Journal of Applied Sciences 4 (3): 498-507, 2004
ISSN 1607-8926

An SQP Method for Solving the Nonlinear Minimax Problem

Zhibin Zhu
Department of Computational Science and Mathematics,
Guilin Institute of Electronic Technology, Guilin,541004, People's Republic of China

Abstract: In this study, a new algorithm is presented to solve the following nonlinear minimax problem

$$\text{minimize } M_f(x) = \max_{1\le j\le m} f_j(x), x\in R^n.$$

This algorithm belongs to the sequential quadratic programming (SQP) type methods. At each iteration, the search direction d is obtained by solving one quadratic programming according to the K-T condition of the minimax problem. When d is equal to zero, then the corresponding iteration point x is a K-T point, otherwise, d is a descent direction. Unlike the SQP type algorithms for nonlinear programming, the direction d doesn't induce any Maratos like effect. A particular linear search with above-mentioned direction assure global convergence as well as superlinear convergence. Numerical results to date suggest the resulting algorithm is effective.

Key words: Nonlinear minimax problem, SQP algorithm, global convergence, superlinear convergence, AMS (MOS) subject classifications, 90C30, 65K10

INTRODUCTION

The minimax optimization problem can be stated as

$$\text{minimize } M_f(x) = \max_{1\le j\le m} f_j(x), x\in R^n, \qquad (1)$$

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

The objective function $M_f(x)$ has discontinuous first partial derivatives at points where two or more of the functions $f_j(x)$ are equal to M_f even if $f_j(x)(j=1\sim m)$ have continuous first partial derivatives. Thus we can not use directly the well known gradient methods to minimize $M_{e_j}(x)$[1]. For excellent works on this topic and its applications was reported[2-7].

The minimax problem (1) is equivalent to the following nonlinear programming:

$$\begin{aligned} &\min \quad t \\ &\text{s.t.} \quad f_j(x) - t \le 0, j\in I = \{1,\ldots,m\}. \end{aligned} \qquad (2)$$

According to (2), the K-T condition of (1) is obtained as follows:

$$\begin{aligned} &\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \begin{pmatrix} \nabla f(x) \\ -e^T \end{pmatrix} u = 0, \\ &u_j(f_j(x) - M_f(x)) = 0, \\ &u_j \ge 0, f_j(x) - M_f(x) \le 0, j = 1\sim m, \end{aligned} \qquad (3)$$

where:

$$e^T = (1,\ldots,1) \in R^m, F(x) = f_1(x), f_2(x),\ldots,f_m(x)^T \in R^m,$$

$$\nabla F(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial f_1}{\partial x_n} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} = (\nabla f_1,\ldots,\nabla f_m) \in R^{n\times m}.$$

The active set is denoted as follows:

$$I(x) = \{j \mid f_j(x) = M_f(x), j\in I\}. \qquad (4)$$

A lot of approaches have been proposed for solving the problem (1), most of them transform the minimax problem into the nonlinear programming problem (2) and solves it by well-established methods. Vardi[8] uses some slack variables to handle inequality constraints of (2), then solves only equality-constrained minimization problem taking advantage of trust-region strategy.

Charalambous and Conn[9] propose an algorithm to solve the minimax problem directly, in which two distinct

Corresponding Author: Zhibin Zhu, Department of Computational Science and Mathematics, Guilin Institute of Electronic Technology, Guilin, 541004, Peoples's Republic of China E-mail: zhu_zhibin@163.com

search directions are necessary to compute: The first, the horizontal direction, attempts to reduce $M_f(\varkappa)$ whilst, at the same time, keeping those functions whose values are close to $M_f(\varkappa)$, approximately equal. The second, the vertical direction, amounts to attempting to decrease the error to within which those functions are equal to $M_f(\varkappa)$ by means of linearization. Then, under assumptions that the minimax solution is unique, that $f_j(j=1\sim m)$ are convex and twice differentiable and that the strict complementarity condition is satisfied, global convergence is proven. However, computational effort is large and the rate of convergence is at most linear.

Due to its superlinear convergence rate, the sequential quadratic programming (SQP) methods are considered among the most effective methods for solving nonlinear programming problems. A new algorithm is proposed by Zhou and Tits[10] with taking advantage of the idea of SQP algorithms to solve the problem (1). There, in order to avoid the Maratos effect, it takes a nonmonotone line search along a direction d which is obtained by solving a QP subproblem. However, it is observed that, with such a line search and the direction d, it may not ensure that the full step of one is accepted. For this reason, it is necessary to perform a nonmonotone arc search along $td+t^2\tilde{d}$ based on a correction $\tilde{d}$ which is obtained by solving another QP subproblem. Moreover, its convergence rate is two-step supperlinear.

