INTERNAL REPORTS IN

SIMULATION, OPTIMIZATION AND CONTROL

No. SOC-3

NONLINEAR MINIMAX OPTIMIZATION AS A SEQUENCE OF LEAST PTH OPTIMIZATION WITH FINITE VALUES OF P

C. Charalambous and J.W. Bandler

June 1973

(Revised March 1974)

FACULTY OF ENGINEERING
McMASTER UNIVERSITY
HAMILTON, ONTARIO, CANADA

NONLINEAR MINIMAX OPTIMIZATION AS A SEQUENCE OF LEAST PTH OPTIMIZATION WITH FINITE VALUES OF p

C. CHARALAMBOUS
Department of Combinatorics and Optimization
University of Waterloo, Waterloo, Canada

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

Following developments in nonlinear least pth optimization by the authors it is possible to derive two new methods of nonlinear minimax optimization. Unlike the Polya algorithm in which a sequence of least pth 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 pth 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 10000.

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.

1. Introduction

Various algorithms have been proposed for solving the discrete nonlinear minimax problem, some of the most relevant of which are due to Waren, Lasdon and Suchman [Ref.1], Osborne and Watson [Ref.2], Bandler, Srinivasan and Charalambous [Ref.3] and Bandler and Charalambous [Ref.4].

The first method transforms the nonlinear minimax optimization problem into a nonlinear programming problem and solves it by well-established methods such as the one by Fiacco and McCormick [Ref.5]. 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 [Ref.6]. A pth 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 pth 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 nonlinear functions

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

where $\phi \triangleq [\phi_1 \ \phi_2 \ \ldots \ \phi_k]^T$ is a k-dimensional column vector containing the k adjustable parameters and $I \triangleq \{1,2, \ldots, m\}$. Let

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

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

2.1 Assumptions

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

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

(c) The functions $f_i(\phi)$ for $i\epsilon I$ belong to class C^1 (once continuously differentiable).

2.2 Definitions

Consider the following objective function

$$U(\phi,\xi) = M(\phi,\xi)\left(\sum_{i\epsilon K}\left(\frac{f_i(\phi) - \xi}{M(\phi,\xi)}\right)^q\right)^{\frac{1}{q}} \text{ for } M(\phi,\xi) \neq 0$$

$$= 0 \quad \text{for } M(\phi,\xi) = 0 \tag{3}$$

where

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

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

where p has a <u>constant</u> value in the range $1<p<\infty$.

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

The objective function given in (3) is a generalization of the usual least pth objective function. Under the assumptions (a) and (b), the continuity of $f_i(\phi)$ for $i\epsilon I$ and because $U(\phi,\xi) \geq M(\phi,\xi)$ (see lemmas 3.1 and 3.4) the objective function $U(\phi,\xi)$ is continuous and has a minimum which is located at a finite point. Also, due to assumption (c), $U(\phi,\xi)$ has continuous first partial derivatives <u>except</u> when both $M(\underset{\sim}{\phi},\xi) = 0$ and two or more of the functions $(f_i(\underset{\sim}{\phi}) - \xi)$ for $i\varepsilon I$ are equal to zero.

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

The symbols ε and η will be used to denote small positive numbers.

3. The New Algorithms

3.1 Algorithm 1

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

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

3. Set

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

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

$$\xi^{r+1} = \xi^r + \lambda^r M(\check{\phi}^r, \xi^r) \tag{8}$$

$$= (1-\lambda^r)\xi^r + \lambda^r M_f(\check{\phi}^r)$$

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 in Refs. 4 and 7 where the value of p must be very large. Algorithm 2 is different from algorithm 1 if $M_f(\check{\phi}) > 0$, otherwise it is the same. The main difference is that in the algorithm 1 we try to push the maximum away from the level ξ^r at the rth iteration (this causes $M(\check{\phi}^r, \xi^r) < 0$, and $\xi^{r+1} < \xi^r$ for $r \geq 2$), while in algorithm 2 we try to predict the value of $M_f(\check{\phi})$ by increasing the value of ξ^r from zero appropriately (this causes, $M(\check{\phi}^r, \xi^r) > 0$, 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 $f_i(\phi)$ for $i\varepsilon I$ and $f_i(\phi) + \beta$ for $i\varepsilon I$ does not change when β is constant it will be possible to use algorithm 2 even when $M_f(\check{\phi}) \leq 0$ but we have to raise all the

$f_i(\phi)$ for $i\varepsilon I$ by an amount $\beta > |M_f(\check{\phi})|$.

The first step of algorithm 1 ($\xi^1 = \min [0, M_f(\phi^o)]$) could be modified to $\xi^1 = M_f(\phi^o)$. A reason for not modifying it is the following. In engineering problems (e.g., filter design (Ref. 4)) the sign of $M_f(\phi)$ indicates whether a particular structure can satisfy certain design specifications. That is, if,

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

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

3.4 Convergence Proofs for Algorithm 1

Lemma 3.1 If $y_i \geq 0$ for $i\varepsilon I$ and $p \geq 1$, then

$$m^{-\frac{1}{p}} \min_{i\varepsilon I} y_i \leq \left(\sum_{i\varepsilon I} y_i^{-p} \right)^{-\frac{1}{p}} \leq \min_{i\varepsilon I} y_i$$

The proof is simple and is omitted.

