# A METHOD OF FEASIBLE DIRECTIONS USING FUNCTION APPROXIMATIONS, WITH APPLICATIONS TO MIN MAX PROBLEMS

by

R. Klessig and E. Polak

Memorandum No. ERL-M287

26 October 1970

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

# A Method of Feasible Directions Using Function Approximations, with Applications to Min Max Problems.

R. Klessig and E. Polak

Department of Electrical Engineering and Computer Sciences
and the Electronics Research Laboratory
University of California, Berkeley, California 94720

Abstract

This paper presents a demonstrably convergent method of feasible directions for solving the problem $\min\{\phi(x) \mid g^i(x) \leq 0 \quad i = 1,2,\ldots,m\}$, which approximates adaptively, both $\phi(x)$ and $\nabla\phi(x)$. There approximations are necessitated by the fact that in certain problems, such as when $\phi(x) = \max\{f(x,y) \mid y \in \Omega_y\}$, a precise evaluation of $\phi(x)$ and $\nabla\phi(x)$ is extremely costly. The adaptive procedure progressively refines the precision of the approximations as an optimum is approached and as a result should be much more efficient than fixed precision algorithms.

It is shown how this new algorithm can be used for solving problems of the form $\min_{y\in\Omega_x} \max_{y\in\Omega_y} f(x,y)$ under the assumption that $\Omega_x = \{x \mid g^j(x) \leq 0,\ j = 1,\ldots,s\} \subset \mathbb{R}^n$, $\Omega_y = \{y \mid \zeta^i(y) \leq 0,\ i = 1,\ldots,t\} \subset \mathbb{R}^m$, with $f$, $g^j$, $\zeta^i$ continuously differentiable, $f(x,\cdot)$ concave, $\zeta^i$ convex for $i = 1,\ldots,t$, and $\Omega_x$, $\Omega_y$ compact.

Research sponsored by the National Aeronautics and Space Administration, Grant NGL-05-003-016(Sup 7&8).

## I. Introduction

One of the major classes of algorithms for solving nonlinear programming problems of the form $\min\{\phi(x) | g(x) \leq 0\}$ (with $\phi$: $\mathbb{R}^n \to \mathbb{R}^1$, g: $\mathbb{R}^n \to \mathbb{R}^m$ continuously differentiable) is the class of methods of feasible directions. This class of algorithms was introduced by Zoutendijk [1], in 1959. Since then a number of more or less directly related algorithms in this class have been proposed by Zoukhovitzkii, Polyak and Primak [2], Topkis and Veinott [3] and Polak [4]. Incidentally, the Frank and Wolfe method [5], the Rosen gradient projection method [6] and the Huard modified method of centers [7] can also be considered as belonging to this class of algorithms. All these algorithms have the following feature in common: to compute $x_{i+1}$ from $x_i$, one must compute both $\phi(x_i)$ and $\nabla\phi(x_i)$. Although usually this results in no difficulty, there are some cases where the need to compute $\phi(x_i)$ (and $\nabla\phi(x_i)$) leads to severe complications. For example, suppose that $\phi(x) \equiv \max_{y \in \Omega_y} f(x,y)$. Then, to compute $\phi(x)$ we must bring in a subprocedure (probably also a method of feasible directions) which constructs a sequence $\{y^j\}$ such that $f(x,y^j) \to \phi(x)$ as $j \to \infty$. Therefore, if viewed constructively, a method of feasible directions cannot be applied to such a problem, since we would have to compute an infinite sequence $\{x_i\}$, each element of which is only obtainable as the limit point of an infinite sequence $\{y_i^j\}$. Even if one adopts a nontheoretical point of view , it is clear that the computation of adequate approximations to $\phi(x_i)$ and to $\nabla\phi(x_i)$ is bound to be extremely time consuming when $\phi(x) \equiv \max_{y \in \Omega_y} f(y,x)$.

We shall show in this paper how one particular method of feasible directions (due to Polak [4]) can be modified and extended so as to eliminate both the theoretical and practical difficulties indicated above. A similar treatment also appears to be possible for some of the other methods of feasible directions. To obtain our new algorithm, we began by extending some of the methods for implementing theoretical algorithms discussed in [8], [9] and [10]. The major source of difficulty in this task stemmed from the fact that we wanted to obtain an algorithm which can be used for min-max problems, i.e. problems in which approximations to $\phi(x)$ (which is to be minimized) must be computed by constructing a sequence $\{y^j\}$ which maximizes $f(x,y)$ (the assumptions in [9] specifically exclude this case). Having obtained a method for implementing theoretical algorithms, which coped with the various difficulties we foresaw, we proceeded to modify the above mentioned method of feasible directions. The final result is a rather complex algorithm. While we probably could have covered our tracks and presented the baffled reader with this complex algorithm without a word of explanation, we feel that the process by which it was constructed is important in itself and should be made available to the reader. Consequently, in the next section, we present a general method for implementation of certain types of theoretical algorithms. After that we shall construct a method of feasible directions using function approximations, and, finally, we shall show that it applies to min-max problems.

## II. A Model for Implementation.

Let $\mathcal{X}$ be a normed linear space and let T be a closed subset of $\mathcal{X}$. Suppose that T contains a set $\Delta$ of desirable points and that we wish to find an $x \in \Delta$. Quite commonly, a theoretical algorithm for finding an $x \in \Delta$ will make use of a search function $A: T \to 2^T$ and of a stop rule (surrogate cost) function $c: T \to \mathbb{R}^1$ and will have the form below.

### 2.1 Algorithm Model

Step 0: Select an $x_0 \in T$ and set $i = 0$.

Step 1: Compute a $y \in A(x_i)$.

Step 2: If $c(y) \geq c(x_i)$, stop; else, set $x_{i+1} = y$ and go to Step 3.

Step 3: Set $i = i+1$ and go to Step 1.

2.2 Theorem (Polak [10]). Suppose that $c(\cdot)$ is continuous on T, and that for every $x \in T$ satisfying $x \notin \Delta$ there exists an $\varepsilon(x) > 0$ and a $\delta(x) > 0$ such that for all $x' \in T$, with $\|x' - x\| \leq \varepsilon(x)$,

$$c(y) - c(x') \leq -\delta(x) \qquad \forall\, y \in A(x'). \tag{2.3}$$

Then, either the sequence $\{x_i\}$ constructed by algorithm (2.1) is finite and its last element is in $\Delta$, or else $\{x_i\}$ is infinite and every accumulation point of $\{x_i\}$ is in $\Delta$. $\square$

When the functions $A(\cdot)$ and $c(\cdot)$ appearing in (1) cannot be evaluated in a reasonable manner, one needs to approximate $A(x)$ and $c(x)$ somehow. In this paper, we shall use sequences $\{A_j(\cdot)\}_{j=0}^{\infty}$ and $\{C_j(\cdot)\}_{j=0}^{\infty}$ of approximating functions, where $A_j: T \to 2^T$ and $c_1: T \to 2^{\mathbb{R}^1}$ for $j = 0,1,2...$

We shall assume that the functions $c(\cdot)$, $C_j(\cdot)$ and $A_j(\cdot)$, and the sets $T$ and $\Delta$ have the following properties.

2.4 Assumptions:

(i) $c(\cdot)$ is continuous on $T$;

(ii) $T$ is compact;

(iii) Given any $x \in T$ satisfying $x \notin \Delta$, there exists an $\varepsilon(x) > 0$, a $\delta(x) > 0$ and an integer $N(x) \geq 0$ such that

2.5 $$c_j(y) - c_j(x') \leq -\delta(x) \qquad \forall\, y \in A_j(x'),\ \forall\, x' \in B(x,\varepsilon(x)),$$
$$\forall\, c_j(x') \in C_j(x'),\ \forall\, c_j(y) \in C_j(y),\ \forall\, j \geq N(x),$$

where

2.6 $$B(x,\varepsilon) \triangleq \{x' \in T \mid \|x' - x\| \leq \varepsilon\} \ ;$$

(iv) Given any integer $j \geq 0$, there exists a $w_j > -\infty$ such that

2.7 $$c_j(x) \geq w_j \qquad \forall\, c_j(x) \in C_j(x),\ \forall\, x \in T.$$

(v) Given any $\gamma > 0$, there exist an integer $M(\gamma) \geq 0$ such that

2.8 $$|c_j(x) - c(x)| \leq \gamma \qquad \forall\, c_j(x) \in C_j(x),\ \forall\, j \geq M(\gamma),\ \forall\, x \in T. \ \square$$

It is not difficult to see that the assumptions in (2.4) are directly related to those in theorem (2.2) and the requirement that the $C_j(\cdot)$ be approximations to $c(\cdot)$ and that the $A_j(\cdot)$ be approximations to $A(\cdot)$. In terms of these new functions algorithm (2.1) expands as follows.

2.9 Algorithm Model

Step 0: Select an $x_0 \in T$; select parameters $\varepsilon_0 > 0$, $\alpha \in (0,1)$, and an integer $j_0 \geq 0$. Set $i = 0$, $j = j_0$, $q(0) = j_0$, and $\varepsilon = \varepsilon_0$.

<u>Step 1</u>: Compute a $c_j(x_i) \in C_j(x_i)$.

<u>Step 2</u>: Compute a $y \in A_j(x_i)$ and a $c_j(y) \in C_j(y)$.

<u>Step 3</u>: If $c_j(y) - c_j(x_i) > -\varepsilon$, set $j = j+1$, $\varepsilon = \alpha\varepsilon$ and go to Step 1; else set $x_{i+1} = y$, $\varepsilon_{i+1} = \varepsilon$, $q(i+1) = j$ and go to Step 4.

<u>Step 4</u>: Set $i = i+1$ and go to Step 2. □

2.10 <u>Comment</u>: The $\varepsilon$-test in Step 3 above serves the purpose of ensuring that the integer $j$ used at $x_i$ was sufficiently large for the approximations $A_j(x_i)$, $C_j(x_i)$, $C_j(y)$, to $A(x_i)$, $c(x_i)$, $c(y)$, to be adequate. It is borrowed from a simpler implementation of (2.1) which only approximates $A(\cdot)$, see (1.3.34) of [10]. □

2.11 <u>Comment</u>: The sequence $\{q(i)\}$ and $\{\varepsilon_i\}$ are defined in (2.9) only because we shall need them later. Note that for $i = 0, 1,2,3,...$

2.12 $$\varepsilon_i = \alpha^{q(i)}\varepsilon_0$$

2.13 $$x_{i+1} \in A_{q(i+1)}(x_i) \text{ . } \square$$

The following lemmas will enable us to state the convergence properties of algorithm (2.9)

2.14 <u>Lemma</u>: Suppose that the algorithm (2.9) jams up at a point $x_i$, cycling indefinitely between Steps 3 and 1. Then $x_i \in \Delta$.

<u>Proof</u>: Suppose that the algorithm (2.9) jams up at $x_i$ and that $x_i \notin \Delta$. Then by (2.4)(iii) there exist an $\varepsilon(x_i) > 0$ a $\delta(x_i) > 0$ and an integer $N(x_i) \geq 0$, such that

2.15 $c_j(y) - c_j(x_i) \le -\delta(x_i) \qquad \forall\, y \in A_j(x_i),\ \forall\, c_j(x_i) \in C_j(x_i),$

$$\forall\, c_j(y) \in C_j(y),\ \forall\, j \ge N(x_i)\ .$$

Since the algorithm is cycling indefinitely between Steps 3 and 1, it must be constructing sequences $\{y_r\}_{r=0}^{\infty}$, $\{c_{q(i)+r}\ (x_i)\}_{i=0}^{\infty}$ and $\{c_{q(i)+r}(y_r)\}_{r=0}^{\infty}$, such that

2.16 $y_r \in A_{q(i)+r}(x_i),\ c_{q(i)+r}(x_i) \in C_{q(i)+r}(x_i),\ c_{q(i)+r}(y_r) \in$

$$C_{q(i)+r}(y_r),\ r = 0,1,2,\ldots$$

and

2.17 $c_{q(i)+r}(y_r) - c_{q(i)+r}(x_i) > -\alpha^{q(i)+r}\varepsilon_0 = -\alpha^r\varepsilon_i,\ r = 0,1,2,\ldots\ .$

However, there exists an integer $p \ge 0$ such that

2.18 $\alpha^{q(i)+p}\varepsilon_0 \le \delta(x_i)$ and $q(i) + p \ge N(x_i)$ .

Consequently, for $r \ge p$, (2.17) contradicts (2.15) and (2.18) and hence we conclude that we must have $x_i \in \Delta$. □

2.19 Lemma: Consider the sequences $\{\varepsilon_i\}$ and $\{q(i)\}$ generated by algorithm (2.9) while constructing a sequence $\{x_i\} \subset T$. If $\{x_i\}$ is infinite, then $q(i) \to \infty$ and $\varepsilon(i) \to 0$ as $i \to \infty$.

Proof: Suppose that $\{x_i\}$ is infinite. Then $\{\varepsilon_i\}$ is an infinite, monotonically decreasing sequence bounded from below by zero. Consequently, $\varepsilon_i \to \varepsilon^* \ge 0$ for $i \to \infty$. Suppose that $\varepsilon^* > 0$. We shall show that this leads to a contradiction.

Since $\varepsilon_i \to \varepsilon^*$ and $\varepsilon^* > 0$, it follows from (2.12) that there exists an integer $N'$ such that for $i \geq N'$ $\varepsilon_i = \varepsilon_{i+1} = \ldots = \varepsilon^*$ and $q(i) = q(i+1) = \ldots . = q^*$. It now follows from the test in Step 2 of (2.9) that for $i \geq N'$, $c_{q(i+1)}(x_i) \in C_{q^*}(x_i)$ and $c_{q(i+1)}(x_{i+1}) \in C_{q^*}(x_{i+1})$. We may therefore write

$$2.20 \qquad c_{q(i+1)}(x_{i+1}) - c_{q(i+1)}(x_i) = c_{q^*}(x_{i+1}) - c_{q^*}(x_i) \leq - \varepsilon_{i+1} = - \varepsilon^* \qquad \forall\, i \geq N',$$

where $c_{q^*}(x_i) = c_{q(i+1)}(x_i)$ and $c_{q^*}(x_{i+1}) = c_{q(i+1)}(x_{i+1})$.