A SQP type algorithm is proposed by Xue[11] to solve the problem (1). However, in order to obtain the one-step superlinear convergence rate, there it is necessary to make two additional assumptions which are too strong: Firstly, the entire sequence generated by the algorithm converges to the K-T point of (1); Secondly, the stepsize is always equal to one after finite iterations.

In this paper, a new SQP algorithm is proposed to solve the problem (1) directly according to the K-T condition (3). This algorithm overcomes the shortcomings just pointed out. It performs a monotone line search along a direction which is obtained by solving only one QP subproblem. With such a particular line search, we point out, because of the intrinsic properties of the minimax problem (1) in contrast with the nonlinear programming, unlike[10] there does not exist any Maratos-like effect and the search direction is not necessary to revise any more. Global convergence is obtained without the above-mentioned strong assumptions in[9]. Under some suitable conditions which are weaker than those in[11] one-step superlinear convergence rate is proven.

Description of algorithm: The following algorithm is proposed for solving problem (1).

Algorithm A

Step 1-Initialization and date: $\varkappa^0 \in R^n$, $B_0 \in R^{n\times n}$ a symmetric positive definite matrix.

$$\alpha \in (0,\frac{1}{2}), k=0;$$

Step 2-Computation of the main search direction d^k: Compute (z^k,d^k) by solving the following quadratic problem at $\varkappa^k$:

$$\begin{aligned} \min \quad & z+\frac{1}{2}d^T B_k d \\ \text{s.t.} \quad & f_j(\varkappa^k)-M_f(\varkappa^k)+\nabla f_j(\varkappa^k)^T d \le z, j\in I. \end{aligned} \tag{5}$$

Let λ^k be the corresponding multipliers vector. If $(z^k,d^k)=(0, 0)$, STOP;

Step 3-The line search: Compute t_k, the first number t of the sequence $\{1,½,¼,.....\}$ satisfying

$$M_f(\varkappa^k+td^k) \le M_f(\varkappa^k)+\alpha t z^k \tag{6}$$

Step 4-Updates: Compute a new symmetric definite positive matrix B_{k+1}. Like Vardi[8] we use Powell's modification[12]

$$B_{k+1}=BFGS(B_k,\Delta\varkappa^k,\bar{y}^k), \tag{7}$$

where:

$$\Delta\varkappa^k=\varkappa^{k+1}-\varkappa^k, y^k=\sum_{j=1}^{m}\lambda_j^k(\nabla f_j(\varkappa^{k+1})-\nabla f_j(\varkappa^k)), \bar{y}^k=\theta y^k+(1-\theta)B_k\Delta\varkappa^k, \tag{8}$$

$$\theta = \begin{cases} 1, & \text{if}(y^k)^T \Delta\varkappa^k > 0.2(\Delta\varkappa^k)^T B_k \Delta\varkappa^k, \\ \dfrac{0.8(\Delta\varkappa^k)^T B_k \Delta\varkappa^k}{(\Delta\varkappa^k)^T B_k \Delta\varkappa^k - (y^k)^T \Delta\varkappa^k} & \text{otherwise} \end{cases}$$

Let $\varkappa^{k+1} = \varkappa^k + t_k d^k, k = k+1$. Go back to step 2.

Global convergence of algorithm: It is first shown that Algorithm A is well defined. The following general assumptions are true throughout the paper.

A1: $f_j(j=1\sim m)$ are continuously differentiable.

A2: For all $\varkappa \in R^n$, the vectors $\left\{ \begin{pmatrix} \nabla f_j(\varkappa) \\ -1 \end{pmatrix}, j \in I(x^k) \right\}$ are linearly independent.

A3: There exist two constants $0<a\le b$, such that $a\|d\|^2 \le d^T B_k d \le b\|d\|^2$, for all k, for all $d \in R^n$.

Lemma 1: The solution (z^k,d^k) of the quadratic subproblem (5) is unique and $z^k+\frac{1}{2}(d^k)^T B_k d^k \le 0$.

Proof: Since (0,0) is a feasible point of (5) and B_k is positive definite, it is clear that the claim holds.

Lemma 2: Let (z^k,d^k) be the solution of the QP (5) at $\varkappa^k$, the following are equivalent:

(i) $(z^k,d^k)=(0,0)$;
(ii) $z^k=0$;
(iii) $d^k=0$;
(iv) $z^k+\frac{1}{2}(d^k)^T B_k d^k=0$.

Proof: (i)⇒(ii). It is clear; (ii)⇒(iii) Since $z^k=0$, then $d^k=0$ by Lemma 1; (iii)⇒(iv). Since $I(\varkappa^k)=\{j \mid f_j(\varkappa^k)=M_f(\varkappa^k), j\in I\}$ is nonempty, let $j\in I(\varkappa^k)$, (5) implies that $z^k\ge 0$. So, $z^k+\frac{1}{2}(d^k)^T B_k d^k \ge 0$, thereby, $z^k+\frac{1}{2}(d^k)^T B_k d^k=0$; (iv)⇒(i). Since $z^k+\frac{1}{2}(d^k)^T B_k d^k=0$, it is clear that(0,0) is unique solution of (5) by Lemma 1, that is to say $(z^k,d^k)=(0,0)$.

