Technical Research Report

SYSTEMS RESEARCH CENTER

Supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), the University of Maryland, Harvard University, and Industry

Nonmonotone Line Search for Minimax Problems

by J.L. Zhou and A.L. Tits

TR 91-65

Nonmonotone Line Search for Minimax Problems[1]

Jian L. Zhou and André L. Tits[2]

Electrical Engineering Department

and

Systems Research Center

University of Maryland, College Park, MD 20742

Abstract. It was recently shown that, in the solution of smooth constrained optimization problems by sequential quadratic programming (SQP), the Maratos effect can be prevented by means of a certain nonmonotone (more precisely, four-step monotone) line search. Using a well known transformation, this scheme can be readily extended to the case of minimax problems. It turns out however that, due to the structure of these problems, one can use a simpler scheme. Such a scheme is proposed and analyzed in this paper. It is also shown that a three-step monotone (rather than four-step monotone) line search, with a relaxed decrease requirement, can be used without losing the theoretical convergence properties. Numerical experiments indicate a significant advantage of the proposed line search over the (monotone) Armijo search.

Key words. Minimax problems, SQP direction, Maratos effect, Superlinear convergence.

---

[1]This research was supported in part by NSF's Engineering Research Centers Program No. NSFD-CDR-88-03012, by NSF grant No. DMC-88-15996 and by a grant from the Westinghouse Corporation.

[2]To whom the correspondence should be addressed.

1. Introduction. Consider the "minimax" problem

$$(P) \quad \text{minimize } f(x) \quad \text{s.t.} \quad x \in \mathbb{R}^n$$

where

$$f(x) = \max_{i=1,\cdots,p} f_i(x)$$

with $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \cdots, p$, smooth.

Several authors have proposed, among other approaches (e.g., [1–3]), extensions of the popular sequential quadratic programming (SQP) scheme (originally proposed for the solution of smooth constrained problems) to the minimax framework (e.g., [4–9]). Global convergence is usually insured by means of a line search, forcing a decrease of f at each iteration. Typically, under mild assumptions, these algorithms exhibit a local superlinear (or two-step superlinear) rate of convergence provided the step size is not truncated by the line search when a solution is approached. Unfortunately, it is known that in general the full step does not yield a decrease of f and thus the line search may prevent superlinear convergence to take place (Maratos-like effect). As pointed out by Womersley and Fletcher [9] and by Conn and Li [3], the watchdog technique [10] and the bending technique [11,12], proposed for circumventing the Maratos effect in the context of smooth constrained optimization, can be easily extended to the minimax framework. Both approaches however have drawbacks. The watchdog technique may result in repeated backtracking in early iterations and the bending technique requires an additional evaluation of f at each iteration.

A few years ago, in the context of Newton's method for smooth unconstrained optimization, Grippo, Lampariello and Lucidi [13] proposed a "nonmonotone" line search according to which the objective function is not forced to decrease at every iteration but merely every M iterations, where M is a freely selected positive integer. They showed that with such a line search global convergence is still guaranteed, and they pointed out that, as the full Newton step can then be taken earlier, convergence may often be sped up. Their numerical tests were indeed very promising. Recently, it was shown that making use of a suitable extension of this scheme to smooth constrained optimization, in the framework of SQP with penalty function-based line search, has the additional advantage of automatically allowing a full step to be taken locally and thus avoiding the Maratos effect [14].

Many of the schemes that have been proposed for the solution of minimax problems can be viewed as follows. First (P) is replaced by the equivalent smooth constrained problem in $(x^0, x^1, \cdots, x^n) \in \mathbb{R}^{n+1}$,

$$(P') \qquad \begin{array}{ll} \text{minimize} & x^0 \\ \text{subject to} & f_i(x) \le x^0 \quad i = 1, \cdots, p, \end{array}$$

and application of a constrained optimization algorithm to this problem is considered. The resulting iteration is then refined to exploit the structure of the problem. In particular, in the case of sequential quadratic programming, refinements include (i) line search on f rather than on a penalty function, (ii) constraints made tight at the end of each iteration, and (iii) estimation of a Hessian of size $n \times n$ instead of $(n+1) \times (n+1)$. The question thus arises here of whether similar refinements on the nonmonotone line search scheme of [14] are viable. Specifically, (i) does a nonmonotone line search in the "max" function f still enforce global convergence? (ii) does such a line search prevent the Maratos effect? It turns out that not only the answer to both questions are positive, but moreover, in the minimax context, one can enforce three-step monotonicity (whereas in the general smooth constrained case, four-step monotonicity is all that can be ensured if the Maratos effect is to be avoided). Finally, apparently even more than in the smooth constrained case, nonmonotone line search in the minimax case leads to significantly improved results on numerical tests. In this paper, a nonmonotone line search based algorithm is described and analyzed; extension to constrained minimax problems is outlined; numerical experiments are discussed.

The balance of the paper is organized as follows. The algorithm is presented in Section 2. Global and local convergence are analyzed in Section 3. Numerical results are presented in Section 4. Section 5 discusses alternative line search rules and Section 6 is devoted to final remarks.

2. The Algorithm. Our algorithm can be viewed as dealing with (P) in the same spirit as in [4,5] and [16]. Specifically, at iteration k, an SQP direction d_k is first computed as the solution of the quadratic problem $QP(x_k, H_k)$ defined for $x_k \in \mathbb{R}^n$ and $H_k \in \mathbb{R}^{n\times n}$ symmetric positive definite by

$$\min_{d\in R^n} \quad \frac{1}{2}\langle d, H_k d\rangle + f'(x_k, d)$$

where

$$f'(x_k, d) = \max_{i=1,\cdots,p} \{f_i(x_k) + \langle \nabla f_i(x_k), d\rangle\} - f(x_k), \tag{1}$$

a first order approximation to $f(x_k + d) - f(x_k)$ at x_k in direction d. It is well known that, under suitable assumptions, the iteration obtained by setting

$$x_{k+1} = x_k + d_k \tag{2}$$

converges superlinearly to a locally optimal solution. It turns out, as will be shown below (Theorem 3.8), that close to a solution, this iteration satisfies

$$f(x_{k+1}) \leq f(x_{k-2}) - \alpha\langle d_k, H_k d_k\rangle \tag{3}$$

where α is any prescribed positive number. This suggests that no Maratos effect would arise if global convergence was enforced by means of a line search criterion requiring that the stepsize t_k satisfy

$$f(x_k + t_k d_k) \leq \max_{\ell=0,1,2} f(x_{k-\ell}) - \alpha t_k \langle d_k, H_k d_k\rangle \tag{4}$$

where the "max" insures that a positive step will always be accepted ((4) is less stringent than an Armijo type criterion). As f is not required to decrease at each iteration, such line search is referred to as a nonmonotone line search. It is known to induce global convergence when f is smooth [13]; we show below (Theorem 3.3) that it still does here. In view of (3), the nonmonotone line search criterion would accept the full step of one provided the "undamped" iteration (2) has been used for the last two iterations. To this end, following [14,15], we propose to initialize this procedure, whenever $t_k = 1$ does not satisfy (4), by performing an arc search based on a correction $\tilde{d}_k$ so that a stepsize t_k is determined to satisfy

$$f(x_k + t_k d_k + t_k^2 \tilde{d}_k) \leq \max_{\ell=0,1,2} f(x_{k-\ell}) - \alpha t_k \langle d_k, H_k d_k \rangle . \tag{5}$$

