A DERIVATIVE-FREE ALGORITHM FOR A CLASS OF INFINITELY CONSTRAINED PROBLEMS

by

R. Trahan and E. Polak

Memorandum No. UCB/ERL M78/75

8 September 1977

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

A DERIVATIVE-FREE ALGORITHM FOR A CLASS OF INFINITELY CONSTRAINED PROBLEMS*

by

R. Trahan 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 derivative-free algorithm, based on local variations and phase I-phase II methods of feasible directions for solving optimization problems with distributed constraints. Such constraints normally arise in the transcription of engineering design specifications.

*Research sponsored by National Science Foundation Grant ENG73-08214-A01 and National Science Foundation (RANN) Grant ENV76-04264.

I. Introduction

It has been pointed out in [1] that an important class of engineering design problems can be transcribed into optimization problems of the form

$$\min\{f^0(z) \mid g^i(z) \le 0,\ j=1,2,\ldots,p;\ f^j(z) \le 0,\ j=1,2,\ldots,m\} \tag{1}$$

where $f^0\colon \mathbb{R}^n \to \mathbb{R}'$ and $g^j\colon \mathbb{R}^n \to \mathbb{R}'$, $j = 1,2,\ldots,p$, are continuously differentiable functions, while the $f^j(\cdot)$ are of the form

$$f^j(z) = \max_{c_0 \in \Omega} \phi^j(z,\omega) \tag{2}$$

where $\phi^j\colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$, with $\phi^j(\cdot,\cdot)$ and $\nabla_z\phi^i(\cdot,\cdot)$, $j = 1,2,\ldots,m$, continuous. The set Ω is assumed to be an interval. The algorithm in [1] and the more refined version of this method given in [15] both require the explicit computation of the gradients $\nabla_z\phi(z,\omega)$, at certain values of ω. Now, quite commonly, to compute $\nabla_z\phi(z,\omega)$, one has to integrate a linearized differential equation, in the usual manner of computing sensitivities. This operation generally turns out to be very expensive and, in the case where the dynamics of a system exhibit hysteresis, quite difficult to program (see [14] for a description of these calculations in steel framed structures). Hence, there is great incentive to construct an algorithm for solving problems of the form (1) and (2), which uses function values only. In the case of unconstrained optimization there are well known direct search methods such as the method of local variations [5,8], the method of Hooke and Jeeves [9], and the method of Rosenbrock [10]. These methods use only local searches based only on function values of the cost, without regard to any derivative information. This is in contrast to descent methods in which gradients, or their approximations, are used. The steepest descent method is one example of a descent method. The above mentioned unconstrained direct search methods cannot be applied to constrained problems because they can jam up at nonstationary points. In Fig. 1, it can be seen that the method of local variations jams up at the nonstationary points x_0 because a feasible step along any coordinate direction is impossible.

In this paper we use a modified local variations method in which we generate spacer steps* using a derivative-free modification of the algorithm in [15]. In this way we ensure that the new algorithm cannot jam up and also that it cannot converge to nonstationary points. The modification of the algorithm in [15] was obtained by following the theory of adaptive approximations given in [7,5].

To be more specific, we use a modified local variations method on (1) until the step length falls below a certain value. Then, using the function values obtained in the last iteration of the local search, the gradients of the constraints and the cost are approximated by difference formulas and a feasible search direction is computed using the same method as in [15]. A step is taken in this direction according to the step length rules in [15]. Using the new point, a local variations type search is again initiated and carried out until the step length once more decreases beyond a certain point, and the whole process is repeated. Hence, the feasible directions steps become spacer steps between local variations steps.

Given a design problem, of the form (1)-(2), in which gradient calculations are impossible or are prohibitively expensive, the new algorithm presented here gives the designer a greater flexibility in determining an

---

*A spacer step is one which is repeated infinitely often and ensures convergence. See section 7.9 in [11].

acceptable design. For example, in earthquake resistant building design problems, such as the one in [3], the designer may be satisfied to simply gain some improvement over his initial design. In problems such as these, the gradient calculations may be very difficult to obtain and the designer may wish to use a simple local variations type method. In this case, various parameters may be set to allow our new algorithm to spend much more time in the local variations steps than in the more complicated feasible directions steps. On the other hand, the algorithm can be forced to take many more feasible directions steps. In practice, it will be desirable to use an algorithm interactively so as to use the local variations option to a greater extent in early iterations and then switch to more feasible directions iterations as greater precision is required towards the end.

II. Definitions and Assumptions

We shall make use of the following notation. Given $z \in \mathbb{R}^n$, we define $\psi : \mathbb{R}^n \to \mathbb{R}$ by

$$\psi(z) \triangleq \max\{g^j(x), j \in \underline{p} \; ; \; f^j(z), \; j \in \underline{m}\} \tag{3}$$

where $\underline{p} \triangleq \{1,2,\ldots,p\}$ and $\underline{m} \triangleq \{1,2,\ldots,m\}$. Then we define $\psi_o : \mathbb{R}^n \to \mathbb{R}$ by

$$\psi_o(z) \triangleq \max\{0,\psi(z)\} \tag{4}$$

For any $\varepsilon \geq 0$, we define the ε-active constraint sets by

$$J_\varepsilon^f(z) \triangleq \{j \in \underline{m} | f^j(z) - \psi_o(z) \geq -\varepsilon\} \tag{5}$$

$$J_\varepsilon^g(z) \triangleq \{j \in \underline{p} | g^j(z) - \psi_o(z) \geq -\varepsilon\} \tag{6}$$

$$\Omega_\varepsilon^j(z) \triangleq \{\omega \in \Omega | \phi^j(z,\omega) - \psi_o(z) \geq -\varepsilon\}, \; j \in \underline{m} \tag{7}$$

By following the same procedure as in Chapter 2, we define an "approximation" of $\Omega_\varepsilon^j(z)$ by

$$\tilde{\Omega}_\varepsilon^j(z) \triangleq \{\omega \in \Omega_\varepsilon^j(z) \mid \omega \text{ is a left local maximizer of } \phi^j(z,\cdot)\} \quad (8)$$

where a point $\bar{\omega} \in \Omega$ is a left local maximizer of $\phi^j(z,\cdot)$ if there exists a $\mu > 0$ such that

$$\phi^j(z,\omega) < \phi^j(z,\bar{\omega}) \qquad \forall \omega \in (\bar{\omega}-\mu,\bar{\omega}) \cap \Omega \quad (9)$$

$$\phi^j(z,\bar{\omega}) \geq \phi^j(z,\omega) \qquad \forall \omega \in (\bar{\omega},\bar{\omega}+\mu) \cap \Omega \quad (10)$$

The following hypotheses are assumed to be true.