Lemma 3: If $(z^k,d^k)=(0,0)$, then $\varkappa^k$ is a K-T point of (1). If $\varkappa^k$ is not a K-T point of (1), then $z^k<0$, $d^k\neq 0$ and

$$\nabla f_j(\varkappa^k)^T d^k \le z^k < 0, j \in (\varkappa^k).$$

Proof: From (5), we have

$$\begin{aligned} &\begin{pmatrix} B_k d^k \\ 1 \end{pmatrix} + \sum_{j=1}^{m} \lambda_j^k \begin{pmatrix} \nabla f_j(\varkappa^k) \\ -1 \end{pmatrix} = 0, \\ &\lambda_j^k (f_j(\varkappa^k) - M_f(\varkappa^k)) + \nabla f_j(\varkappa^k)^T d^k - z^k) = 0, \\ &\lambda_j^k \ge 0, f_j(\varkappa^k) - M_f(\varkappa^k) + \nabla f_j(\varkappa^k)^T d^k - z^k \le 0, j \in I. \end{aligned} \tag{9}$$

If $(z^k,d^k)=(0,0)$, then

$$\begin{aligned} &\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \sum_{j=1}^{m} \lambda_j^k \begin{pmatrix} \nabla f_j(q^k) \\ -1 \end{pmatrix} = 0, \\ &\lambda_j^k (f_j(\varkappa^k) - M_f(\varkappa^k)) = 0 \\ &\lambda_j^k \ge 0, f_j(\varkappa^k) - M_f(\varkappa^k) \le 0, j \in I. \end{aligned} \tag{10}$$

It shows that $\varkappa^k$ is a K-T point.
If $\varkappa^k$ is not a K-T point, then $(z^k,d^k)\neq(0,0)$. So,

$$z^k+\tfrac{1}{2}(d^k)^T B_k d^k<0,\ z^k<-\tfrac{1}{2}(d^k)^T B_k d^k\le 0,$$

and

$$\nabla f_j(x^k)^T d^k \le z^k<0, j\in I(x^k).$$

The claim holds.

Lemma 4: Algorithm A is well defined.

Proof: We only need to show that the line search yields a step $t_k=(\tfrac{1}{2})^i$ for some finite $i=i(k)$. Denote

$$a_k \triangleq f_j(x^k+td^k)-M_f(x^k)-\alpha t z^k=f_j(x^k)-M_f(x^k)+t\nabla(x^k)^T d^k-\alpha t z^k+o(t),\ j\in I.$$

For $j\in I\backslash I(x^k)$, since $f_j(x^k)<$, $M_f(x^k)$ continuity of f_j implies that there exists some $\bar{t}_j>0$, such that $a_k\le 0$.

For $j\in I(x^k)$, from (5), the fact $\alpha<1$ implies that there exists some $\bar{t}_j>0$, such that $a_k\le 0$.

Define $\bar{t}=\min\{\bar{t}_j,\ j\in I\}$. It is clear that the line search condition (6) is satisfied for all t in $[0,\bar{t}]$.

In the sequel, we'll prove that any accumulation point x^* of $\{x^k\}$ generated by the algorithm must be a K-T point of the problem (1). Since there are only finitely many choices for sets $I(x^k)\subseteq I$, we might as well assume that there exists a subsequence K, such that

$$x^k\to x^*,\ B_k\to B_*,\ d^k\to d^*,\ z^k\to z^*,\ \lambda^k\to\lambda^*,\ I_k=I(x^k)\equiv I_*, k\in K, \tag{11}$$

where I_* is a constant set.

Lemma 5: If $x^k\to x^*$, $B_k\to B_*$, $k\in K$, then $d^k\to 0$, $k\in K$.

Proof: First of all, in view of (6) and Lemma 3, it is evident that $\{M_f(x^k)\}$ is monotonous decreasing. Hence, considering to $\{x^k\}_{k\in K}$ $x^*$and continuity of $M_f(x)$ it holds that

$$M_f(x^k)\to M_f(x^*),\ k\to\infty \tag{12}$$

Suppose by contradiction that $d^*\neq 0$, then $z^*<0$. Since

$$f_j(x^k)-M_f(x^k)+\nabla f_j(x^k)^T d^k\le z^k,\ j\in I,$$

Let $k\in K$, $k\to\infty$, then

$$f_j(x^*)-M_f(x^*)+\nabla f_j(x^*)^T d^k\le z^*,\ j\in I.$$

The corresponding QP problem (5) at x^* is

$$\begin{aligned} &\min \quad z+\frac{1}{2}d^T B_* d \\ &\text{s.t.}\quad f_i(x^*)-M_f(x^*)+\nabla f_i(x^*)^T d\le z, j\in I. \end{aligned} \tag{13}$$