$\tilde{d}_k$ will be chosen in such a way as to guarantee that (i) $t_k = 1$ is accepted in (5) for k large enough, where $\alpha \in (0, \frac{1}{2})$; and (ii) $d_k + \tilde{d}_k$ converges to d_k in order to preserve the properties of the quasi-Newton direction. Such $\tilde{d}_k$ can be chosen, for instance, as the solution $\tilde{d}$ of the quadratic program $\widetilde{QP}(x_k, d_k, H_k)$ given by

$$\min_{\tilde{d} \in R^n} \quad \frac{1}{2} \langle (d_k + \tilde{d}), H_k(d_k + \tilde{d}) \rangle + \tilde{f}'(x_k + d_k, x_k, \tilde{d}) \tag{6}$$

if $\|\tilde{d}\| \leq \|d_k\|$, and zero otherwise. In (6),

$$\tilde{f}'(x_k + d_k, x_k, \tilde{d}) = \max_{i=1,\cdots,p} \{ f_i(x_k + d_k) + \langle \nabla f_i(x_k), \tilde{d} \rangle \} - f(x_k + d_k).$$

It is shown below (Proposition 3.4) that $\tilde{d}_k$ obtained from (6) is always suitable for k large enough.

Algorithm NLS.

Parameters. $\alpha \in (0, \frac{1}{2})$, $\beta \in (0, 1)$.

Data. $x_0 \in \mathbb{R}^n$, $H_0 = H_0^T > 0$.

Step 0. Initialization. Set $k = 0$, and $x_{-2} = x_{-1} = x_0$.

Step 1. Computation of search direction and stepsize.

i. Compute d_k by solving the quadratic program $QP(x_k, H_k)$. If $\|d_k\| = 0$, stop.

ii. If

$$f(x_k + d_k) \leq \max_{\ell=0,1,2} f(x_{k-\ell}) - \alpha \langle d_k, H_k d_k \rangle, \tag{7}$$

set $t_k = 1$, $\tilde{d}_k = 0$ and go to *Step 2.*

iii. Compute $\tilde{d}_k$ by solving the quadratic program $\widetilde{QP}(x_k, d_k, H_k)$. If $\|\tilde{d}_k\| > \|d_k\|$, set $\tilde{d}_k = 0$.

iv. Compute t_k, the first number t in the sequence $\{1, \beta, \beta^2, \cdots\}$ satisfying

$$f(x_k + td_k + t^2\tilde{d}_k) \leq \max_{\ell=0,1,2} \{f(x_{k-\ell})\} - \alpha t \langle d_k, H_k d_k \rangle. \tag{8}$$

Step 2. Updates.

Set

$$x_{k+1} = x_k + t_k d_k + t_k^2 \tilde{d}_k.$$

Compute a new symmetric positive definite approximation H_{k+1} to the Hessian of the Lagrangian. Increase k by 1. Go back to *Step 1*.

□

Remark 2.1. Without *Step 1 ii*, the algorithm is a simple combination of Han's method (except that t_k is determined differently) and a second order correction to obtain superlinear convergence.

3. Convergence analysis. Given $x \in \mathbb{R}^n$, the set of active functions at x is defined by

$$I(x) = \{i : f_i(x) = f(x)\}.$$

The following standard assumptions are made throughout the analysis.

A1. The functions f_i, $i = 1, \cdots, p$, are continuously differentiable.

A2. For any $x_0 \in \mathbb{R}^n$, the set $\Omega = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ is compact.

A3. There exist $\sigma_1, \sigma_2 > 0$ such that

$$\sigma_1 \|x\|^2 \leq \langle x, H_k x \rangle \leq \sigma_2 \|x\|^2, \quad \forall x \in \mathbb{R}^n, \ \forall k.$$

For problem (P), the Lagrangian is defined by

$$L(x, \lambda) = \sum_{i=1}^{p} \lambda_i f_i(x).$$

A point $x^* \in X$ is *stationary* for (P) (see, e.g., [5,7]) if there exist $\lambda_i^* \geq 0, i = 1, \cdots, p$, such that

$$\begin{aligned} &\nabla_x L(x^*, \lambda^*) = 0 \ \ \& \ \ \sum_{i=1}^{p} \lambda_i^* = 1 \\ &\lambda_i^* = 0 \ \ \forall i \ \text{ s.t. } \ f_i(x^*) < f(x^*). \end{aligned} \tag{9}$$

It is clear that any local solution of (P) is stationary. The first order necessary conditions of optimality for $QP(x_k, H_k)$ can be expressed as follows. If d_k solves $QP(x_k, H_k)$, there exist $\lambda_{k,i} \geq 0, i = 1, \cdots, p$, such that

$$\begin{aligned} H_k d_k + \nabla_x L(x_k, \lambda_k) = 0 \;\; \& \;\; \sum_{i=1}^{p} \lambda_{k,i} = 1 \\ \lambda_{k,i} = 0 \; \forall i \;\; \text{s.t.} \;\; f_i(x_k) + \langle \nabla f_i(x_k), d_k \rangle < \max_{i=1,\cdots,p} \{ f_i(x_k) + \langle \nabla f_i(x_k), d_k \rangle \} \, . \end{aligned} \tag{10}$$

Due to the equivalence of (P) and the smooth constrained optimization problem (P'), some of the proofs are fairly standard and are either given in the Appendix or altogether left out.

3.1. Global convergence. In view of A1, A2 and A3, $QP(x_k, H_k)$ and $\widetilde{QP}(x_k, d_k, H_k)$ have unique and bounded (as k goes to ∞) solutions d_k and $\tilde{d}_k$ respectively. The following lemma shows that d_k is a direction of descent for $f(x)$ at x_k (see, e.g., [5]).

Lemma 3.1. The directional derivative $Df(x_k, d_k)$ of $f(x)$ at x_k along d_k satisfies

$$Df(x_k, d_k) \leq -\langle d_k, H_k d_k \rangle \qquad \forall k,$$

and d_k is zero if and only if x_k is a stationary point. □

In view of A3, of the continuity of $f(x)$, and of the boundedness of $\tilde{d}_k$, it follows from Lemma 3.1 that the line search is well defined. Therefore, unless the algorithm stops at *Step 1 i* at a stationary point, it constructs an infinite sequence $\{x_k\}$. In the sequel, we assume the latter.

The following property, which holds even though monotone line search is not enforced, is a key to global convergence. Although the underlying ideas of the proof are analogous to those used by Grippo *et al.* in the smooth unconstrained case [13], the details of the extension to the present situation are nontrivial.

Lemma 3.2. The sequence $\{x_k\}$ is bounded and the sequences $\{t_k d_k\}$ and $\{x_{k+1} - x_k\}$ both converge to zero.