Assumption 1. $f^o(\cdot)$ and $g^j(\cdot)$, $j \in \underline{p}$ are continuously differentiable; $\phi^j(\cdot,\cdot)$ and $\nabla_z\phi^j(\cdot,\cdot)$, $j \in \underline{m}$ are continuous. ¤

Assumption 2. For all $z \in \mathbb{R}^n$, for all $j \in \underline{m}$, $\Omega_o^j(z)$ is a finite set. ¤

Assumption 3. For all $z \in \mathbb{R}^n$, $\varepsilon > 0$, and $j \in J_\varepsilon^f(z)$, $\tilde{\Omega}_\varepsilon^j(z)$ is a finite set. ¤

Assumption 4. For all $z \in \mathbb{R}^n$, $\{\nabla_z\phi^j(z,\omega),\ \omega \in \Omega_o^j(z),\ j \in J_o^f(z);\ \nabla g^j(z),\ j \in J_o^g(z)\}$ is a set of positive linearly independent vectors.⁺ ¤

Because we shall make use of a feasible directions type algorithm, we must have some form of gradient information for the cost and each constraint. Unlike the algorithm in Chapter 2, we do not require exact gradient calculations but only an approximation to each gradient. For any $z \in \mathbb{R}^n$, $\tau \geq 0$, let $\bar{\nabla}f^o(z;\tau)$, $\bar{\nabla}g^j(z;\tau)$, $j \in \underline{p}$, $\bar{\nabla}\phi^j(z,\omega;\tau)$, $\omega \in \Omega$, $j \in \underline{m}$ be approximations to the gradients of the cost and constraints

⁺We say a set of vectors $\{\eta_j\}_{j=1}^n$ is positive linearly independent if the zero-vector is not contained in the convex hull of $\{\eta_j\}_{j=1}^n$. This assumption is related to the Kuhn-Tucker constraint qualification [12].

respectively, in which τ indicates the precision of the approximation. Regarding this precision, we assume the following hypotheses hold.

Assumption 5. Given any compact set $C \subset \mathbb{R}^n$, and any $\mu > 0$, there exists a $\bar{\tau} > 0$ such that for all $z \in C$, and for all $\tau \in [0,\bar{\tau}]$,

$$\|\bar{\nabla}\phi^j(z,\omega;\tau) - \nabla_z\phi^j(z,\omega)\| \le \mu \qquad \forall\omega \in \Omega,\ j \in \underline{m} \tag{11}$$

$$\|\bar{\nabla}g^j(z;\tau) - \nabla g^j(z)\| \le \mu \qquad j \in \underline{p} \tag{12}$$

$$\|\bar{\nabla}f^o(z;\tau) - \nabla f^o(z)\| \le \mu \tag{13}$$

¤

Assumption 6. For all $\tau \ge 0$, $\bar{\nabla}f^o(\cdot;\tau)$, $\bar{\nabla}g^j(\cdot;\tau)$, $j \in \underline{p}$ and $\bar{\nabla}\phi^j(\cdot,\cdot;\tau)$, $j \in \underline{m}$, are continuous functions. ¤

An example of an approximation which satisfies the above assumptions is the simple difference formula, which, for $\bar{\nabla}f^o(z;\tau)$, would be

$$\bar{\nabla}f^o(z;\tau) = \begin{pmatrix} (f^o(z+\tau_1 e_1)-f^o(z))/\tau_1 \\ (f^o(z+\tau_2 e_2)-f^o(z))/\tau_2 \\ \vdots \\ (f^o(z+\tau_n e_n)-f^o(z))/\tau_n \end{pmatrix} \tag{14}$$

where e_j is the jth column of the nxn identity matrix and $\tau_j \in (0,\tau]$, $j = 1,2,\ldots,n$, $\tau > 0$ are positive constants. For $\tau = 0$, we simply define $\bar{\nabla}f^o(z;0) \triangleq \nabla f^o(z)$.

For notational convenience we define

$$\bar{D}f^j_\varepsilon(z,h;\tau) \triangleq \begin{cases} \max\limits_{\omega\in\tilde{\Omega}^j_\varepsilon(z)} \langle \bar{\nabla}\phi^j(z,\omega;\tau),h\rangle & \text{if } j \in J^f_\varepsilon(z) \\ -\infty & \text{otherwise} \end{cases} \tag{15}$$

$$\bar{D}g^j_\varepsilon(z,h;\tau) \triangleq \begin{cases} \langle \bar{\nabla} g^j(z;\tau),h\rangle & \text{if } j \in J^g_\varepsilon(z) \\ -\infty & \text{otherwise} \end{cases} \tag{16}$$

where, $z \in \mathbb{R}^n$, $h \in \mathbb{R}^n$, $\tau \geq 0$, and $\varepsilon \geq 0$.

In the descent part of the new algorithm to be presented here we make use of the following functions which define linear programming (LP) and quadratic programming (QP) problems respectively. For any $z \in \mathbb{R}^n$, $\varepsilon \geq 0$, and $\tau \geq 0$, define

$$\bar{\theta}^1_\varepsilon(z;\tau) \triangleq \min_{\|h\|_\infty \leq 1} \max\{\langle \bar{\nabla} f^o(z;\tau),h\rangle - \gamma\psi_o(z);$$

$$\bar{D}f^j_\varepsilon(z,h;\tau),\ j \in \underline{m}; \bar{D}g^j_\varepsilon(z,h;\tau),\ j \in \underline{p}\} \tag{17}$$

$$\bar{\theta}^2_\varepsilon(z;\tau) \triangleq \min\{ \frac{1}{2} \|h\|^2 + \max\{\langle \bar{\nabla} f^o(z;\tau),h\rangle - \gamma\psi_o(z);$$

$$\bar{D}f^j_\varepsilon(z,h;\tau),\ j \in \underline{m};\ \bar{D}g^j_\varepsilon(z,h;\tau),\ j \in \underline{p}\}\} \tag{18}$$

where $\gamma \geq 1$ is a constant. Let $\bar{h}^1_\varepsilon(z;\tau)$ ($\bar{h}^2_\varepsilon(z;\tau)$) denote a solution of the program defined by $\bar{\theta}^1_\varepsilon(z;\tau)$ ($\bar{\theta}^2_\varepsilon(z;\tau)$). Note that because of Assumption 3, the programs defined by (17) and (18) are easily solved by conventional LP and QP computer codes.

