi_1 \Delta_k$ , where  $\Psi_k^*(d_k)$  is a (1.4) evolved. The following forms:*

$$\begin{aligned} \min \quad & \Psi_k^*(d) = g_k^T d + \frac{1}{2} d^T H_k d, \\ \text{s.t.} \quad & C_i(x) + A_i^T(x)d = 0, \quad i \in E = \{1, 2, \dots, m_e\}, \\ & C_i(x) + A_i^T d \leq 0, \quad i \in I^* = \{m_e + 1, m_e + 2, \dots, m\}, \\ & \|d\|_\infty \leq \Delta_k. \end{aligned} \tag{1.11}$$

Without loss of generality, we assume that  $\|C_i(x)\| \leq M_2$ ,  $\|\nabla^2 C_i(x)\| \leq M_2$  for all  $x \in S$  and  $i = 1, 2, \dots, m$ , where  $M_2 > 0$  is a constant. We now consider  $C_i(x_k^+)$ . From Taylor expansion, when (1.11) is consistent, there is

$$C_i(x_k^+)^+ \leq (C_i(x_k) + A_i^T(x_k)d)^+ + \frac{1}{2}n^2M_2\Delta_k^2.$$

Thus,  $h(x_k^+) \leq \frac{1}{2}n^2M_2\Delta_k^2$ . As for (1.11), Fletcher et al. do very good analysis in [13].

Hence, if  $\Delta_k \leq (1 - \eta_3)\xi_1 / 3nM_2$ , then

$$P(x_k) - P(x_k^+) \geq -\eta_3\Psi_k^*(d),$$

where  $\eta_3 > \eta$ .

If  $h(x_k) > 0$  and  $\Delta_k \leq \sqrt{\frac{2\beta h(x_k)}{n^2M_2}}$ , then  $h(x_k^+) \leq \beta h(x_k)$ .

**Theorem 2.** *Let Assumption A hold and let  $\{x_k\} \in S$  be generated by the algorithm which has accumulation to be a feasible point of problem (1.2) at which MFCQ holds. Then,  $\{x_k\}$  has an accumulation which is a KKT point.*

**Proof.** If Algorithm A terminates finitely and  $d_k = 0$  for some  $k$ , a KKT point is obtained and the result holds apparently.

Now, we assume Algorithm A terminates infinitely. If the result were false, according to Lemma 1, the result in [14] and the above analysis, there would exist an integer  $k_0$  such  $h(x_k) < \xi_2$  is small enough,  $P(x_k) - P(x_k + d_k) \geq \eta[\Psi_k(0) - \Psi_k(d_k)]$  for all  $k > k_0$ . Hence, the Restoration Algorithm does not appear when  $k > k_0$ . For convenience, we denote

$$K_2 = \{k | k > k_0, P(x_k) - P(x_k + d_k) \geq \eta[\Psi_k(0) - \Psi_k(d_k)]\}.$$

On the other hand, from Assumption A, we have

$$\begin{aligned} \infty &> \sum_{K_2} [P(x_k) - P(x_{k+1})] \geq \sum_{K_2} [-\eta_3 \Psi_k^*(d_k)] \\ &\geq \sum_{K_2} \left( \frac{1}{3} \eta_3 \xi_1 \Delta_k \right). \end{aligned}$$

Therefore,  $\lim_{k \rightarrow \infty} \Delta_k = 0$  for  $k \in K_2$  because  $h(x_k) \rightarrow 0$ . In terms of the definition of  $\Delta_k$ , we hence have  $\lim_{k \rightarrow \infty} \Delta_k = 0$  for all  $k$ .

There thus exists  $\Delta_k$  such that

$$\Delta_k^{-\alpha_2} > \frac{6}{\eta_3 \xi_1 \alpha_1} (\max\{(1 - \beta), \gamma\}).$$

For sufficiently large  $k$  and  $2\alpha_1 \Delta_k^{1+\alpha_2} \geq h(x_k^+)$ , where  $\alpha_1, \alpha_2$  are constants satisfies  $0 < \alpha_1 < \alpha_2 < 1$ .

