INT. J. SYSTEMS SCI., 1976, VOL. 7, NO. 4, 377–391

# Non-linear minimax optimization as a sequence of least $p$th optimization with finite values of $p$

C. CHARALAMBOUS† and J. W. BANDLER‡

Following developments in non-linear least $p$th optimization by the authors it is possible to derive two new methods of non-linear minimax optimization. Unlike the Polya algorithm in which a sequence of least $p$th optimizations as $p \to \infty$ is taken our methods do not require the value of $p$ to tend to infinity. Instead we construct a sequence of least $p$th optimization problems with a finite value of $p$. It is shown that this sequence will converge to a minimax solution. Two interesting minimax problems were constructed which illustrate some of the theoretical ideas. Further numerical evidence is presented on the modelling of a fourth-order system by a second-order model with values of $p$ varying between 2 and 10 000.

## 1. Introduction

Various algorithms have been proposed for solving the discrete non-linear minimax problem, some of the most relevant of which are due to Waren *et al.* (1967), Osborne and Watson (1969), Bandler *et al.* (1972) and Bandler and Charalambous (1972).

The first method transforms the non-linear minimax optimization problem into a non-linear programming problem and solves it by well-established methods such as the one by Fiacco and McCormick (1964). The second method deals with minimax formulations by following two steps—a linear programming part which provides a given step in the parameter space, followed by a linear search along the direction of the step. The third method uses gradient information of one or more of the functions to get a downhill direction by solving a suitable linear programming problem. A linear search follows to find the minimum in that direction, and the procedure is repeated. The last method is a generalization of the Polya algorithm (Rice 1969). A $p$th norm-like function is formed which has the property that, if $p = \infty$, the function is equal to the maximum of the set of functions which we want to minimize.

In this paper two new algorithms are presented in which a sequence of least $p$th optimization problems is constructed with a constant value of $p$ in the range $1 < p < \infty$. It is shown that this sequence will converge to a minimax solution. Numerical evidence is presented to show that the scheme works well in practice.

## 2. The problem

Consider a system of $m$ real non-linear functions

$$f_i(\boldsymbol{\phi}), \quad i \in I \tag{1}$$

where $\boldsymbol{\phi} \triangleq [\phi_1 \; \phi_2 \; \ldots \; \phi_k]^{\mathrm{T}}$ is a $k$-dimensional column vector containing the

Received 19 November 1974.

† Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Canada.

‡ Group on Simulation, Optimization and Control and Department of Electrical Engineering, McMaster University, Hamilton, Canada.

$k$ adjustable parameters and $I \triangleq \{1, 2, \ldots, m\}$. Let

$$M_f(\boldsymbol{\phi}) = \max_{i \in I} f_i(\boldsymbol{\phi}) \tag{2}$$

The problem of minimax optimization of system (1) consists of finding a point $\check{\boldsymbol{\phi}}$ such that $M_f(\check{\boldsymbol{\phi}}) \leqslant M_f(\boldsymbol{\phi})$ for all points $\boldsymbol{\phi}$ at least in the neighbourhood of $\check{\boldsymbol{\phi}}$.

### 2.1. *Assumptions*

(*a*) We assume that $M_f(\boldsymbol{\phi})$ is bounded below, i.e. we assume the existence of greatest lower bound $M_f(\check{\boldsymbol{\phi}})$ such that $M_f(\boldsymbol{\phi}) \geqslant M_f(\check{\boldsymbol{\phi}}) > -\infty$.

(*b*) The set $S \triangleq \{\boldsymbol{\phi} \mid M_f(\boldsymbol{\phi}) \leqslant c\}$ is bounded for every finite value of $c$. This ensures that any local minimum is located at a finite point.

(*c*) The functions $f_i(\boldsymbol{\phi})$ for $i \in I$ belong to class $C^1$ (once continuously differentiable).

### 2.2. *Definitions*

Consider the following objective function

$$\begin{aligned} U(\boldsymbol{\phi}, \xi) &= M(\boldsymbol{\phi}, \xi) \left[ \sum_{i \in K} \left( \frac{f_i(\boldsymbol{\phi}) - \xi}{M(\boldsymbol{\phi}, \xi)} \right)^q \right]^{1/q} \quad \text{for } M(\boldsymbol{\phi}, \xi) \neq 0 \\ &= 0 \quad \text{for } M(\boldsymbol{\phi}, \xi) = 0 \end{aligned} \tag{3}$$

where

$$M(\boldsymbol{\phi}, \xi) \triangleq \max_{i \in I} (f_i(\boldsymbol{\phi}) - \xi) = M_f(\boldsymbol{\phi}) - \xi \tag{4}$$

$$q = p \times \operatorname{sign} M(\boldsymbol{\phi}, \xi) \tag{5}$$

where $p$ has a *constant* value in the range $1 < p < \infty$.

$$K = \begin{cases} J(\boldsymbol{\phi}, \xi) \triangleq \{i \mid f_i(\boldsymbol{\phi}) - \xi \geqslant 0, \quad i \in I\} & \text{if } M(\boldsymbol{\phi}, \xi) > 0 \\ I \quad \text{if } M(\boldsymbol{\phi}, \xi) < 0 & \end{cases} \tag{6}$$

The objective function given in (3) is a generalization of the usual least $p$th objective function. Under the assumptions (*a*) and (*b*), the continuity of $f_i(\boldsymbol{\phi})$ for $i \in I$ and because $U(\boldsymbol{\phi}, \xi) \geqslant M(\boldsymbol{\phi}, \xi)$ (see Lemmas 3.1 and 3.4) the objective function $U(\boldsymbol{\phi}, \xi)$ is continuous and has a minimum which is located at a finite point. Also, due to assumption (*c*), $U(\boldsymbol{\phi}, \xi)$ has continuous first partial derivatives *except* when both $M(\boldsymbol{\phi}, \xi) = 0$ and two or more of the functions $(f_i(\boldsymbol{\phi}) - \xi)$ for $i \in I$ are equal to zero.

The reason why all the functions $(f_i(\boldsymbol{\phi}) - \xi)$ for $i \in K$ are normalized with respect to their maximum is to avoid numerical difficulties arising from the use of large values of $p$.