In the algorithm to be presented here, we use a direct search (a modified local variations method) in which we must ensure that the step length can be made arbitrarily small in a finite number of iterations. Therefore, we require the following to be true.

Assumption 7. For all $z_o \in \mathbb{R}^n$, the set $C(z_o) \triangleq \{z \in \mathbb{R}^n | \psi_o(z) \leq \psi_o(z_o)\}$ is bounded. ¤

III. An Algorithm Model

Because we are considering the use of approximations for the gradients, we require a more complicated algorithm model than that in Chapter 2. Our new requirements, therefore, are (i) the algorithm must be parameterized by an accuracy parameter, and (ii) the combined phase I - phase II feature of the previous model in Chapter 2 should be retained. These considerations motivated the development of a new algorithm model. This new model is a modification of 1.3.34 in [5] in which a combined phase I - phase II feature has been incorporated.

Given a set $F \subset \mathbb{R}^n$, a set of desirable points $\Delta \subset F$, and cost functions $C_1 : F \to \mathbb{R}$ and $C_2 : F^c \to \mathbb{R}$, we consider the following algorithm model in which the map $A : \mathbb{R}_+ \times \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is used.

<u>Algorithm Model</u>

Data: $\tau_o > 0$, $\alpha > 0$, $z_o \in \mathbb{R}^n$.

<u>Step 0</u>: Set $i = 0$.

<u>Step 1</u>: Set $\tau = \tau_o$.

<u>Step 2</u>: Compute a $y \in A(\tau, z_i)$.

<u>Step 3</u>: (i) For $z_i \in F$, if $C_1(y) - C_1(z_i) \leq -\alpha\tau$, set $z_{i+1} = y$, $\tau_i = \tau$, $i = i+1$, and go to step 2; else, go to step 4. (ii) For $z_i \in F^c$, if $C_2(y) - C_2(z_i) \leq -\alpha\tau$, set $z_{i+1} = y$, $\tau_i = \tau$, $i = i+1$, and go to step 2; else, go to step 4.

<u>Step 4</u>: Set $\tau = \tau/2$ and go to step 2. ⌑

We assume that the following hypotheses regarding the map $A(\cdot,\cdot)$, and the cost functions $C_1(\cdot)$ and $C_2(\cdot)$, are true.

<u>Assumption 8</u>. For $A : \mathbb{R}_+ \times \mathbb{R}^n \to 2^{\mathbb{R}^n}$, $F \subset \mathbb{R}^n$, $C_1 : F \to \mathbb{R}$, $C_2 : F^c \to \mathbb{R}$, $\Delta \subset \mathbb{R}^n$,

(i) $C_1(\cdot)$ and $C_2(\cdot)$ are continuous.

(ii) $A(\tau, F) \subset F$ for all $\tau > 0$.

(iii) for every $z \notin \Delta$, there exist a $\mu(z) > 0$, a $\rho(z) > 0$ and a $\tau(z) > 0$ such that

$$C_1(z'') - C_1(z') \le -\mu(z) \qquad \forall z' \in B(z,\rho(z)) \cap F$$
$$\forall z'' \in A(\tau,z')$$
$$\forall \tau \in (0,\tau(z)] \tag{19}$$

$$C_2(z'') - C_2(z') \le -\mu(z) \qquad \forall z' \in B(z,\rho(z)) \cap F^c$$
$$\forall z'' \in A(\tau,z')$$
$$\forall \tau \in (0,\tau(z)] \tag{20}$$

where $B(z,\rho) \triangleq \{z' \in \mathbb{R}^n \mid \|z-z'\| \le \rho\}$. ¤

In the proof of the convergence theorem,we shall make use of the following lemma.

Lemma 1. Suppose Assumption 8 is satisfied. If the Algorithm Model constructs an infinite sequence $\{z_i\}_{i=0}^{\infty}$, either $\tau_i \to 0$ or $\{z_i\}_{i=0}^{\infty}$ has no accumulation points.

Proof. Let $\hat{z}$ be an accumulation point of $\{z_i\}$; i. e., $z_i \overset{K}{\to} \hat{z}$ for some $K \subset \mathbb{Z}_+$ where $\mathbb{Z}_+ \triangleq \{0,1,2,\ldots\}$. Because $\{\tau_i\}$ is a monotonically decreasing sequence it must converge. Suppose $\tau_i \to \hat{\tau} > 0$. Then, by construction, $\tau_i = \hat{\tau} > 0$ for all $i \ge \hat{i}$ for some $\hat{i} \in \mathbb{Z}_+$.

Case 1. $z_i \in F^c$ for all $i \in \mathbb{Z}_+$. For all $i \ge \hat{i}$

$$C_2(z_{i+1}) - C_2(z_i) \le -\alpha\tau \tag{21}$$

and, hence,

$$\lim_{i\to\infty} [C_2(z_{i+1}) - C_2(z_i)] \le -\alpha\hat{\tau} < 0 \tag{22}$$

Since $\{C_2(z_i)\}$ is a monotonically decreasing sequence, it must converge; i.e. $C_2(z_i) \to C_2(\hat{z})$. Consequently, $\lim_{i\to\infty} [C_2(z_{i+1}) - C_2(z_i)] = 0$, which is a contradiction to (22).