Therefore, we must have $c_{q^*}(x_i) \to -\infty$ as $i \to \infty$. But, by (2.7) $c_{q^*}(x_i) \geq \omega_{q^*} > -\infty$, and hence we have a contradiction. Therefore $\varepsilon^* = 0$. Finally, since $\varepsilon_i \to 0$ as $i \to \infty$, it follows from (2.12) that $q(i) \to \infty$ as $i \to \infty$. $\square$

2.21 <u>Lemma</u>: Suppose that algorithm (2.9) constructs an infinite sequence $\{x_i\}_{i=0}^{\infty}$. Let $\Lambda$ denote the set of accumulation points of $\{x_i\}_{i=0}^{\infty}$. Then, given any $\gamma > 0$, there exists an integer $P(\gamma)$ such that

$$2.22 \qquad \min\{\|x_i - x^*\| \,|\, x^* \in \Lambda\} \leq \gamma \qquad \forall\, i \geq P(\gamma).$$

<u>Proof</u>: Since $\{x_i\}_{i=0}^{\infty} \subset T$ and $T$ is compact by (2.4)(ii), $\Lambda$ is a nonempty, compact set. Hence the min in (2.22) is well defined. Now suppose that there is no integer $P(\gamma)$ for which (2.22) holds. Then there must exist a subsequence $\{x_i\}_{i \in K}$, $K \subset \{0,1,\ldots\}$ such that

$$2.23 \qquad \min\{\|x_i - x^*\| \,|\, x^* \in \Lambda\} > \gamma \qquad \forall\, i \in K$$

But $\{x_i\}_{i \in K}$ is a compact sequence and hence there exists a subsequence $\{x_i\}_{i \in K_1}$, with $K_1 \subset K$ such that $x_i \to \tilde{x}$ as $i \to \infty$, for $i \in K_1$. By the definition of $\Lambda$, $\tilde{x} \in \Lambda$, and, since $x_i \to \tilde{x}$ as $i \to \infty$ for $i \in K_1$, (2.23) cannot hold for all $i \in K_1 \subset K$. Consequently (2.22) must hold. $\square$

2.24 <u>Theorem</u>: Algorithm (2.9) will either jam up at a point $x_i$, cycling indefinitely between Steps 3 and 1, in which case $x_i \in \Delta$, or else, it will construct an infinite sequence $\{x_i\}$ which has at least one accumulation point in $\Delta$.

<u>Proof</u>: The first part of the theorem was established in lemma (2.14). Hence, suppose that $\{x_i\}$ is infinite. To obtain a contradiction, suppose that $\Lambda \cap \Delta = \phi$, where $\Lambda$ is the set of accumulation points of $\{x_i\}$. Since $T$ is compact, $\Lambda$ is a nonempty compact set, and hence (because we have assumed that $\Lambda \cap \Delta = \phi$) it follows from (2.4)(iii) that there exist an $\varepsilon_\Lambda > 0$, a $\delta_\Lambda > 0$ and an integer $N_\Lambda \geq 0$ such that

2.25 $$c_j(y) - c_j(x') \leq -\delta_\Lambda \qquad \forall\, y \in A_j(x'),\ \forall\, x' \in B(x^*, \varepsilon_\Lambda),$$
$$\forall\, c_j(y) \in C_j(y),\ \forall\, c_j(x') \in C_j(x'),\ \forall\, x^* \in \Lambda,\ \forall\, j \geq N_\Lambda .$$

Let $P(\varepsilon_\Lambda)$ be defined as in (2.22) (for $\gamma = \varepsilon_\Lambda$). Then, since $q(i) \to \infty$ as $i \to \infty$ by lemma (2.19), there exists an integer $N_1 \geq P(\varepsilon_\Lambda)$ such that $q(i) \geq N_\Lambda$ for all $i \geq N_1$, and hence

2.26 $$c_{q(i+1)}(x_{i+1}) - c_{q(i+1)}(x_i) \leq -\delta_\Lambda \qquad \forall\, c_{q(i+1)}(x_i) \in C_{q(i+1)}(x_i),$$
$$\forall\, c_{q(i+1)}(x_{i+1}) \in C_{q(i+1)}(x_{i+1}) \quad \forall\, i \geq N_1 .$$

Now, from (2.4)(v)(see (2.8)), we conclude that there exists an integer

$N_2 \geq N_1$ such that

$$2.27 \qquad |c_j(x_i) - c(x_i)| \leq \delta_\Lambda/4 \qquad \forall c_j(x_i) \in C_j(x_i),\ \forall\ i \geq N_2,\ \forall\ j \geq q(i)\ .$$

Hence, since $q(i+1) \geq q(i)$, for $i = 0,1,2,\ldots,$

$$2.28 \qquad c_{q(i)}(x_i) \geq c(x_i) - \delta_\Lambda/4 \geq c_{q(i+1)}(x_i) - \delta_\Lambda/2 \qquad \forall\ c_{q(i)}(x_i) \in C_{q(i)}(x_i),$$
$$\forall\ c_{q(i+1)}(x_i) \in C_{q(i+1)}(x_i),\ \forall\ i \geq N_2\ .$$

Combining (2.28) with (2.26), we now get

$$2.29 \qquad c_{q(i)}(x_i) \geq c_{q(i+1)}(x_i) - \delta_\Lambda/2 \geq c_{q(i+1)}(x_{i+1}) + \delta_{\Lambda/2}$$
$$\forall\ c_{q(i)}(x_i) \in C_{q(i)}(x_i),\ \forall\ c_{q(i+1)}(x_i) \in C_{q(i+1)}(x_i),$$
$$\forall\ c_{q(i+1)}(x_{i+1}) \in C_{q(i+1)}(x_{i+1}),\ \forall\ i \geq N_2,$$

and therefore we must have $c_{q(i)}(x_i) \to -\infty$ as $i \to \infty$, for any $c_{q(i)}(x_i) \in C_{q(i)}(x_i)$, $i = 0,1,\ldots$ .

Now let $K \subset \{0,1,2,\ldots\}$ be such that $x_i \to x^* \in \Lambda$ as $i \to \infty$, $i \in K$. Then, by (2.4)(v), and lemma (2.19), there exists an integer $N_3 \geq 0$ such that

$$2.30 \qquad |c_{q(i)}(x_i) - c(x_i)| \leq |c(x^*)|/4 \qquad \forall\ c_{q(i)}(x_i) \in C_{q(i)}(x_i),$$
$$\forall\ i \geq N_3,\ i \in K,$$

and also, since $c(\cdot)$ is continuous,

$$2.31 \qquad |c(x_i) - c(x^*)| \leq |c(x^*)|/4 \qquad \forall\ i \geq N_3,\ i \in K,$$

where $c(x^*) > -\infty$ because $c(\cdot)$ is continuous on T. Combining (2.30) and

(2.31), we obtain,

$$2.32 \qquad c_{q(i)}(x_i) \geq c(x^*) - |c(x^*)|/2 \geq -\frac{3}{2}\,|c(x^*)| > -\infty\,,$$

$$\forall\; c_{q(i)}(x_i) \in C_{q(i)}(x_i),\ \forall\; i \geq N_3,\ i \in K,$$

which contradicts our previous conclusion that

$c_{q(i)}(x_i) \to -\infty$ as $i \to \infty$, for any $c_{q(i)}(x_i) \in C_{q(i)}(x_i)$, based on the hypothesis that $\Lambda \cap \Delta = \phi$. Hence $\Lambda \cap \Delta \neq \phi$ and we are done. □

Theorem (2.24) states that when the sequence $\{x_i\}$ is infinite, it must have at least one accumulation point in $\Delta$, the set of desirable points. Clearly, if $x_i \to x^*$ as $i \to \infty$, $x^* \in \Delta$. The reader may well wonder as to the value of algorithm (2.9) when the sequences it constructs have more than one accumulation point. Although at present, we cannot make a general statement, we can assert that it is sometimes possible to add to an algorithm of the form of (2.9) a simple subprocedure which sifts out a subsequence, <u>all</u> of whose accumulation points are in $\Delta$. In such a case, we obtain an algorithm of value. In particular, we shall see that the above assertion applies to the algorithm which we shall develop in the next section.

With these preliminaries out of the way, we shall now construct a new method of feasible directions, using function approximations.

## III. An Implementation of the Polak Method of Feasible Directions

Consider the problem

$$3.1 \qquad \min\{\phi(x)\,|\,g(x) \leqq 0\}\,,$$

where $\phi$: $\mathbb{R}^n \to \mathbb{R}^1$ and $g$: $\mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable functions. Let $\Omega_x \subset \mathbb{R}^n$ be defined by

3.2 $$\Omega_x = \{x | g(x) \leqq 0\} .$$

Now, for any $x \in \Omega_x$ and for any $\varepsilon \geq 0$, let the index set $I_x(x,\varepsilon) \subset \{1,2,3,..,m\}$ be defined by

3.3 $$I_x(x,\varepsilon) = \{q \in \{1,2,..,m\} | g^q(x) \geq -\varepsilon\} ,$$

let $S \subset \mathbb{R}^n$ be defined by

3.4 $$S = \{h \in \mathbb{R}^n | \|h\|_\infty \leq 1\} ,$$

and let $\Theta$: $\Omega_x \times \mathbb{R}^+ \to \mathbb{R}^1$ be defined by

$$\Theta(x,\varepsilon) = \min_{h \in S} \max\{\langle \nabla\phi(x),h \rangle ; \langle \nabla g^q(x),h \rangle , q \in I_x(x,\varepsilon)\} .$$

Note that $\Theta(x,\varepsilon) \leq 0$ for all $x \in \Omega_x$, for all $\varepsilon \geq 0$. We can now state a well known result.

3.5 <u>Proposition</u>: Suppose that $\hat{x} \in \Omega_x$ solves (3.1), i.e. $\phi(\hat{x}) = \min\{\phi(x) | x \in \Omega_x\}$ . Then for every $\varepsilon \geq 0$ $\Theta(\hat{x},\varepsilon) = 0$. $\square$

The Polak method of feasible directions is an algorithm for finding points $\hat{x}$ in $\Omega$ such that

3.6 $$\Theta(\hat{x},0) = 0,$$

i.e., it finds points in $\Omega$ which satisfy the optimality condition (3.5).

3.7 <u>The Polak Method of Feasible Directions (see (4.3.47) in [10])</u>.

<u>Step 0</u>: Compute an $x_0 \in \Omega_x$; select parameters $\varepsilon_0^1 > 0$; $\alpha_1 \in (0,1)$, set $i = 0$.

Step 1: Set $\varepsilon^1 = \varepsilon_0^1$ .

Step 2: Compute $\Theta(x_i,\varepsilon^1)$ and an $h(x_i,\varepsilon^1) \in S$ such that

3.8 $$\Theta(x_i,\varepsilon^1) = \max\{\langle \nabla\phi(x_i),h(x_i,\varepsilon^1) \rangle ; \langle \nabla g^q(x_i),h(x_i,\varepsilon^1) \rangle , \quad q \in I_x(x_i,\varepsilon^1)\} .$$

Step 3: If $\Theta(x_i,\varepsilon^1) = 0$, compute $\Theta(x_i,0)$ and go to Step 4; else, go to Step 5.

Step 4: If $\Theta(x_i,0) = 0$, set $x_{i+1} = x_i$ and stop; else, set $\varepsilon^1 = \alpha_1\varepsilon^1$ and go to Step 2.

Step 5: If $\Theta(x_i,\varepsilon^1) \leq - \varepsilon^1$, go to Step 6, else set $\varepsilon^1 = \alpha_1\varepsilon^1$ and go to Step 2.

Step 6: Set $\lambda = 1$.

Step 7: Compute

3.9 $$G \triangleq g(x_i + \lambda h(x_i,\varepsilon^1))$$

Step 8: If $G \leqq 0$, go to Step 9; else set $\lambda = \frac{\lambda}{2}$ and go to Step 7.

Step 9: Compute

3.10 $$\psi \triangleq \phi(x_i+\lambda h(x_i,\varepsilon^1)) - \phi(x_i) - \frac{\lambda}{2} \langle \nabla\phi(x_i),h(x_i,\varepsilon^1) \rangle .$$

Step 10: If $\psi > 0$, set $\lambda = \frac{\lambda}{2}$ and go to Step 7; else, set $x_{i+1} = x_i + \lambda h(x_i,\varepsilon_1)$ and go to Step 11.

Step 11: Set $i = i+1$ and go to Step 1.

The above algorithm is derived from a Zoutendijk method of feasible directions [1] by means of two simple modifications. The first modification is to substitute the Armijo step size rule $\psi \leq 0$ (see [11] ) for

Zoutendijk's step size rule $\min\{\phi(x_i+\lambda h(x_i,\varepsilon^1))|\lambda \geq 0\}$ . The second is to make the method time invariant by returning from Step 11 to Step 1 (Zoutendijk's method returns to Step 2). With Armijo's rule, the step size $\lambda$ can be calculated in a finite manner; with Zoutendijk's it cannot. The reason for the second modification is that it results in an algorithm which can be examined within the framework of theorem (2.2). Zoutendijk's original method requires much more complex machinery for analyses. Computationally, algorithm (3.7) is sometimes superior and sometimes inferior to Zoutendijk's method (assuming, of course, that the Armijo step size rule is also used to modify Zoutendijk's method). In practice, for best results, one would tend to alternate between the time varying and time invariant methods in the course of a long computation.

Now suppose that to compute $\phi(x)$ and $\nabla\phi(x)$ we must use a subprocedure which constructs two sequences, $\{\phi_j(x)\}_{j=0}^{\infty}$ , $\{\nabla_j\phi(x)\}_{j=0}^{\infty}$, such that $\phi_j(x) \to \phi(x)$ and $\nabla_j\phi(x) \to \nabla\phi(x)$ as $j \to \infty$. In constructing an algorithm which truncates these sequences we shall need the following hypotheses to hold (c.f (2.4)).

3.11 Assumptions:

i) The set $\Omega_x$ in (3.2) is compact.

ii) For $j = 0,1,2,\ldots$, $\Phi_j$: $\mathbb{R}^n \to 2^{\mathbb{R}^1}$, $\nabla_j\Phi$: $\mathbb{R}^n \to 2^{\mathbb{R}^n}$ are functions such that given any $\gamma > 0$ there exists an integer $M(\gamma) \geq 0$ such that

3.12 $$|\phi_j(x) - \phi(x)| \leq \gamma \quad \forall\, x \in \Omega_x,\ \forall\, \phi_j(x) \in \Phi_j(x),\ \forall\, j \geq M(\gamma),$$

3.13 $$\|\nabla_j\phi(x) - \nabla\phi(x)\| \leq \gamma \quad \forall\, x \in \Omega_x,\ \forall\, \nabla_j\phi(x) \in \nabla_j\Phi(x),\ \forall\, j \geq M(\gamma).$$

iii) Given any integer $j \geq 0$, there exists a $w_j > -\infty$ such that

3.14 $$\phi_j(x) \geq w_j \quad \forall\, \phi_j(x) \in \Phi_j(x),\ \forall\, x \in \Omega_x \quad \square$$

3.15 Definition: We define $\tilde{\Theta}$: $\Omega_x \times \mathbb{R}^n \times \mathbb{R}^+ \to \mathbb{R}^1$ and H: $\Omega_x \times \mathbb{R}^n \times \mathbb{R}^+ \to 2^S$ as

3.16 $$\tilde{\Theta}(x,u,\varepsilon) = \min_{h \in S} \max\{\langle u,h \rangle ; \langle \nabla g^q(x),h \rangle ,\ q \in I_x(x,\varepsilon)\}$$

and

3.17 $$\tilde{H}(x,u,\varepsilon) = \{ h \in S \,|\, \tilde{\Theta}(x,u,\varepsilon) = \max\{\langle u,h \rangle ; \langle \nabla g^q(x),h \rangle ,\ q \in I_x(x,\varepsilon)\}\} \quad \square .$$

Note: $\tilde{\Theta}(x,u,\varepsilon)$ and a vector $h \in \tilde{H}(x,u,\varepsilon)$ can be computed by solving a linear programing problem (see Sec. 4.3 in [10]).