It is clear that (z^*, d^*) is a feasible solution of (13) and B_* is positive definite. So, we might as well $(\tilde{z},\tilde{d})$ let be the unique solution of (13). Since

$$z^k+\frac{1}{2}(d^k)^T B_k d^k=\min\left\{z+\frac{1}{2}d^T B_k d \mid f_i(x^k)-M_f(x^k)+\nabla f_j(x^k)^T d\le z, j\in I\right\}$$

as $k\in K$, $k\to\infty$ we obtain that

$$z^* + \frac{1}{2}(d^*)^T B_* d^* = \min\left\{z + \frac{1}{2}d^T B_* d \mid f_j(x^*) + \nabla f_j(x^*)^T d \le z, j \in I\right\}$$
$$= \tilde{z} + \frac{1}{2}\tilde{d}^T B_* \tilde{d}.$$

So, it can be seen that $\tilde{z} = z^* < 0, \tilde{d} = d^* \neq 0$ thereby,

$$\nabla f_j(x^*)^T d^* \le z^* < 0,\ j \in I(x^*),$$

and for $k \in K$, k large enough, we have

$$z^k \le \frac{1}{2} z^* < 0, \nabla f_j(x^k)^T d^k \le \frac{1}{2}, \nabla f_j(x^*)^T d^* < 0, j \in I(x^*) \tag{14}$$

From (14), similar to Lemma 4, we can conclude that the step-size t_k obtained by the linear search is bounded away from zero on K, i.e.,

$$t_k \ge t_* = \inf\{t_k, k \in K\},\ k \in K$$

So, from (6), (12), we get

$$0 = \lim_{k \in K}(M_f(x^{k+1}) - M_f(x^k))\ \lim_{k \in K} \alpha t_k z^k \le \frac{1}{2}\alpha t_* z^* < 0$$

It is contradiction, which shows that $d^k \to 0$, $k \in K$.

According to Lemma 3, Lemma 5 and the fact that (z^*, d^*) is the solution of (13) , we can get the following convergent theorem.

Theorem 1: The algorithm A either stops at the K-Tpoint $\{x^k\}$ of (1) in finite iteration, or generates an infinite sequence $\{x^k\}$ whose all accumulation points are K-T point of (1).

Rate of convergence: Now we strengthen the regularity assumptions on the functions involved. Assumption A1 are replaced by:

A1': The functions $f_j(j=1\sim m)$ are two times continuously differentiable.
We also make the following additional assumptions:
A4: The sequence generated by the algorithm possesses an accumulation point a (in view of Theorem 1, a K-T point).
A5: $B_k \to B_*$, $k \to \infty$

A6: The matrix $\sum_{j=1}^{m} u_j^* \nabla^2 f_j(x^*) = \sum_{j \in I(x^*)} u_j^* \nabla^2 f_j(x^*)$ is nonsingular and it holds that

$$d^T \left(\sum_{j=1}^{m} u_j^* \nabla^2 f_j(x^*)\right) d > 0, \forall 0 \neq d \in Y(x^*, u^*),$$

where, $u^* = (u_j^*, j \in I)$ is the corresponding K-T multipliers of x^* and

$$Y(x^*, u^*) = \left\{d \in R^n \mid \nabla f_j(x^*)^T d = 0, j \in I(x^*), u_j^* > 0\right\}$$

A7: The strict complementary slackness are satisfied at the K-T point pair (x^*, u^*), i.e., $u_j^* > 0, j \in I(x^*)$

Theorem 2: The Kuhn-Tucker point x^* is isolated.

Proof: This proof is similar to that of[13] but with some technical differences since assumption A2 about the linear independence constraint qualification is different with that of the nonlinear programming and the details are omitted.

Lemma 6: The entire sequence $\{x^k\}$ converges to the K-T point x^*, i.e., $x^k \to x^*$, $k \to \infty$.