Case 2. $z_{i'} \in F$ for some $i' \geq \hat{i}$. By Assumption 8(ii), $z_i \in F$ for all $i \geq i'$. Thus, for all $i \geq i'$

$$C_1(z_{i+1}) - C_1(z_i) \leq -\alpha\hat{\tau} \tag{23}$$

and, hence,

$$\lim_{i\to\infty} [C_1(z_{i+1}) - C_1(z_i)] \leq -\alpha\hat{\tau} < 0 \tag{24}$$

Since $\{C_1(z_i)\}$ is a monotonically decreasing sequence, it must converge, i.e. $C_1(z_i) \to C_1(\hat{z})$. Consequently $\lim_{i\to\infty} [C_1(z_{i+1}) - C_1(z_i)] = 0$ which is a contradiction to (24).

Thus, in either case we obtain a contradiction and we conclude that $\tau_i \to 0$. ¤

We now state the main convergence result for the Algorithm Model.

Theorem 1. Given a sequence $\{z_i\}$ constructed by the Algorithm Model, if it is finite then its last element is desirable, or else it is infinite and every accumulation point of $\{z_i\}$ is desirable.

Proof: Suppose the sequence $\{z_i\}$ is finite and $z_{\hat{i}}$ is the last element. The algorithm then constructs an infinite sequence $\{y_j\}$ where $y_j \in A(\frac{\tau_{\hat{i}-1}}{2^j}, z_{\hat{i}})$, $j = 0,1,2,\ldots$. If $z_{\hat{i}} \in F$, then $C_1(y_j) - C_1(z_{\hat{i}}) > -\alpha \frac{\tau_{\hat{i}-1}}{2^j}$, $j = 0,1,2,\ldots$. If $z_{\hat{i}} \in F^c$, then $C_2(y_j) - C_2(z_{\hat{i}}) > -\alpha\frac{\tau_{\hat{i}-1}}{2^j}$, $j = 0,1,2,\ldots$. Suppose, contrary to what is to be proved, that $z_{\hat{i}} \notin \Delta$. By Assumption 8

(iii) there exist a $\mu(z_{\hat{i}}) > 0$, $\rho(z_{\hat{i}})$, and a $\tau(z_{\hat{i}}) > 0$ such that (19) and (20) hold. Let $\hat{j} \in \mathbb{Z}_+$ be such that $\frac{\tau_{\hat{i}-1}}{2^{\hat{j}}} \leq \min\{\frac{\mu(z_{\hat{i}})}{\alpha}, \tau(z_{\hat{i}})\}$. Hence, if $z_{\hat{i}} \in F$,

$$C_1(y_j) - C_1(z_{\hat{i}}) \leq -\mu(z_i) \leq -\alpha \frac{\tau_{\hat{i}-1}}{2^j}, \quad \forall_j \geq \hat{j} \tag{25}$$

or if $z_{\hat{i}} \in F^c$

$$C_2(y_j) - C_2(z_{\hat{i}}) \leq -\mu(z_{\hat{i}}) \leq -\alpha \frac{\tau_{\hat{i}-1}}{2^j}, \quad \forall j \geq \hat{j} \tag{26}$$

This is a contradiction, and we conclude that $z_{\hat{i}} \in \Delta$; i.e. $z_{\hat{i}}$ is desirable. Now suppose $\{z_i\}$ is infinite and has an accumulation point $\hat{z}$. Let $K \subset \mathbb{Z}_+$ define a subsequence such that $z_i \xrightarrow{K} \hat{z}$. Suppose that $\hat{z} \notin \Delta$. By Assumption 8(iii) there exist a $\mu(\hat{z}) > 0$, a $\rho(\hat{z}) > 0$, and a $\tau(\hat{z}) > 0$ such that (19) and (20) hold. By Lemma 1, $\tau_i \to 0$ and therefore, there exists a $k_1 \in K$ such that $\tau_i \leq \tau(\hat{z})$ for all $i \geq k_1$. Let $\hat{k} \geq k_1$ be such that $z_i \in B(\hat{z},\rho(\hat{z}))$ for all $i \geq \hat{k}$, $i \in K$. We now consider two cases.

<u>Case 1</u>. $z_i \in F^c$ for all $i \in \mathbb{Z}_+$. From (20) we obtain

$$C_2(z_{i+1}) - C_2(z_i) \leq -\mu(\hat{z}) \qquad \forall i \geq \hat{k},\ i \in K \tag{27}$$

and hence, $\lim_{i \in K} [C_2(z_{i+1}) - C_2(z_i)] \leq -\mu(\hat{z})$. But $\{C_2(z_i)\}_{i=0}^{\infty}$ is a monotonically decreasing sequence and, therefore, it must converge. Hence, $\lim_{i\to\infty} [C_2(z_{i+1}) - C_2(z_i)] = 0$. Consequently, $\lim_{i \in K} [C_2(z_{i+1}) - C_2(z_i)] = 0$ which is a contradiction.

<u>Case 2</u>. $z_{\hat{i}} \in F$ for some $\hat{i} \in \mathbb{Z}_+$. By Assumption 8(ii), $z_i \in F$ for all $i \geq \hat{i}$. Hence, $z_i \in F$, for all $i \geq \max\{\hat{i},\hat{k}\}$ and we obtain from (19),

$$C_1(z_{i+1}) - C_1(z_i) \leq -\mu(\hat{z}), \qquad \forall i \geq \max\{\hat{i},\hat{k}\},\ i \in K \tag{28}$$

Hence, $\lim_{i \in K} [C_1(z_{i+1}) - C_1(z_i)] \leq -\mu(\hat{z})$. But $\{C_1(z_i)\}_{i=\hat{i}}^{\infty}$ is a monotonically decreasing sequence and, therefore, must converge. Hence, $\lim_{i \to \infty} [C_1(z_{i+1}) - C_1(z_i)] = 0$. Consequently, $\lim_{i \in K} [C_1(z_{i+1}) - C_1(z_i)] = 0$ which is a contradiction.

Because we obtain a contradiction in either case we conclude that $\hat{z} \in \Delta$; i.e. $\hat{z}$ is desirable. ¤

IV. The Algorithm

As explained in Section I, the algorithm to be presented here is a hybrid direct search - feasible directions descent method. The direct search part will be given here as two subprocedures, each of which is called from the main algorithm which is a feasible directions type method. The first subprocedure is a modified local variations method in which a feasibility condition is imposed at each iteration. The second subprocedure is a local variations method applied to the unconstrained problem of decreasing $\psi_o(z)$ whenever $\psi_o(z) > 0$.