We shall now modify algorithm (3.7) so as to make it correspond to algorithm Model (2.9), and, in addition, we shall add a sifting subprocedure to extract a subsequence $\{x_i\}_{i \in K}$ all of whose accumulation points $x^*$ will be shown to satisfy $\Theta(x^*,0) = 0$. For the sake of convenience, we break up the following algorithm into two subprocedures.

3.18 Implementation of Algorithm (3.7)

Subprocedure I: Method of Feasible Directions with Approximations

Begin: Step 0: Select parameters $\varepsilon_0^1 > 0$, $\varepsilon_0^2 > 0$, $\varepsilon_0^3 > 0$, $\lambda_{min} \in (0,1]$, $\alpha_1 \in (0,1)$, $\alpha_2 \in (0,1)$, $\alpha_3 \in (0,1)$ and an integer $j_0 \geq 0$; compute an $x_0 \in \Omega_x$; set $i = 0$, $j = j_0$, $k = 0$, $\varepsilon^2 = \varepsilon_0^2$, $\varepsilon^3 = \varepsilon_0^3$ .

Step 1: Set $\varepsilon^1 = \varepsilon_0^1$ .

Step 2: Compute a $\phi_j(x_i) \in \Phi_j(x_i)$ and a $\nabla_j\phi(x_i) \in \nabla_j\Phi(x_i)$.†

Step 3: Compute $\tilde{\Theta}(x_i,\nabla_j\phi(x_i),\varepsilon^1)$ and a vector $h(x_i,\nabla_j\phi(x_i),\varepsilon^1) \in \tilde{H}(x_i,\nabla_j\phi(x_i),\varepsilon^1)$.

Step 4: If $\tilde{\Theta}(x_i,\nabla_j\phi(x_i),\varepsilon^1) = 0$, compute $\tilde{\Theta}(x_i,\nabla_j\phi(x_i),0)$ and go to Step 5; else, go to Step 6.

Step 5: If $\tilde{\Theta}(x_i,\nabla_j\phi(x_i),0) = 0$, set $x' = x_i$, set $\phi_j(x') = \phi_j(x_i)$ and go to Step 14; else set $\varepsilon^1 = \alpha_1\varepsilon^1$ and go to Step 3.

Step 6: If $\tilde{\Theta}(x_i,\nabla_j\phi(x_i),\varepsilon^1) \leq -\varepsilon^1$, go to Step 7; else, set $\varepsilon^1 = \alpha_1\varepsilon^1$ and go to Step 3.

Step 7: Set $\lambda = 1$.

Step 8: Compute $G = g(x_i+\lambda h(x_i,\nabla_j\phi(x_i),\varepsilon^1))$.

Step 9: If $G \leq 0$, go to Step 10; else, set $\lambda = \lambda/2$ and go to Step 8.

Step 10: Compute a $\phi_j(x_i+\lambda h(x_i,\nabla_j\phi(x_i),\varepsilon^1)) \in \Phi_j(x_i+\lambda h(x_i,\nabla_j\phi(x_i),\varepsilon^1))$.

Step 11: Compute $D = \phi_j(x_i+\lambda h(x_i,\nabla_j\phi(x_i),\varepsilon^1)) - \phi_j(x_i) - \frac{\lambda}{2}\langle \nabla_j\phi(x_i),h(x_i,\nabla_j\phi(x_i),\varepsilon^1)\rangle$.

Step 12: If $D > 0$ go to Step 13; else set $x' = x_i+\lambda h(x_i,\nabla_j\phi(x_i),\varepsilon^1)$, set $\phi_j(x') = \phi_j(x_i+\lambda h(x_i,\nabla_j\phi(x_i),\varepsilon^1))$ and go to Step 14.

Step 13: If $\lambda \geq \lambda_{min}/2^j$, set $\lambda = \lambda/2$ and go to Step 8; else set $x' = x_i$, set $\phi_j(x') = \phi_j(x_i)$ and go to Step 14.

Step 14: If $\phi_j(x') - \phi_j(x_i) \leq -\varepsilon^2$, go to Step 15; else, set $j = j+1$, set $\varepsilon^2 = \alpha_2\varepsilon^2$ and go to Step 1.

Step 15: Set $x_{i+1} = x'$, set $q(i+1) = j$, $\varepsilon^2_{i+1} = \varepsilon^2$.

†Note that a $\nabla_j\phi(x_i) \in \nabla_j\Phi(x_i)$ may already be available because of its computation in Step 17 and hence need not be recomputed.

Comment: Do not compute $q(i+1)$ and $\varepsilon^2_{i+1}$. These quantities are introduced only for the convenience of the proofs to follow.

End: Step 16: Set $i = i+1$.

Subprocedure II: Sieve

Begin: Step 17: Compute a $\nabla_j\phi(x') \in \nabla_j\Phi(x')$.

Step 18: Compute $\tilde{\Theta}(x', \nabla_j\phi(x'),\varepsilon^3)$.

Step 19: If $\tilde{\Theta}(x',\nabla_j\phi(x'),\varepsilon^3) \geq -\varepsilon^3$, go to Step 20; else, go to Step 1.

Step 20: Set $z_k = x'$, set $\varepsilon^3_k = \varepsilon^3$, set $p(k) = q(i)$.

Comment: Do not compute $\varepsilon^3_k$ and $p(k)$. These quantities are introduced only for the convenience of the proofs to follow.

End. Step 21: Set $\varepsilon^3 = \alpha_3\varepsilon^3$, set $k = k+1$, and go to Step 1. $\square$

We shall now show that Subprocedure I (Steps 0 to 16) of algorithm (3.18) corresponds to the model (2.9), with the functions $A_j(\cdot)$ being defined by the Steps 1 to 13 of (3.18), and with $\Phi_j(\cdot)$, $\varepsilon^2$ and $\alpha_2$ in (3.18) taking the place of $C_j(\cdot)$, $\varepsilon$ and $\alpha$ in (2.9). The additional parameters in Step 0 of (3.18) are used either to define the $A_j(\cdot)$ or in the sifting Subprocedure II, defined by Steps 17 to 21 of (3.18).

First, we must show that the maps $A_j(\cdot)$ are well defined by Steps 1 to 13 of (3.18), i.e. that Subprocedure I of (3.18) cannot jam up before reaching Step 14. We shall do this in the following lemmas.

3.19 Proposition: For any $x \in \Omega_x$, there exists a $\rho(x) > 0$ such that

3.20 $$I_x(x,\varepsilon) = I_x(x,0) \qquad \forall\ \varepsilon \in [0,\rho(x)],$$

3.21 $$\tilde{\Theta}(x,u,\varepsilon) = \tilde{\Theta}(x,u,0) \qquad \forall\ \varepsilon \in [0,\rho(x)],\ \forall\ u \in \mathbb{R}^n.$$

3.22 __Lemma__: Subprocedure I of algorithm (3.18) cannot cycle indefinitely in the loop defined by Steps 3 through 6.

__Proof__: Suppose that $\tilde{\Theta}(x_i)u,0) = 0$ for some $u \in \mathbb{R}^n$. Then, since $I_x(x_i,0) \subset I_x(x_i,\varepsilon^1)$ for all $\varepsilon^1 \geq 0$, we must have

3.23 $$0 = \tilde{\Theta}(x_i,u,0) \leq \tilde{\Theta}(x_i,u,\varepsilon^1) \leq 0,$$

and hence $\tilde{\Theta}(x_i,u,\varepsilon^1) = 0$. So that when $\tilde{\Theta}(x_i,\nabla_j\phi(x_i),0) = 0$, algorithm (3.18) proceeds from Step 3 to Step 4 to Step 5 and hence to Step 14. Now suppose that $\tilde{\Theta}(x_i,\nabla_j\phi(x_i),0) < 0$. It then follows from proposition (3.19) that when $\varepsilon$ has become reduced to the point where $\varepsilon^1 \leq \min\{\rho(x_i), -\tilde{\Theta}(x_i,\nabla_j\phi(x_i),0)\}$, which is a finite process, we shall have $\tilde{\Theta}(x_i,\nabla_j\phi(x_i),\varepsilon^1) \leq -\varepsilon^1$ and algorithm will proceed from Step 6 to Step 7. Consequently, algorithm (3.18) cannot jam up in the loop defined by Steps 3 to 6. □

3.24 __Proposition__: Suppose that $x_i \in \Omega_x$, $u \in \mathbb{R}^n$, $\varepsilon^1 > 0$, and $j \geq 0$ are such that $\tilde{\Theta}(x_i,u,\varepsilon^1) \leq -\varepsilon^1$. Then there exists a $\lambda(x_i,\varepsilon^1) > 0$ such that

3.25 $$g(x_i+\lambda h) \leqq 0 \qquad \forall\ \lambda \in [0,\lambda(x_i,\varepsilon^1)],\ \forall\ h \in \tilde{H}(x_i,u,\varepsilon^1).$$ □

3.26 __Lemma__: Suprocedure I of algorithm (3.18) cannot cycle indefinitely in the loop defined by Steps 8 and 9.

__Proof__: If in the execution of algorithm (3.18) Step 8 has been reached, then by the test in Step 6, we must have $\tilde{\Theta}(x_i, \nabla_j \phi(x_i), \varepsilon^1) \leq -\varepsilon^1$, with $\varepsilon^1 > 0$. It therefore follows from proposition (3.24) that, after a finite number of halvings, $\lambda$ will be sufficiently small to satisfy $G \leq 0$ in Step 9. Hence algorithm (3.18) cannot jam up in the loop defined by Steps 8 and 9. $\square$

We have thus established that Steps 1 to 13 of algorithm (3.18) define a map $A_j: \Omega_x \to 2^{\Omega_x}$ ($x' \in A_j(x_i)$ with $x'$ defined in Step 5, in Step 12, or in Step 13, as may be appropriate). To show that Steps 0 to 16 of algorithm (3.18) correspond to algorithm model (2.9), we introduce the following correspondences.

3.27 Correspondence Table:

| Algorithm Model (2.9) | Subprocedure I of (3.18) |
|---|---|
| $\mathcal{X}$ | $\mathbb{R}^n$ |
| $T$ | $\Omega_x$ |
| $A_j(\cdot)$ | Steps 1-13 |
| $C_j(\cdot)$ | $\Phi_j(\cdot)$ |
| $c(\cdot)$ | $\phi(\cdot)$ |
| $\Delta$ | $\{x \in \Omega_x \mid \Theta(x,0) = 0\}$ . |

To conclude that theorem (2.24) applies to Subprocedure I of algorithm (3.18), we must show that the assumptions (2.4)(i)-(v) are satisfied. It follows directly from (3.1) and (3.11) that the assumptions (2.4)(i), (2.4)(ii), (2.4)(iv) and (2.4)(v) are satisfied. It remains to show that assumption (2.4)(iii) is satisfied. This will require several lemmas.

Definition: For any $x \in \Omega_x$ and any $\varepsilon > 0$, we define

3.28 $$B_x(x,\varepsilon) \triangleq \{x' \in \Omega_x \mid \|x' - x\| \leq \varepsilon\}. \quad \square$$

3.29 Proposition: Given any $x \in \Omega_x$, there exists a $\rho(x) > 0$ such that

3.30 $$I_x(x',\varepsilon) \subset I_x(x,0) \qquad \forall\, x' \in B_x(x,\rho(x)),\ \forall\, \varepsilon \in [0,\rho(x)]. \quad \square$$

3.31 Lemma: Given any $x \in \Omega_x$ and any $\gamma > 0$, there exists a $\rho(x,\gamma) > 0$ such that

3.32 $$\Theta(x',\varepsilon) \leq \Theta(x,0) + \gamma \qquad \forall\, x' \in B_x(x,\rho(x,\gamma)),\ \forall\, \varepsilon \in [0,\rho(x,\gamma)].$$

Proof: Let $x \in \Omega_x$ and $\gamma > 0$ be given. We define m: $\mathbb{R}^n \to \mathbb{R}^1$ by

3.33 $$m(z) = \min_{h \in S} \max\{\langle \nabla\phi(z),h \rangle ; \langle \nabla g^q(z),h \rangle,\ q \in I_x(x,0)\}.$$

Clearly $m(\cdot)$ is continuous and

3.34 $$m(x) = \Theta(x,0).$$

Now, since $m(\cdot)$ is continuous and because of (3.30), there exists a $\rho(x,\gamma) > 0$ such that $I_x(x',\varepsilon) \subset I_x(x,0)$ and $m(x') \leq m(x) + \gamma$ for all $x' \in B_x(x,\rho(x,\gamma))$, for all $\varepsilon \in [0,\rho(x,\gamma)]$. Consequently,

3.35 $$\Theta(x',\varepsilon) \leq m(x') \leq m(x) + \gamma = \Theta(x,0) + \gamma, \qquad \forall\, x' \in B_x(x,\rho(x,\gamma)),\ \forall\, \varepsilon \in [0,\rho(x,\gamma)],$$

and hence we are done. $\square$

3.36 Corollary: Given any $x \in \Omega_x$ and any $\gamma > 0$, there exist a $\rho(x,\gamma) > 0$ and an integer $M'(\gamma)$ such that

3.37 $$\tilde{\Theta}(x',u,\varepsilon) \le \Theta(x,0) + \gamma \qquad \forall\, x' \in B_x(x,\rho(x,\gamma)),\ \forall\, u \in \nabla_j\Phi(x'),$$
$$\forall\, j \ge M'(\gamma),\ \forall\, \varepsilon \in [0,\rho(x,\gamma)].$$

Proof: Since S is compact, it follows from (3.13) that there exists an integer $M'(\gamma)$ such that

3.38 $$|\langle u,h \rangle - \langle \nabla\phi(x'),h \rangle| \le \gamma/2 \qquad \forall\, x' \in \Omega_x,\ \forall\, h \in S,$$
$$\forall\, u \in \nabla_j\Phi(x'),\ \forall\, j \ge M'(\gamma).$$

Hence,

3.39 $$\tilde{\Theta}(x',u,\varepsilon) \le \Theta(x',\varepsilon) + \gamma/2 \qquad \forall\, x' \in \Omega_x,\ \forall\, u \in \nabla_j\Phi(x'),$$
$$\forall\, \varepsilon \ge 0,\ \forall\, j \ge M'(\gamma).$$

Finally, utilizing (3.39) and (3.32), where we replace $\gamma$ by $\gamma/2$, we obtain (3.37). □