Lemma 3.2 Let y_i for $i\varepsilon I$ be a set of real numbers and $x \geq \max_{i\varepsilon I} y_i$. Then

$$U(x) = -\left(\sum_{i\varepsilon I} (x-y_i)^{-p} \right)^{-\frac{1}{p}} , \quad p \geq 1$$

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

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

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

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

$$- U((1-\lambda)\, x^{(1)} + \lambda x^{(2)}) = \left(\sum_{i\varepsilon I} ((1-\lambda)\, x^{(1)} + \lambda x^{(2)} - y_i)^{-p} \right)^{-\frac{1}{p}}$$

$$= \left(\sum_{i\varepsilon I} ((1-\lambda)(x^{(1)} - y_i) + \lambda(x^{(2)} - y_i))^{-p} \right)^{-\frac{1}{p}}$$

$$\geq (1-\lambda)\left(\sum_{i\varepsilon I} (x^{(1)} - y_i)^{-p} \right)^{-\frac{1}{p}} + \lambda \left(\sum_{i\varepsilon I} (x^{(2)} - y_i)^{-p} \right)^{-\frac{1}{p}}$$

See Ref. 8 for the last inequality. Therefore, convexity follows.

<u>Lemma 3.3</u> For $r \geq 2$, $|U(\check{\phi}^r, \xi^r)| \geq |U(\check{\phi}^{r+1}, \xi^{r+1})|$.

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

$$U(\check{\phi}^r, \xi^r) = -\left(\sum_{i\varepsilon I} (\xi^r - f_i(\check{\phi}^r))^{-p} \right)^{-\frac{1}{p}}$$

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

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

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

because $\xi^{r+1} \leq \xi^r$ and due to lemma 3.2.

<u>Theorem 3.1</u> $|U(\check{\phi}^r, \xi^r)| \to 0$ as $r \to \infty$.

<u>Proof</u> For $r \geq 2$

$$|U(\check{\phi}^r, \xi^r)| = \left(\sum_{i\varepsilon I} (\xi^r - f_i(\check{\phi}^r))^{-p} \right)^{-\frac{1}{p}}$$

$$\leq \min_{i\varepsilon I} (\xi^r - f_i(\check{\phi}^r)) = \xi^r - \max_{i\varepsilon I} f_i(\check{\phi}^r)$$

(From lemma 3.1)

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

Therefore,

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

$$\leq \xi^2 = M_f(\check{\phi}^1)$$

Therefore,

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

<u>Theorem 3.2</u> As $r \to \infty$, $M_f(\check{\phi}^r) \to M_f(\check{\phi})$.

<u>Proof</u> Assume that as $r \to \infty$, $M_f(\check{\phi}^r) \to L_f \geq M_f(\check{\phi})$. We must show that $L_f = M_f(\check{\phi})$.

Assume $L_f > M_f(\check{\phi})$. Because $M_f(\phi)$ is continuous it is possible to find a point $\bar{\phi}$ such that

$$M_f(\check{\phi}) < M_f(\bar{\phi}) < L_f \tag{11}$$

In other words

$$f_i(\bar{\phi}) - L_f < 0, \ i \in I$$

Since

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

$$\lim_{r\to\infty} \xi^r = L_f$$

But

$$U(\bar{\phi}, L_f) = -\left(\sum_{i \in I} (L_f - f_i(\bar{\phi}))^{-p} \right)^{-\frac{1}{p}} < 0$$

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

<u>Theorem 3.3</u> As $r \to \infty$, the necessary conditions for a minimax optimum are satisfied [Ref. 9], that is,

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

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

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

where

$$\hat{J} \triangleq \{ i \mid f_i(\check{\phi}^\infty) = M_f(\check{\phi}^\infty) \ , \ i \in I \} \tag{12d}$$

and

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

<u>Proof</u> Since a necessary condition that a point be a local minimum of an unconstrained function is that the first partial derivatives vanish then for $r \geq 2$,

$$\underset{\sim}{\nabla} U(\underset{\sim}{\check{\phi}}^r, \xi^r) = \left(\sum_{i\varepsilon I} (\xi^r - f_i(\underset{\sim}{\check{\phi}}^r))^{-p} \right)^{-\frac{1}{p}-1} \cdot$$

$$\sum_{i\varepsilon I} (\xi^r - f_i(\underset{\sim}{\check{\phi}}^r))^{-p-1} \underset{\sim}{\nabla} f_i(\underset{\sim}{\check{\phi}}^r)$$

$$= A \sum_{i\varepsilon I} \left(\frac{\xi^r - f_i(\underset{\sim}{\check{\phi}}^r)}{\xi^r - \xi^{r+1}} \right)^{-p-1} \underset{\sim}{\nabla} f_i(\underset{\sim}{\check{\phi}}^r)$$

$$= \underset{\sim}{0}$$

where

$$A = \left(\sum_{i\varepsilon I} \left(\frac{\xi^r - f_i(\underset{\sim}{\check{\phi}}^r)}{\xi^r - \xi^{r+1}} \right)^{-p} \right)^{-\frac{1}{p}-1} \tag{14}$$

Since $\xi^r - \xi^{r+1} = \min\limits_{i\varepsilon I}(\xi^r - f_i(\underset{\sim}{\check{\phi}}^r))$, $A \neq 0$ and therefore,

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

Let

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

then

$$\sum_{i \varepsilon I} \mu_i^r \underset{\sim}{\nabla} f_i(\check{\underset{\sim}{\phi}}^r) = \underset{\sim}{0} \tag{17}$$

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

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

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

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

and

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

Therefore,

$$\sum_{i \varepsilon \hat{J}} u_i \underset{\sim}{\nabla} f_i(\check{\underset{\sim}{\phi}}^{\infty}) = \underset{\sim}{0} \tag{21}$$

3.5 Convergence Proofs for Algorithm 2

Lemma 3.4 Let y_i for $i\varepsilon I$ be a set of real numbers such that $\max_{i\varepsilon I} y_i \geq 0$, then

$$\max_{i\varepsilon I} y_i \leq \left(\sum_{i\varepsilon L} y_i^p\right)^{\frac{1}{p}} \leq m^{\frac{1}{p}} \max_{i\varepsilon I} y_i \quad , \; p \geq 1$$

where

$$L \triangleq \{i | y_i \geq 0, \; i\varepsilon I\}$$

The proof in simple and is omitted.