For notational convenience we define $d_1 \triangleq e_1$, $d_2 \triangleq -e_1$, $d_3 \triangleq e_2$, $d_4 \triangleq -e_2, \ldots, d_{2n-1} \triangleq e_n$, $d_{2n} \triangleq -e_n$, and we denote the feasible region by $F \triangleq \{z \in \mathbb{R}^n | \psi_o(z) = 0\}$.

Subprocedure 1.

Input: x_o, τ; Output: $S_1(x_o, \tau)$.

Data: $\hat{\rho} > 0$.

Step 0: Set $i = 0$, $x = x_o$, $\rho_o = \tau\hat{\rho}$.

Step 1: Set $j = 1$.

Step 2: Compute $f^o(x+\rho_i d_j)$ and $\psi_o(x+\rho_i d_j)$.

Step 3: If $f^o(x+\rho_i d_j) < f^o(x)$ and $\psi_o(x+\rho_i d_j) = 0$, set $x = x+\rho_i d_j$ and go to step 1; else, go to step 4.

Step 4: If $j < 2n$, set $j = j+1$ and go to step 2; else, go to step 5.

Step 5: If $\rho_i \le \tau$, set $S_1(x_o,\tau) = x$ and stop; else, set $x_{i+1} = x$, $\rho_{i+1} = \rho_i/2$, $i = i+1$ and go to step 1. ¤

It is assumed that $\psi_o(x_o) = 0$ whenever $S_1(x_o,\tau)$ is to be computed. See Fig. 2 for an example of the application of Subprocedure 1.

Subprocedure 2.

Input: x_o, τ; Output: $S_2(x_o,\tau)$.

Data: $\hat{\rho} > 0$.

Step 0: Set $i = 0$, $x = x_o$, $\rho_o = \tau\hat{\rho}$.

Step 1: Set $j = 1$.

Step 2: Compute $\psi_o(x+\rho_i d_j)$.

Step 3: If $\psi_o(x+\rho_i d_j) < \psi_o(x)$ set $x = x+\rho_i d_j$, and go to step 1; else, go to step 4.

Step 4: If $j < 2n$, set $j = j+1$ and go to step 2; else, go to step 5.

Step 5: If $\rho_i \le \tau$, set $S_2(x_o,\tau) = x$ and stop; else, set $x_{i+1} = x$, $\rho_{i+1} = \rho_i/2$, $i = i+1$, and go to step 1. ¤

Although it is not indicated, it is necessary to store the function values in step 2 of either subprocedure, whenever $\rho_i \le \tau$. Then it is possible to use these values in computing the approximations to the gradients of each function.$^+$

The following lemma is an obvious consequence of Assumption 7.

Lemma 2. Given $x_o \in \mathbb{R}^n$ and $\tau > 0$, Subprocedure 1 (2) will construct a point $S_1(x_o,\tau)(S_2(x_o,\tau))$ after a finite number of iterations. ¤

We can now state the main algorithm in which the subprocedures are used in combination with a feasible directions, descent type, method.

$^+$In the implementable version of this algorithm only a discrete subset of Ω would be used as in Chapter 2. Then only a finite number of function values of $\phi^j(z,\omega)$ would be stored. In this chapter, however, we consider only the conceptual computation over Ω. The algorithm here can be modified in an obvious way employing the same method as in Chapter 2.

The Algorithm

Data: $\alpha_1 \in (0,1)$, $\alpha_2 \in (0,1)$, $\beta \in (0,1)$, $\gamma \geq 1$, $\delta > 0$, $\varepsilon_o > 0$, $\tau_o > 0$, $\lambda_{min} \in (0,1]$, $\lambda_o > \lambda_{min}$, $y_o \in \mathbb{R}^n$, $\pi \in \{1,2\}$.

Step 0: Set $i = 0$, $\tau = \tau_o$.

Step 1: If $y_o \in F$ compute $z_o = S_1(y_o,\tau)$. If $y_o \in F^c$ compute $z_o = S_2(y_o,\tau)$. Store the function values computed at the points $z_o + \tau_i^j e_j$, $j = 1,2,\ldots,n.$*

Step 2: Set $\varepsilon = \varepsilon_o$.

Step 3: Compute $\bar{\nabla} f^o(z_i;\tau)$, $\bar{\nabla} g^j(z_i;\tau)$, $j \in J_\varepsilon^g(z)$, $\bar{\nabla}\phi^j(z_i,\omega;\tau)$, $\omega \in \tilde{\Omega}_\varepsilon^j(z_i)$; $j \in J_\varepsilon^f(z_i)$, using the function values computed in the last subprocedure call.

Step 4: Compute $\bar{\theta}_\varepsilon^\pi(z_i;\tau)$ and $\bar{h}_\varepsilon^\pi(z_i;\tau)$.

Step 5: If $\bar{\theta}_\varepsilon^\pi(z_i;\tau) = 0$, compute $\bar{\theta}_o^\pi(z_i;\tau)$ and go to step 6; else, go to step 7.

Step 6: If $\bar{\theta}_o^\pi(z_i;\tau) = 0$, set $y = z_i$ and go to step 15; else, set $\varepsilon = \varepsilon/2$ and go to step 3.

Step 7: If $\bar{\theta}_\varepsilon^\pi(z_i;\tau) \leq -\delta\varepsilon$, (set $\varepsilon(z_i)=\varepsilon$)+ go to step 8; else, set $\varepsilon = \varepsilon/2$ and go to step 3.

Step 8: Set $\lambda = \lambda_o$.

Step 9: If $\psi_o(z_i) > 0$, go to step 11.

Step 10: If

$$f^o(z_i+\lambda\bar{h}_\varepsilon^\pi(z_i;\tau)) - f^o(z_i) \leq -\alpha_1\delta\lambda\varepsilon \tag{29}$$

$$g^j(z_i+\lambda\bar{h}_\varepsilon^\pi(z_i;\tau)) \leq 0 \quad j \in \underline{p} \tag{30}$$

$$f^j(z_i+\lambda\bar{h}_\varepsilon^\pi(z_i;\tau)) \leq 0 \quad j \in \underline{m} \tag{31}$$

go to step 13; else, set $\lambda = \beta\lambda$ and go to step 12.