3.40 Lemma: Suppose that $x \in \Omega_x$ satisfies $\Theta(x,0) < 0$. Then there exists an $\varepsilon(x) > 0$ and an integer $N(x) \ge 0$ such that for all $x_i \in B_x(x,\varepsilon(x))$ and for all integers $j \ge N(x)$, algorithm (3.18) satisfies $\tilde{\Theta}(x_i,\nabla_j\phi(x_i),\varepsilon^1) \le -\varepsilon^1$ in Step 6, and reaches Step 7, with $\varepsilon^1$ satisfying

3.41 $$\varepsilon^1 > \varepsilon(x)$$

Proof: Suppose that $x \in \Omega_x$ is such that $\Theta(x,0) < 0$. Then, by corollary (3.36), there exist an $\rho(x) > 0$ and an integer $N(x) \ge$ such that

3.42 $$\tilde{\Theta}(x_i,u,\varepsilon^1) \le \frac{1}{2}\,\Theta(x,0) < 0 \qquad \forall\, x_i \in B_x(x,\rho(x)),$$
$$\forall\, u \in \nabla_j\Phi(x_i),\ \forall\, \varepsilon^1 \in [0,\rho(x)],\ \forall\, j \ge N(x).$$

Let $\bar{\varepsilon}(x) = \min\{\rho(x), -\frac{1}{2}\Theta(x,0)\}$. Then, by (3.42),

3.43 $$\tilde{\Theta}(x_i,u,\varepsilon^1) \leq -\varepsilon^1 \qquad \forall\, x_i \in B_x(x,\bar{\varepsilon}(x)),\ \forall\, u \in \nabla_j\Phi(x_i),$$
$$\forall\, \varepsilon^1 \in [0,\bar{\varepsilon}(x)],\ \forall\, j \geq N(x).$$

Since Step 6 of algorithm (3.18) requires that (3.43) be satisfied with $\varepsilon^1 = \alpha_1^p\, \varepsilon_0^1$, for some integer $p \geq 0$, we see that if we set $\varepsilon(x) = \alpha_1\bar{\varepsilon}(x)$, then (3.43) can always be satisfied with $\varepsilon^1 = \alpha_1^p\, \varepsilon_0^1 > \varepsilon(x)$, for some integer p, and hence we are done. □

3.44 <u>Corollary</u>: Suppose that $x \in \Omega_x$ satisfies $\Theta(x,0) < 0$, and suppose that $\varepsilon(x) > 0$ and the integer $N(x) \geq 0$ are such that the conclusion of lemma (3.40) holds. Then there exists an integer $\ell(x) \geq 0$ such that

3.45 $$g(x_i + (\tfrac{1}{2})^p h) \leqq 0 \qquad \forall\, x_i \in B_x(x,\varepsilon(x)),\ \forall\, h \in \tilde{H}(x_i,u,\varepsilon^1(x_i,u)),$$
$$\forall\, u \in \nabla_j\Phi(x_i),\ \forall\, j \geq N(x),$$
$$p = \ell(x), \ell(x)+1, \ell(x)+2, \ldots$$

where $\varepsilon^1(x_i,u)$ is the value of $\varepsilon^1$ at which algorithm (3.18) passes from Step 6 to Step 7, for the computed $u \in \nabla_j\Phi(x_i)$.

<u>Proof</u>: By lemma (3.40), for $j \geq N(x)$ and $x_i \in B_x(x,\varepsilon(x))$, $\varepsilon^1(x_i,u) > \varepsilon(x) > 0$ $\forall\, u \in \nabla_j\Phi(x_i)$. Let $x_i \in B(x,\varepsilon(x))$ and $u \in \nabla_j\Phi(x_i)$ be arbitrary. Then since the algorithm (3.18) ensures that $\tilde{\Theta}(x_i,u,\varepsilon^1(x_i,u)) \leq -\varepsilon^1(x_i u)$, and $\varepsilon^1(x_i,u) > \varepsilon(x)$, we must have either $\langle \nabla g^q(x_i),h \rangle \leq -\varepsilon(x)$ for all $h \in \tilde{H}(x_i,u,\varepsilon^1(x_i,u))$, or else $g^q(x_i) \leq -\varepsilon(x)$, $q \in \{1,2,\ldots,m\}$. Since $B_x(x,\varepsilon(x))$ and S are both compact and the functions $g^q(\cdot)$ are continuously differentiable, the existence of an integer $\ell(x) \geq 0$ for which (3.45)

holds now follows directly. (cf. (3.24)) $\square$

3.46 Theorem: Suppose that $x \in \Omega_x$ satisfies $\Theta(x,0) < 0$. Then there exist an $\varepsilon(x) > 0$, an integer $N'(x) \geq 0$ and an integer $\ell'(x) \geq 0$ such that

3.47 $$\phi_j(x_i + (\tfrac{1}{2})^{\ell'(x)} h) - \phi_j(x_i) - (\tfrac{1}{2})^{\ell'(x)+1} \langle \nabla_j \phi(x_i), h \rangle \leq 0$$

3.48 $$g(x_i + (\tfrac{1}{2})^{\ell'(x)} h) \leq 0$$

$$\begin{aligned}
&\forall\, x_i \in B_x(x,\varepsilon(x)),\\
&\forall\, \phi_j(x_i + (\tfrac{1}{2})^{\ell'(x)} h) \in \Phi_j(x_i + (\tfrac{1}{2})^{\ell'(x)} h),\\
&\forall\, \phi_j(x_i) \in \Phi_j(x_i),\\
&\forall\, h \in \tilde{H}(x_i, u, \varepsilon^1(x_i,u)),\\
&\forall\, u \in \nabla_j \Phi(x_i),\\
&\forall\, j \geq N'(x),
\end{aligned}$$

where $\varepsilon^1(x_i,u)$ is the value of $\varepsilon^1$ at which the test $\tilde{\Theta}(x_i, \nabla_j \phi(x_i), \varepsilon^1) \leq -\varepsilon^1$ is satisfied in Step 6 of algorithm (3.18).

Proof: Suppose $\Theta(x,0) < 0$. Then by lemma (3.40), there exist an $\varepsilon(x) > 0$ and in $N(x) \geq 0$ such that (3.41) holds. Now by the mean value theorem, for any $h \in S$, $u \in \mathbb{R}^n$, and $\lambda \geq 0$,

3.49 $$\begin{aligned}
&\phi(x_i + \lambda h) - \phi(x_i) - \tfrac{1}{2}\lambda \langle u,h \rangle \\
&\quad = \lambda[\langle \nabla\phi(x_i + \tilde{\lambda} h), h \rangle - \tfrac{1}{2}\langle \nabla u, h) \rangle],
\end{aligned}$$

where $\tilde{\lambda} \in [0,\lambda]$.

Since $B_x(x,\varepsilon(x))$ and $S$ are compact, it follows from (3.11)(ii) that there exist an integer $N''(x) \geq N(x)$ and a $\lambda'(x) > 0$ such that

$$3.50 \qquad \langle \nabla\phi(x_i+\tilde{\lambda}h),h \rangle \leq \langle u,h \rangle + \frac{1}{8}\varepsilon(x) \qquad \forall\, x_i \in B_x(x,\varepsilon(x)),$$
$$\forall\, u \in \nabla_j\Phi(x_i),\ \forall\, h \in \tilde{H}(x_i,u,\varepsilon^1(x_i,u)),$$
$$\forall\, \tilde{\lambda} \in [0,\lambda'(x)],\ \forall\, j \geq N''(x).$$

Since for $u \in \nabla_j\Phi(x_i)$, $\langle u,h \rangle \leq -\varepsilon^1(x_i,u) \leq -\varepsilon(x)$ for all $h \in \tilde{H}(x_i,u,\varepsilon^1(x_i,u))$, (3.49) and (3.50) imply that

$$3.51 \qquad \phi(x_i+\lambda h) - \phi(x_i) - \frac{1}{2}\lambda\langle u,h \rangle$$
$$\leq \lambda[\langle u,h \rangle + \frac{1}{8}\varepsilon(x) - \frac{1}{2}\langle u,h \rangle]$$
$$\leq -\frac{3}{8}\lambda\varepsilon(x) \qquad \forall\, x_i \in B_x(x,\varepsilon(x)),\ \forall\, h \in \tilde{H}(x_i,u,\varepsilon^1(x_i,u)),$$
$$\forall\, u \in \nabla_j\Phi(x_i),\ \forall\, \lambda \in [0,\lambda'(x)],\ \forall\, j \geq N''(x).$$

Now, because of the manner in which $\varepsilon(x) > 0$ and $N''(x) \geq 0$ were chosen, it follows from corollary (3.44) that there exists an integer $\ell(x) \geq 0$ such that

$$3.52 \qquad g(x_i + (\tfrac{1}{2})^p h) \leqq 0 \qquad \forall\, x_i \in B_x(x_i,\varepsilon(x)),\ \forall\, h \in \tilde{H}(x_i,u,\varepsilon^1(x_i,u,)),$$
$$\forall\, u \in \nabla_j\Phi(x_i),\ \forall\, j \geq N''(x),\ \forall\, p \geq \ell(x),$$

where p is assumed to be an integer.

Let $\ell'(x)$ be the smallest integer satisfying $(\frac{1}{2})^{\ell'(x)} \leq \lambda'(x)$ and $\ell'(x) \geq \ell(x)$. Then, by (3.11)(ii), there exists an integer $N'(x) \geq N''(x)$, such that $|\phi_j(x) - \phi(x)| \leq (\frac{1}{2})^{\ell'(x)}\frac{\varepsilon(x)}{8}$ for all $x \in \Omega_x$, for all $\phi_j(x) \in \Phi_j(x)$, for all $j \geq N'(x)$, and hence, from (3.51), for $\lambda = (\frac{1}{2})^{\ell'(x)}$, we obtain

$$3.53 \qquad \phi_j(x_i + (\tfrac{1}{2})^{\ell'(x)} h)) - \phi_j(x_i) - (\tfrac{1}{2})^{\ell'(x)+1}\langle u,h \rangle \leq$$

$$\leq \left(\frac{1}{2}\right)^{\ell'(x)} \left[\frac{\varepsilon(x)}{4} - \frac{3}{8}\varepsilon(x)\right] = -\left(\frac{1}{2}\right)^{\ell'(x)} \frac{\varepsilon(x)}{8} < 0.$$

$\forall\, u \in \nabla_j \Phi(x_i)$, $\forall\, \phi_j(x_i) \in \Phi_j(x_i)$,

$\forall\, \phi_j(x_i + (\frac{1}{2})^{\ell'(x)} h) \in \Phi_j(x_i + (\frac{1}{2})^{\ell'(x)} h)$, $\forall\, x_i \in B_x(x,\varepsilon(x))$,

$\forall\, h \in \tilde{H}(x_i,u,\varepsilon^1(x_i,u))$, $\forall\, j \geq N'(x)$.

Hence (3.47) holds. Since $\ell'(x) \geq \ell(x)$, it follows from (3.52) that (3.48) also holds, and so we are done. □

3.54 Corollary: Suppose that $x \in \Omega_x$ satisfies $\Theta(x,0) < 0$. Then there exists an $\varepsilon(x) > 0$, a $\delta(x) > 0$ and an integer $N'(x) \geq$ such that

3.55 $$\phi_j(x_{i+1}) - \phi_j(x_i) \leq -\delta(x) \qquad \forall\, \phi_j(x_i) \in \Phi_j(x_i),$$
$$\forall\, \phi_j(x_{i+1}) \in \Phi_j(x_{i+1}),\ \forall\, x_i \in B_x(x,\varepsilon(x)),\ \forall\, j \geq N'(x),$$

and for all $x_{i+1} = x_i + \lambda h$, $h \in \tilde{H}(x_i,u,\varepsilon^1(x_i,u)$, which algorithm (3.18) can construct from the given $x_i$, where $\varepsilon'(x_i,u)$ is the value of $\varepsilon^1$ for which the test $\tilde{\Theta}(x_i,u,\varepsilon^1(x_i,u)) \leq -\varepsilon^1(x_i,u)$ is satisfied in Step 6.

Proof: Let $\varepsilon(x) > 0$, $N'(x) \geq N(x) \geq 0$ and $\ell'(x)$ be such that (3.40), (3.47) and (3.48) hold. Then, clearly, for all $x_i \in B_x(x,\varepsilon(x))$, for all $j \geq N'(x)$, algorithm (3.18) will construct $x_{i+1} = x_i + \lambda h$, with $h \in \tilde{H}(x_i,u,\varepsilon^1(x_i,u))$ and $\lambda \triangleq (\frac{1}{2})^{\ell_i} \geq (\frac{1}{2})^{\ell'(x)}$, $u \in \nabla_j \Phi(x_i)$. Consequently, we must have

3.56 $$\phi_j(x_{i+1}) - \phi_j(x_i) \leq \left(\frac{1}{2}\right)^{\ell_i} \langle u,h \rangle \leq -\left(\frac{1}{2}\right)^{\ell_i} \varepsilon^1(x_i,u)$$
$$\leq -\left(\frac{1}{2}\right)^{\ell'(x)} \varepsilon(x) \triangleq -\delta(x)$$

$$\forall\, x_i \in B(x,\varepsilon(x)),\ \forall\, u \in \nabla_j\Phi(x_i),\ \forall\, \phi_j(x_i) \in \Phi_j(x_i),$$
$$\forall\, \phi_j(x_{i+1}) \in \Phi_j(x_{i+1}),\ \forall\, j \geq N'(x);$$

and hence we are done. □

3.57 <u>Theorem</u>: Subprocedure I of algorithm (3.18) satisfies the assumptions (2.4)(i)-(v).

<u>Proof</u>: That the assumptions (2.4)(i), (2.4)(ii), (2.4)(iv) and (2.4)(v) are satisfied follows directly from (3.11) and the correspondence table (3.27). That assumption (2.4)(iii) is satisfied follows from corollary (3.54) and the correspondence table (3.27). □

In view of Theorems (3.57), (2.24) and the correspondence table (3.27) the following is obvious.

3.58 <u>Corollary</u>: Subprocedure I of algorithm (3.18) will either jam up at a point $x_i$, cycling indefinitely in the loop defined by Steps 1 to 14, in which case $x_i$ satisfies the optimality condition $\Theta(x_i,0) = 0$, or else it will construct an infinite sequence $\{x_i\}$ which has at least one accumulation point $x^*$ satisfying $\Theta(x^*,0) = 0$. □

We shall now establish the convergence properties of the sequence $\{z_k\}$ sieved out by Subprocedure II of algorithm (3.18) from an infinite sequence $\{x_i\}$ constructed by Subprocedure I of (3.18). For this purpose we shall need the following propositions, the proofs of which we omit, either because they are obvious or because they can easily be established by following the reasoning used for analogous results in the first part of this section.