Lemma 3.5 Let $U(\check{\phi}^r, \xi^r)$, $U(\check{\phi}^{r+1}, \xi^{r+1}) > 0$. Then

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

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

$$U(\check{\phi}^r, \xi^r) = \left(\sum_{i\varepsilon J(\check{\phi}^r, \xi^r)} (f_i(\check{\phi}^r) - \xi^r)^p \right)^{\frac{1}{p}}$$

$$\geq \left(\sum_{i\varepsilon J(\check{\phi}^r, \xi^{r+1})} (f_i(\check{\phi}^r) - \xi^{r+1})^p \right)^{\frac{1}{p}}$$

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

$$\geq \left(\sum_{i\varepsilon J(\check{\phi}^{r+1}, \xi^{r+1})} (f_i(\check{\phi}^{r+1}) - \xi^{r+1})^p \right)^{\frac{1}{p}}$$

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

$$= U(\overset{\vee}{\underset{\sim}{\phi}}^{r+1}, \xi^{r+1})$$

<u>Theorem 3.4</u> $U(\overset{\vee}{\underset{\sim}{\phi}}^{r}, \xi^{r}) \to 0$ as $r \to \infty$.

<u>Proof</u> Here all we have to consider is the case in which the λ values are such that $U(\overset{\vee}{\underset{\sim}{\phi}}^{r}, \xi^{r}) > 0$, because if $U(\overset{\vee}{\underset{\sim}{\phi}}^{r}, \xi^{r}) < 0$, the proof is given in theorem 3.2.

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

$$U(\overset{\vee}{\underset{\sim}{\phi}}^{1}, \xi^{1}) \leq m^{\frac{1}{p}} M_f^1 \quad , \quad \text{where } M_f^r = M_f(\overset{\vee}{\underset{\sim}{\phi}}^{r})$$

$$U(\overset{\vee}{\underset{\sim}{\phi}}^{2}, \xi^{2}) \leq m^{\frac{1}{p}} (M_f^2 - \lambda^1 M_f^1) \leq m^{\frac{1}{p}} (M_f^2 - \lambda M_f^1)$$

Similarly,

$$U(\overset{\vee}{\underset{\sim}{\phi}}^{r}, \xi^{r}) \leq m^{\frac{1}{p}} (M_f^r - \lambda (1-\lambda)^{r-2} M_f^1 - \ldots - \lambda M_f^{r-1})$$

Therefore,

$$\sum_r U(\overset{\vee}{\underset{\sim}{\phi}}^{r}, \xi^{r}) \leq m^{\frac{1}{p}} (M_f^r + (1-\lambda) M_f^{r-1} + \ldots + (1-\lambda)^{r-3} M_f^3 + (1-\lambda)^{r-2} M_f^2$$

$$+ (1-\lambda)^{r-1} M_f^1)$$

Due to the fact that $\xi^r < M_f(\overset{\vee}{\underset{\sim}{\phi}})$ and because of lemma 3.5, M_f^1 , M_f^2 , ... , $M_f^r \leq \alpha < \infty$

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

Therefore,

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

Theorem 3.5 As $r \to \infty$, $M_f(\check{\phi}^r) \to M_f(\check{\phi})$.

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

$$\xi^{r+1} = M_f(\check{\phi}^r) + \varepsilon \tag{24}$$

where ε is a small number.

Problem 1 Minimize the maximum of the following three functions,

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

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

Starting from the point $[2\ 2]^T$ $(M_f(\phi^o) = 20)$ and using p=2 throughout algorithm 1 in conjunction with the Fletcher optimization subroutine [Ref. 10] generated the sequence shown in table 1. Note that $M_f(\check{\phi}^r)$ asymptotically approaches the value 2 and that after 7 steps our optimum agrees to 6 significant figures with the minimax optimum. The value of ε used is 10^{-8}.

Fig. 2 shows contours of U for $p = 2$ and $\xi = 0$. Fig. 3 shows contours for $p = 2$ and $\xi = 2.3574 + \varepsilon (= M_f(\check{\phi}^1) + \varepsilon)$ and Fig. 4 shows contours for $p = 2$ and $\xi = 2.0361 + \varepsilon (= M_f(\check{\phi}^2) + \varepsilon)$.

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 λ used throughout was 0.5. Observe that ξ^r increases from zero and $M_f(\check{\phi}^r)$ decreases and both of them tend asymptotically to 2. Also, the optimum parameter vector tends to $[1\ 1]^T$.

If $\lambda^{(1)} = \dfrac{M_f(\check{\phi})}{M_f(\check{\phi}^1)} = \dfrac{2}{2.3574} = 0.8484$, in other words $\xi^2 = M_f(\check{\phi}) = 2$, then we reach the minimax optimum in 2 steps. This was verified with algorithm 2.