---

*Whenever S_2 is called, the cost function must also be computed at the points $z_i + \tau_i^j e_j$, $j = 1,2,\ldots,n$.

+Do not store $\varepsilon(z_i)$. This quantity is used only in the proofs.

Step 11: If

$$\psi(z_i+\lambda\bar{h}^{\pi}_{\varepsilon}(z_i;\tau)) - \psi(z_i) \leq - \alpha\delta\lambda\varepsilon \tag{32}$$

go to step 13; else, set $\lambda = \beta\lambda$ and go to step 12.

Step 12: If $\lambda \geq \tau\cdot\lambda_{min}$, go to step 9; else, set $y = z_i$ and go to step 15.

Step 13: Set $z_i' = z_i + \lambda\bar{h}^{\pi}_{\varepsilon}(z_i;\tau)$.

Step 14: If $z_i \in F$, compute $y = S_1(z_i',\tau)$. If $z_i \in F^c$, compute $y = S_2(z_i',\tau)$. Store the function values computed at the points $z_i + \tau_i^j e_j$, $j = 1,2,\ldots,n$.

Step 15: For $z_i \in F$, if $f^o(y) - f^o(z_i) \leq -\alpha_2\tau$, go to step 16; else, set $\tau = \tau/2$ and go to step 2. For $z_i \in F^c$, if $\psi(y) - \psi(z_i) \leq -\alpha_2\tau$, go to step 16; else, set $\tau = \tau/2$ and go to step 2.

Step 16: Set $z_{i+1} = y$, $\tau_{i+1} = \tau$, $i = i+1$, and go to step 2. ⌑

Steps 2 through 14 of this algorithm define a map $A : \mathbb{R}_+ \times \mathbb{R}^n \to 2^{\mathbb{R}^n}$ and step 10 provides the property that $A(\tau,F) \subset F$ for all $\tau > 0$. We define the two cost functions $C_1 = f^o$ and $C_2 \triangleq \psi$ on F and F^c respectively. With these quantities defined, it is clear that this algorithm is of the same form as the Algorithm Model of the previous section. In order to make use of the algorithm model it is necessary to also define the set of desirable points Δ and then we must show that Assumptions 8 (i),(ii), and (iii) hold. By the definition of C_1 and C_2 it is obvious that Assumption 8(i) holds, and since $A(\tau,F) \subset F$ for all $\tau > 0$, Assumption 8(ii) is satisfied. In order to define Δ we state the following lemma which is an obvious extension of Proposition 1 in [1].

Lemma 3. If $z \in F$ is optimal for (1) then $\bar{\theta}^{\pi}_o(z;0) = \theta^{\pi}_o(z) = 0$, $\pi \in \{1,2\}$, where we define for $z \in \mathbb{R}^n$, $\varepsilon \geq 0$, $\gamma \geq 1$

$$\theta^1_\varepsilon(z) \triangleq \min_{\|h\|_\infty \leq 1} \max\{\langle \nabla f^o(z), h\rangle - \gamma\psi_o(z);$$

$$Df^j_\varepsilon(z,h),\ j \in \underline{m};\ Dg^j_\varepsilon(z,h),\ j \in \underline{p}\} \tag{33}$$

$$\theta^2_\varepsilon(z) \triangleq \min\{\tfrac{1}{2}\|h\|^2 + \max\{\langle \nabla f^o(z), h\rangle - \gamma\psi_o(z);$$

$$Df^j_\varepsilon(z,h),\ j \in \underline{m};\ Dg^j_\varepsilon(z,h),\ j \in \underline{p}\}\} \tag{34}$$

where

$$Df^j_\varepsilon(z,h) \triangleq \begin{cases} \max\limits_{\omega \in \tilde{\Omega}^j_\varepsilon(z)} \langle \nabla_z \phi^j(z,\omega), h\rangle & \text{if } j \in J^f_\varepsilon(z) \\ -\infty & \text{otherwise} \end{cases} \tag{35}$$

$$Dg^j_\varepsilon(z,h) \triangleq \begin{cases} \langle \nabla g^j(z), h\rangle & \text{if } j \in J^g_\varepsilon(z) \\ -\infty & \text{otherwise} \end{cases} \tag{36}$$

Furthermore, for all $z \in F^c$, $\theta^\pi_o(z) < 0$ for $\pi \in \{1,2\}$. ¤

As a consequence of Lemma 3 we define the set of desirable points as

$$\Delta \triangleq \{z \in \mathbb{R}^n \mid \bar{\theta}^\pi_o(z;0) = 0,\ \pi \in \{1,2\}\} \tag{37}$$

Note that $\bar{\theta}^1_o(z;0) = 0$ if and only if $\bar{\theta}^2_o(z;0) = 0$.

In order to show that the map A is well defined, we will require the following result which is proved in Appendix A.

Proposition 1. For all $z \in \mathbb{R}^n$, $\tau > 0$, $\pi \in \{1,2\}$, such that $\bar{\theta}^\pi_o(z;\tau) < 0$, there exists an $\bar{\varepsilon} > 0$ such that $\bar{\theta}^\pi_\varepsilon(z;\tau) \leq \bar{\theta}^\pi_o(z;\tau) < 0$ for all $\varepsilon \in [0,\bar{\varepsilon}]$. ¤

The following lemma is a consequence of Proposition 1 and it demonstrates that the feasible directions part of the algorithm is well defined.

<u>Lemma 4</u>. The algorithm cannot cycle indefinitely in the loop defined by steps 3 through 6 or in the loop defined by steps 3 through 7.