The symbols $\epsilon$ and $\eta$ will be used to denote small positive numbers.

## 3. The new algorithms

### 3.1. *Algorithm* 1

(1) Assume a starting point $\boldsymbol{\phi}^0$ is given; set $\xi^1 = \min [0, M_f(\boldsymbol{\phi}^0)]$, $r = 1$ and select a value of $p > 1$.

(2) Minimize with respect to $\boldsymbol{\phi}$ the objective function $U(\boldsymbol{\phi}, \xi)$ for $\xi = \xi^r$. Let $\check{\boldsymbol{\phi}}^r$ denote the optimum parameter vector of $U(\boldsymbol{\phi}, \xi)$ at the $r$th optimization.

(3) Set

$$\xi^{r+1} = M_f(\check{\boldsymbol{\phi}}^r) \tag{7}$$

(4) Convergence criterion: if $|\xi^{r+1} - \xi^r| < \eta$ stop; otherwise set $r = r + 1$ and go to 2.

3.2. *Algorithm* 2

(1) As in Algorithm 1.

(2) As in Algorithm 1.

(3) If $M(\check{\boldsymbol{\phi}}^r, \xi^r) < 0$ remain with Algorithm 1; otherwise set

$$\begin{aligned} \xi^{r+1} &= \xi^r + \lambda^r M(\check{\boldsymbol{\phi}}^r, \xi^r) \\ &= (1 - \lambda^r)\xi^r + \lambda^r M_f(\check{\boldsymbol{\phi}}^r) \end{aligned} \tag{8}$$

where

$$0 < \lambda^r < 1 \tag{9}$$

(4) As in Algorithm 1.

3.3. *Comments*

It is important to note that for both algorithms the value of $p$ is kept *constant* in the range $1 < p < \infty$, unlike the algorithm presented by Bandler and Charalambous (1972, 1973) where the value of $p$ must be very large. Algorithm 2 is different from Algorithm 1 if $M_f(\check{\boldsymbol{\phi}}) > 0$, otherwise it is the same. The main difference is that in the Algorithm 1 we try to push the maximum away from the level $\xi^r$ at the $r$th iteration (this causes $M(\check{\boldsymbol{\phi}}^r, \xi^r) < 0$, and $\xi^{r+1} < \xi^r$ for $r \geqslant 2$), while in Algorithm 2 we try to predict the value of $M_f(\check{\boldsymbol{\phi}})$ by increasing the value of $\xi^r$ from zero appropriately (this causes, $M(\check{\boldsymbol{\phi}}^r, \xi^r) > 0$, and $\xi^{r+1} > \xi^r$ as long as we stay with Algorithm 2). Due to the fact that the minimax solution of the set of functions $f_i(\boldsymbol{\phi})$ for $i \in I$ and $f_i(\boldsymbol{\phi}) + \beta$ for $i \in I$ does not change when $\beta$ is constant it will be possible to use Algorithm 2 even when $M_f(\check{\boldsymbol{\phi}}) \leqslant 0$ but we have to raise all the $f_i(\boldsymbol{\phi})$ for $i \in I$ by an amount $\beta > |M_f(\check{\boldsymbol{\phi}})|$.

The first step of Algorithm 1 ($\xi^1 = \min [0, M_f(\boldsymbol{\phi}^0)]$) could be modified to $\xi^1 = M_f(\boldsymbol{\phi}^0)$. A reason for not modifying it is the following. In engineering problems (e.g. filter design (Bandler and Charalambous 1972)) the sign of $M_f(\boldsymbol{\phi})$ indicates whether a particular structure can satisfy certain design specifications. That is, if,

$$M_f(\boldsymbol{\phi}) \begin{cases} > 0 & \text{the specifications are violated} \\ = 0 & \text{the specifications are just met} \\ < 0 & \text{the specifications are satisfied} \end{cases}$$

By using $\xi^1 = \min [0, M_f(\boldsymbol{\phi}^0)]$ the first optimum of $U(\boldsymbol{\phi}, \xi)$ (i.e. $\check{\boldsymbol{\phi}}^1$) yields the above.

3.4. *Convergence proofs for Algorithm* 1

*Lemma* 3.1

If $y_i \geqslant 0$ for $i \in I$ and $p \geqslant 1$, then

$$m^{-(1/p)} \min_{i \in I} y_i \leqslant \left( \sum_{i \in I} y_i^{-p} \right)^{-(1/p)} \leqslant \min_{i \in I} y_i$$

The proof is simple and is omitted.

*Lemma* 3.2

Let $y_i$ for $i \in I$ be a set of real numbers and $x \geqslant \max_{i \in I} y_i$. Then

$$U(x) = -\left(\sum_{i \in I} (x - y_i)^{-p}\right)^{-(1/p)}, \quad p \geqslant 1$$

decreases as $x$ increases and, moreover, it is convex.

*Proof*

$$\frac{dU(x)}{dx} = -\left(\sum_{i \in I} (x - y_i)^{-p}\right)^{-(1/p)-1} \sum_{i \in I} (x - y_i)^{-p-1} < 0 \quad \text{for } x > \max_{i \in I} y_i$$

Note that the maximum value of $U(x)$ is zero.

Let $x^{(1)}$ and $x^{(2)}$ be two distinct points such that $x^{(1)}, x^{(2)} \geqslant \max_{i \in I} y_i$ and $0 \leqslant \lambda \leqslant 1$. Then,

$$\begin{aligned} -U((1-\lambda)x^{(1)} + \lambda x^{(2)}) &= \left(\sum_{i \in I} ((1-\lambda)x^{(1)} + \lambda x^{(2)} - y_i)^{-p}\right)^{-(1/p)} \\ &= \left(\sum_{i \in I} ((1-\lambda)(x^{(1)} - y_i) + \lambda(x^{(2)} - y_i))^{-p}\right)^{-(1/p)} \\ &\geqslant (1-\lambda)\left(\sum_{i \in I} (x^{(1)} - y_i)^{-p}\right)^{-(1/p)} \\ &\qquad + \lambda\left(\sum_{i \in I} (x^{(2)} - y_i)^{-p}\right)^{-(1/p)} \end{aligned}$$

See Hardy *et al.* (1934) for the last inequality. Therefore, convexity follows.