Problem 2 Find the minimax optimum of the following three functions,

$$f_1 = \phi_1^2 + \phi_2^4$$

$$f_2 = (2-\phi_1)^2 + (2-\phi_2)^2 \qquad (26)$$

$$f_3 = 2\exp(-\phi_1 + \phi_2)$$

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 $\phi_1 = 1.13904$, $\phi_2 = 0.89956$, where $f_1 = f_2 = 1.95222$ and $f_3 = 1.57408$. (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 $\varepsilon = 10^{-8}$. For the second algorithm $\lambda = 0.5$. From table 3 it can be seen that after 6 steps algorithm

1 reaches the minimax optimum very accurately. It is also interesting to note again from table 4 how ξ^r increases from zero and $M_f(\check{\phi}^r)$ decreases asymptotically to $M_f(\check{\phi})$. The value of p = 2 and $\lambda = 0.5$ were choser 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

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

by

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

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

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

$$e_i(\underset{\sim}{\phi}) \triangleq F(\underset{\sim}{\phi},t_i) - S(t_i) \quad , i\varepsilon I \tag{31}$$

where $I = \{1, 2, \ldots, 51\}$. Therefore, our aim is to find a point $\check{\underset{\sim}{\phi}}$ such that $\max_{i\varepsilon I}|e_i(\check{\underset{\sim}{\phi}})| \leq \max_{i\varepsilon I}|e_i(\underset{\sim}{\phi})|$. The minimax optimum is at $[0.68442 \;\; \pm 0.95409 \;\; 0.12286]^T$ and the maximum value of the absolute error is 0.79471×10^{-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 10000 individually the results shown in tables 5, 6 and 7 were obtained. Table 5 shows how many function evaluations are required for $M_f(\underset{\sim}{\phi})$ to be equal to 0.79471×10^{-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 evaluations. Table 6 shows the values of $M_1^1, M_2^1, \ldots, M_5^1$ where $M_j^r = \{|e_i^r| \;\big|\; |e_i^r| > |e_\ell^r| \;;\; |\ell - i| = 1 \;;\; i, \ell\varepsilon I\}$ for a different values of p.

Table 7 shows the number of function evaluations for $M_f(\underset{\sim}{\phi})$ to be equal to 0.79471×10^{-2} for different values of p, by using algorithm 2. The value of λ 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 nonlinear 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 [Ref. 10] and Charalambous [Ref. 12].

Recently, the authors have transformed the nonlinear programming problem into an unconstrained minimax problem which under certain conditions has the same optimum as the original problem [Ref. 11]. The two methods presented can thus be used to solve the nonlinear programming problem, and also constrained minimax problems which may be converted to the nonlinear programming formulation.

References

1. WAREN, A.D., LASDON, L.S., and SUCHMAN, D.F., Optimization in Engineering Design, Proceedings of the IEEE, Vol. 55, No. 11, 1967.

2. OSBORNE, M.R., and WATSON, G.A., An Algorithm for Minimax Approximation in the Nonlinear Case, Computer Journal, Vol. 12, No. 1, 1969.

3. BANDLER, J.W., SRINIVASAN, T.V., and CHARALAMBOUS, C., Minimax Optimization of Networks by Grazor Search, IEEE Transactions on Microwave Theory and Techniques, Vol. MTT-20, No. 9, 1972.

4. BANDLER, J.W., and CHARALAMBOUS, C., Practical Least pth Optimization of Networks, IEEE Transactions on Microwave Theory and Techniques, Vol. MTT-20, No. 12, 1972.

5. FIACCO, A.V., and McCORMICK, G.P., The Sequential Unconstrained Minimization Technique for Nonlinear Programming, A Primal-Dual Method, Management Science, Vol. 10, No. 2, 1964.

6. RICE, J.R., The Approximation of Functions Vol. II, Addison-Wesley, Reading, Mass., 1969.

7. BANDLER, J.W., and CHARALAMBOUS, C., On Conditions for Optimality in Least pth Approximation with $p \to \infty$, Journal of Optimization Theory and Applications, Vol. 11, No. 5, 1973.

8. HARDY, G.H., LITTLEWOOD, J.E., and PÓLYA, G., Inequalities, Cambridge University Press, 1934.

9. BANDLER, J.W., Conditions for a Minimax Optimum, IEEE Transactions on Circuit Theory, Vol. CT-18, No. 4, 1971.

10. FLETCHER, R., A New Approach to Variable Metric Algorithms, Computer Journal, Vol. 13, No. 3, 1970.

11. BANDLER, J.W., and CHARALAMBOUS, C., Nonlinear Programming Using Minimax Techniques, Journal of Optimization Theory and Applications, Vol. 13, No. 6, 1974.