**Definition**: For any $x \in \Omega_x$ and for any $\varepsilon > 0$, we define (i) the index set $\hat{I}(x,\varepsilon)$ by

3.59 $$\hat{I}(x,\varepsilon) = \{q \in \{1,2,\dots,m\} \mid g^q(x) > -\varepsilon\},$$

(ii) the function $\hat{\theta}$: $\mathbb{R}^n \times \mathbb{R}^+ \to \mathbb{R}^1$ by

3.60 $$\hat{\theta}(x,\varepsilon) = \min_{h \in S} \max\{\langle \nabla\phi(x),h \rangle ; \langle \nabla g^q(x),h \rangle ,q \in \hat{I}(x,\varepsilon)\},$$

(iii) the function $\overline{\theta}$: $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^+ \to \mathbb{R}^1$ by

3.61 $$\overline{\theta}(x,u,\varepsilon) = \min_{h \in S} \max\{\langle u,h \rangle ; \langle \nabla g^q(x),h \rangle , q \in \hat{I}(x,\varepsilon)\},$$

(iv) the function $\overline{H}$: $\mathbb{R}^n \times \mathbb{R}^+ \to 2^S$ by

3.62 $$\overline{H}(x,u,\varepsilon) = \{h \in S \mid \overline{\theta}(x,u,\varepsilon) = \max\{\langle u,h \rangle ; \langle \nabla g^q(x),h \rangle , q \in \hat{I}(x,\varepsilon)\}\}. \quad \square$$

3.63 **Proposition**: For every $x \in \Omega_x$ and every $\varepsilon \geq 0$,

3.64 $$\hat{I}(x,\varepsilon) \subset I_x(x,\varepsilon).$$

3.65 $$\hat{\theta}(x,\varepsilon) \leq \theta(x,\varepsilon)$$

3.66 $$\overline{\theta}(x,u,\varepsilon) \leq \tilde{\theta}(x,u,\varepsilon) \qquad \forall\ u \in \nabla_j\Phi(x),\ j = 0,1,2,\dots \quad \square$$

3.67 **Proposition**: Given any $x \in \Omega_x$ and $\varepsilon \geq 0$, there exists a $\rho(x,\varepsilon) > 0$ such that

3.68 $$\hat{I}(x',\varepsilon) \supset \hat{I}(x,\varepsilon) \quad \forall\ x' \in B_x(x,\rho(x,\varepsilon)) . \square$$

3.69 **Proposition**: Given any $x \in \Omega_x$, any $\varepsilon > 0$ and any $\gamma > 0$, there exists a $\rho(x,\varepsilon) > 0$ such that

3.70 $$\hat{\theta}(x',\varepsilon) \geq \hat{\theta}(x,\varepsilon) - \gamma \quad \forall\, x' \in B_x(x,\rho(x,\varepsilon)) \;. \square$$

3.71 Corollary: Given any $x \in \Omega_x$, any $\varepsilon > 0$ and any $\gamma > 0$, there exists a $\sigma(x,\varepsilon) > 0$ and an integer $J(x,\varepsilon) \geq 0$ such that

3.72 $$\overline{\theta}(x',u,\varepsilon) \geq \hat{\theta}(x,\varepsilon) - \gamma \quad \forall\, x' \in B_x(x,\sigma(x,\varepsilon)), \forall\, u \in \nabla_j\Phi(x'), \quad \forall\, j \geq J(x,\varepsilon) \;. \square$$

3.73 Lemma: Suppose that the sequence $\{x_i\}$ generated by subprocedure I of algorithm (3.18) is infinite. Then the sequence $\{z_k\}$ sieved out by Subprocedure II of algorithm (3.18) is also infinite.

Proof: We see that according to Steps 19 and 20 of (3.18), Subprocedure II sets $z_k = x_i$ and $k = k+1$, whenever $\tilde{\theta}(x_i,u,\varepsilon^3) \geq -\varepsilon^3$, with $u \in \nabla_{q(i)}\Phi(x_i)$, where $\varepsilon^3 = \alpha^k\varepsilon_0^3$. Consequently, to establish the lemma, it suffices to show that for any $\overline{\varepsilon}^3 > 0$ there exists a subsequence $\{x_i\}_{i \in K(\overline{\varepsilon}^3)} \subset \{x_i\}$ such that

3.74 $$\tilde{\theta}(x_i,u_i,\overline{\varepsilon}^3) \geq -\overline{\varepsilon}^3 \quad \forall\, u_i \in \nabla_{q(i)}\Phi(x_i), \ \forall\, i \in K(\overline{\varepsilon}^3).$$

We recall that according to lemma (2.19), we must have $q(i) \to \infty$ as $i \to \infty$, since $\{x_i\}$ is infinite. Next, according to corollary (3.58), there exists a subsequence $\{x_i\}_{i \in K_1}$ such that $x_i \to x^*$ as $i \to \infty$, $i \in K_1$, and $\theta(x^*,0) = 0$. Since $\hat{I}(x^*,\varepsilon) \supset I_x(x^*,0)$ for all $\varepsilon > 0$, we conclude that

3.75 $$0 \geq \hat{\theta}(x^*,\varepsilon) \geq \theta(x^*,0) = 0 \quad \forall\, \varepsilon > 0,$$

i.e. that $\hat{\theta}(x,\varepsilon) = 0$ for all $\varepsilon > 0$. Let $\overline{\varepsilon}^3 > 0$ be arbitrary. Since

$x_i \to x^*$, as $i \to \infty$ for $i \in K_1$, it follows from corollary (3.71) (and because of the fact that $I(x,\varepsilon) \supset \hat{I}(x,\varepsilon)\ \forall\ x \in \Omega_x$, $\forall\ \varepsilon > 0$) that there exists an integer $\tilde{J}(x^*,\bar{\varepsilon}^3)$ such that

3.76 $$\tilde{\theta}(x_i,u_i,\bar{\varepsilon}^3) \geq \bar{\theta}(x_i,u_i,\bar{\varepsilon}^3) \geq \hat{\theta}(x^*,\bar{\varepsilon}^3) - \bar{\varepsilon}^3 = -\bar{\varepsilon}^3$$
$$\forall\ u_i \in \nabla_{q(i)}\Phi(x_i),\ \forall\ i \geq \tilde{J}(x^*,\bar{\varepsilon}^3) \text{ and } i \in K_1.$$

Let $K(\bar{\varepsilon}^3) = \{\ i \in K_1 | i \geq \tilde{J}(x^*,\bar{\varepsilon}^3)\}$. Then we see that (3.74) holds for this index set $K(\bar{\varepsilon}^3)$, and we are done. □

3.77 Theorem: Suppose that Subprocedure I of algorithm (3.18) generates an infinite sequence $\{x_i\}$. Then every accumulation point of the sequence $\{z_k\}$ constructed by Subprocedure II of algorithm (3.18) belongs to the set $\{z \in \Omega_x | \Theta(z,0) = 0\}$.

Proof: Suppose that $z_k \to z^*$ as $k \to \infty$ for $k \in K$. Since by lemma (2.19) $q(i) \to \infty$ as $i \to \infty$, and $\{z_k\}$ is infinite, $p(k) \to \infty$ as $k \to \infty$ and $\varepsilon_k^3 \to 0$ as $k \to \infty$, where $p(k)$, $\varepsilon_k^3$ are as defined in Step 20 of (3.18). Hence, from corollary (3.36) we conclude that

3.78 $$\lim_{\substack{k\to\infty \\ k\in K}} \tilde{\Theta}(z_k,u_k,\varepsilon_k^3) \leq \Theta(z^*,0) \leq 0 \qquad \forall\ u_k \in \nabla_{p(k)}\Phi(z_k).$$

However, by construction, there exists a sequence $\{u_k\}_{k=0}^{\infty}$, ($u_k = \nabla_j\phi(x_i)$ for some j and i) such that $u_k \in \nabla_{p(k)}\Phi(z_k)$ and

3.79 $$0 \geq \tilde{\Theta}(z_k,u_k,\varepsilon_k^3) \geq -\varepsilon_k^3 \qquad k = 0,1,2,\ldots$$

and hence $\lim_{\substack{k\to\infty \\ k\in K}} \tilde{\Theta}(z_k,u_k,\varepsilon_k^3) = 0$. Substituting into (3.78), we find that

$\Theta(z^*,0) = 0$, and we are done. $\square$

We can summarize our preceding results as follows.

3.80 Theorem: Algorithm (3.18) will either jam up at a point $x_i$, cycling indefinitely in the loop defined by Steps 1 to 14, in which case $x_i$ satisfies the optimality condition $\Theta(x_i,0) = 0$, or else, it will construct an infinite sequence $\{z_k\}$ every accumulation point of which belongs to the set $\{z^* \in \Omega_x | \Theta(z^*,0) = 0\}$. $\square$

IV. Solution of Min Max Problems Under Strict Concavity Assumption.

Let $\Omega_x \subset \mathbb{R}^n$ and $\Omega_y \subset \mathbb{R}^m$ be two compact sets defined by

4.1 $$\Omega_x = \{x \in \mathbb{R}^n | g(x) \leqq 0\}$$

4.2 $$\Omega_y = \{y \in \mathbb{R}^n | \zeta(y) \leqq 0\},$$

where $g: \mathbb{R}^n \to \mathbb{R}^s$ and $\zeta: \mathbb{R}^m \to \mathbb{R}^t$ are continuously differentiable. We shall suppose that $\Omega_y$ is convex with interior. Furthermore, let $f: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^1$ be a continuously differentiable function such that $f(x,\cdot)$ is strictly concave for all $x \in V$, where $V$ is an open set containing $\Omega_x$. Finally, we define the function $\phi: V \to \mathbb{R}^1$ by

4.3 $$\phi(x) = \max_{y \in \Omega_y} f(x,y) .$$

We shall now show how algorithm (3.18) can be used to solve the problem

$$\min\{\phi(x) | x \in \Omega_x\},$$

i.e. how it can be used to find an $\hat{x} \in \Omega_x$ and a $\hat{y} \in \Omega_y$ such that

4.4 $$f(\hat{x},\hat{y}) = \min_{x\in\Omega_x} \max_{y\in\Omega_y} f(x,y).$$

To apply algorithm (3.18) to (4.3), we must specify a procedure for computing the approximations $\phi_j(x_i)$ and $\nabla_j\phi(x_i)$ which (3.18) uses. For this purpose, we shall need the following results. Let $\xi$: $V \to \Omega_y$ be defined by

4.5 $$f(x,\xi(x)) = \max_{y\in\Omega_y} f(x,y)$$

<u>Proposition</u>: Under the assumptions stated, the function $\xi(\cdot)$ is well defined and is continuous. $\square$

4.6 <u>Theorem</u>: The function $\phi(\cdot)$ is continuously differentiable on V and

4.7 $$\nabla\phi(x) = \nabla_x f(x,\xi(x)). \square$$

We shall now show that given an $x_i \in \Omega_x$ and an integer j we may compute the approximations $\phi_j(x_i)$ to $\phi(x_i)$ and u to $\nabla\phi(x_i)$ in (3.18) by any subprocedure of the form below, provided that it satisfies the assumptions of the theorem to follow.

4.8 <u>Approximations Subprocedure</u>: $\tilde{A}$: $\Omega_x \times \Omega_y \to 2^{\Omega_y}$; an $\eta_0 \in \Omega_y$, an $x \in \Omega_x$ and a positive integer j must be supplied.

<u>Step 0</u>: Set $\ell = 0$

<u>Step 1</u>: Compute a $\eta_{\ell+1} \in \tilde{A}(x,\eta_\ell)$.

<u>Step 2</u>: If $f(x,\eta_{\ell+1}) \leq f(x,\eta_\ell)$, set $\eta = \eta_\ell$ and go to Step 4; else, go to Step 3.

**Step 3:** If $\ell+1 = j$ set $\eta = \eta_{\ell+1}$ and go to Step 4; else set $\ell = \ell+1$ and go to Step 1.

**Step 4:** Set $\phi_j(x) = f(x,\eta)$, set $\nabla_j\phi(x) = \nabla_x f(x,\eta)$ and stop. $\square$

4.9 **Theorem:** For $j = 0,1,2,\ldots$, and any $x \in \Omega_x$, let $\Phi_j(x_i)$ and $\nabla_j\Phi(x_i)$ be the sets consisting of all $\phi_j(x_i)$, $\nabla_j\phi(x_i)$, respectively, which subprocedure (4.8) can construct starting from arbitrary points $\eta_0 \in \Omega_y$. Suppose that given any $(x,y) \in \Omega_x \times \Omega_y$, with $y \neq \xi(x)$, there exist an $\varepsilon(x,y) > 0$ and a $\delta(x,y) > 0$ such that

4.10 $$f(x',y'') - f(x',y') \geq \delta(x,y) \qquad \forall\, x' \in B_x(x,\varepsilon(x,y)),$$
$$\forall\, y' \in B_y(y,\varepsilon(x,y)),\ \forall\, y'' \in \tilde{A}(x',y')$$

(where $B_y(y,\varepsilon) = \{y' \in \Omega_y \mid \|y' - y\| \leq \varepsilon\}$.

Then, given any $\gamma > 0$ there exists an integer $M(\gamma) \geq 0$ such that

4.11 $$|\phi_j(x') - \phi(x')| \leq \gamma \qquad \forall\, x' \in \Omega_x,\ \forall\, \phi_j(x') \in \Phi_j(x'),\ \forall\, j \geq M(\gamma),$$

4.12 $$\|\nabla_j\phi(x') - \nabla\phi(x')\| \leq \gamma \qquad \forall\, x' \in \Omega_x,\ \forall\, \nabla_j\phi(x') \in \nabla_j\Phi(x'),\ \forall\, j \geq M(\gamma).$$

**Proof:** For any $x \in \Omega_x$, any $\eta_0 \in \Omega_y$, and any positive integer $j$, let $\Gamma(x,\eta_0,j) \subset \Omega_y$ denote the set of points $\eta$ which can be computed by (4.8), at which (4.8) stops, given these parameters. Because of the way $\phi_j(x)$ and $\nabla_j\phi(x)$ are defined in Step 4 of (4.8) and because $\Omega_x$ is compact, it is clear that to prove the theorem it is enough to show that given any $x \in \Omega_x$, any $\gamma' > 0$, there exist an $\varepsilon'(x) > 0$ and an integer $M'(\gamma')$ such that

4.13 $$\|\eta - \xi(x')\| \leq \gamma' \quad \forall\, x' \in B_x(x,\varepsilon'(x)),\ \forall\, \eta \in \Gamma(x',\eta_0,j),$$
$$\forall\, \eta_0 \in \Omega_y,\ \forall\, j \geq M'(\gamma').$$

Since (4.13) follows directly from theorem (A.2) in the Appendix, we are done. $\square$

As an approximation subprocedure which satisfies the assumptions of the preceding theorem, we can use the following adaptation of algorithm (3.7).

4.14 Approximations Subprocedure: An $\eta_0 \in \Omega_y$, an $x \in \Omega_x$, a positive integer j, and two parameters $\varepsilon_0 > 0$ and $\alpha \in (0,1)$ must be supplied.

Step 0: Set $\ell = 0$.

Step 1: Set $\varepsilon = \varepsilon_0$.