Proof. Clearly $f(x_k) \leq f(x_0)$ for all k. Thus the boundedness of $\{x_k\}$ follows from A2. Now, for k given, let $\ell(k)$ be an index such that

$$f(x_{\ell(k)}) = \max_{\ell=0,1,2} f(x_{k-\ell}) = \max_{\ell=k-2,k-1,k} f(x_\ell).$$

We first show that, for some $f^* \in R$,

$$f(x_{\ell(k)}) \to f^* \quad \text{as} \quad k \to \infty. \tag{11}$$

For this, note that, in view of the definition of $\ell(k)$,

$$\begin{aligned} f(x_{\ell(k+1)}) &= \max_{\ell=k-1,k,k+1} f(x_\ell) \\ &\leq \max_{\ell=k-2,\cdots,k+1} f(x_\ell) \\ &= \max\{f(x_{\ell(k)}), f(x_{k+1})\} \\ &= f(x_{\ell(k)}) \end{aligned}$$

since, in view of the construction of x_{k+1} in the algorithm, $f(x_{k+1}) \leq f(x_{\ell(k)})$. Thus $f(x_{\ell(k)})$ is nonincreasing. Since $x_k \in \Omega$ for all k, (11) then follows from A1 and A2.

Second we show that, for any integer j, the following implications hold:

$$f(x_{\ell(k)-j}) \to f^* \;\text{ as }\; k \to \infty \;\;\Rightarrow\;\; t_{\ell(k)-(j+1)} d_{\ell(k)-(j+1)} \to 0 \;\text{ as }\; k \to \infty \tag{12}$$

and

$$f(x_{\ell(k)-j}) \to f^* \;\text{ as }\; k \to \infty \;\;\Rightarrow\;\; x_{\ell(k)-j} - x_{\ell(k)-(j+1)} \to 0 \;\text{ as }\; k \to \infty. \tag{13}$$

(Throughout the remainder of this proof, k is taken large enough for the indexes to make sense.) Indeed from the construction of x_{k+1} and in view of A3, we have

$$\begin{aligned} f(x_{\ell(k)-j}) &\leq f(x_{\ell(\ell(k)-(j+1))}) - \alpha t_{\ell(k)-(j+1)} \langle d_{\ell(k)-(j+1)}, H_{\ell(k)-(j+1)} d_{\ell(k)-(j+1)} \rangle \\ &\leq f(x_{\ell(\ell(k)-(j+1))}) - \alpha \sigma_1 t_{\ell(k)-(j+1)} \|d_{\ell(k)-(j+1)}\|^2. \end{aligned}$$

In view of (11), the left hand side of (12) implies

$$f^* \leq f^* - \lim_{k\to\infty} \alpha \sigma_1 t_{\ell(k)-(j+1)} \|d_{\ell(k)-(j+1)}\|^2.$$

Thus

$$t_{\ell(k)-(j+1)} \|d_{\ell(k)-(j+1)}\|^2 \to 0 \;\text{ as }\; k \to \infty.$$

Since t_k is bounded, (12) follows. Since $\|\tilde{d}_k\| \leq \|d_k\|$ and $|t_k| \leq 1 \; \forall k$,

$$\|x_{\ell(k)-j} - x_{\ell(k)-(j+1)}\| \leq 2 t_{\ell(k)-(j+1)} \|d_{\ell(k)-(j+1)}\|$$

and (13) also follows.

Third we show by induction on j that, if j is any nonnegative integer,

$$f(x_{\ell(k)-j}) \to f^* \quad \text{as} \quad k \to \infty. \tag{14}$$

In view of (11), (14) holds for $j = 0$. Suppose it holds for some $\hat{\jmath}$. Then, from (13),

$$x_{\ell(k)-\hat{\jmath}} - x_{\ell(k)-(\hat{\jmath}+1)} \to 0 \text{ as } k \to \infty.$$

Since $\{x_k\}$ is bounded, continuity of f and the induction hypothesis imply

$$f(x_{\ell(k)-(\hat{\jmath}+1)}) \to f^* \quad \text{as} \quad k \to \infty$$

and this completes the proof of (14).

The proof of the lemma can now be readily completed. Indeed, (12), (13) and (14) imply that, for any nonnegative integer j,

$$t_{\ell(k)-(j+1)} d_{\ell(k)-(j+1)} \to 0 \quad \text{as} \quad k \to \infty \tag{15}$$

$$x_{\ell(k)-j} - x_{\ell(k)-(j+1)} \to 0 \quad \text{as} \quad k \to \infty. \tag{16}$$

From the fact (see definition of $\ell(k)$) that, for all k,

$$\ell(k) - 1 \leq (k-1) \leq \ell(k-1) + 2$$

and

$$\ell(k) \leq k,$$

it follows that

$$\ell(k-1) \leq \ell(k) \leq k \leq \ell(k-1) + 3$$

and thus the three subsequences (15) (resp. (16)) corresponding to $j = 0, 1, 2$ cover the entire sequence $\{t_k d_k\}$ (resp. $\{x_{k+1} - x_k\}$), so that

$$t_k d_k \to 0 \quad \text{as} \quad k \to \infty$$

$$x_{k+1} - x_k \to 0 \quad \text{as} \quad k \to \infty.$$

□

Theorem 3.3. Let x^* be an accumulation point of the sequence generated by the algorithm and $\{x_k\}_{k\in K}$ be any subsequence converging to x^*. Then, x^* is a stationary point of (P) and the sequence $\{d_k\}_{k\in K}$ converges to zero. □

3.2. Superlinear convergence. Assumption A1 is replaced by

A1'. The functions f_i, $i = 1, \cdots, p$, are three times continuously differentiable.

Let x^* be an accumulation point of $\{x_k\}$ and let $\lambda_i^*, i = 1, \cdots, p$, be the corresponding multipliers. The following assumptions are used in the analysis of local convergence.

A4. At x^*, any scalars $\lambda_i, i \in I(x^*)$, satisfying

$$\sum_{i\in I(x^*)} \lambda_i \nabla f_i(x^*) = 0 \quad \& \quad \sum_{i\in I(x^*)} \lambda_i = 0$$

must all be zero.

A5. The second order sufficiency conditions with strict complementary slackness are satisfied at x^*, i.e., $\lambda_i^* > 0 \ \forall i \in I(x^*)$ and

$$\langle h, \nabla^2_{xx} L(x^*, \lambda^*) h \rangle > 0 \qquad \forall h \in S^*, h \neq 0,$$

with

$$S^* = \{h : \langle h, \nabla f_i(x^*) \rangle = \langle h, \nabla f_j(x^*) \rangle \ \ \forall i, j \in I(x^*)\}.$$

Proposition 3.4. (i) The entire sequence $\{x_k\}$ converges to x^* and the entire sequence $\{d_k\}$ converges to zero; (ii) the multiplier vector λ_k associated with the solution d_k of $QP(d_k, H_k)$ converges to λ^* and, for k large enough,

$$\{i : \lambda_{k,i} > 0\} = I(x^*); \tag{17}$$

(iii)

$$\|\tilde{d}_k\| = O(\|d_k\|^2). \tag{18}$$

□

Now, without loss of generality, assume that $I(x^*) = \{1, \cdots, m\}$ for some m and define, for any $j \in I(x^*)$, $\bar{f}_j(x) = [f_i(x) - f_j(x) : \forall i \in I(x^*)\backslash\{j\}]^T$.