12. CHARALAMBOUS, C., Unconstrained Optimization Based on Homogeneous Models, Mathematical Programming, Vol. 5, 1973.

Table 1. Problem 1 using Algorithm 1

Steps (r)	$\check{\phi}_1^r$	$\check{\phi}_2^r$	$M_f(\check{\phi}^r)$
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 2. Problem 1 using Algorithm 2

Steps (r)	$\check{\phi}_1^r$	$\check{\phi}_2^r$	$M_f(\check{\underset{\sim}{\phi}}^r)$	ξ^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 3. Problem 2 using Algorithm 1

Steps (r)	$\check{\phi}_1^r$	$\check{\phi}_2^r$	$M_f(\check{\underset{\sim}{\phi}}^r)$
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 4. Problem 2 using Algorithm 2

Steps (r)	$\check{\phi}_1^r$	$\check{\phi}_2^r$	$M_f(\check{\underset{\sim}{\phi}}^r)$	ξ^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 5. Results of Algorithm 1

Parameters	Starting Point	$M_f(\underset{\sim}{\phi}^o)$
α	1.0	
β	1.0	0.26289
c	1.0	

Value of p	Number of function evaluations for $M_f(\underset{\sim}{\phi})$ to reach 0.79471×10^{-2}
2	213
4	161
6	166
10	142
100	187
1000	144
10000	302
Average function evaluations	188

Table 6. Values of M_1^1, M_2^1, ..., M_5^1

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
1000	0.79508	0.79466	0.79412	0.79385	0.05079
10000	0.79474	0.79470	0.79464	0.79462	0.05090

Table 7. Results of Algorithm 2

Parameters	Starting point	$M_f(\underset{\sim}{\phi}^o)$
α	1.0	0.26289
β	1.0	
c	1.0	

Value of p	Number of function evaluations for for $M_f(\underset{\sim}{\phi})$ to reach 0.79471×10^{-2}
2	159
4	188
6	157
10	143
100	184
1000	148
10000	289
Average function evaluations	182

Figure Captions

Fig. 1. Contours of $M_f(\underset{\sim}{\phi})$ for Problem 1.

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

Fig. 3. Contours of $U(\underset{\sim}{\phi},\xi)$ for Problem 1 with $\xi = 2.3574 + \varepsilon$.

Fig. 4. Contours of $U(\underset{\sim}{\phi},\xi)$ for Problem 1 with $\xi = 2.0361 + \varepsilon$.

Fig. 5. Contours of $M_f(\underset{\sim}{\phi})$ for Problem 2.