Step 2: Compute $I_y(\eta_\ell,\varepsilon) \triangleq \{p \in \{1,2,..,t\} \mid \zeta^p(\eta_\ell) \geq -\varepsilon\}$,

4.15 $$\Theta_1(x,\eta_\ell,\varepsilon) \triangleq \min_{h\in S} \max\{\langle \nabla_y f(x,\eta_\ell),h\rangle;\ \langle \nabla\zeta^p(\eta_\ell),h\rangle,\ p \in I_y(\eta_\ell,\varepsilon)\}$$

and a vector $\bar{h}(x,\eta_\ell,\varepsilon) \in S$ which solves (4.15).

Step 3: If $\Theta_1(x,\eta_\ell,\varepsilon) = 0$, compute $\Theta_1(x,\eta_\ell,0)$ and go to Step 4; else go to Step 5.

Step 4: If $\Theta_1(x,\eta_\ell,0) = 0$, set $\eta_{\ell+1} = \eta_\ell$ and go to Step 8; else, set $\varepsilon = \alpha\varepsilon$ and go to Step 2.

Step 5: If $\Theta_1(x,\eta_\ell,\varepsilon) \leq -\varepsilon$, go to Step 6; else, set $\varepsilon = \alpha\varepsilon$ and go to Step 2.

Step 6: Compute the smallest integer $r \geq 0$ such that

4.16 $$\zeta(\eta_\ell + (\tfrac{1}{2})^r \bar{h}(x,\eta_\ell,\varepsilon)) \leqq 0$$

and

$$f(x,\eta_\ell + (\tfrac{1}{2})^r \bar{h}(x,\eta_\ell,\varepsilon)) - f(x,\eta_\ell) - (\tfrac{1}{2})^{r+1} \langle \nabla_y f(x,\eta_\ell), \bar{h}(x,\eta_\ell,\varepsilon) \rangle \leq 0.$$

<u>Step 7</u>: Set $\eta_{\ell+1} = \eta_\ell + (\frac{1}{2})^r \bar{h}(x,\eta_\ell,\varepsilon)$.

<u>Step 8</u>: If $\ell+1 < j$, set $\ell = \ell+1$ and go to Step 1; else, set $\eta = \eta_{\ell+1}$, set $v = f(x,\eta_{\ell+1})$, set $u = \nabla f_x(x,\eta_{\ell+1})$, and stop. □

4.17 <u>Theorem</u>: The map $\tilde{A}(\cdot,\cdot)$ defined by Steps 1 to 7 of Suprocedure (4.14) satisfies assumptions (4.10) of theorem (4.9). □

We omit a proof of this theorem since it is rather lengthy and since it can be constructed in a rather straightforward manner either by utilizing arguments similar to the ones used to establish corollary (3.54), or by extending the arguments used to establish theorem (31) in [4].

When subprocedure (4.14) is incorporated into algorithm (3.18), we obtain the following algorithm for solving the problem defined by (4.4), under the assumptions stated at the beginning of this section.

4.18 <u>Algorithm for Min Max Problems I</u>.

<u>Step 0</u>: Compute an $x_0 \in \Omega_x$, a $y_0 \in \Omega_y$. Select parameters $\lambda_{min} \in (0,1]$, $\varepsilon_0 > 0$, $\varepsilon_0^1 > 0$, $\varepsilon_0^2 > 0$, $\varepsilon_0^3 > 0$, $\alpha \in (0,1)$, $\alpha_1 \in (0,1)$, $\alpha_2 \in (0,1)$, $\alpha_3 \in (0,1)$, and an integer $j_0 \geq 0$. Set $i = 0$, $k = 0$, $j = j_0$; set $\varepsilon^2 = \varepsilon_0^2$, $\varepsilon^3 = \varepsilon_0^3$.

<u>Step 1</u>: Set $y' = y_i$

<u>Step 2</u>: Set $\varepsilon^1 = \varepsilon_0^1$.

Step 3: Set $x = x_i$, Set $\eta_0 = y'$ and use Subprocedure (4.14) to compute a vector $\eta \in \Omega_y$, a $v \in \Phi_j(x_i)$ and a vector $u \in \nabla_j\Phi(x_i)$.

Step 4: Set $\phi_j(x_i) = v$, $\nabla_j\phi(x_i) = u$ and $y' = \eta$.

Step 5: Compute $\tilde{\Theta}(x_i,\nabla_j\phi(x_i),\varepsilon^1)$ and a vector $h \in \tilde{H}(x_i,\nabla_j\phi(x_i),\varepsilon^1)$.

Step 6: If $\tilde{\Theta}(x_i,\nabla_j\phi(x_i),\varepsilon^1) = 0$, compute $\tilde{\Theta}(x_i,\nabla_j\phi(x_i),0)$ and go to Step 7; else, go to Step 8.

Step 7: If $\tilde{\Theta}(x_i,\nabla_j\phi(x_i),0) = 0$, set $x' = x_i$, set $\phi_j(x') = \phi_j(x_i)$, set $\nabla_j\phi(x') = \nabla_j\phi(x_i)$ and go to Step 17; else, set $\varepsilon^1 = \alpha_1\varepsilon^1$ and go to Step 5.

Step 8: If $\tilde{\Theta}(x_i,\nabla_j\phi(x_i),\varepsilon^1) \leq -\varepsilon^1$, go to Step 9; else set $\varepsilon^1 = \alpha_1\varepsilon^1$ and go to Step 5.

Step 9: Set $\lambda = 1$.

Step 10: Compute $G = g(x_i+\lambda h)$.

Step 11: If $G \leq 0$, go to Step 12; else, set $\lambda = \lambda/2$ and go to Step 10.

Step 12: Set $x = x_i+\lambda h$, set $\eta_0 = y'$ and use Subprocedure (4.14) to compute a vector $\eta \in \Omega_y$ a $v \in \Phi_j(x)$ and a $u \in \nabla_j\Phi(x)$.

Step 13: Set $\phi_j(x_i+\lambda h) = v$, $\nabla_j\phi(x_i+\lambda h) = u$, and $y'' = \eta$.

Step 14: Compute $D = \phi_j(x_i+\lambda h) - \phi_j(x_i) - \frac{\lambda}{2}\langle \nabla_j\phi(x_i),h \rangle$.

Step 15: If $D > 0$ go to Step 16; else, set $x' = x_i+\lambda h$, set $\phi_j(x') = \phi_j(x_i+\lambda h)$, set $\nabla_j\phi(x') = \nabla_j\phi(x_i+\lambda h)$ and go to Step 17.

Step 16: If $\lambda \geq \lambda_{min}/2^j$, set $\lambda = \lambda/2$ and go to Step 10; else set $x' = x_i$, set $\phi_j(x') = \phi_j(x_i)$, set $\nabla_j\phi(x') = \nabla_j\phi(x_i)$ and go to Step 17.

Step 17: If $\phi_j(x') - \phi_j(x_i) \leq -\varepsilon^2$, go to Step 18; else, set $j = j+1$, set $\varepsilon^2 = \alpha_2\varepsilon^2$ and go to Step 2.

Step 18: Set $x_{i+1} = x'$, set $y_{i+1} = y''$.

Step 19: Set $i = i+1$

Step 20: Compute $\tilde{\Theta}(x',\nabla_j\phi(x'),\varepsilon^3)$.

Step 21: If $\tilde{\Theta}(x',\nabla_j\phi(x'),\varepsilon^3) \geq -\varepsilon^3$, go to Step 22; else go to Step 1.

Step 22: Set $z_k = x'$, set $\bar{y}_k = y_i$.

Step 23: Set $\varepsilon^3 = \alpha_3\varepsilon^3$, set $k = k+1$, and go to Step 1.

4.19 Theorem: Algorithm (4.18) will eigher jam up at a point $x_i$, cycling indefinitely in the loop defined by Steps 2 to 17, while constructing an infinite sequence $\{y_j''\} \subset \Omega_y$, in which case $x_i$ satisfies the optimality condition $\tilde{\Theta}(x_i, \nabla_x f(x_i,\xi(x_i)),0) = 0$ and $y_j'' \to \xi(x_i)$ as $j \to \infty$, or algorithm (4.18) will construct infinite sequences $\{z_k\}$ and $\{\bar{y}_k\}$ such that if $z_k \to z^*$ and $\bar{y}_k \to y^*$ as $k \to \infty$ for $k \in K \subset \{0,1,2,\ldots\}$, then $y^* = \xi(z^*)$ and $\tilde{\Theta}(x^*, \nabla_x f(x^*,y^*),0) = 0$. □

Proof: By theorem (3.80), if algorithm (4.18) jams up at an $x_i$, then $x_i$ solves problem (4.3). Also by theorem (3.80), if the sequence $\{z_k\}$ is infinite, then every accumulation point of that sequence solves problem (4.3). It now follows from theorem (4.9) and the fact that $j \to \infty$ that $y_j'' \to \xi(x_i)$ as $j \to \infty$, when (4.18) jams up at $x_i$, and that $y^* = \xi(z^*)$ when $\{z_k\}$ is infinite. □

## V. Solution of Min Max Problems Under Concavity Assumption

As in the preceding section, we assume that we are given three continuously differentiable functions, $f\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^1$, $g\colon \mathbb{R}^n \to \mathbb{R}^s$ and $\zeta\colon \mathbb{R}^m \to \mathbb{R}^t$. We shall suppose that $\Omega_x$ and $\Omega_y$ are defined as in (4.1) and (4.2) respectively, and are compact, that $\Omega_y$ is convex with interior, and that for every $x \in V$, $f(x,\cdot)$ is concave, where V is an open set containing $\Omega_x$. Finally, we define $\phi$: $V \to \mathbb{R}^1$ as in (4.3), and address ourselves to the problem (4.4).

Since $f(x,\cdot)$ is no longer assumed to be <u>strictly</u> concave, as in the preceding section, theorem (4.6) does not apply, and hence we must modify our approach. In the present case, the following result holds.

5.1 <u>Theorem</u> (Danskin [12]): Let $\phi(\cdot)$ be defined as in (4.3), with $f(\cdot,\cdot)$ continuously differentiable and $f(x,\cdot)$ concave for all $x \in V$, then $\phi(\cdot)$ is directionally differentiable on V and, given any $x \in V$ and any $h \in \mathbb{R}^n$, its directional derivative at x, in the direction h, is given by

5.2 $$\phi'(x,h) \triangleq \lim_{\alpha \downarrow 0} \frac{\phi(x+\alpha h)-\phi(x)}{\alpha} = \max_{y \in Y(x)} \langle \nabla_x f(x,y), h \rangle ,$$

where

5.3 $$Y(x) = \{y \in \Omega_y \mid f(x,y) = \phi(x)\}. \square$$

The following result is easy to establish.

5.4 <u>Theorem</u>: Suppose that $\hat{x} \in \Omega_x$ satisfies $\phi(\hat{x}) = \min_{x \in \Omega_x} \phi(x)$, with $\phi(\cdot)$ defined as in (4.3), and suppose that the assumptions stated in (5.1) are satisfied, then

5.5 $$\underset{\sim}{\Theta}(\hat{x}) \triangleq \min_{h \in S} \max\{\max_{y \in Y(\hat{x})} \langle \nabla_x f(\hat{x},y), h \rangle ; \langle \nabla g^q(\hat{x}), h \rangle , \quad q \in I_x(\hat{x},0)\} = 0,$$

where S, $I_x(\cdot,\cdot)$ are defined as in (3.3) and (3.4), respectively. $\square$

Because $\phi(\cdot)$ can no longer be assumed to have a gradient, algorithm (4.18) cannot be used directly for finding points $\hat{x} \in \Omega_x$ which satisfy (5.5). However, as we shall soon show, it is possible to combine parts of algorithm (4.18) with a penalty function method, to produce a new algorithm which can be used for finding points $\hat{x} \in \Omega_x$ such that $\underset{\sim}{\Theta}(\hat{x}) = 0$.

5.6 <u>Definition</u>: Let $f^p$: $\mathbb{R}^n \times \mathbb{R}^m \times [0,1] \to \mathbb{R}^1$ be defined by

$$f^p(x,y,\omega) = f(x,y) - \frac{\omega}{2} \|y\|^2,$$

and let $\phi^p$: $V \times [0,1] \to \mathbb{R}^1$ be defined by

$$\phi^p(x,\omega) = \max_{y \in \Omega_y} f^p(x,y,\omega).$$

Then for every $\omega \in (0,1]$, $f_p(x,\cdot,\omega)$ is strictly concave and hence $\nabla_x \phi^p(\cdot,\omega)$ exists and is continuous for all $\omega \in (0,1]$. Our penalty function method depends on the following assumption and result.

5.7 <u>Assumption</u>: Suppose that for $j = 0,1,2,\ldots$, $\Phi_j^p$: $\Omega_x \times (0,1] \to 2^{\mathbb{R}^1}$ and $\nabla_j \Phi^p$: $\Omega_x \times (0,1] \to 2^{\mathbb{R}^n}$ are such that given any $\gamma > 0$, there exists an integer $N(\gamma) \geq 0$ such that

5.8 $$|\phi_j^p(x,\omega) - \phi^p(x,\omega)| \leq \gamma \qquad \forall\, j \geq N(\gamma),\ \forall\, x \in \Omega_x,\ \forall\, \omega \in (0,1],\ \forall\, \phi_j^p(x,\omega) \in \Phi_j^p(x,\omega),$$

5.9 $$\|\nabla_j \phi^p(x,\omega) - \nabla_x \phi^p(x,\omega)\| \leq \gamma \qquad \forall\, j \geq N(\gamma),\ \forall\, x \in \Omega_x,\ \forall\, \omega \in (0,1],\ \forall\, \nabla_j \phi^p(x,\omega) \in \nabla_j \Phi^p(x,\omega).$$

5.10 <u>Theorem</u>: Suppose that $\{x_i\}_{i=0}^{\infty} \subset \Omega_x$, $\{\omega_i\}_{i=0}^{\infty} \subset (0,1]$ and $\{\varepsilon_i\}_{i=0}^{\infty} \subset (0,\infty)$, are such that $x_i \to x^*$, $\omega_i \downarrow 0$ and $\varepsilon_i \downarrow 0$ as $i \to \infty$.

Also suppose that for $i = 0,1,2,\ldots$, $j(i)$ is an integer such that $j(i+1) \geq j(i) + 1$, $i = 0,1,2\ldots$ . If for some $\nabla_{j(i)} \phi^p(x_i,\omega_i) \in \nabla_{j(i)}\Phi^p(x_i,\omega_i)$, $i = 0,1,2,\ldots$,

5.11 $$\tilde{\theta}(x_i,\nabla_{j(i)}\phi^p(x_i,\omega_i),\varepsilon_i) \geq -\varepsilon_i,\ i = 0,1,2,\ldots,$$

then

5.12 $$\underset{\sim}{\theta}(x^*) = 0,$$

where $\underset{\sim}{\theta}(\cdot)$ is defined as in (5.5). □

The proof of this theorem requires the following lemmas.