Proof: From (5) and (6), we have that

$$M_f(x^{k+1}) \le M_f(x^k) + \alpha t_k z^k \le M_f(x^k) - \frac{1}{2}a\alpha t_k \|d^k\|^2$$

So, from (12), it holds that

$$\lim_{k\to\infty} \| x^{k+1} - x^k \| = \lim_{k\to\infty} t_k \|d^k\|^2 = 0$$

Thereby, according to assumptions A6, A7, Theorem 2 and proposition by Painer and Tits[14] we have $x^k \to x^*$, $K \to \infty$

Lemma 7: For $K \to \infty$, we have

$$d^k \to 0, \lambda^k \to u^*, L_k \equiv I(x^*), z^k = O(\|d^k\|)$$

where:

$$L_k = \{j \mid f_j(x^k) - M_f(x^k) + \nabla f_j(x^k)^T d^k - z^k = 0\} \tag{18}$$

Proof: According to Lemma 5 and $x^k \to x^*$, $B_k \to B_*$, $K \to \infty$, it holds that

$$d^k \to 0,\ z^k \to 0,\ k \to \infty.$$

Since (x^*, u^*) is a K-T point pair of (1), we have

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \begin{pmatrix} \nabla F(x^*) \\ -e^T \end{pmatrix} u^* = 0,$$
$$u_j^*(f_j(x^*) - M_f(x^*)) = 0, u_j^* \ge 0, j \in I$$

Denote

$$N_* = \begin{pmatrix} \nabla F(x^*) \\ -e^T \end{pmatrix}, D_* = \mathrm{diag}((f_j(x^*) - M f(x^*))^2, j \in I$$

From A2, it is clear that

$$(N_*^T N_* + D_*) \text{is nonsingular},$$

and

$$u^* = -(N_*^T N_* + D_*)^{-1} N_*^T \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{19}$$

At the same time, denote

$$N_k = \begin{pmatrix} \nabla f(x^k) \\ -e^T \end{pmatrix}, D_k = \mathrm{diag}((f_i(x^k) - M f(x^k))^2, j \in I$$
$$R_k = N_k^T N_k + D_k + \mathrm{diag}((f_j(x^k) - M_f(x^k))\ (\nabla f_j(x^k)^T d^k - z^k), j \in I).$$

From (9), we have

$$\begin{pmatrix} B_k d^k \\ -e^T \end{pmatrix} + \begin{pmatrix} \nabla f(x^k) \\ -e^T \end{pmatrix} \lambda^k = 0,$$
$$(D_k = \mathrm{diag}((f_j(x^k) - M_f(x^k))(\nabla f_j(x^k)^T d^k - z^k), j \in I))\lambda^k = 0,$$

thereby,

$$R_k \lambda^k = -N_k^T \begin{pmatrix} B_k d^k \\ 1 \end{pmatrix} \tag{20}$$

Since

$$R_k \to N_*^T N_* + D_*,$$

it is obvious , for k large enough, that

$$R_k \text{ is nonsingular and } R_k^{-1} \to (N_*^T N_* + D_*)^{-1}.$$

So,

$$\lambda^k = -R_k^{-1} N_k^T \begin{pmatrix} B_k d^k \\ 1 \end{pmatrix} \to -(N_*^T N_* + D_*)^{-1} N_*^T \begin{pmatrix} 0 \\ 1 \end{pmatrix} = u^*.$$

In addition, according to $d^k \to 0$, $z^k \to 0$ and (18), it is clear that $L_k \subseteq I(x^*)$.
On the contrary, for $j \in I(x^*)$, assumption A7 implies that $u_j^* > 0$. So, $\lambda_j^k > 0$ for k large enough, thereby, it holds that

$$f_j(x^k) - M_f(x^k) + \nabla f_j(x^k)^K d^k - z^k = 0, \text{i .e., } j \in L_k.$$

So, $L_k \equiv I(x^*)$.

Thereby, from $I(x^k) \subseteq I_* \equiv L_k$ and A2, it is clear that $Z^k = O(\|d^k\|)$.
In order to obtain superlinear convergence, we make the following assumption:

A8: $\nabla^2 f_j(x)$ (j=1~m) are Lipschitz continuous on some ball $B(x^*, \epsilon)$of radius $\epsilon > 0$ about x^*, i.e., there exist some $v_j > 0$, such that

$$\|\nabla^2 f_j(y) - \nabla^2 f_j(x)\| \le v_j \|y - x\|, \; j \in I, \; \forall x, y \in B(x^*, \varepsilon)$$

Lemma 8: Under assumptions A1-A8, $\|x^k + d^k - x^*\| = o(\|x^k - x^*\|)$ if and only if the following condition holds:

A9: The sequence of matrices $\{B_k\}$satisfies

$$\left\| \left(B_k - \sum_{j \in I(x^*)} 1_j^k \nabla^2 f_j(x^k) \right) d^k \right\| = o\left(\left\|d^k\right\|\right) \Leftrightarrow \left\| \left(B_k - \sum_{j=1}^{m} 1_j^k \nabla^2 f_j(x^k) \right) d^k \right\| = o\left(\left\|d^k\right\|\right) \Leftrightarrow \left\| \left(B_k - \sum_{j=1}^{m} u_j^* \nabla^2 f_j(x^*) \right) d^k \right\| = o\left(\left\|d^k\right\|\right)$$

Proof: This proof is similar to the proof of Theorem by Xue[11].