A6. H_k approximates the Hessian of the Lagrangian at x^* in the sense that

$$\frac{\|P_k\{H_k - \nabla^2_{xx}L(x^*, \lambda^*)\}P_k d_k\|}{\|d_k\|} \to 0 \text{ as } k \to \infty, \tag{19}$$

where the matrices P_k are defined by

$$P_k = I - R_k(R_k^T R_k)^{-1} R_k^T$$

with $R_k = \frac{\partial \bar{f}_1^T}{\partial x}(x_k)$ (in view of A4, $R_k^T R_k$ is invertible for k large enough).

Remark 3.1. Note that elementary column operations on R_k do not affect P_k. Thus, P_k is unchanged if R_k is replaced by $\frac{\partial \bar{f}_j^T}{\partial x}(x_k)$ for any $j \in I(x^*)$.

Assumption A6 has been observed to often hold, e.g., under some conditions, when H_k is updated using Powell's modification of the BFGS formula (see [17]). In the presence of the strong properties stated in Proposition 3.4, it ensures that the iteration is close enough to the Newton iteration that a full step is eventually accepted by the line search.

Proposition 3.5. For k large enough, $t_k = 1$. □

Next, because the correction $\tilde{d}_k$ is small (see (18)), A6 implies two-step superlinear convergence in the present context, as it does when the unperturbed SQP iteration is used (see, e.g., [18,19]).

Theorem 3.6. Under the stated assumptions, the convergence rate is two-step superlinear, *i.e.,*

$$\lim_{k\to\infty} \frac{\|x_{k+2} - x^*\|}{\|x_k - x^*\|} = 0.$$

Moreover,

$$\|x_{k+1} - x_k\| = O(\|x_k - x^*\|) \quad \& \quad \|x_{k+1} - x^*\| = O(\|x_k - x^*\|).$$

□

Finally, and most importantly, as mentioned in the introduction, two-step superlinear convergence implies that, for k large enough,

$$f(x_k + d_k) \le f(x_{k-2}) - \alpha\langle d_k, H_k d_k\rangle$$

and thus *Step 1 iii* of the algorithm is eventually bypassed. (In [14] and [15] in the context of smooth constrained minimization, with a stricter descent requirement, a related inequality was shown to hold in four steps, i.e., with $k-3$ instead of $k-2$.) The proof of this results involves the following lemma which is a simple extension of a result shown in [10] in the case of smooth constrained optimization problems.

Lemma 3.7. There exists $c_1 > 0$ such that, for all x close to x^*,

$$f(x) - f(x^*) \geq c_1 \|x - x^*\|^2 .$$

□

Theorem 3.8. For k large enough, $x_k + d_k$ is always accepted and *Step 1 iii* (computation of $\tilde{d}_k$) is not performed.

Proof. As suggested above, we show that, for k large enough,

$$f_i(x_k + d_k) \leq f(x_{k-2}) - \alpha \langle d_k, H_k d_k \rangle \quad \forall i = 1, \cdots, p . \tag{20}$$

In view of the continuity of f and of Proposition 3.4(i), it follows that, for k large enough,

$$f(x_k + d_k) = \max_{i \in I(x^*)} f_i(x_k + d_k) . \tag{21}$$

Therefore, it suffices to prove (20) for all $i \in I(x^*)$. Let $i, j \in I(x^*)$. Expanding $f_i(x_k + d_k)$ around x^* gives, in view of A1' and (9),

$$f_i(x_k + d_k) = f_i(x^*) + \langle \nabla f_i(x^*), x_k + d_k - x^* \rangle + O(\|x_k + d_k - x^*\|^2) \tag{22}$$

$$= f_i(x^*) - \sum_{\ell \in I(x^*)} \lambda_\ell^* \langle \nabla f_\ell(x^*) - \nabla f_i(x^*), x_k + d_k - x^* \rangle + O(\|x_k + d_k - x^*\|^2) . \tag{23}$$

Since $f_i(x^*) = f_j(x^*)$, (22) implies

$$f_j(x_k + d_k) - f_i(x_k + d_k) = \langle \nabla f_i(x^*) - \nabla f_j(x^*), x_k + d_k - x^* \rangle + O(\|x_k + d_k - x^*\|^2) . \tag{24}$$

On the other hand, expanding $f_i(x_k + d_k)$ around x_k gives

$$f_i(x_k + d_k) = f_i(x_k) + \langle \nabla f_i(x_k), d_k \rangle + O(\|d_k\|^2)$$

which implies, in view of (10) and A1',

$$f_j(x_k + d_k) - f_i(x_k + d_k) = O(\|d_k\|^2) \quad \forall i, j \in I(x^*). \tag{25}$$

By substituting (25) and (24) in (23), we obtain

$$f_i(x_k + d_k) = f_i(x^*) + O(\|d_k\|^2) + O(\|x_k + d_k - x^*\|^2) \quad \forall i \in I(x^*).$$

Therefore, in view of (21), A3, Lemma 3.7 and Theorem 3.6, the above expression implies

$$\begin{aligned} f(x_k + d_k) &= f(x^*) + O(\|d_k\|^2) + O(\|x_k + d_k - x^*\|^2) \\ &= f(x^*) - \alpha\langle d_k, H_k d_k\rangle + O(\|d_k\|^2) + O(\|x_k + d_k - x^*\|^2) \\ &\leq f(x_{k-2}) - \alpha\langle d_k, H_k d_k\rangle - c_1\|x_{k-2} - x^*\|^2 + O(\|d_k\|^2) + O(\|x_k + d_k - x^*\|^2) \\ &= f(x_{k-2}) - \alpha\langle d_k, H_k d_k\rangle - c_1\|x_{k-2} - x^*\|^2 + o(\|x_{k-2} - x^*\|^2). \end{aligned} \tag{26}$$

Therefore (20) holds. □

4. Numerical experiments. As can been seen easily, the presence of linear constraints does not increase the complexity of the algorithm and a set of linearly constrained minimax problems has been included in our test. An efficient implementation of the algorithm described in this paper has been incorporated into a more general code (FSQP Version 2.1 [20]). In this implementation, $\alpha = 0.1$, $\beta = 0.5$, and H_k is updated by means of the BFGS formula with Powell's modification [17], with $H_0 = I$ the identity matrix.

Results obtained on selected minimax problems are summarized in Table 1. All computations were performed on a SUN 4/SPARC station 1. Gradients were computed by finite differences (for the ith component, the perturbation parameter was $2 \times 10^{-8} \max\{1, |x_k^i|\}$). Problems BARD, DAVD2, F&R, HETTICH, and WATS are from [21]; CB2, CB3, R-S, WONG and COLV are from [22, Examples 5.1-5]; MAD1 to MAD8 are from [23, Examples 1-8]. Some of these test problems allow one to freely select the number of variables; problems WATS-6 and WATS-20 correspond to 6 and 20 variables, respectively, and MAD8-10, MAD8-30 and MAD8-50 to 10, 30 and 50 variables respectively. Problems BARD down to WONG are unconstrained and MAD1 down to MAD8 are linearly constrained minimax problems. In Table 1, the performance of Algorithm NLS is compared with that of the same algorithm with an Armijo type line search (ALS)[3] and with that of algorithms proposed in [3] (CL) and [23] (MS). To make such comparison meaningful, we attempted to best approximate the stopping rule used in each of the references. Thus (*i*) for problems BARD down to WONG, execution was terminated when $\|d_k\|$ was smaller than the corresponding value of ϵ in the EPS column, and (*ii*) for problems MAD1 down to MAD8-50, execution was terminated when $\|d_k\|$ was smaller than $\|x_k\|$ times the corresponding value of ϵ in the EPS column. As pointed out by Madsen and Schjær-Jacobsen, all their problems cited here except MAD-2 satisfy Haar's condition.