*Lemma* 3.3

For $r \geqslant 2$, $|U(\check{\boldsymbol{\phi}}^r, \xi^r)| \geqslant |U(\check{\boldsymbol{\phi}}^{r+1}, \xi^{r+1})|$.

*Proof*

For $r \geqslant 2$ we have $M(\check{\boldsymbol{\phi}}^r, \xi^r) < 0$ and therefore $q = -p$. In this case

$$\begin{aligned} U(\check{\boldsymbol{\phi}}^r, \xi^r) &= -\left(\sum_{i \in I} (\xi^r - f_i(\check{\boldsymbol{\phi}}^r))^{-p}\right)^{-(1/p)} \\ &\leqslant U(\check{\boldsymbol{\phi}}^{r+1}, \xi^r) \end{aligned}$$

(because $\check{\boldsymbol{\phi}}^r$ is the optimum parameter vector of $U(\boldsymbol{\phi}, \xi)$ with respect to the level $\xi^r$)

$$\leqslant U(\check{\boldsymbol{\phi}}^{r+1}, \xi^{r+1})$$

because $\xi^{r+1} \leqslant \xi^r$ and due to Lemma 3.2.

*Theorem* 3.1

$|U(\check{\boldsymbol{\phi}}^r, \xi^r)| \to 0$ as $r \to \infty$.

*Proof*

For $r \geqslant 2$

$$\begin{aligned} |U(\check{\boldsymbol{\phi}}^r, \xi^r)| &= \left(\sum_{i \in I} (\xi^r - f_i(\check{\boldsymbol{\phi}}^r))^{-p}\right)^{-(1/p)} \\ &\leqslant \min_{i \in I} (\xi^r - f_i(\check{\boldsymbol{\phi}}^r)) = \xi^r - \max_{i \in I} f_i(\check{\boldsymbol{\phi}}^r) \end{aligned}$$

(from Lemma 3.1)

$$= \xi^r - \xi^{r+1}$$

Therefore,

$$\lim_{i\to\infty} \sum_{r=2}^{i} |U(\check{\boldsymbol{\phi}}^i, \xi^i)| \leqslant \lim_{i\to\infty} (\xi^2 - \xi^{i+1})$$

$$\leqslant \xi^2 = M_{\mathrm{f}}(\check{\boldsymbol{\phi}}^1)$$

Therefore,

$$\lim_{r\to\infty} |U(\check{\boldsymbol{\phi}}^r, \xi^r)| \to 0 \tag{10}$$

*Theorem* 3.2

As $r\to\infty$, $M_{\mathrm{f}}(\check{\boldsymbol{\phi}}^r)\to M_{\mathrm{f}}(\check{\boldsymbol{\phi}})$.

*Proof*

Assume that as $r\to\infty$, $M_{\mathrm{f}}(\check{\boldsymbol{\phi}}^r)\to L_{\mathrm{f}} \geqslant M_{\mathrm{f}}(\check{\boldsymbol{\phi}})$. We must show that $L_{\mathrm{f}} = M_{\mathrm{f}}(\check{\boldsymbol{\phi}})$. Assume $L_{\mathrm{f}} > M_{\mathrm{f}}(\check{\boldsymbol{\phi}})$. Because $M_{\mathrm{f}}(\boldsymbol{\phi})$ is continuous it is possible to find a point $\overline{\boldsymbol{\phi}}$ such that

$$M_{\mathrm{f}}(\check{\boldsymbol{\phi}}) < M_{\mathrm{f}}(\overline{\boldsymbol{\phi}}) < L_{\mathrm{f}} \tag{11}$$

In other words

$$f_i(\overline{\boldsymbol{\phi}}) - L_{\mathrm{f}} < 0, \quad i \in I$$

Since

$$U(\check{\boldsymbol{\phi}}^r, \xi^r) = 0 \quad \text{as } r\to\infty$$

$$\lim_{r\to\infty} \xi^r = L_{\mathrm{f}}$$

But

$$U(\overline{\boldsymbol{\phi}}, L_{\mathrm{f}}) = -\Big( \sum_{i\in I} (L_{\mathrm{f}} - f_i(\overline{\boldsymbol{\phi}}))^{-p}\Big)^{-(1/p)} < 0$$

This contradicts the fact that $\check{\boldsymbol{\phi}}^r$ minimizes $U$ for $r\to\infty$ with respect to $L_{\mathrm{f}}$.

*Theorem* 3.3

As $r\to\infty$, the necessary conditions for a minimax optimum are satisfied (Bandler 1971), that is,

$$\sum_{i\in\hat{J}} u_i \nabla f_i(\check{\boldsymbol{\phi}}^\infty) = \mathbf{0} \tag{12 a}$$

$$u_i \geqslant 0, \quad i \in \hat{J} \tag{12 b}$$

$$\sum_{i\in\hat{J}} u_i > 0 \tag{12 c}$$

where

$$\hat{J} \triangleq \{i \mid f_i(\check{\boldsymbol{\phi}}^\infty) = M_{\mathrm{f}}(\check{\boldsymbol{\phi}}^\infty), \quad i \in I\} \tag{12 d}$$

and

$$\nabla \triangleq \left[\frac{\partial}{\partial\phi_1} \frac{\partial}{\partial\phi_2} \cdots \frac{\partial}{\partial\phi_k}\right]^{\mathrm{T}} \tag{13}$$

*Proof*

Since a necessary condition that a point be a local minimum of an unconstrained function is that the first partial derivatives vanish then for $r \geqslant 2$,

$$\nabla U(\check{\boldsymbol{\phi}}^r, \xi^r) = \Big( \sum_{i\in I} (\xi^r - f_i(\check{\boldsymbol{\phi}}^r))^{-p}\Big)^{-(1/p)-1} \cdot \sum_{i\in I} (\xi^r - f_i(\check{\boldsymbol{\phi}}^r))^{-p-1} \nabla f_i(\check{\boldsymbol{\phi}}^r)$$

$$= A \sum_{i \in I} \left( \frac{\xi^r - f_i(\check{\boldsymbol{\phi}}^r)}{\xi^r - \xi^{r+1}} \right)^{-p-1} \nabla f_i(\check{\boldsymbol{\phi}}^r)$$
$$= \mathbf{0}$$

where

$$A = \left[ \sum_{i \in I} \left( \frac{\xi^r - f_i(\check{\boldsymbol{\phi}}^r)}{\xi^r - \xi^{r+1}} \right)^{-p} \right]^{-(1/p)-1} \tag{14}$$