Lemma 9: If assumptions A1-A9 are true, then, for k large enough, $t_k \equiv 1$.

Proof: Firstly, we prove, for k large enough, that

$$f_j(x^k + d^k) \le M_f(x^k) + a\, z^k, \forall j \in I. \tag{21}$$

Let

$$s_j = f_j(x^k + d^k) - M_f(x^k) - \alpha z^k = f_j(x^k) - M_f(x^k) + \nabla f_j(x^k)^T d^k - \alpha z^k + o(\|d^k\|), \; j \in I.$$

Denote

$$I((x^k + d^k) = \{j \mid f_j(x^k + d^k) = M_f(x^k + d^k)\}$$

it is obvious, for k large enough, that $I(x^k + d^k) \subseteq I_*$.

For $j \in I \backslash I(x^*)$, the facts $f_j(x^*) < M_f(x^*)$ and $d^k \to 0$, $z^k = O(\|d^k\|)$ imply that $s_j \le 0$, (for k large enough).
For $j \in I \backslash I_*(x^k + d^k)$, from (5), the facts $\alpha < 1$, $d^k \to 0$, $z^k = O(\|d^k\|)$ imply that $s_j \le 0$, (for k large enough).
From above analysis, in order to prove that (21) are true, we only prove, for k large enough, that

$$f_j(x^k + d^k) = M_f(x^k + d^k) \le M_f(x^k) + \alpha z x^k, \forall_j \in I(x^k + d^k).$$

i.e.,

$$\sum_{j\in I(x^k+d^k)} f_j(x^k+d^k) = \sum_{j\in I(x^k+d^k)} M_f(x^k+d^k) \le \sum_{j\in I(x^k+d^k)} (M_f(x^k+\alpha z^k),$$

While, it is always proven that

$$\sum_{j\in I_*\setminus I(x^k+d^k)} f_j(x^k+d^k) \le \sum_{j\in I_*\setminus I(x^k+d^k)} (M_f(x^k)+\alpha z^k),$$

Thereby, it only needs to prove that

$$\sum_{j\in I(x^*)} f_j(x^k+d^k) \le \sum_{j\in I(x^*)} (M_f(x^k)+\alpha z^k),$$

i.e.,

$$\sum_{j\in I(x^*)} \lambda_j^k f_j(x^k+d^k) \le \sum_{j\in I(x^*)} \lambda_j^k (M_f(x^k)+\alpha z^k),$$

From (9) and $L_k \equiv I(x^*)$, we have

$$\sum_{j\in I(x^*)} \lambda_j^k \nabla f_j(x^k) = -B_k d^k, \quad \sum_{j\in I(x^*)} \lambda_j^k = 1. \tag{22}$$

Denote

$$\begin{aligned}
s &\triangleq \sum_{j\in I(x^*)} \lambda_j^k f_j(x^k+d^k) - \sum_{j\in I(x^*)} \lambda_j^k (M_f(x^k) - \alpha \sum_{j\in I(x^*)} \lambda_j^k z^k \\
&= \sum_{j\in I(x^*)} \lambda_j^k (f_j(x^k) - M_f(x^k) + \nabla f_j(x^k)^T d^k) + \frac{1}{2}(d^k)^T \left(\sum_{j\in I(x^*)} \lambda_j^k \nabla^2 f_j(x^k)\right) d^k \\
&\quad -\alpha \sum_{j\in I(x^*)} \lambda_j^k (f_j(x^k) - M_f(x^k) + \nabla f_j(x^k)^T d^k) + o(\|d^k\|^2) \\
&= (1-\alpha) \sum_{j\in I(x^*)} \lambda_j^k (f_j(x^k) - M_f(x^k)) + (1-\alpha) \sum_{j\in I(x^*)} \lambda_j^k \nabla f_j(x^k)^T d^k \\
&\quad + \frac{1}{2}(d^k)^T \left(\sum_{j\in I(x^*)} \lambda_j^k \nabla^2 f_j(x^k)\right) d^k + o(\|d^k\|^2)
\end{aligned}$$

from (22) and A8, we get

$$\begin{aligned}
&= (1-\alpha) \sum_{j\in I(x^*)} \lambda_j^k (z_j(x^k) - M_f(x^k)) + (\frac{1}{2}-\alpha) \sum_{j\in I(x^*)} \lambda_j^k \nabla f_j(x^k)^T d^k \\
&\quad + \frac{1}{2}(d^k)^T \left(\sum_{j\in I(x^*)} \lambda_j^k \nabla^2 f_j(x^k) - B_k\right) d^k + o(\|d^k\|^2) \\
&= \frac{1}{2} \sum_{j\in I(x^*)} \lambda_j^k (f_j(x^k) - M_f(x^k)) + (\frac{1}{2}-\alpha) \sum_{j\in I(x^*)} \lambda_j^k (f_j(x^k) - M_f(x^k) + \nabla f_j(x^k)^T d^k) \\
&\quad + \frac{1}{2}(d^k)^T \left(\sum_{j\in I(x^*)} \lambda_j^k \nabla^2 f_j(x^k) - B_k\right) d^k + o(\|d^k\|^2) \\
&\le (\frac{1}{2}-\alpha) \sum_{j\in I(x^*)} \lambda_j^k + o(\|d^k\|^2) \\
&\le \frac{1}{2}\alpha\ (\alpha - \frac{1}{2}) \|d^k\|^2 + o(\|d^k\|^2).
\end{aligned}$$