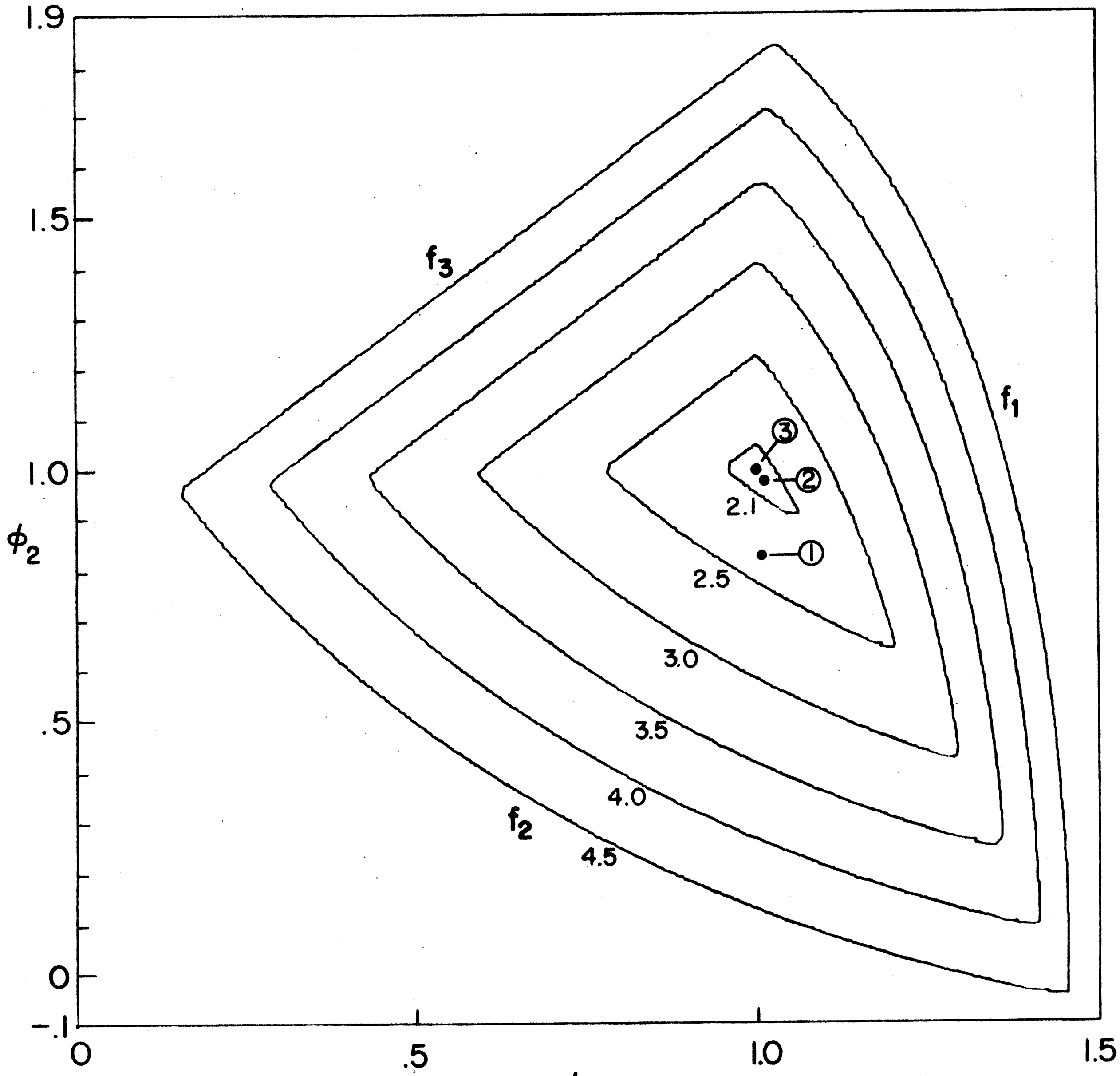

Fig. 1

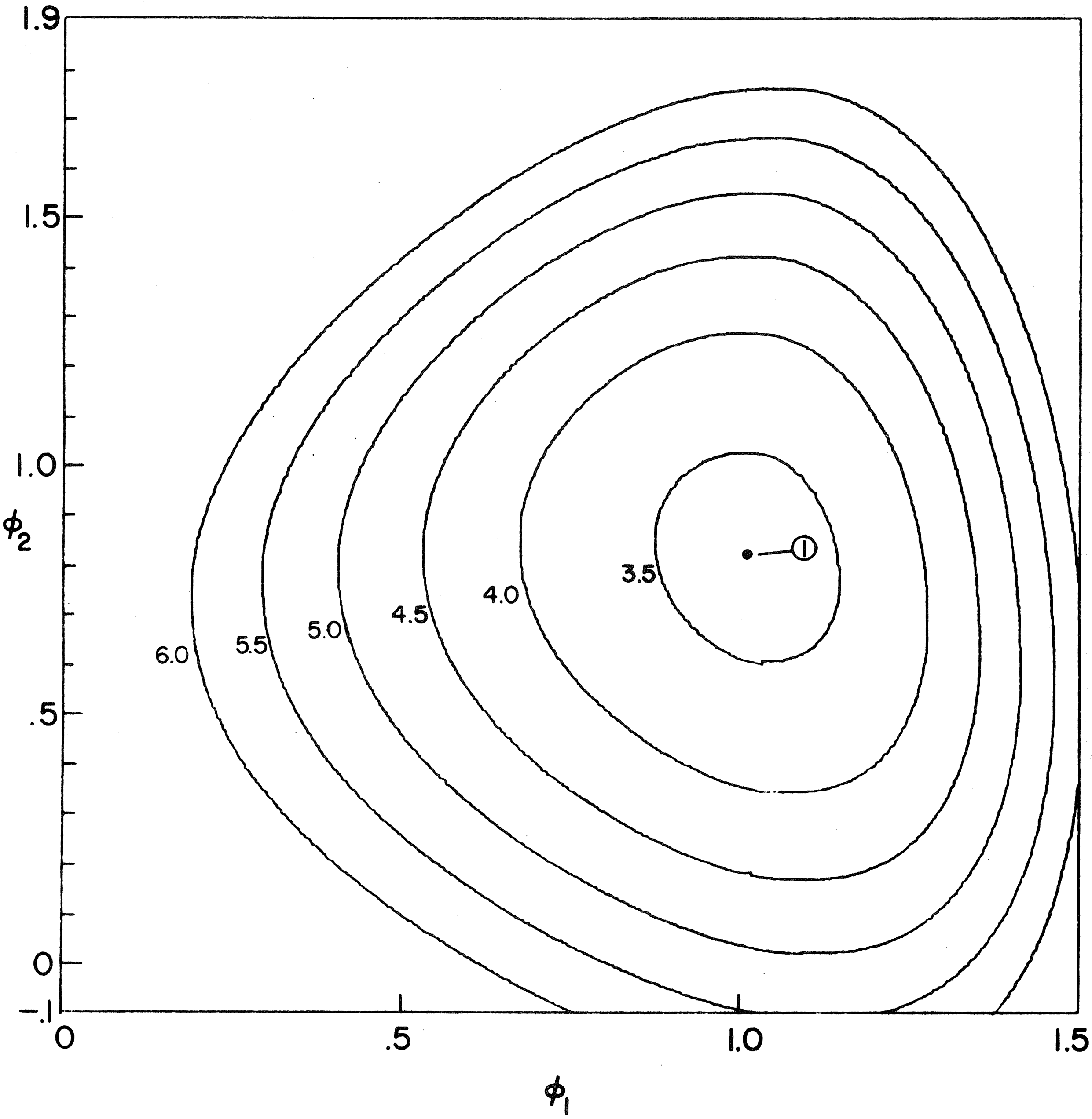

Fig 2