Since $\xi^r - \xi^{r+1} = \min_{i \in I} (\xi^r - f_i(\check{\boldsymbol{\phi}}^r))$, $A \neq 0$ and therefore,

$$\sum_{i \in I} \left( \frac{\xi^r - f_i(\check{\boldsymbol{\phi}}^r)}{\xi^r - \xi^{r+1}} \right)^{-p-1} \nabla f_i(\check{\boldsymbol{\phi}}^r) = \mathbf{0} \tag{15}$$

Let

$$\mu_i^r = \left( \frac{\xi^r - \xi^{r+1}}{\xi^r - f_i(\check{\boldsymbol{\phi}}^r)} \right)^{p+1}, \quad i \in I \tag{16}$$

then

$$\sum_{i \in I} \mu_i^r \nabla f_i(\check{\boldsymbol{\phi}}^r) = \mathbf{0} \tag{17}$$

Note that $0 \leqslant \mu_i^r \leqslant 1$ for $i \in I$ and at least one of them is equal to one. Let

$$u_i = \lim_{r \to \infty} \mu_i^r, \quad i \in I \tag{18}$$

then it is clear from Theorem 3.1 that $\lim_{r \to \infty} (\xi^r - \xi^{r+1}) \to 0$, therefore

$$u_i \begin{cases} = 0, & i \notin \hat{J} \\ \geqslant 0, & i \in \hat{J} \end{cases} \tag{19}$$

and

$$\sum_{i \in \hat{J}} u_i > 0 \tag{20}$$

Therefore,

$$\sum_{i \in \hat{J}} u_i \nabla f_i(\check{\boldsymbol{\phi}}^\infty) = \mathbf{0} \tag{21}$$

3.5. *Convergence proofs for Algorithm* 2

*Lemma* 3.4

Let $y_i$ for $i \in I$ be a set of real numbers such that $\max_{i \in I} y_i \geqslant 0$, then

$$\max_{i \in I} y_i \leqslant \left( \sum_{i \in L} y_i^p \right)^{1/p} \leqslant m^{1/p} \max_{i \in I} y_i, \quad p \geqslant 1$$

where

$$L \triangleq \{ i \mid y_i \geqslant 0, \quad i \in I \}$$

The proof is simple and is omitted.

*Lemma* 3.5

Let $U(\check{\boldsymbol{\phi}}^r, \xi^r)$, $U(\check{\boldsymbol{\phi}}^{r+1}, \xi^{r+1}) > 0$. Then

$$U(\check{\boldsymbol{\phi}}^{r+1}, \xi^{r+1}) \leqslant U(\check{\boldsymbol{\phi}}^r, \xi^r)$$

*Proof*

For the case considered $M(\check{\boldsymbol{\phi}}^r, \xi^r) > 0$, $M(\check{\boldsymbol{\phi}}^{r+1}, \xi^{r+1}) > 0$ and therefore $q = p$. In this case

$$U(\breve{\phi}^r, \xi^r) = \Big(\sum_{i \in J(\breve{\phi}^r, \xi^r)} (f_i(\breve{\phi}^r) - \xi^r)^p\Big)^{1/p}$$

$$\geqslant \Big(\sum_{i \in J(\breve{\phi}^r, \xi^{r+1})} (f_i(\breve{\phi}^r) - \xi^{r+1})^p\Big)^{1/p}$$

(the inequality is due to the fact that $\xi^{r+1} \geqslant \xi^r$ and $J(\breve{\phi}^r, \xi^{r+1}) \subseteq J(\breve{\phi}^r, \xi^{r+1})$)

$$\geqslant \Big(\sum_{i \in J(\breve{\phi}^{r+1}, \xi^{r+1})} (f_i(\breve{\phi}^{r+1}) - \xi^{r+1})^p\Big)^{1/p}$$

(because $\breve{\phi}^{r+1}$ is the optimum parameter vector with respect to the level $\xi^{r+1}$).

$$= U(\breve{\phi}^{r+1}, \xi^{r+1})$$

*Theorem* 3.4

$U(\breve{\phi}^r, \xi^r) \to 0$ as $r \to \infty$.

*Proof*

Here all we have to consider is the case in which the $\lambda$ values are such that $U(\breve{\phi}^r, \xi^r) > 0$, because if $U(\breve{\phi}^r, \xi^r) < 0$, the proof is given in Theorem 3.2.

Let $\lambda = \min \lambda^r$, then from Lemma 3.4

$$U(\breve{\phi}^1, \xi^1) \leqslant m^{1/p} M_f^{\,1}, \quad \text{where } M_f^{\,r} = M_f(\breve{\phi}^r)$$

$$U(\breve{\phi}^2, \xi^2) \leqslant m^{1/p}(M_f^{\,2} - \lambda^1 M_f^{\,1}) \leqslant m^{1/p}(M_f^{\,2} - \lambda M_f^{\,1})$$

Similarly,

$$U(\breve{\phi}^r, \xi^r) \leqslant m^{1/p}(M_f^{\,r} - \lambda(1-\lambda)^{r-2} M_f^{\,1} - \ldots - \lambda M_f^{\,r-1})$$

Therefore,

$$\sum_r U(\breve{\phi}^r, \xi^r) \leqslant m^{1/p}(M_f^{\,r} + (1-\lambda) M_f^{\,r-1} + \ldots + (1-\lambda)^{r-3} M_f^{\,3} + (1-\lambda)^{r-2} M_f^{\,2} + (1-\lambda)^{r-1} M_f^{\,1})$$

Due to the fact that $\xi^r < M_f(\breve{\phi})$ and because of Lemma 3.5, $M_f^{\,1}, M_f^{\,2}, \ldots, M_f^{\,r} \leqslant \alpha < \infty$,

$$\lim_{i \to \infty} \sum_{r=1}^{i} U(\breve{\phi}^i, \xi^i) \leqslant m^{1/p} \frac{\alpha}{\lambda} < \infty \tag{22}$$

Therefore,

$$\lim_{r \to \infty} U(\breve{\phi}^r, \xi^r) \to 0 \tag{23}$$

*Theorem* 3.5

As $r \to \infty$, $M_f(\breve{\phi}^r) \to M_f(\breve{\phi})$.