5.13 <u>Lemma</u>: Suppose that $\{x_i\}_{i=0}^{\infty} \subset \Omega_x$ and $\{\omega_i\}_{i=0}^{\infty} \subset (0,1]$ are such that $x_i \to x^*$ and $\omega_i \downarrow 0$ as $i \to \infty$. For $i = 0,1,2,\ldots$, let $\xi^p(x_i,\omega_i) \subset \Omega_y$ be such that

5.14 $$f^p(x_i,\xi^p(x_i,\omega_i),\omega_i) = \max_{y\in\Omega_y} f^p(x_i,y,\omega_i),$$

Then

5.15 $$\lim_{i\to\infty} d(\xi^p(x_i,\omega_i),\ Y(x^*)\} = 0,$$

where $Y(x^*)$ is defined as in (5.3) and $d(\xi,Y(x^*)) = \min_{y\in Y(x^*)} \|\xi-y\|$.

<u>Proof</u>: First we note that $Y(x^*)$ is compact, and hence that the quantities $d(\xi^p(x_i,\omega_i),\ Y(x^*))$ are well defined.

Since $f^p(\cdot,\cdot,\cdot)$ is continuous, it is uniformly continuous on the compact set $\Omega_x \times Y(x^*) \times [0,1]$. Consequently, since $x_i \to x^*$ and $\omega_i \downarrow 0$, given any $\mu > 0$, there exists an integer $N \geq 0$ such that

5.16 $$f^p(x_i,y,\omega_i) \geq f^p(x^*,y,0) - \mu \quad \forall\ y \in Y(x^*),\ \forall\ i \geq N.$$

It now follows from the definition of the $\xi^p(x_i,\omega_i)$ and (5.16) that

5.17 $$f^p(x_i,\xi^p(x_i,\omega_i),\omega_i) \geq f^p(x_i,y,\omega_i) \geq f^p(x^*,y,0) - \mu \quad \forall\, y \in Y(x^*),\ \forall\, i \geq N.$$

Since $\mu > 0$ in (5.17) is arbitrary, and since $f^p(\cdot,\cdot,\cdot)$ is continuous, (5.17) yields,

5.18 $$\liminf_{i\to\infty} f^p(x_i,\xi^p(x_i,\omega_i),\omega_i) \geq f^p(x^*,y,0) \quad \forall\, y \in Y(x^*).$$

Since (5.18) implies that any accumulation point $\xi^*$ of $\xi^p(x_o,\omega_i)$ satisfies $\xi^* \in Y(x^*)$, (5.15) must hold. □

5.19 <u>Lemma</u>: Suppose that $\{x_i\}_{i=0}^{\infty} \subset \Omega_x$ and $\{\omega_i\}_{i=0}^{\infty} \subset (0,1]$ are such that $x_i \to x^*$ and $\omega_i \downarrow 0$ as $i \to \infty$. Then, given any $\gamma > 0$, there exists an integer $N \geq 0$ such that

5.20 $$\langle \nabla_j\phi^p(x_i,\omega_i),h \rangle \leq \phi'(x^*,h) + \gamma \qquad \forall\, j \geq N, \quad \forall\, i \geq N,\ \forall h \in S,\ \forall\, \nabla_j\phi^p(x_i,\omega_i) \in \nabla_j\Phi^p(x_i,\omega_i),$$

where $\phi'(\cdot,\cdot)$ is as in (5.2).

<u>Proof</u>: First, we note that for $\omega_i > 0$,

5.21 $$\nabla_x\phi^p(x_i,\omega_i) = \nabla_x f(x_i,\xi^p(x_i,\omega_i)),$$

where $\xi^p(x_i,\omega_i)$ is defined as in (5.14). Next, since $\nabla_x f(\cdot,\cdot)$ is uniformly continuous on the compact set $\Omega_x \times \Omega_y$, given $\gamma > 0$, there exists a $\delta > 0$ such that

5.22 $$\|\nabla_x f(x',y') - \nabla_x f(x,y)\| \leq \frac{\gamma}{2\sqrt{n}} \qquad \forall\, x \in \Omega_x,\ \forall\, y \in \Omega_y,$$
$$\forall\, x' \in B_x(x,\delta),\ \forall\, y' \in B_y(y,\delta).$$

Also, because of (5.9), there exists an integer $N' \geq 0$, such that

5.23 $$\|\nabla_j \phi^P(x_i,\omega_i) - \nabla_x f(x_i,\xi^P(x_i,\omega_i))\| \leq \frac{\gamma}{2\sqrt{n}},$$
$$\forall\, \nabla_j \phi^P(x_i,\omega_i) \in \nabla_j \Phi^P(x_i,\omega_i), \forall\, j \geq N',\ \forall\, i \geq 0.$$

Now, from lemma (5.13) we conclude that there exists a sequence $\{y_i\}_{i=0}^{\infty}$ $\subset Y(x^*)$ such that

5.24 $$\lim_{i\to\infty} \|\xi^P(x_i,\omega_i) - y_i\| = 0$$

Since $\nabla_x f(\cdot,\cdot)$ is uniformly continuous an $\Omega_x \times \Omega_y$, because of (5.22) and (5.24), there exists an integer $N \geq N'$ such that

5.25 $$\|\nabla_x f(x_i,\xi^P(x_i,\omega_i)) - \nabla_x f(x^*,y_i)\| \leq \frac{\gamma}{2\sqrt{n}} \qquad \forall\, i \geq N.$$

Consequently because of (5.23) and (5.25) and since $\|h\|_\infty \leq \sqrt{n}$ for all $h \in S$,

5.26 $$|\langle \nabla_j \phi^P(x_i,\omega_i),h\rangle - \langle \nabla_x f(x^*,y_i),h\rangle| \leq$$
$$\|\nabla_j \phi^P(x_i,\omega_i) - \nabla_x f(x^*,y_i)\| \sqrt{n} \leq$$
$$\leq \{\|\nabla_j \phi^P(x_i,\omega_i) - \nabla_x f(x_i,\xi^P(x_i,\omega_i)\| + \| \nabla_x f(x_i,\xi^P(x_i,\omega_i)) -$$
$$- \nabla_x f(x^*,y_i)\|\}\sqrt{n} \leq \gamma \qquad \forall\, h \in S,\ \forall\, \nabla_j \phi^P(x_i,\omega_i) \in \nabla_j \Phi^P(x_i,\omega_i),$$
$$\forall\, i \geq N,\ \forall\, j \geq N.$$

Hence,

5.27 $$\langle \nabla_j \phi^P(x_i,\omega_i),h\rangle \leq \langle \nabla_x f(x^*,y_i),h\rangle + \gamma$$

$$\leq \max_{y \in Y(x^*)} \langle \nabla_x f(x^*,y),h \rangle + \gamma = \phi'(x^*,h) + \gamma$$

$$\forall\, h \in S,\ \forall\, i \geq N,\ \forall\, j \geq N,$$

which completes our proof. $\square$

Proof of theorem (5.10): Suppose that (5.11) holds. Then, making use of proposition (3.29), of lemma (5.19), and of the continuity of $\nabla g$, we conclude that, for the given sequence $\{\nabla_{j(i)}\phi^p(x_i,\omega_i)\}$, given any $\gamma > 0$, there exists an integer $N'' \geq 0$ such that

5.28
$$\begin{aligned} -\varepsilon_i &\leq \min_{h \in S} \max\{\langle \nabla_{j(i)}\phi^p(x_i,\omega_i),h \rangle ; \langle \nabla g^q(x_i),h \rangle,\ q \in I_x(x_i,\varepsilon_i)\} \\ &\leq \min_{h \in S} \max\{\langle \nabla_{j(i)}\phi^p(x_i,\omega_i),h; \langle \nabla g^q(x_i),h \rangle,\ q \in I_x(x^*,0)\} \\ &\leq \min_{h \in S} \max\{\phi'(x^*,h); \langle \nabla g^q(x_i),h \rangle,\ q \in I_x(x^*,0)\} + \gamma/2 \\ &\leq \min_{h \in S} \max\{\phi'(x^*,h), \langle \nabla g^q(x^*), h \rangle,\ q \in I_x(x^*,0)\} + \gamma \\ &= \underset{\sim}{\Theta}(x^*) + \gamma \qquad \forall\, i \geq N''. \end{aligned}$$

Consequently,

5.29
$$\lim_{i \to \infty} -\varepsilon_i = 0 \leq \underset{\sim}{\Theta}(x^*) \leq 0,$$

and hence we are done. $\square$

The following algorithm constructs sequences $\{x_i\}$, $\{\omega_i\}$, $\{\nabla_j\phi^p(x_i,\omega_i)\}$ and $\{\varepsilon_i\}$ which contain subsequences satisfying the assumptions of theorem (5.10).

5.30 Algorithm for Min Max Problems II.

Step 0: Compute an $x_0 \in \Omega_x$ and a $y_0 \in \Omega_y$. Select parameters $\lambda_{min} \in (0,1]$,

$\varepsilon_0 > 0$, $\varepsilon_0^1 > 0$, $\varepsilon_0^2 > 0$, $\varepsilon_0^3 > 0$, $\omega_0 > 0$, $\alpha \in (0,1)$, $\alpha_1 \in (0,1)$, $\alpha_2 \in (0,1)$, $\alpha_3 \in (0,1)$, $\beta \in (0,1)$ and integers $j_0 \geq 0$, $k_0 \geq 0$, and $e \geq 0$.

Step 1: Set $\varepsilon^3 = \varepsilon_0^3$, $\omega = \omega_0$ and $k = k_0$.

Step 2: Set $i = 0$ and set $q(i) = k_0$.

Step 3: Set $j = j_0$.

Step 4: Set $y' = y_i$.

Step 5: Set $\varepsilon^1 = \varepsilon_0^1$.

Step 6: Replace $f(\cdot,\cdot)$ by $f^P(\cdot,\cdot,\omega)$, set $x = x_i$, set $\eta_0 = y'$ and use (4.14) to compute a vector $\eta \in \Omega_y$, a $v \in \phi_j^P(x_i,\omega)$, and a $u \in \nabla_j \Phi^P(x_i,\omega)$.

Step 7: Set $\phi_j^P(x_i,\omega) = v$, $\nabla_j \phi^P(x_i,\omega) = u$ and $y' = \eta$.

Step 8: Compute $\tilde{\Theta}(x_i,\nabla_j \phi^P(x_i,\omega),\varepsilon^1)$ and an $h \in \tilde{H}(x_i,\nabla_j \phi^P(x_i,\omega), \varepsilon^1)$

Step 9: If $\tilde{\Theta}(x_i,\nabla_j \phi^P(x_i,\omega),\varepsilon^1) = 0$, compute $\tilde{\Theta}(x_i,\nabla_j \phi^P(x_i,\omega),0)$ and go to Step 10; else go to Step 11.

Step 10: If $\tilde{\Theta}(x_i,\nabla_j \phi^P(x_i,\omega),0) = 0$, set $x' = x_i$, set $\phi_j^P(x',\omega) = \phi_j^P(x_i,\omega)$, set $\nabla_j \phi^P(x',\omega) = \nabla_j \phi^P(x_i,\omega)$, and go to Step 20; else set $\varepsilon^1 = \alpha_1 \varepsilon^1$ and go to Step 8.

Step 11: If $\tilde{\Theta}(x_i,\nabla_j \phi^P(x_i,\omega), \varepsilon^1) \leq -\varepsilon^1$, go to Step 12; else set $\varepsilon^1 = \alpha_1 \varepsilon^1$ and go to Step 8.

Step 12: Set $\lambda = 1$.

Step 13: Compute $G = g(x_i+\lambda h)$.

Step 14: If $G \leq 0$, go to Step 15; else set $\lambda = \lambda/2$ and go to Step 13.

Step 15: Replace $f(\cdot,\cdot)$ by $f^P(\cdot,\cdot,\omega)$, set $x = x_i+\lambda h$, set $\eta_0 = y'$ and use (4.14) to compute a vector $\eta \in \Omega_y$ a $v \in \phi_j^P(x_i,\omega)$, and a $u \in \nabla_j \Phi^P(x_i,\omega)$.

Step 16: Set $\phi_j^p(x_i+\lambda h,\omega) = v$, $\nabla_j\phi^p(x_i+\lambda h,\omega) = u$, and $y'' = \eta$.

Step 17: Compute $D = \phi_j^p(x_i+\lambda h,\omega) - \phi_j^p(x_i,\omega) - \frac{1}{2}\lambda \langle \nabla_j\phi^p(x_i,\omega),h \rangle$.

Step 18: If $D > 0$, go to Step 19; else set $x' = x_i+\lambda h$, set $\phi_j^p(x',\omega) = \phi_j^p(x_i+\lambda h,\omega)$, set $\nabla_j\phi^p(x',\omega) = \nabla_j\phi^p(x_i+\lambda h,\omega)$, and go to Step 20.

Step 19: If $\lambda \geq \lambda_{min}/2^j$, set $\lambda = \lambda/2$ and go to Step 13; else set $x' = x_i$, set $\phi_j^p(x',\omega) = \phi_j^p(x_i,\omega)$, set $\nabla_j\phi^p(x',\omega) = \nabla_j\phi^p(x_i,\omega)$, and go to Step 20.

Step 20: If $\phi_j^p(x',\omega) - \phi_j^p(x_i,\omega) \leq \varepsilon^2$, go to Step 22; else set $j = j+1$, set $\varepsilon^2 = \alpha_2\varepsilon^2$, and go to Step 21.

Step 21: If $j - q(i) \geq e$, go to Step 22; else go to Step 5.

Comment: The test in Step 21 is needed because algorithm (4.18) can jam up at a point $x_i$ such that $\tilde{\Theta}(x_i,\nabla_x f(x_i,\xi^p(x_i,\omega)),0) = 0$.

Step 22: Set $x_{i+1} = x'$, set $y_{i+1} = y''$, and set $q(i+1) = j$.

Step 23: Set $i = i+1$.

Step 24: If $j \geq k$, go to Step 25; else go to Step 4.

Step 25: Set $\nabla_j\phi^p(x_i,\omega) = \nabla_j\phi^p(x',\omega)$.

Step 26: Compute $\tilde{\Theta}(x_i,\nabla_j\phi^p(x_i,\omega),\varepsilon^3)$.

Step 27: If $\tilde{\Theta}(x_i,\nabla_j\phi^p(x_i,\omega),\varepsilon^3) \geq -\varepsilon^3$, go to Step 28; else go to Step 4.

Step 28: Set $z_k = x_i$ and set $\overline{y}_k = y_i$.