The following observations can be made. First, NLS performs much better than ALS in terms of the number of function evaluations. Second, it compares well with other algorithms. WATS-20 is peculiar since from iteration 20 on, the 14 significant digits printed out by FSQP do not change. On the MAD problems for which the Haar condition holds, the performance of NLS appears to be comparable to that of the algorithm of [23].

5. Alternative line search rules. Clearly all the theoretical results would hold if line search (8) were replaced by monotone line search. Yet, in conjunction of (7) which is essential, it is natural to use the former. (Note that Grippo *et al.* used it merely to perform larger steps.) It is easy to check that if $\ell = 0, 1, 2$ were replaced by $\ell = 0, \cdots, M$ for some arbitrary positive integer $M \geq 2$, Lemma 3.2 would still be true and global convergence would still be guaranteed; $M \geq 2$ is needed for Theorem 3.8 to hold. ($M \equiv 0$ corresponds to a monotone line search as used in [4–7]; as discussed in the introduction, Theorem 3.8 would not hold in this case.)

A line search requiring a decrease by an amount proportional to $-\langle d_k, H_k d_k \rangle$ was first used in [5] for minimax problems and Han argued there that a larger step would be allowed

[3]FSQP gives the user the option to choose either NLS or ALS with bending (i.e., replace $\ell = 0, 1, 2$ by $\ell = 0$).

than if $f'(x_k, d_k)$ defined by (1) were used. There is no known theoretical advantage however when a monotone line search is used. In our context, if $f'(x_k, d_k)$ were used in the line search (see [15]), a similar analysis could be carried out with the difference that M has to be at least 3 instead of 2 in order for Theorem 3.8 to hold.

6. Concluding remarks. We have described and analyzed an SQP based algorithm for unconstrained nonlinear minimax problems with nonmonotone line search. It is proved that the Maratos-like effect can be avoided while auxiliary function evaluations are performed only during early iterations. Extension to the linearly constrained case presents no difficulty, but an assumption of linear independence of gradients of active constraints has to be imposed on the analysis of global convergence to ensure that multipliers associated with constraints are bounded. For nonlinearly constrained minimax problems, either the algorithm given in [14] could be invoked with suitable modifications concerning our max function $f(x)$ if feasibility of successive iterates is not required, or the algorithm in [15] could be invoked, as has already been suggested there, if feasibility is required at each iteration starting, from an initial feasible point (nonlinear equality constraints are not allowed). The analyses of such algorithms can be easily carried out by combining the results in this paper and results in [14] or results in [15]. In fact, Algorithm NLS has been combined with that in [15] and has been successfully implemented in FSQP [24] to solve nonlinearly constrained minimax problems. Table 2 contains some numerical results. These problems are obtained from problems 43, 84, 113 and 117 in [25] by removing certain constraints and including instead additional objectives of the form $f_i(x) = f(x) + \alpha_i g_j(x)$ where the α_i's are positive scalars and $g_j(x) \leq 0$. Specifically, P43M is constructed from problem 43 by taking out the first two constraints and including two corresponding objectives with $\alpha_i = 15$ for both; P84M similarly corresponds to problem 84 without constraints 5 and 6 but with two corresponding additional objectives, with $\alpha_i = 20$ for both; for P113M the first three linear constraints from problem 113 are turned into objectives, with $\alpha_i = 10$ for all three; for P117M, the first two nonlinear constraints are turned into objectives, again with $\alpha_i = 10$ for both. `NNL` denotes the number of nonlinear constraints. NG denotes the number of individual constraint evaluations. All other notations are the same as in Table 1. It is apparent that nonmonotone line search significantly decreases the number of evaluations of both objective functions and constraints.

Appendix. Proofs of Theorem 3.3, Proposition 3.4 and Proposition 3.5.

Proof of Theorem 3.3. We first show that $\{d_k\}$ converges to zero on K. Proceeding by contradiction, we suppose there exists an infinite subset $K' \subset K$ such that $\inf_{k \subset K'} \|d_k\| > 0$, i.e., $\exists \underline{d} > 0$ s.t. $\|d_k\| \geq \underline{d}, k \in K'$. We show that there exists $\underline{t} > 0$ independent of k such that line search (7) or (8) is always satisfied for some $t_k \geq \underline{t}$ for all $k \in K'$. Expanding f_i at x_k gives

$$f_i(x_k + td_k + t^2\tilde{d}_k) = f_i(x_k) + \langle \nabla f_i(x_k), td_k + t^2\tilde{d}_k \rangle + o(td_k + t^2\tilde{d}_k).$$

Thus, in view of (10) and the boundedness of d_k and $\tilde{d}_k$, we have, for $t \in [0,1]$ and $i = 1, \cdots, p$,

$$\begin{aligned} f_i(x_k + td_k + t^2\tilde{d}_k) &= f_i(x_k) + t\langle \nabla f_i(x_k), d_k \rangle + o(t) \\ &\leq f_i(x_k) + t\left\{ \max_{i=1,\cdots,p} \{f_i(x_k) + \langle \nabla f_i(x_k), d_k \rangle\} - f_i(x_k) \right\} + o(t) \\ &= (1-t)f_i(x_k) + t\sum_{i=1}^{p} \lambda_{k,i} f_i(x_k) + t\sum_{i=1}^{p} \lambda_{k,i} \langle \nabla f_i(x_k), d_k \rangle + o(t) \\ &= (1-t)f_i(x_k) + t\sum_{i=1}^{p} \lambda_{k,i} f_i(x_k) - t\langle d_k, H_k d_k \rangle + o(t) \\ &\leq f(x_k) - t\langle d_k, H_k d_k \rangle + o(t) \\ &\leq f(x_k) - \alpha t\langle d_k, H_k d_k \rangle - t(1-\alpha)\langle d_k, H_k d_k \rangle + o(t). \end{aligned}$$

In view of A3 and the contradiction assumption, it follows that

$$f_i(x_k + td_k + t^2\tilde{d}_k) \leq f(x_k) - \alpha t\langle d_k, H_k d_k \rangle - t(1-\alpha)\sigma_1 \underline{d}^2 + o(t).$$

Therefore, since $\alpha < 1$, there exist $\underline{t}_i > 0$ independent of k such that, for all $t \in [0, \underline{t}_i]$,

$$f_i(x_k + td_d + t^2\tilde{d}_k) \leq f(x_k) - \alpha t\langle d_k, H_k d_k \rangle, \quad i = 1, \cdots, p.$$

If we choose $\underline{t} = \min_{i=1,\cdots,p} \underline{t}_i$, then, for all $k \in K'$ at which a stepsize t_k is obtained via a line search, $t_k \geq \underline{t}$. Therefore, $\{t_k d_k\}$ is uniformly bounded from below on K' by $\underline{td}$, a contradiction to Lemma 3.2. Thus, $\{d_k\}$ converges to zero on K.