*Proof*

The proof is similar to that of Theorem 3.2.

*Theorem* 3.6

As $r \to \infty$ the necessary conditions for a minimax optimum are satisfied. The proof is similar to that of Theorem 3.3.

3.6. *Examples*

Two problems are going to be considered to illustrate some of the theoretical

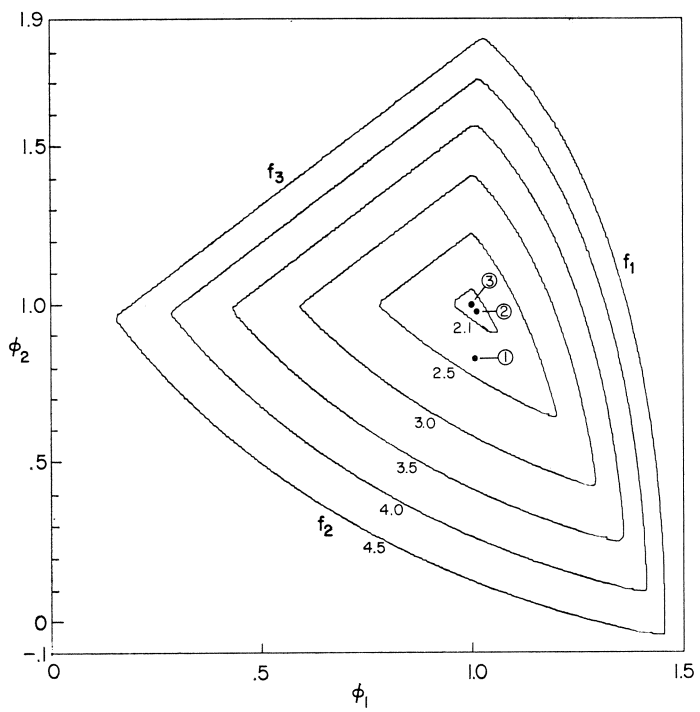


Figure 1. Contours of $M_f(\boldsymbol{\phi})$ for Problem 1.

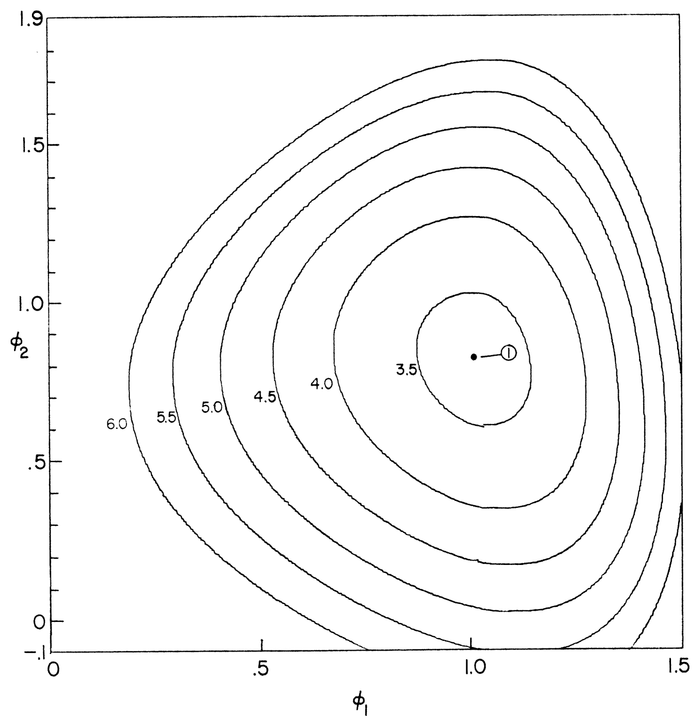


Figure 2. Contours of $U(\boldsymbol{\phi}, \xi)$ for Problem 1 with $\xi=0$.

ideas. To overcome the difficulty of discontinuous derivatives which might arise when $M(\boldsymbol{\phi}, \xi)=0$, we replace step 3 of Algorithm 1 by

$$\xi^{r+1}=M_{\mathrm{f}}(\check{\boldsymbol{\phi}}^r)+\epsilon \tag{24}$$

where $\epsilon$ is a small number.

*Problem* 1

Minimize the maximum of the following three functions,

$$\left.\begin{aligned} f_1&=\phi_1{}^4+\phi_2{}^2\\ f_2&=(2-\phi_1)^2+(2-\phi_2)^2\\ f_3&=2\exp(-\phi_1+\phi_2)\end{aligned}\right\} \tag{25}$$

The optimum minimax value of 2 occurs at $\phi_1=\phi_2=1$. This point satisfies the necessary conditions for a minimax optimum. Figure 1 shows contours for this problem.

Starting from the point $[2\ 2]^{\mathrm{T}}$ ($M_{\mathrm{f}}(\boldsymbol{\phi}^0)=20$) and using $p=2$ throughout Algorithm 1 in conjunction with the Fletcher optimization subroutine (Fletcher 1970) generated the sequence shown in Table 1. Note that $M_{\mathrm{f}}(\check{\boldsymbol{\phi}}^r)$ asymptotically approaches the value 2 and that after seven steps our optimum agrees to six significant figures with the minimax optimum. The value of $\epsilon$ used is $10^{-8}$.

| Steps ($r$) | $\check{\phi}_1{}^r$ | $\check{\phi}_2{}^r$ | $M_{\mathrm{f}}(\check{\boldsymbol{\phi}}^r)$ |
|---|---|---|---|
| 1 | 1·01702 | 0·82055 | 2·35736 |
| 2 | 1·01129 | 0·97115 | 2·03608 |
| 3 | 1·00153 | 0·99654 | 2·00388 |
| 4 | 1·00017 | 0·99962 | 2·00042 |
| 5 | 1·00002 | 0·99996 | 2·00003 |
| 6 | 1·00000 | 0·99999 | 2·00001 |
| 7 | 1·00000 | 1·00000 | 2·00000 |

Table 1. Problem 1 using algorithm 1.