Since $a\in(0,½)$, it holds that $s\le 0$. So, (21) are true, for k large enough. Thereby, we have, for k large enough, that

$$M_f(x^k+d^k) \le M_f(x^k)+\alpha z^k$$

i.e., $t_k \equiv 1$. The claim holds.

From Lemma 8 and Lemma 9, we can get the following theorem:

Theorem 3: Under all above-mentioned assumptions, the algorithm is superlinearly convergent, i.e., $\|x^{k+1}-x^*\|=o(\|x^k-x^*\|)$.

Numerical experiments: We carry out numerical experiments based on the algorithm. The results show that the algorithm is effective.

During the numerical experiments $\alpha=0.25$ and $B_0=I$, the $n\times n$ unit matrix. B_k is updated by the BFGS formula (8). In following Tables, IP is the initial point, NIT is the number of iterations of algorithm, AS is the approximate solution.

Problem 1: ([9]). Minimize the maximum of the following three functions:

$$f_1(x)=x_1^4+x_2^2,$$
$$f_2(x)=(2-x_1)^2+(2-x_2)^2$$
$$f_3(x)=2\exp(x_1-x_2)^2.$$

Problem 2: ([11]). Minimize the maximum of the following six functions:

$$f_1(x)=x_1^2+x_2^2+x_3^2-1,$$
$$f_2(x)=x_1^2+x_2^2+(x_3-2)^2,$$
$$f_3(x)=x_1+x_2+x_3-1,$$
$$f_4(x)=x_1+x_2-x_3+1,$$
$$f_5(x)=2x_1^3+6x_2^2+2(5x_3-x_1+1)^2,$$
$$f_6(x)=x_1^2-9x_3.$$

Problem 3: Rosen-Suzuki Problem ([9,11]). Minimize the maximum of the following functions:

$$f_1(x)=x_1^2+x_2^2+2x_3^2+x_4^2-5x_1-5x_2-21x_3+7x_4,$$
$$f_2(x)=f_1(x)+10(x_1^2+x_2^2+x_3^2+x_4^2+x_1-x_2+x_3-x_4-8),$$
$$f_3(x)=f_1(x)+10(x_1^2+2x_2^2+x_3^2+2x_4^2-x_1-x_4-10),$$
$$f_4(x)=f_1(x)+10(x_1^2+x_2^2+x_3^2+2x_1-x_2-x_4-5).$$

Problem 4: Wong problem 1 ([9,11]). Minimize the maximum of the following functions:

$$f_1(x)=(x_1-10)^2+5(x_2-12)^2+x_3^4+3(x_4-11)^2+10x_5^6,$$
$$+7x_6^2+x_7^4-4x_6x_7+10x_6-8x_7,$$
$$f_2(x)=f_1(x)+10(2x_1^2+3x_2^4+4x_4^2+x_3+5x_5-127),$$
$$f_3(x)=f_1(x)+10(10x_3^2+7x_1+3x_2+x_4-x_5-282),$$
$$f_4(x)=f_1(x)+10(x_2^2+6x_6^2+23x_1-8x_7-196),$$
$$f_5(x)=f_1(x)+10(4x_1^2+x_2^2+2x_3^2-3x_1x_2+5x_6-11x_7).$$

Problem 5: Wong problem 2 ([11]). Minimize the maximum of the following functions:

$$f_1(x)=x_1^2+x_2^2+(x_3-10)^2+4(x_4-5)^2+(x_5-3)^2+2(x_6-1)^2+5x_7^2$$
$$+7(x_8-11)^2+2(x_9-10)^2-(x_{10}-7)^2+x_1x_2-14x_1-16x_2+45,$$
$$f_2(x)=f_1(x)+10(3(x_1-2)^2+4(x_2-3)^2+2x_3^2-7x_4-120),$$
$$f_3(x)=f_1(x)+10(5x_1^2+(x_3-6)^2+8x_2-2x_4-40),$$
$$f_4(x)=f_1(x)+10(0.5(x_1-8)^2+2(x_2-4)^2+3x_5^2-x_6-30),$$