Step 29: Set $\omega = \beta\omega$, set $\varepsilon^3 = \alpha_3\varepsilon^3$, set $k = k+1$ and go to Step 3. □

Note: Instead of the test in Step 21, it is computationally more efficient to test whether $\varepsilon^1 \leq \varepsilon^3$ and $j \geq k$ whenever $\tilde{\Theta}(x_i,\nabla_j\phi^p(x_i,\omega),\varepsilon^1) = 0$. Our reason for not including this test in our algorithm is that it introduces additional loops which make the resultant algorithm considerably harder to understand than algorithm (5.30). $\square$

Proceding as in the proof of theorem (A.3), it is possible to show that Suprocedure (4.14), as used in the above algorithm constructs elements in sets $\Phi_j^p(x,\omega)$ and $\nabla_j\Phi^p(x,\omega)$ satisfying the hypotheses (5.8) and (5.9). Because of this and because of theorem (5.10), algorithm (5.30) has the following property.

5.32 <u>Theorem</u>: The sequences $\{z_k\}$ and $\{y_k\}$ constructed by algorithm (5.30) are infinite. If for $k \in K \subset \{0,1,2,\ldots\}$, $z_k \to z^*$ and $y_k \to y^*$ as $k \to \infty$, then $\underset{\sim}{\theta}(z^*) = 0$ and $f(z^*,y^*) = \max_{y\in\Omega_y} f(z^*,y)$. $\square$

Since proving theorem (5.32) would amount to substantially retracing the steps followed in proving theorem (4.19), we shall content ourselves with a brief outline of the arguments to be used in proving (5.32). First, it is necessary (and rather easy) to establish a result similar to theorem (4.17). That is, assuming that f is replaced by $f^p$ in Subprocedure (4.14) to define a map $\tilde{A}^p : \Omega_x \times \Omega_y \times [0,1] \to 2^{\Omega_y}$, it is necessary to show that if $(x,y,\omega) \in \Omega_x \times \Omega_y \times [0,1]$ is such that y does not solve $\max\{f^p(x,y',\omega) : y' \in \Omega_y\}$, then there exist an $\varepsilon(x,y,\omega) > 0$ and a $\delta(x,y,\omega) > 0$ such that

5.33
$$\begin{aligned} f^p(x',y'',\omega') - f^p(x',y',\omega') &\geq \delta(x,y,\omega) \\ &\forall\, x' \in B_x(x,\varepsilon(x,y,\omega)) \\ &\forall\, y' \in B_y(x,\varepsilon(x,y,\omega)) \\ &\forall\, \omega' \in B(\omega,\varepsilon(x,y,\omega)) \cap [0,1], \\ &\forall\, y'' \in \tilde{A}^p(x',y',\omega'). \end{aligned}$$

This fact can then be used to establish a result analogous to (A.3), and thus it can easily be shown that Subprocedure (4.14), as used in algorithm (5.30), yields approximations satisfying (5.8) and (5.9). It then follows from theorem (4.14) that the sequences $\{z_k\}$ and $\{\bar{y}_k\}$ must be infinite. Also, theorem (5.10) applies to yield that $\underset{\sim}{\Theta}(z^*) = 0$. Finally, we can deduce that $f(z^*,\bar{y}^*) = \max_{y\in\Omega y} f(z^*,y)$ from (5.8).

## Appendix: A Property of Subprocedure 4.8.

If we ignore the instruction in Step 4 of (4.8) "Set $\nabla_j\phi(x_i) = \nabla_x f(x_i,\eta_{\ell+1})$", we see that (4.8) is simply an algorithm for solving the problem

A.1 $$\max\{f(x,y)\,|\,y \in \Omega_y\}$$

for a given value of x. As we shall now see, its most crucial property holds even when $f(x,\cdot)$ is <u>not strictly concave</u>, for all x of interest. We shall only assume that $f(x,\cdot)$ is concave.

Let $\Omega_x \subset \mathbb{R}^n$, $\Omega_y \subset \mathbb{R}^m$ be two compact sets defined as in (4.1), (4.2) respectively, with g: $\mathbb{R}^n \to \mathbb{R}^3$, $\zeta$: $\mathbb{R}^m \to \mathbb{R}^t$ continuously differentiable, and let f: $\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^1$ be continuously differentiable. Suppose that $\Omega_y$ is convex with interior and that $f(x,\cdot)$ is concave for all $x \in V$ where V is an open set containing $\Omega_x$.

For any $x \in \Omega_x$, any $\eta_0 \in \Omega_y$, and any integer $j \geq 0$, let $\Gamma(x,y_0,j) \subset \Omega_y$ denote the set of points $\eta$ which can be computed by Subprocedure (4.8), at which Subprocedure (4.8) stops, (i.e. given $x,\eta_0,j$, $\Gamma(x,\eta_0,j)$ is the set of all points $\eta$ which could conceivably be used in setting $\phi_j(x_i) = f(x,\eta)$).

For any $x \in \Omega_y$, let Y: $\Omega_x \to 2^{\Omega_y}$ be defined by

A.2 $$Y(x) = \{y \in \Omega_y\,|\,f(x,y) = \max_{\eta\in\Omega_y} f(x,\eta)\}.$$

A.3 <u>Theorem</u>: Suppose that given any $(x,y) \in \Omega_x \times \Omega_y$, with $y \notin Y(x)$, there exists an $\varepsilon(x,y) > 0$ and a $\delta(x,y) > 0$ such that

A.4 $$f(x',y'') - f(x',y') \geq \delta(x,y) \qquad \forall\, x' \in B_x(x,\varepsilon(x,y)),$$
$$\forall\, y' \in B_y(y,\varepsilon(x,y)),\ \forall\, y'' \in \tilde{A}(x',y').$$

Then, for any $x \in \Omega_x$ and for any $\gamma' > 0$, there exists an $\varepsilon'(x) > 0$ and an integer $M'(x,\gamma') \geq 0$ such that

A.5 $$\|\eta - \eta'\| \leq \gamma' \quad \forall\, \eta \in Y(x_i),\ \forall\, x_i \in B_x(x,\varepsilon'(x)),$$
$$\forall\, \eta' \in \Gamma(x',\eta_0,j),\ \forall\, \eta_0 \in \Omega_y,\ \forall\, j \geq M'(x,\gamma').$$

Proof: Let $L\colon\ \Omega_x \times \mathbb{R}^t \to 2^{\Omega_y}$ be defined by

A.6 $$L(x,\alpha) = \{y \in \Omega_y \,|\, f(x,y) \geq \phi(x)-\alpha\},$$

where, for $x \in \Omega_x$,

A.7 $$\phi(x) \triangleq \max_{y\in\Omega_y} f(x,y).$$

We begin by showing that, given any $x \in \Omega_x$ and any $\gamma' > 0$, there exists an $\alpha(x) > 0$ such that

A.8 $$L(x',\alpha(x)) \subset \bigcup_{y\in Y(x)} B_y^0(y,\gamma'), \qquad \forall\, x' \in B_x(x,\alpha(x)),$$

where $B_y^0(y,\gamma') = \{y' \in \Omega_y \,|\, \|y'-y\| < \gamma'\}$. Suppose, therefore, that for some $x \in \Omega_x$ there is no $\alpha(x) > 0$ for which (A.8) holds. Then we can construct sequences $\{x_i\} \subset \Omega_x$, $\{y_i\} \subset \Omega_y$ and $\{\alpha_i\} \subset \mathbb{R}^1$ such that $x_i \to x$, $y_i \to y^*$ and $\alpha_i \downarrow 0$ as $i \to \infty$, satisfying

A.9 $$\min_{y\in Y(x)} \|y_i-y\| \geq \gamma',\ i = 0,1,2,\ldots$$

and

A.10 $$f(x_i,y_i) \geq \phi(x_i)-\alpha_i,\ i = 0,1,2,\ldots$$

(Since $Y(x)$ is compact, (A.9) is well defined). Consequently, (A.9) implies that $y^* \notin Y(x)$. However, since $x_i \to x$, $y_i \to y^*$, $\alpha_i \to 0$ and $\phi(\cdot)$ and $f(\cdot,\cdot)$ are continuous, (A.10) implies that $f(x,y^*) = \phi(x)$, i.e. that $y^* \in Y(x)$, contradicting (A.9). Consequently, (A.7) must be true. Next, we show that given any $x \in \Omega_x$ and any $\alpha > 0$, there exists $\rho(\alpha) > 0$ and a $\sigma(\alpha) > 0$ such that

A.11 $$\bigcup_{y \in Y(x)} B_y^0(y,\rho(\alpha)) \subset L(x',\alpha), \qquad \forall\, x' \in B_x(x,\sigma(\alpha)).$$

Thus, we assume that $x \in \Omega_x$ and $\alpha > 0$ are given. Since $f(\cdot,\cdot)$ is uniformly continuous on $\Omega_x \times \Omega_y$, there exists a $\rho(\alpha) \in (0,\gamma']$ such that

A.12 $$|f(x,y') - f(x,y)| < \frac{\alpha}{2} \qquad \forall\, y' \in B_y^0(y,\rho(\alpha)),\ \forall\, y \in \Omega_y ,$$

and hence

A.13 $$f(x,y') > \phi(x) - \frac{\alpha}{2} \qquad \forall\, y' \in \bigcup_{y \in Y(x)} B_y^0(y,\rho(\alpha)),\ \forall\, y \in \Omega_y .$$

Since $\overline{\bigcup_{y \in Y(x)} B_y^0(y,\rho(\alpha))}$ and $\Omega_x$ are compact, the function $\beta\colon \Omega_x \to \mathbb{R}^1$, defined by

A.14 $$\beta(x') = \min\{f(x',y) \mid y \in \overline{\bigcup_{y \in Y(x)} B_y^0(y,\rho(\alpha))}\}$$

is uniformly continuous on $\Omega_x$. Hence, and because from (A.13) $\beta(x) \geq \phi(x) - \frac{\alpha}{2}$, there exists a $\sigma_1(\alpha) > 0$ such that

A.15 $$\beta(x') \geq \phi(x) - \frac{2}{3}\alpha \qquad \forall\, x' \in B_x(x,\sigma(\alpha)).$$

Since $\phi(\cdot)$ is continuous, there exists a $\sigma(\alpha) \in (0,\sigma_1(\alpha)]$ such that

A.16 $$\phi(x) \geq \phi(x') - \frac{1}{3}\alpha \qquad \forall\, x' \in B_x(x,\sigma_1(\alpha)).$$

Consequently,

A.17 $$f(x',y') \geq \beta(x') \geq \phi(x') - \alpha \qquad \forall\, x' \in B_x(x,\sigma(x)),$$
$$\forall\, y' \in \bigcup_{y\in Y(x)} B_y^0(y,\rho(\alpha))$$

But (A.17) implies (A.11), and hence (A.11) must be true.

Now, for a given x, let $\alpha(x)$ be such that (A.8) holds and let $\bar{\rho} \triangleq \rho(\alpha(x)) \in (0,\gamma']$, $\bar{\sigma} \triangleq \sigma(\alpha(x))$ be such that (A.11) holds for $\alpha = \alpha(x)$.

Let $\bar{B} = \overline{(\text{compl} \bigcup_{y\in Y(x)} B_y^0(y,\bar{\rho}))}$, then, since $\bar{B}$ is compact, it follows from (4.10) that there exist an $\varepsilon'(x) \in (0,\bar{\sigma}]$ and a $\delta(x) > 0$ such that

A.18 $$f(x',\eta_{\ell+1}) - f(x',\eta_\ell) \geq \delta(x) \quad \forall\, x' \in B_x(x,\varepsilon'(x)),$$
$$\forall\, \eta_\ell \in \bar{B}, \quad \forall\, \eta_{\ell+1} \in \tilde{A}(x',y').$$

Let $m = \min_{x\in\Omega_x} \min_{y\in\Omega_y} f(x,y)$ and let $M = \max_{x\in\Omega_y} \max_{y\in\Omega_y} f(x,y)$. Let $x' \in B(x,\varepsilon'(x))$ and $\eta_0 \in \Omega_y$ be arbitrary. If $\eta_0 \in \bigcup_{y\in Y(x)} B_y^0(y,\bar{\rho})$, then, since by construction in (4.8), $f(x',\eta_{\ell+1}) \geq f(x',\eta_\ell)$, for $\ell = 0,1,2,\ldots$, (where x is replaced by x'), we see, making use of (A.11), that

A.19 $$\eta_\ell \in L(x',\alpha(x)) \subset \bigcup_{y\in Y(x)} B^0(y,\gamma') \qquad \forall\, \ell \geq 0 \quad .$$

Hence suppose that $\eta_0 \in \bar{B}$. Then, if we let $M'(x,\gamma')$ be the smallest integer such that $M'(x,\gamma)\delta(x) \geq (M-m)$, we find that because of (A.18), (A.19) must hold for all $\ell \geq M'(x,\gamma)$. Consequently, with $\varepsilon'(x) > 0$ and $M'(x,\gamma) \geq 0$ defined as above, we see that (A.5) holds. $\square$

## Conclusion

We have shown in this paper that, when well known methods of feasible directions cannot practically be applied to certain problems because of the great cost of precise function and derivative approximations, it is possible to insert into such methods stable and efficient approximation procedures which do not disrupt the convergence properties of the original algorithm. We have also examined the exact nature of the calculations to be performed when such an algorithm with approximations is to be applied to a constrained min max problem. The approximation procedures described in this paper are quite general and it may be hoped that they will find their way into many algorithms when frequent precise function and derivative calculations are not practically feasible.

References

[1] G. Zontendijk, Methods of feasible directions, Elsevier Publishing Co., 1960.

[2] S. I. Zoukhovitzkii, R. A. Polyak, and M. E. Primak, An algorithm for solution of convex programming problems, DAN USSR, vol. 153, no. 5, 1963, pp. 991-1000.

[3] D. M. Topkis and A. Veinott, On the convergence of some feasible directions algoritms for nonlinear programming, J. SIAM Control, vol. 5, no. 2, 1967, pp. 268-279.

[4] E. Polak, Of the convergence of optimization algorithms, Revue Francaise d'Informatique et de Recherche Operationelle, no. 16R-1, 1969, pp. 17-34.

[5] M. Frank and P. Wolfe, An algorithm for quadratic programming, Naval Research Logistics Quarterly, vol. 3, 1956, pp. 95-110.

[6] J. B. Rosen, The gradient projection method for nonlinear programming, J. SIAM, vol. 8, no. 1, 1960, pp. 181-217 and J. SIAM, vol. 9, no. 4, 1961, pp. 514-532.

[7] P. Huard, The method of centers, in "Nonlinear Programming: A Course," North-Holland Publishing Co., 1965.

[8] E. Polak, On the implementation of conceptual algorithms, Proceedings of the Nonlinear Programming Symposium, U. of Wisconsin, Madison, Wisconsin, May 4-6, 1970.

[9] G. Meyer and E. Polak, Abstract models for the synthesis of optimization algorithms, U. of California, Berkeley, Electronics Research Laboratory Memorandum No. ERL-268, October, 1969.

[10] E. Polak, Computational methods in optimization: a unified approach, Academic Press, 1971.

[11] L. Armijo, Minimization of functions having Lipschitz continuous first partial derviatives, Pacific J. Math., vol. 16, no. 1, 1966, pp. 1-3.

[12] J. Danskin, The theory of max-min, Springer-Verlag, 1967.