Figure 2 shows contours of $U$ for $p=2$ and $\xi=0$. Figure 3 shows contours for $p=2$ and $\xi=2{\cdot}3574+\epsilon(=M_{\mathrm{f}}(\check{\boldsymbol{\phi}}^1)+\epsilon)$ and Fig. 4 shows contours for $p=2$ and $\xi=2{\cdot}0361+\epsilon(=M_{\mathrm{f}}(\check{\boldsymbol{\phi}}^2)+\epsilon)$.

The first three optima are shown in Fig. 1 as ①, ② and ③, respectively. The defined objective function (3) has the property of smoothing the minimax contours. This can be seen from Figs. 2, 3 and 4 where the partial derivatives of $U$ are continuous (except when $M=0$ and two or more maxima are equal), unlike the minimax contours which are discontinuous when two or more maxima are equal.

Starting from the same point as in Algorithm 1 and using the same value of $p$ Algorithm 2 in conjunction with the Fletcher optimization subroutine generated the sequence shown in Table 2. The value of $\lambda$ used throughout was 0·5. Observe that $\xi^r$ increases from zero and $M_{\mathrm{f}}(\check{\boldsymbol{\phi}}^r)$ decreases and both of them tend asymptotically to 2. Also, the optimum parameter vector tends to $[1\ 1]^{\mathrm{T}}$.

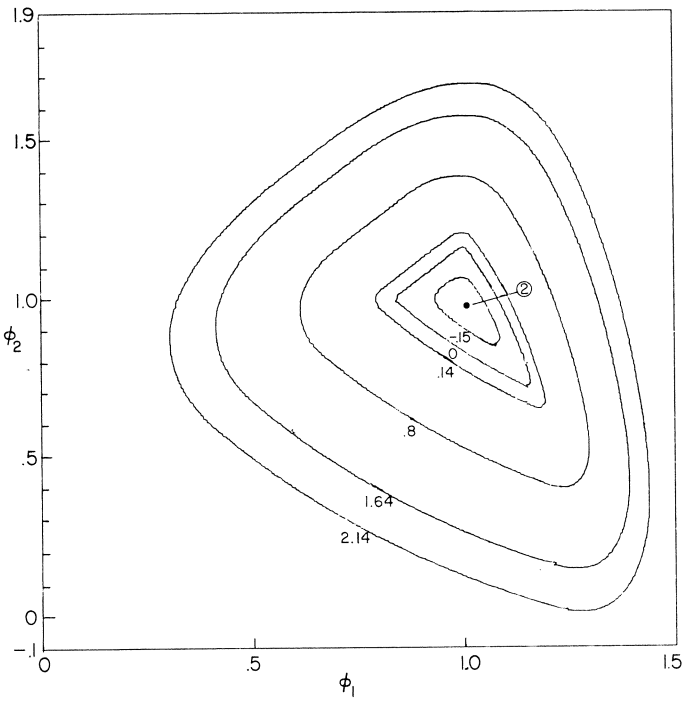


Figure 3. Contours of $U(\boldsymbol{\phi}, \xi)$ for Problem 1 with $\xi = 2{\cdot}3574 + \epsilon$.

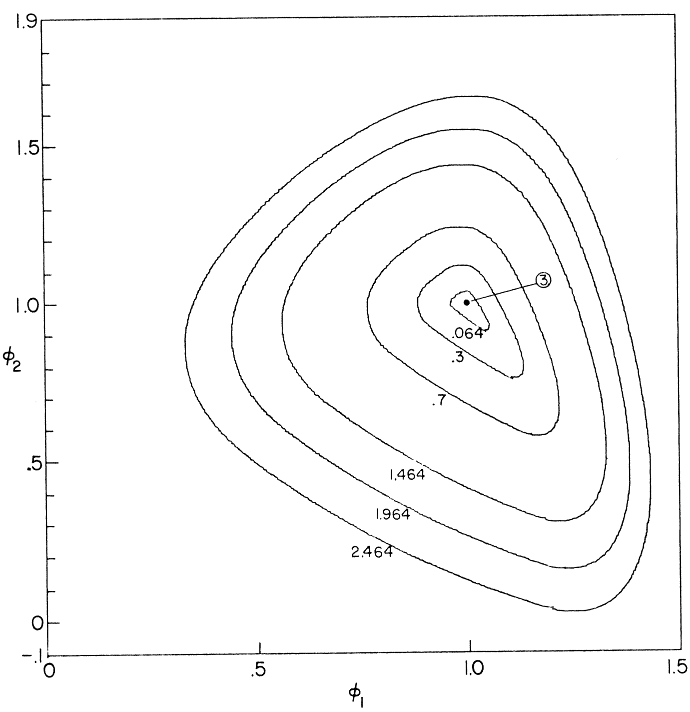


Figure 4. Contours of $U(\boldsymbol{\phi}, \xi)$ for Problem 1 with $\xi = 2{\cdot}0361 + \epsilon$.

| Steps ($r$) | $\check{\phi}_1^r$ | $\check{\phi}_2^r$ | $M_f(\check{\boldsymbol{\phi}}^r)$ | $\xi^r$ |
|---|---|---|---|---|
| 1 | 1·01702 | 0·82055 | 2·3574 | 0 |
| 2 | 1·02148 | 0·88911 | 2·1916 | 1·1787 |
| 3 | 1·01481 | 0·94482 | 2·0840 | 1·6851 |
| 4 | 1·00705 | 0·97709 | 2·0323 | 1·8846 |
| 5 | 1·00280 | 0·99134 | 2·0118 | 1·9584 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 9 | 1·00005 | 0·99985 | 2·0002 | 1·9993 |

Table 2. Problem 1 using algorithm 2.

If $\lambda^{(1)} = [M_f(\check{\boldsymbol{\phi}})/M_f(\check{\boldsymbol{\phi}}^1)] = 2/2{\cdot}3574 = 0{\cdot}8484$, in other words $\xi^2 = M_f(\check{\boldsymbol{\phi}}) = 2$, then we reach the minimax optimum in two steps. This was verified with Algorithm 2.