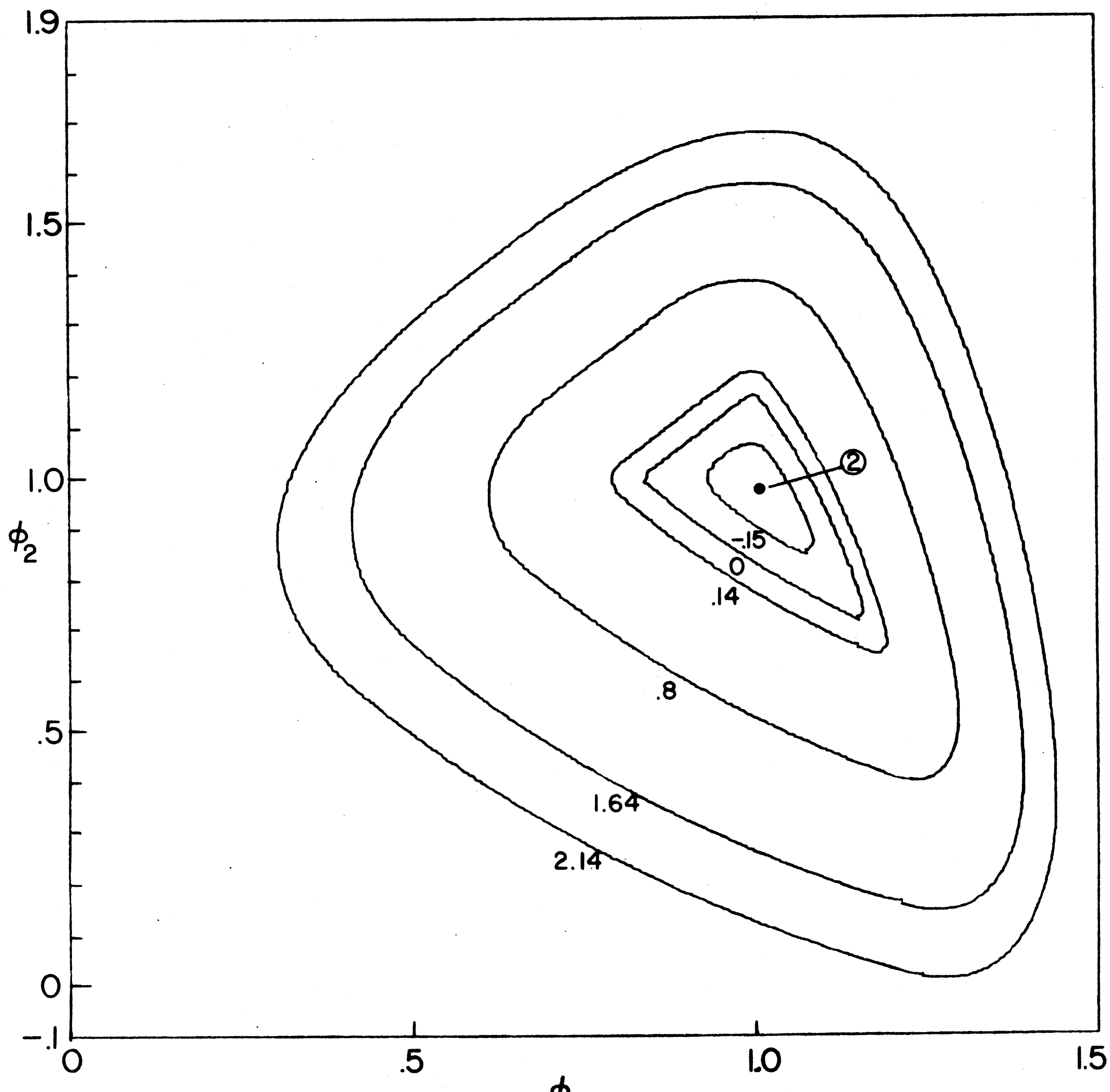

Fig. 3

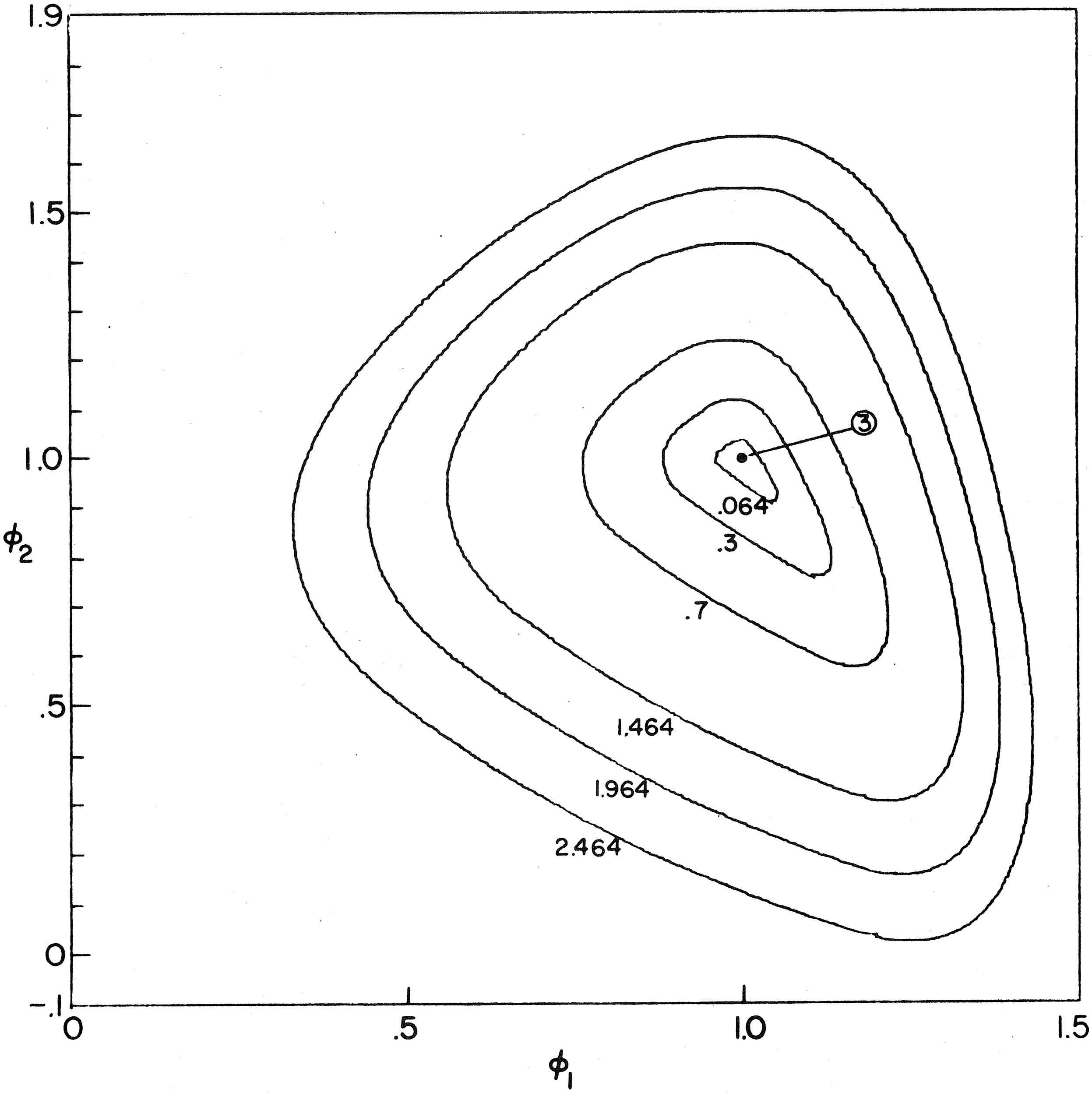

Fig 4

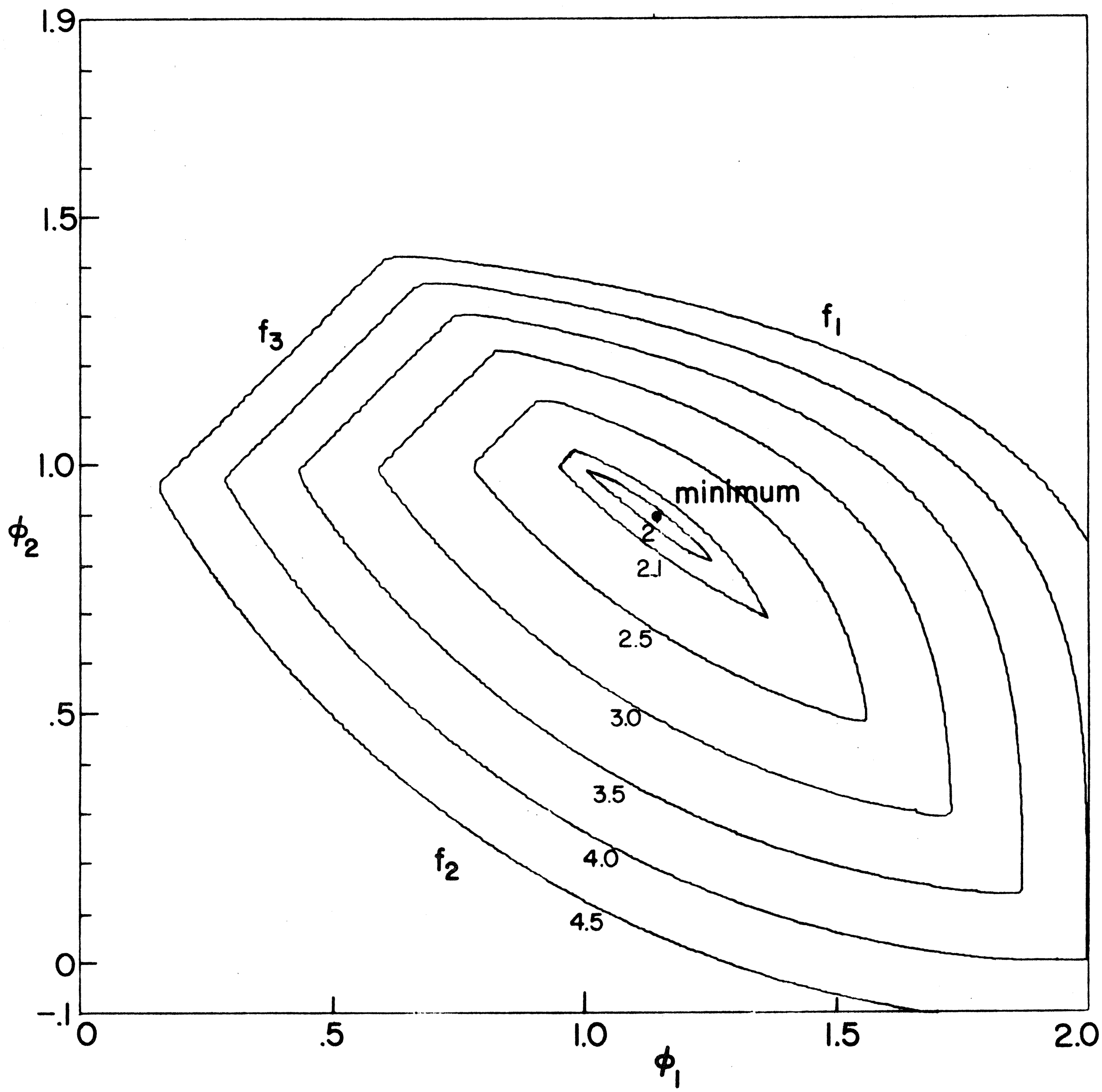
1.9
1.5
1.0
.5
0
-.1
ϕ_2
0
.5
1.0
1.5
2.0
ϕ_1
f_1
f_2
f_3
minimum
2
2.1
2.5
3.0
3.5
4.0
4.5

Fig. 5