<u>Proof</u>. Let $\pi \in \{1,2\}$. For all $z \in \mathbb{R}^n$, $\varepsilon \geq 0$, $\tau > 0$,

$$\bar{\theta}^{\pi}_{o}(z;\tau) \leq \bar{\theta}^{\pi}_{\varepsilon}(z;\tau) \leq 0 \tag{38}$$

If $\bar{\theta}^{\pi}_{o}(z_i;\tau_i) = 0$, we have from (38) that $\bar{\theta}^{\pi}_{\varepsilon_o}(z_i;\tau_i) = 0$ and the algorithm proceeds through steps 3 through 6 and then to step 15. Hence, the algorithm cannot cycle indefinitely in the given loops if $\bar{\theta}^{\pi}_{o}(z_i;\tau_i) = 0$.

Suppose that $\bar{\theta}^{\pi}_{o}(z_i;\tau_i) < 0$. Then by Proposition 1 there exists an $\bar{\varepsilon} > 0$ such that

$$\bar{\theta}^{\pi}_{\varepsilon}(z_i;\tau_i) \leq \frac{1}{2}\,\bar{\theta}^{\pi}_{o}(z_i;\tau_i) \qquad \forall \varepsilon \in [0,\bar{\varepsilon}] \tag{39}$$

Let $\hat{k} \geq 0$ be the smallest nonnegative integer such that

$$\hat{\varepsilon} \triangleq \varepsilon_o 2^{-\hat{k}} \leq \min\{\bar{\varepsilon}, -\frac{1}{2\delta}\,\bar{\theta}^{\pi}_{o}(z_i;\tau_i)\} \tag{40}$$

We then obtain from (39)

$$\bar{\theta}^{\pi}_{\hat{\varepsilon}}(z_i;\tau_i) \leq -\frac{1}{2}\,\bar{\theta}^{\pi}_{o}(z_i;\tau_i) \leq -\delta\hat{\varepsilon} \tag{41}$$

Thus, the algorithm will construct $\varepsilon(z_i) \geq \hat{\varepsilon}$ such that $\bar{\theta}^{\pi}_{\varepsilon(z_i)}(z_i;\tau_i) \leq -\delta\varepsilon(z_i)$ in a finite number of cycles between steps 3 and 7. ¤

The following lemma shows that Assumption 8 (iii) holds. The proof is contained in Appendix A.

<u>Lemma 5</u>. For all $z \in \mathbb{R}^n$ such that $\bar{\theta}^{\pi}_{o}(z;0) < 0$, $\pi \in \{1,2\}$, there exist $\mu > 0$, $\rho > 0$, and $\bar{\tau} > 0$ such that

$$f^o(z'') - f^o(z') \le -\mu \qquad \forall z' \in B(z,\rho) \cap F$$
$$\forall z'' \in A(z';\tau)$$
$$\forall \tau \in (0,\bar{\tau}] \tag{42}$$

$$\psi(z'') - \psi(z') \le -\mu \qquad \forall z' \in B(z,\rho) \cap F^c$$
$$\forall z'' \in A(z';\tau)$$
$$\forall \tau \in (0,\bar{\tau}] \tag{43}$$

¤

We can now state the main convergence result, which, from the above discussion, follows directly from the use of Theorem 1.

Theorem 2. If the algorithm constructs a sequence $\{z_i\}$ which is finite then the last point constructed is desirable. If the sequence is infinite then every accumulation point is desirable. ¤

V. Scaling

When using numerical approximations of derivatives it may be necessary to take special precautions to ensure sufficient accuracy. Curtis and Reid in [13] proposed a scaling method for choosing appropriate step lengths when using difference formulas for first derivative approximations. Their method is one in which a balance is maintained between rounding and truncation error estimates.

We shall briefly describe the method in [13] by considering a single function $f : \mathbb{R} \to \mathbb{R}$ for which we wish to approximate $\frac{df}{dx}$. As an estimate of the truncation error we define

$$\Delta_t = \left| \frac{f(x+h)-f(x-h)}{2h} - \frac{f(x+h)-f(x)}{h} \right| \tag{44}$$

As an estimate of the round-off error we define

$$\Delta_r = h \max\left\{\left|\frac{f(x+h)-f(x)}{h}\right|, \left|\frac{f(x-h)-f(x)}{h}\right|\right\} \tag{45}$$

The justification of these approximations is given in [13] and will not be repeated here. It should be noted, however, that these formulas are based on heuristic arguments; i.e. the quantities Δ_t and Δ_r are really only upper bounds on the truncation error and round-off error respectively. According to Curtis and Reid these bounds are very conservative.

The balance of Δ_t and Δ_r is maintained by examining the ratio

$$u = \frac{\Delta_t}{\Delta_r}$$

The aim is to keep u in the range $[u_{min},u_{max}]$. If, after calculating u for a given h, $u \notin [u_{min},u_{max}]$, a new vector h' is computed, where

$$h' = h\sqrt{u_{aim}/\max(u,1)} \tag{46}$$

and u_{aim} is some selected value in the range $[u_{min},u_{max}]$. Finally, the value of h' is restricted to a range $[h_{min},h_{max}]$.

In the algorithm of Section IV, we approximate the derivatives (gradients) by using difference formulas in which the functions are evaluated along each coordinate axis. In Subprocedures 1 and 2, the steps along the coordinate axes are given by the value of ρ_i. The normal procedure could be to simply use the last value of $\rho_i \leq \tau$ for the step length in the derivative approximation formulas. One can then insert a test to check if the ratio $u = \Delta_t/\Delta_r$ is within a prescribed range for each component of the approximated gradient. If so, then no further calculations would be needed for the gradient approximations. If u is not in the prescribed range for any component, additional function evaluations would be needed with a step length, given by (46), along the direction corresponding to that component. Note that a separate test is performed for each coordinate direction.

The scaling technique proposed in [13] can be implemented in the algorithm of Section IV with no effect on the convergence results. On some well-scaled problems it may be unnecessary to include the additional coding for this scaling. Depending on the required overall accuracy of the final result the user of this algorithm should decide whether his problem is scaled well enough.

VI. Conclusions

In this chapter we have presented a new, derivative-free algorithm for solving a certain type of infinitely constrained optimization problem. This new algorithm, unlike most other algorithms for constrained optimization, uses a hybrid direct search-descent method. It gives the user a greater flexibility when it is used in computer aided design applications since it requires no explicit derivative calculations. Also, if the algorithm is used interactively, the designer has the additional flexibility of controlling whether the algorithm uses the direct search more or less than the descent part of the algorithm.