*Problem* 2

Find the minimax optimum of the following three functions,

$$\left.\begin{aligned} f_1 &= \phi_1^2 + \phi_2^4 \\ f_2 &= (2-\phi_1)^2 + (2-\phi_2)^2 \\ f_3 &= 2\exp(-\phi_1 + \phi_2) \end{aligned}\right\} \tag{26}$$

When $\phi_1 = \phi_2 = 1$, $f_1 = f_2 = f_3 = 2$ but this point is not a minimax optimum because the necessary conditions for a minimax optimum are not satisfied. The minimax optimum is defined by the functions $f_1$ and $f_2$ at $\phi_1 = 1{\cdot}13904$,

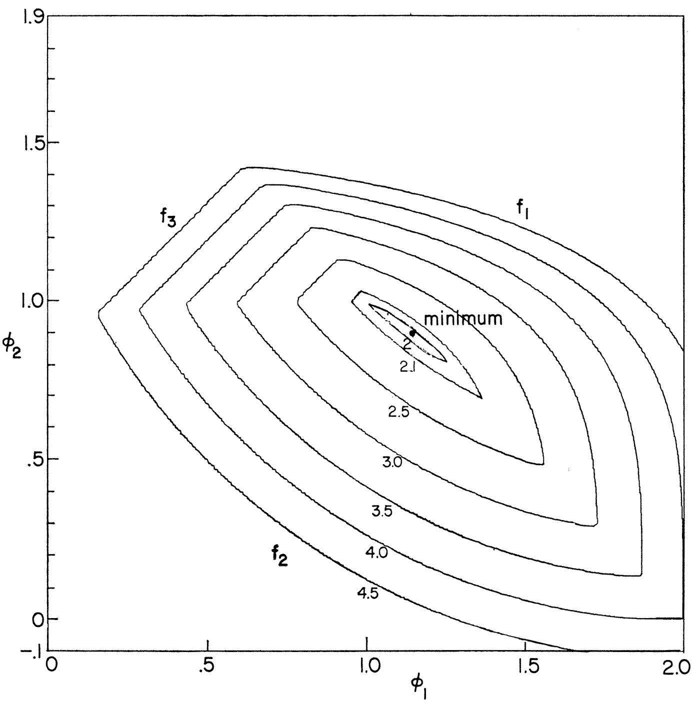


Figure 5. Contours of $M_f(\boldsymbol{\phi})$ for Problem 2.

| Steps ($r$) | $\check{\phi}_1^r$ | $\check{\phi}_2^r$ | $M_f(\check{\boldsymbol{\phi}}^r)$ |
|---|---|---|---|
| 1 | 1·24176 | 0·77401 | 2·07800 |
| 2 | 1·14118 | 0·89563 | 1·95721 |
| 3 | 1·13896 | 0·89953 | 1·95242 |
| 4 | 1·13904 | 0·89956 | 1·952233 |
| 5 | 1·13904 | 0·89956 | 1·952226 |
| 6 | 1·13904 | 0·89956 | 1·95222 |

Table 3. Problem 2 using algorithm 1.

| Steps ($r$) | $\check{\phi}_1^r$ | $\check{\phi}_2^r$ | $M_f(\check{\boldsymbol{\phi}}^r)$ | $\xi^r$ |
|---|---|---|---|---|
| 1 | 1·24176 | 0·77401 | 2·07800 | 0 |
| 2 | 1·19897 | 0·82093 | 2·03184 | 1·03900 |
| 3 | 1·13557 | 0·88307 | 1·99477 | 1·53542 |
| 4 | 1·13153 | 0·89596 | 1·97314 | 1·76510 |
| 5 | 1·13561 | 0·89791 | 1·96177 | 1·86912 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 10 | 1·13898 | 0·89953 | 1·95239 | 1·95161 |

Table 4. Problem 2 using algorithm 2.

$\phi_2 = 0{\cdot}89956$, where $f_1 = f_2 = 1{\cdot}95222$ and $f_3 = 1{\cdot}57408$ (see Fig. 5). This point satisfies the necessary conditions for a minimax optimum. Using both algorithms this point was reached.

Tables 3 and 4 show the progress of Algorithms 1 and 2, respectively, from the starting point $[2\ 2]^T$. For both algorithms $p = 2$ and $\epsilon = 10^{-8}$. For the second algorithm $\lambda = 0{\cdot}5$. From Table 3 it can be seen that after six steps Algorithm 1 reaches the minimax optimum very accurately. It is also interesting to note again from Table 4 how $\xi^r$ increases from zero and $M_f(\check{\boldsymbol{\phi}}^r)$ decreases asymptotically to $M_f(\check{\boldsymbol{\phi}})$. The value of $p = 2$ and $\lambda = 0{\cdot}5$ were chosen so as to better illustrate the progress of the algorithms.

## 4. Example

Here we want to find a second-order model of a fourth-order system, when the input to the system is an impulse, in the minimax sense. The transfer function of the system is

$$G(s) = \frac{(s+4)}{(s+1)(s^2+4s+8)(s+5)} \tag{27}$$

and of the model it is

$$H(s) = \frac{c}{(s+\alpha)^2+\beta^2} \tag{28}$$

Therefore, we want to approximate

$$S(t) = \tfrac{3}{20} \exp(-t) + \tfrac{1}{52} \exp(-5t) - \frac{\exp(-2t)}{65} (3 \sin 2t + 11 \cos 2t) \tag{29}$$

by

$$F(\boldsymbol{\phi}, t) = \frac{c}{\beta} \exp(-\alpha t) \sin \beta t \tag{30}$$

where $S(t)=\mathscr{L}^{-1}G(s)$, $F(\boldsymbol{\phi}, t)=\mathscr{L}^{-1}H(\boldsymbol{\phi}, s)$ and $\boldsymbol{\phi}=[\alpha\ \beta\ c]^T$.