Now, since $\{\lambda_k\}$ is in the compact set

$$\Lambda = \{\lambda \in \mathbb{R}^p : \sum_{i=1}^{p} \lambda_i = 1 \;\; \& \;\; \lambda_i \geq 0 \;\; \forall i = 1, \cdots, p\},$$

there exist $K' \subset K$ and $\lambda^* \in \Lambda$ such that $\{\lambda_k\}$ converges to λ^* on K'. Taking the limit on K' in (10), in view of A3, it follows that $\{x_k, d_k, \lambda_k\}$ converges on K' to $(x^*, 0, \lambda^*)$ and (x^*, λ^*) satisfies (9). Therefore, x^* is stationary. □

Proof of Proposition 3.4. The argument for (i) and (ii) is standard and thus is left out. For (iii), it can be shown in view of (i) and (ii) and the stated assumptions that, for k large enough, it holds that

$$\{i : \tilde{\lambda}_{k,i} > 0\} = I(x^*) \tag{A.1}$$

with $(\tilde{d}_k, \tilde{\lambda}_k)$ the solution of $\widetilde{QP}(x_k, d_k, H_k)$.

In view of (17) and (10), the unique solution (d_k, λ_k) of $QP(x_k, H_k)$ is also the unique solution of the linear system in (d, λ)

$$\begin{cases} H_k d + \sum_{i \in I(x^*)} \lambda_i \nabla f_i(x_k) = 0 \\ \sum_{i \in I(x^*)} \lambda_i = 1, \quad \lambda_i = 0 \; \forall i \notin I(x^*) \\ f_i(x_k) + \langle \nabla f_i(x_k), d \rangle = f_j(x_k) + \langle \nabla f_j(x_k), d \rangle, \quad \forall i, j \in I(x^*). \end{cases} \tag{A.2}$$

Similarly, $(\tilde{d}_k, \tilde{\lambda}_k)$ is also the unique solution of the linear system in $(\tilde{d}, \tilde{\lambda})$

$$\begin{cases} H_k(d_k + \tilde{d}) + \sum_{i \in I(x^*)} \tilde{\lambda}_i \nabla f_i(x_k) = 0 \\ \sum_{i \in I(x^*)} \tilde{\lambda}_i = 1, \quad \tilde{\lambda}_i = 0 \; \forall i \notin I(x^*) \\ f_i(x_k + d_k) + \langle \nabla f_i(x_k), \tilde{d} \rangle = f_j(x_k + d_k) + \langle \nabla f_j(x_k), \tilde{d} \rangle, \quad \forall i, j \in I(x^*). \end{cases} \tag{A.3}$$

By expanding $f_i(x_k + d_k), i \in I(x^*)$, to second order around x_k, (A.3) is equivalent to the linear system

$$\begin{cases} H_k(d_k + \tilde{d}) + \sum_{i \in I(x^*)} \tilde{\lambda}_i \nabla f_i(x_k) = 0 \\ \sum_{i \in I(x^*)} \tilde{\lambda}_i = 1, \quad \tilde{\lambda}_i = 0 \; \forall i \notin I(x^*) \\ f_i(x_k) + \langle \nabla f_i(x_k), d_k + \tilde{d} \rangle + O(\|d_k\|^2) = f_j(x_k) + \langle \nabla f_j(x_k), d_k + \tilde{d} \rangle + O(\|d_k\|^2), \\ \hfill \forall i, j \in I(x^*). \end{cases} \tag{A.4}$$

The only difference between (A.2) and (A.4) viewed as systems of equations in unknown (d, λ) and $(d_k + \tilde{d}, \tilde{\lambda})$ respectively is a perturbation of order $O(\|d_k\|^2)$. In view of A4, claim (iii) then follows from the implicit function theorem. □

The following lemma is used to facilitate the proof of Proposition 3.5.

Lemma A.1. The SQP direction d_k admits the following decomposition

$$d_k = P_k d_k + d'_k \tag{A.5}$$

where $d'_k = R_k(R_k^T R_k)^{-1}\bar{f}_1(x_k)$. Also, there exists $c_2 > 0$ such that, for k large enough,

$$\|\bar{f}_j(x_k)\| \geq c_2\|d'_k\|, \quad j = 1, \cdots, m. \tag{A.6}$$

Furthermore, there exists $c_3 > 0$ such that, for k large enough and for all $j_k \in I(x^*)$ such that $f_{j_k}(x_k) = f(x_k)$,

$$\langle \bar{f}_{j_k}(x_k), \bar{\lambda}_k \rangle \leq -c_3\|d'_k\| \tag{A.7}$$

where $\bar{\lambda}_k = [\lambda_{k,i} : \forall i \in I(x^*)\backslash\{j_k\}]^T$, with components in the same order as those of $\bar{f}_{j_k}(x_k)$.

Proof. In view of Proposition 3.4(ii), d_k that solves $QP(x_k, H_k)$ satisfies, for k large enough, the following set of linear equations (since $I(x^*) = \{1, \cdots, m\}$)

$$f_i(x_k) + \langle \nabla f_i(x_k), d_k \rangle = f_1(x_k) + \langle \nabla f_1(x_k), d_k \rangle, \quad i = 2, \cdots, m$$

which implies

$$R_k^T d_k = \bar{f}_1(x_k).$$

Since, from the definition of P_k in A6, we have

$$P_k d_k = d_k - R_k(R_k^T R_k)^{-1} R_k^T d_k,$$

(A.5) follows. Since, from assumptions A1, A2 and A6, R_k and $(R_k^T R_k)^{-1}$ are bounded for large k, in view of Remark 3.1, (A.6) follows directly from the definition of d'_k.

Now for any $j_k \in I(x^*)$ such that $f_{j_k}(x_k) = f(x_k)$, $\bar{f}_{j_k}(x_k) \leq 0$. Also Proposition 3.4$(ii)$ implies there exists $\underline{\lambda} > 0$ such that, for k large enough,

$$\min_{j \in I(x^*)} \lambda_{k,j} \geq \underline{\lambda}.$$

Therefore, in view of (A.6), we have

$$\begin{aligned}\langle \bar{f}_{j_k}(x_k), \bar{\lambda}_k \rangle &\leq -\underline{\lambda} \|\bar{f}_{j_k}(x_k)\| \\ &\leq -\underline{\lambda} c_2 \|d_k'\|\end{aligned}$$

and (A.7) follows. □

Proof of Proposition 3.5. Throughout the proof, the phrase "for k large enough" is implicit. We show that

$$f(x_k + d_k + \tilde{d}_k) \leq f(x_k) - \alpha \langle d_k, H_k d_k \rangle, \tag{A.8}$$

which clearly implies the claim. In view of (A.1), it follows that

$$f_i(x_k + d_k) + \langle \nabla f_i(x_k), \tilde{d}_k \rangle = f_j(x_k + d_k) + \langle \nabla f_j(x_k), \tilde{d}_k \rangle \quad \forall i, j \in I(x^*) .$$

This implies, in view of A1' and (18), that

$$f_i(x_k + d_k + \tilde{d}_k) = f_j(x_k + d_k + \tilde{d}_k) + O(\|d_k\|^3) \quad \forall i, j \in I(x^*) ,$$

which in turn implies

$$f(x_k + d_k + \tilde{d}_k) = f_i(x_k + d_k + \tilde{d}_k) + O(\|d_k\|^3) \quad \forall i \in I(x^*) .$$