$2\alpha_1 \Delta_k^{1+\alpha_2} \geq h(x_k^+)$  is rational from Taylor expansion of  $C(x_k^+)$ . In terms of  $h(x_k^+) \leq \frac{1}{2} n^2 M_2 \Delta_k^2$  and  $h(x_k) \leq \alpha_1 \Delta_k^{1+\alpha_2}$  from Restoration Algorithm,  $2\alpha_1 \Delta_k^{1+\alpha_2} \geq h(x_k^+)$  holds when  $\Delta \leq \left( \frac{4\alpha_1}{n^2 M_2} \right)^{\frac{1}{1-\alpha_2}}$ . We consider the trial step  $x_k^+$ .

Form the above analysis, when  $\Delta$  is small enough to satisfy the condition

$$\begin{aligned} \Delta_k &\leq \min \left\{ \xi_3, \frac{(1 - \eta_3) \xi_1}{3nM_2}, \left( \frac{\eta_3 \xi_1}{3\alpha_1} \right), \sqrt{\frac{2\beta h(x_k)}{n^2 M_2}} \right\}. \\ P(x_k) - P(x_k^+) &\geq \frac{1}{3} \eta_3 \xi_1 \Delta_k \\ &= \frac{1}{3} \eta_3 \xi_1 \Delta_k^{-\alpha_2} \Delta_k^{1+\alpha_2} \end{aligned}$$

$$\begin{aligned} &\geq (\max\{(1 - \beta), \gamma\})2\alpha_1\Delta_k^{1+\alpha_2} \\ &\geq \max\{(1 - \beta), \gamma\}h(x_k^+), \end{aligned}$$

which means that  $x_k^+$  will be accepted by the filter.

Then  $\Delta_k$  will not be reduce when is small enough and  $k > k_0$ , which contradicts the fact that  $\lim_{k \rightarrow \infty} \Delta_k = 0$ . The result therefore holds and the proof is complete.

### Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 1101115), the Natural Science Foundation of Hebei Province (No. A2014201033), the Key Research Foundation of Education Bureau of Hebei Province (No. ZD2015069), and Post-graduate's Innovation Fund Project of Hebei University (No. X2016060).

### References

- [1] S. P. Han, Variable metric methods for minimizing a class of non-differentiable functions, *Math. Program* 20 (1981), 1-13.
- [2] E. Polak, D. Q. Mayne and J. E. Higgins, Superlinearly convergent algorithm for min-max problems, *J. Optim. Theory Appl.* 69 (1991), 407-439.
- [3] A. Vardi, New Minimax algorithm, *J. Optim. Theory Appl.* 75 (1992), 613-634.
- [4] Zhibin Zhu, Xiang Cai and Jinbao Jian, An improve SQP algorithm for solving minimax problems, *Applied Mathematics Letters* 22 (2009), 464-469.
- [5] Fusheng Wang and Yanping Wang, Nonmonotone algorithm for minimax optimization problems, *Applied Mathematics and Computation* 217 (2011), 6296-6308.
- [6] I. Husain and Z. Jabeen, Continuous-time fractional minmax programming, *Math. Comput. Modelling* 42 (2005), 701-710.
- [7] J. L. Zhou and A. L. Tits, Nonmonotone line search for minimax problems, *Journal of Optimization Theory and Applications* 76 (1993), 455-476.
- [8] R. Fletcher and S. Leyffer, Nonlinear programming without a penalty function, *Mathematical Programming* 91 (2002), 239-269.

- [9] G. Zhou, A modified SQP method and its global convergence, *J. Glob. Optim.* 11 (1997), 193-205.
- [10] N. Gould and P. Toint, Global convergence of a non-monotone trust-region SQP-filter algorithm for nonlinear programming, *Nonconvex Optim. Appl.* 82 (2006), 125-150.
- [11] M. Ulbrich, S. Ulbrich and L. N. Vicente, programming, *Mathematical Programming* 100 (2004), 279-410.
- [12] Pu-yan Nie and Chang-feng Ma, A trust region filter method for general non-linear programming, *Applied Mathematics and Computation* 172 (2006), 1000-1017.
- [13] R. Fletcher, S. Leyffer and P. L. Toint, On the global convergence of a SQP-filter algorithm, *SIAM Journal on Optimization* 13 (2002), 44-59.
- [14] Y. X. Yuan, On the convergence of a new trust region algorithm, *Numerische Mathematik* 70 (1995), 515-539.