The problem was discretized into 51 uniformly spaced points in the time interval 0 to 10 sec. Let

$$e_i(\boldsymbol{\phi}) \triangleq F(\boldsymbol{\phi}, t_i) - S(t_i), \quad i \in I \tag{31}$$

where $I=\{1, 2, \ldots, 51\}$. Therefore, our aim is to find a point $\check{\boldsymbol{\phi}}$ such that $\max_{i \in I} |e_i(\check{\boldsymbol{\phi}})| \leqslant \max_{i \in I} |e_i(\boldsymbol{\phi})|$. The minimax optimum is at

$$[0{\cdot}68442\ \pm 0{\cdot}95409\ 0{\cdot}12286]^T$$

and the maximum value of the absolute error is $0{\cdot}79471 \times 10^{-2}$. Using both the algorithms in conjunction with the optimization subroutine due to Fletcher, starting from the point $[1\ 1\ 1]^T$, and using the values $p=2$, 4, 6, 10, 100, 1000 and 10 000 individually the results shown in Tables 5, 6 and 7 were obtained. Table 5 shows how many function evaluations are required for $M_f(\boldsymbol{\phi})$ to be

| Parameters | Starting point | $M_f(\boldsymbol{\phi}^0)$ |
|---|---|---|
| $\alpha$ | 1·0 | |
| $\beta$ | 1·0 | 0·26289 |
| $c$ | 1·0 | |
| Value of $p$ | Number of function evaluations for $M_f(\boldsymbol{\phi})$ to reach $0{\cdot}79471 \times 10^{-2}$ | |
| 2 | 213 | |
| 4 | 161 | |
| 6 | 166 | |
| 10 | 142 | |
| 100 | 187 | |
| 1 000 | 144 | |
| 10 000 | 302 | |
| Average function evaluations | 188 | |

Table 5. Results of algorithm 1.

| Value of $p$ | $M_1{}^1 \times 10^2$ | $M_2{}^1 \times 10^2$ | $M_3{}^1 \times 10^2$ | $M_4{}^1 \times 10^2$ | $M_5{}^1 \times 10^2$ |
|---|---|---|---|---|---|
| 2 | 1·2880 | 0·66348 | 0·38106 | 0·27946 | 0·00013 |
| 4 | 1·0194 | 0·72517 | 0·54438 | 0·47185 | 0·00779 |
| 6 | 0·92477 | 0·77198 | 0·62354 | 0·56909 | 0·01657 |
| 10 | 0·85921 | 0·79648 | 0·69289 | 0·65879 | 0·02840 |
| 100 | 0·79886 | 0·79438 | 0·78710 | 0·78646 | 0·04934 |
| 1 000 | 0·79508 | 0·79466 | 0·79412 | 0·79385 | 0·05079 |
| 10 000 | 0·79474 | 0·79470 | 0·79464 | 0·79462 | 0·05090 |

Table 6. Values of $M_1{}^1, M_2{}^1, \ldots, M_5{}^1$.

equal to $0{\cdot}79471 \times 10^{-2}$ for different values of $p$ by using Algorithm 1. Note that a very small or a very large value of $p$ takes relatively more function

| Parameters | Starting point | $M_f(\boldsymbol{\phi}^0)$ |
|---|---|---|
| $\alpha$ | 1·0 | |
| $\beta$ | 1·0 | 0·26289 |
| $c$ | 1·0 | |
| Value of $p$ | Number of function evaluations for $M_f(\boldsymbol{\phi})$ to reach $0.79471 \times 10^{-2}$ | |
| 2 | 159 | |
| 4 | 188 | |
| 6 | 157 | |
| 10 | 143 | |
| 100 | 184 | |
| 1 000 | 148 | |
| 10 000 | 289 | |
| Average function evaluations | 182 | |

Table 7. Results of algorithm 2.

evaluations. Table 6 shows the values of $M_1^1, M_2^1, \ldots, M_5^1$ where

$$M_j^r = \{|e_i^r| \mid |e_i^r| > |e_l^r| \; ; \quad |l-i| = 1 \; ; \quad i, l \in I\}$$

for different values of $p$.

Table 7 shows the number of function evaluations for $M_f(\boldsymbol{\phi})$ to be equal to $0{\cdot}79471 \times 10^{-2}$ for different values of $p$, by using Algorithm 2. The value of $\lambda$ used was $(M_1^1 + M_4^1)/(2M_1^1)$. As can be seen again, if $p$ is very large the convergence slows down. For both algorithms the value of $p = 10$ was the best. From the average function evaluations it can be seen that both algorithms behave similarly.

## 5. Conclusions

Two new methods for non-linear minimax optimization are presented. The new methods abandon the linear programming subproblem which many of the other methods require. An advantage of these methods is that it is possible to use very efficient gradient methods such as the recent minimization algorithms by Fletcher (1970) and Charalambous (1973).

Recently, the authors have transformed the non-linear programming problem into an unconstrained minimax problem which under certain conditions has the same optimum as the original problem (Bandler and Charalambous 1974). The two methods presented can thus be used to solve the non-linear programming problem, and also constrained minimax problems which may be converted to the non-linear programming formulation.

Acknowledgments

The authors wish to thank Dr. T. M. K. Davison and Dr. P. C. Chakravarti of McMaster University and Dr. A. Conn of the University of Waterloo for useful discussions. This work was supported by the National Research Council of Canada under grants A7239 and C154, and by a scholarship to C. Charalambous.

## References

Bandler, J. W., 1971, *I.E.E.E. Trans. Circuit Theory*, **18,** 476.

Bandler, J. W., and Charalambous, C., 1972, *I.E.E.E. Trans. microw. Theory Techn.*, **20,** No. 12, 834; 1973, *J. optim. Theory Applic.*, **11,** 556; 1974, *J. optim. Theory Applic.*, **13,** 607.

Bandler, J. W., Srinivasan, T. V., and Charalambous, C., 1972, *I.E.E.E. Trans. microw. Theory Techn.*, **20,** 596.

Charalambous, C., 1973, *Math. Progm.*, **5,** 189.

Fiacco, A. V., and McCormick, G. P., 1964, *Mgmt Sci.*, **10,** 360.

Fletcher, R., 1970, *Comput. J.*, **13,** 317.

Hardy, G. H., Littlewood, J. E., and Pólya, G., 1934, *Inequalities* (Cambridge University Press).

Osborne, M. R., and Watson, G. A., 1969, *Comput. J.*, **12,** 63.

Rice, J. R., 1969, *The Approximation of Functions*, Vol. 2 (Reading, Massachusetts : Addison-Wesley).

Waren, A. D., Lasdon, L. S., and Suchman, D. F., 1967, *Proc. Instn elect. electron. Engng*, **55,** 1885.