Multiplying both sides of this equation by the corresponding $\lambda_{k,i}$ and summing up over all $i \in I(x^*)$ yields

$$\begin{aligned}f(x_k + d_k + \tilde{d}_k) &= \sum_{i \in I(x^*)} \lambda_{k,i} f_i(x_k + d_k + \tilde{d}_k) + O(\|d_k\|^3) \\ &= L(x_k + d_k + \tilde{d}_k, \lambda_k) + O(\|d_k\|^3) .\end{aligned}$$

Expanding L around x_k gives

$$f(x_k + d_k + \tilde{d}_k) = L(x_k, \lambda_k) + \langle \nabla_x L(x_k, \lambda_k), d_k + \tilde{d}_k \rangle + \frac{1}{2} \langle d_k, \nabla^2_{xx} L(x_k, \lambda_k) d_k \rangle + O(\|d_k\|^3) .$$

Since, for any $j \in I(x^*)$,

$$\begin{aligned}L(x_k, \lambda_k) &= \sum_{i \in I(x^*)} \lambda_{k,i} f_i(x_k) = f_j(x_k) + \sum_{i \in I(x^*)} \lambda_{k,i} \{ f_i(x_k) - f_j(x_k) \} \\ &= f_j(x_k) + \langle \bar{f}_j(x_k), \bar{\lambda}_k \rangle\end{aligned}$$

and, in view of (10) and (18),

$$\langle \nabla_x L(x_k, \lambda_k), d_k + \tilde{d}_k \rangle = -\langle d_k, H_k d_k \rangle + O(\|d_k\|^3)\,,$$

the above expression becomes

$$f(x_k + d_k + \tilde{d}_k) = f_{j_k}(x_k) + \langle \bar{f}_{j_k}(x_k), \bar{\lambda}_k \rangle - \langle d_k, H_k d_k \rangle + \frac{1}{2}\langle d_k, \nabla^2_{xx} L(x_k, \lambda_k) d_k \rangle + O(\|d_k\|^3)\,,$$

with j_k such that $f_{j_k}(x_k) = f(x_k)$. It follows that, in view of Lemma A.1,

$$\begin{aligned} f(x_k + d_k + \tilde{d}_k) &\leq f(x_k) - c_3\|d'_k\| - \tfrac{1}{2}\langle d_k, H_k d_k \rangle + \tfrac{1}{2}\langle d_k, \{\nabla^2_{xx} L(x_k, \lambda_k) - H_k\} d_k \rangle + O(\|d_k\|^3) \\ &= f(x_k) - \alpha\langle d_k, H_k d_k \rangle - c_3\|d'_k\| - (\tfrac{1}{2} - \alpha)\langle d_k, H_k d_k \rangle \\ &\quad + \tfrac{1}{2}\langle d_k, P_k\{\nabla^2_{xx} L(x_k, \lambda_k) - H_k\} P_k d_k \rangle + o(\|d'_k\|) + O(\|d_k\|^3). \end{aligned}$$

Therefore, (A.8) follows in view of (19) and Assumption A3, since $\alpha \in (0, \frac{1}{2})$ and $c_3 > 0$. □

References

[1] R.A. Polyak, "Smooth Optimization Methods for Minimax Problems," *SIAM J. Control Optim.* 26 (1988), 1274–1286.

[2] E. Polak, D.Q. Mayne & J.E. Higgins, "A Superlinearly Convergent Algorithm for Min-max Problems," *Proceedings of the 28th IEEE Conference on Decision and Control* (December 1989).

[3] A.R. Conn & Y. Li, "An Efficient Algorithm for Nonlinear Minimax Problems," University of Waterloo, Research Report CS-88-41, Waterloo, Ontario, N2L 3G1 Canada, November, 1989 .

[4] S.P. Han, "Superlinear Convergence of a Minimax Method," Department of Computer Science, Cornell University, TR78-336, 1978.

[5] S.P. Han, "Variable Metric Methods for Minimizing a Class of Nondifferentiable Functions," *Math. Programming* 20 (1981), 1–13.

[6] A.R. Conn, "An Efficient Second Order Method to Solve the (Constrained) Minimax Problem," Department of Combinatorics and Optimization, Research Report CORR 79-5, University of Waterloo, Ontario, 1979.

[7] W. Murray & M.L. Overton, "A Projected Lagrangian Algorithm for Nonlinear Minimax Optimization," *SIAM J. Sci. Stat. Comput.* 1 (September, 1980).

[8] M.L. Overton, "Algorithms for Nonlinear ℓ_1 and ℓ_∞ Fitting," in *Nonlinear Optimization 1981*, M.J.D. Powell, ed., Acadamic Press, London, 1982, 213–221.

[9] R.S. Womersley & R. Fletcher, "An Algorithm for Composite Nonsmooth Optimization Problems," *J. Optim. Theory Appl.* 48 (1986), 493–523.

[10] R.M. Chamberlain, M.J.D. Powell, C. Lemarechal & H.C. Pedersen, "The Watchdog Technique for Forcing Convergence in Algorithms for Constrained Optimization," *Math. Programming Stud.* 16 (1982), 1–17.

[11] R. Fletcher, "Second Order Corrections for Non-differentiable Optimization," in *Numerical Analysis, Dundee, 1981, Lecture Notes in Mathematics 912*, G.A. Watson, ed., Springer-Verlag, 1981, 85–114.

[12] D. Q. Mayne & E. Polak, "A Superlinearly Convergent Algorithm for Constrained Optimization Problems," *Math. Programming Stud.* 16 (1982), 45–61.

[13] L. Grippo, F. Lampariello & S. Lucidi, "A Nonmonotone Line Search Technique for Newton's Method," *SIAM J. Numer. Anal.* 23 (1986), 707–716.

[14] E.R. Panier & A.L. Tits, "Avoiding the Maratos Effect by Means of a Nonmonotone Line Search. I. General Constrained Problems," *SIAM J. Numer. Anal.* 28 (August 1991, to appear).

[15] J.F. Bonnans, E.R. Panier, A.L. Tits & J. Zhou, "Avoiding the Maratos Effect by Means of a Nonmonotone Line Search. II. Inequality Constrained Problems – Feasible Iterates," Systems Research Center, University of Maryland, Technical Report SRC-TR-89-42r1, College Park, MD 20742, 1989.

[16] E.R. Panier & A.L. Tits, "A Superlinearly Convergent Method of Feasible Directions for Optimization Problems Arising in the Design of Engineering Systems," in *Proc. of the Seventh International Conference on Analysis and Optimization of Systems — Antibes, June 25-27, 1986*, A. Bensoussan & J.L. Lions, eds., Lecture Notes in Control and Information Sciences #83, Springer Verlag, Berlin-Heidelberg-New York-Tokyo, June 1986, 65–73.

[17] M.J.D. Powell, "A Fast Algorithm for Nonlinearly Constrained Optimization Calculations," in *Numerical Analysis, Dundee, 1977, Lecture Notes in Mathematics 630*, G.A. Watson, ed., Springer-Verlag, 1978, 144–157.