$$f_5(\varkappa)=f_1(\varkappa)+10(\varkappa_1^2+2(\varkappa_2-2)^2-2\varkappa_1\varkappa_2+14\varkappa_5-6\varkappa_6),$$
$$f_6(\varkappa)=f_1(\varkappa)+10(4\varkappa_1+5\varkappa_2-3\varkappa_7+9\varkappa_8-105),$$
$$f_7(\varkappa)=f_1(\varkappa)+10(10\varkappa_1-8\varkappa_2-17\varkappa_7+2\varkappa_8),$$
$$f_8(\varkappa)=f_1(\varkappa)+10(-3\varkappa_1+6\varkappa_2+12(\varkappa_9-8)^2-7\varkappa_{10}),$$
$$f_9(\varkappa)=f_1(\varkappa)+10(-8\varkappa_1+2\varkappa_2+5\varkappa_9-2\varkappa_{10}-12).$$

In this paper, detailed information about solutions to above mentioned problems is listed as follows:

Table 1: The information of the solution to Problem 1

IP	NIT	AS	$M_f(x)$	I(x)
1		1.1390376415		
	12		1.9522244802	1,2
-0.1		0.8995599528		

Table 2: The information of the solution to Problem 2

IP	NIT	AS	$M_f(x)$	I(x)
1		0.3282597013		
1	15	0.0000352743	3.5997197635	2,5
1		0.1313200675		

Table 3: The information of the solution to Problem 3

IP	NIT	AS	$M_f(x)$	I(x)
0		0.0007775312		
0		1.0002280777		
	22		-43.9945737628	1,2,4
0		2.0001527822		
0		-0.9976044295		

Table 4: The information of the solution to Problem 4

IP	NIT	AS	$M_f(x)$	I(x)
(1,2,0,		(2.3303430966,1.9513781547,		
		-0.4774965234,		
4,0,	35	4.3657068909,-0.6245328378,	680.6330291228	1,2,5
1,1)		1.0381072527,1.5942513585)		

Table 5: The information of the solution to Problem 5

IP	NIT	AS	$M_f(x)$	I(x)
(2,3,5,		(2.1719581349, 2.3635995920,		1,2,3,
		8.7739159253,		
5,1,2,		5.0959902926, 0.9905624265,		
		1.4305925107,		
	54		24.311298294	5,6,
7,3,6,		1.3217510025, 9.8287025385,		7,9
10)		8.2798843768, 8.3760722933)		

ACKNOWLEDGMENT

The author would like to thank the refree for his helpful comments. This work was supported in part by the NNSF (103G1003) and GIETSF (D20350) of China.

REFERENCES

1. Zang, I., 1980. A smoothing-out technique for min-max optimization. Math. Program., 19: 61-77.
2. Bandler, J.W. and C. Charalambous, 1974. Nonlinear programming using minimax techniques. JOTA., 13: 607-619.
3. Charalambous, C. and J.W. Bandler, 1976. Nonlinear minimax optimization as a sequence of least pth optimization with finite values of p. Intl. J. Systems Sci., 7: 377-391.
4. Osborne, M.R. and G.A.Watson, 1969. An algorithm for minimax approximation in the nonlinear case. Comput. J, 12: 63-68.
5. Bandler, J.W., T.V. Srinivasan and C. Charalambous, 1972. Minimax optimization of networks by grazor search. IEEE Trans. Micowave Theorey Tech. MTT., 20: 596-604.
6. Barrientos, O., 1998. A global regularizaiton method for solving the finite min-max problem. Computational Optimization and Applications,11: 277-295.
7. Polak, E., D.Q. Mayne and J.E. Higgins, 1991. Superlinearly convergent algorithm for min-max problems. JOTA., 69: 407-439.
8. Vardi, A., 1992. New minimax algorithm. JOTA., 75: 613-634.
9. Charalambous, C. and A.R. Conn, 1978. An efficient method to solve the minimax problem directly. SIAM J. Numerical Analysis, 15: 162-187.
10. Zhou, J.L. and A.L. Tits, 1993. Nonmonotone Line Search for Minimax Problems. JOTA., 76: 455-476.
11. Xue, Y., 2002. The sequential quadratic programming method for solving minimax problem. J. Sys. Sci. Math. Sci., 22: 355-364.
12. Powell, M.J.D., 1978. The convergence of variable metric methods for nonlinearly constrained optimization calculations. Nonlinear Programming 3, R.R. Meyer and S.M. Robinson (Eds.). Academic Press, New York.
13. Robinson, S.M., 1972. A quadratically convergent algorithm for general nonlinear programming problems. Math. Program., 3: 145-156.
14. Painier, E.R. and A.L. Tits, 1987. A Superlinearly Convergent Feasible Method for the Solution of Inequality Constrained Optimization Problems. SIAM J. Control and Opti., 25: 934-950.