[18] M.J.D. Powell, "The Convergence of Variable Metric Methods for Nonlinearly Constrained Optimization Calculations," in *Nonlinear Programming 3*, O.L. Mangasarian, R.R. Meyer & S.M. Robinson, eds., Academic Press, New York, 1978, 27–63.

[19] J. Stoer, "The Convergence of Sequential Quadratic Programming Methods for Solving Nonlinear Programs," in *Recent Advances in Communication and Control Theory*, R.E. Kalman, G.I. Marchuk, A.E. Ruberti & A.J. Viterbi, eds., Optimization Software, Inc., New York, N.Y., 1987, 412–421.

[20] J. Zhou & A.L. Tits, "User's Guide for FSQP Version 2.1 – A Fortran Code for Solving Optimization Problems, Possibly Minimax, with General Inequality Constraints and Linear Equality Constraints, Generating Feasible Iterates," Systems Research Center, University of Maryland, SRC TR-90-60r1a, College Park, MD 20742, 1991.

[21] G.A. Watson, "The Minimax Solution of an Overdetermined System of Non-linear Equations," *J. Inst. Math. Appl.* 23 (1979), 167–180.

[22] C. Charalambous & A.R. Conn, "An Efficient Method to Solve the Minimax Problem Directly," *SIAM J. Numer. Anal.* 15 (1978), 162–187.

[23] K. Madsen & H. Schjær-Jacobsen, "Linearly Constrained Minimax Optimization," *Math. Programming* 14 (1978), 208–223.

[24] J. Zhou & A.L. Tits, "User's Guide for FSQP Version 2.0 – A Fortran Code for Solving Optimization Problems, Possibly Minimax, with General Inequality Constraints and Linear Equality Constraints, Generating Feasible Iterates," Systems Research Center, University of Maryland, SRC TR-90-60r1, College Park, MD 20742, 1990.

[25] W. Hock & K. Schittkowski, *Test Examples for Nonlinear Programming Codes*, Lecture Notes in Economics and Mathematical Systems (187), Springer Verlag, 1981.

PROB	CODE	NOBJ	NMF	ITER	OBJMAX	KKT	EPS
BARD	CL	15	10	*	*	*	.50E-05
	ALS		15	8	.508163265E−01	.63E-10	.50E-05
	NLS		7	7	.508168686E−01	.42E-05	.50E-05
CB2	CL	3	11	*	*	*	.50E-05
	ALS		11	6	.195222453E+01	.10E-06	.50E-05
	NLS		6	6	.195222453E+01	.82E-06	.50E-05
CB3	CL	3	6	*	*	*	.50E-05
	ALS		5	3	.200000000E+01	.75E-06	.50E-05
	NLS		5	5	.200000000E+01	.94E-09	.50E-05
COLV	CL	6	49	*	*	*	.50E-05
	ALS		21	21	.323486790E+02	.29E-05	.50E-05
	NLS		17	17	.323486790E+02	.14E-05	.50E-05
DAVD2	CL	20	20	*	*	*	.50E-05
	ALS		20	10	.115706440E+03	.59E-06	.50E-05
	NLS		11	10	.115706440E+03	.93E-06	.50E-05
F&R	CL	2	11	*	*	*	.50E-05
	ALS		17	9	.494895210E+01	.24E-06	.50E-05
	NLS		10	10	.494895210E+01	.21E-06	.50E-05
HETTICH	CL	5	11	*	*	*	.50E-05
	ALS		19	10	.245935695E−02	.28E-05	.50E-05
	NLS		11	10	.245939485E−02	.19E-05	.50E-05
R-S	CL	4	12	*	*	*	.50E-05
	ALS		22	9	−.440000000E+02	.13E-05	.50E-05
	NLS		16	10	−.440000000E+02	.99E-07	.50E-05
WATS-6	CL	31	24	*	*	*	.50E-05
	ALS		23	12	.127170954E−01	.14E-05	.50E-05
	NLS		14	13	.127170913E−01	.31E-08	.50E-05
WATS-20	CL	31	22	*	*	*	.50E-05
	ALS		106	42	.138908355E−07	.35E-06	.50E-05
	NLS		45	43	.141191856E−07	.17E-06	.50E-05
WONG	CL	5			*	*	.50E-05
	ALS		67	20	.680630057E+03	.12E-05	.50E-05
	NLS		49	26	.680630057E+03	.42E-05	.50E-05
MAD1	MS	3	*	8	*	*	.10E-11
	ALS		9	5	−.389659516E+00	.35E-16	.10E-11
	NLS		6	6	−.389659516E+00	.89E-10	.10E-11
MAD2	MS	3	*	*	*	*	.10E-11
	ALS		21	11	−.330357143E+00	.13E-10	.10E-11
	NLS		19	18	−.330357143E+00	.81E-10	.10E-11
MAD4	MS	3	*	8	*	*	.10E-11
	ALS		11	6	−.448910786E+00	.90E-16	.10E-11
	NLS		8	8	−.448910786E+00	.90E-16	.10E-11
MAD5	MS	3	*	8	*	*	.10E-11
	ALS		13	7	−.100000000E+01	.16E-16	.10E-11
	NLS		8	8	−.100000000E+01	.35E-13	.10E-11
MAD6	MS	163	8	*	.113105 E+00	*	.10E-11
	ALS		11	6	.113104635E+00	.20E-10	.10E-11
	NLS		8	8	.113104727E+00	.72E-15	.10E-11
MAD8-10	MS	18	18	*	*	*	.10E-11
	ALS		19	10	.381173963E+00	.99E-12	.10E-11
	NLS		14	14	.381173963E+00	.22E-15	.10E-11
MAD8-30	MS	58	17	*	*	*	.10E-11
	ALS		30	15	.547620496E+00	.21E-15	.10E-11
	NLS		20	18	.547620496E+00	.21E-10	.10E-11
MAD8-50	MS	98	18	*	*	*	.10E-11
	ALS		39	20	.579276202E+00	.20E-15	.10E-11
	NLS		21	21	.579276202E+00	.22E-13	.10E-11

NOBJ: number of objective functions.

NMF: number of evaluations of the max function.

ITER: number of iterations.

OBJMAX: (absolute) max value of the objective functions.

KKT: norm of KKT vector (the gradient of the Lagrangian) at the final iterate.

Table 1

PROB	CODE	NOBJ	NNL	NMF	NG	ITER	OBJMAX	KKT	EPS
P43M	FSQP-ALS	3	1	27	36	14	−.440000000E+02	.30E-06	.50E-05
	FSQP-NLS			20	25	16	−.440000000E+02	.39E-05	.50E-05
P84M	FSQP-ALS	3	4	7	28	4	−.528033513E+07	.0	.50E-05
	FSQP-NLS			3	12	3	−.528033513E+07	.37E-03	.50E-05
P113M	FSQP-ALS	4	5	25	142	13	.243062091E+02	.31E-05	.50E-05
	FSQP-NLS			21	115	15	.243062091E+02	.31E-05	.50E-05
P117M	FSQP-ALS	3	3	48	124	21	.323486790E+02	.46E-05	.50E-05
	FSQP-NLS			19	54	17	.323486790E+02	.26E-04	.50E-05

Table 2