This new algorithm may also be used without the direct search subprocedures. In this case, the algorithm becomes a derivative-free implementation of the algorithm contained in [15].

Appendix A.

In this appendix we shall present proofs for Proposition 1 and Lemma 5. Without loss of generality, we shall assume that $\pi=1$, $p=1$, and $m=1$.

Proposition 1. For all $z \in \mathbb{R}^n$, $\tau > 0$, $\pi \in \{1,2\}$, such that $\bar{\theta}_o^\pi(z;\tau) < 0$, there exists an $\bar{\varepsilon} > 0$ such that $\bar{\theta}_\varepsilon^\pi(z;\tau) \leq \bar{\theta}_o^\pi(z;\tau) < 0$ for all $\varepsilon \in [0,\bar{\varepsilon}]$.

Proof: Given $z \in \mathbb{R}^n$ and $\tau > 0$ such that $\bar{\theta}_o^1(z;\tau) < 0$, there exists a $\mu_1 > 0$ such that

$$\max_{\omega \in N_{\mu_1}(\Omega_o(z))} \langle \bar{\nabla}\phi(z,\omega;\tau),h\rangle - \bar{D}f_o(z,h;\tau) \leq -\frac{1}{2}\bar{\theta}_o^1(z;\tau) \qquad \forall h \in S \tag{A1}$$

where $S \triangleq \{h \in \mathbb{R}^n | \|h\|_\infty \leq 1\}$. This follows directly from Assumption 1. By Proposition 2.1[+] there exists an $\varepsilon_1 > 0$ such that for all $\varepsilon \in [0,\varepsilon_1]$

$$\tilde{\Omega}_\varepsilon(z) \subset N_{\mu_1}(\Omega_o(z)) \tag{A2}$$

Thus, for all $\varepsilon \in [0,\varepsilon_1]$

$$\bar{D}f_\varepsilon(z,h;\tau) \leq \bar{D}f_o(z,h;\tau) - \frac{1}{2}\bar{\theta}_o^1(z;\tau) \qquad \forall h \in S \tag{A3}$$

If $\Omega_o(z) = \phi$ there exists an $\varepsilon_1 > 0$ such that $\Omega_\varepsilon(z) = \phi$ for all $\varepsilon \in [0,\varepsilon_1]$. In this case $\bar{D}f_o(z,h;\tau) = \bar{D}f_\varepsilon(z,h;\tau) = -\infty$, for all $\varepsilon \in [0,\varepsilon_1]$. Let $\bar{\varepsilon} \in (0,\varepsilon_1]$ be such that $J_\varepsilon^g(z) = J_o^g(z)$ for all $\varepsilon \in [0,\bar{\varepsilon}]$. The desired result now follows directly from the definition of $\theta_\varepsilon^1(z;\tau)$. ¤

[+]We denote a proposition, lemma, etc. from another chapter by indicating the chapter number followed by the proposition number, lemma number, etc.

Before proving Lemma 5 we require several preliminary results.

Proposition 2. For any $z \in \mathbb{R}^n$ such that $\Omega_o(z) \neq \phi$, and for any $\mu > 0$, there exist a $\rho > 0$, $\bar{\varepsilon} > 0$ and a $\bar{\tau} > 0$ such that

$$\bar{D}f_\varepsilon(z',h;\tau) \leq Df(z,h) + \mu \quad \forall h \in S$$
$$\forall z' \in B(z,\rho), \quad \forall \tau \in [0,\bar{\tau}], \quad \forall \varepsilon \in [0,\bar{\varepsilon}] \tag{A4}$$

Proof: By Proposition 2.1 there exist a $\rho > 0$ and an $\bar{\varepsilon} > 0$ such that

$$\tilde{D}f_\varepsilon(z',h) \leq Df(z,h) + \frac{\mu}{2} \qquad \forall h \in S, \quad \forall z' \in B(z,\rho)$$
$$\forall \varepsilon \in [0,\bar{\varepsilon}] \tag{A5}$$

By Assumption 5, since $B(z,\rho)$ is compact, there exists a $\bar{\tau} > 0$ such that

$$\|\bar{\nabla}\phi(z',\omega;\tau) - \nabla_z\phi(z',\omega)\| \leq \frac{\mu}{2\sqrt{n}} \qquad \forall \omega \in \Omega$$
$$\forall \tau \in [0,\bar{\tau}]$$
$$\forall z' \in B(z,\rho) \tag{A6}$$

Hence,

$$\langle \bar{\nabla}\phi(z',\omega;\tau),h\rangle - \langle \nabla_z\phi(z',\omega),h\rangle \leq \frac{\mu}{2} \qquad \forall \omega \in \Omega$$
$$\forall \tau \in [0,\bar{\tau}]$$
$$\forall z' \in B(z,\rho) \tag{A7}$$

and it follows that for all $\varepsilon \geq 0$,

$$\bar{D}f_\varepsilon(z',h;\tau) \leq \tilde{D}f_\varepsilon(z',h) + \frac{\mu}{2} \qquad \forall z' \in B(z,\rho)$$
$$\forall \tau \in [0,\bar{\tau}] \tag{A8}$$

Combining (A5) and (A8) we obtain

$$\bar{D}f_\varepsilon(z',h;\tau) \leq Df(z,h) + \mu \qquad \forall z' \in B(z,\rho)$$
$$\forall \tau \in [0,\bar{\tau}]$$
$$\forall \varepsilon \in [0,\bar{\varepsilon}] \quad \square \tag{A9}$$

Proposition 3. For any $z \in \mathbb{R}^n$ such that $\Omega_0(z) \neq \phi$ and for any $\varepsilon > 0$, $\mu > 0$, there exist a $\rho > 0$, and a $\bar{\tau} > 0$ such that for all $z' \in B(z,\rho)$

$$Df(z,h) \leq \bar{D}f_\varepsilon(z',h;\tau) + \mu \qquad \forall h \in S, \quad \forall \tau \in [0,\bar{\tau}] \tag{A10}$$

Proof: By Proposition 2.2, given $z \in \mathbb{R}^n$ such that $\Omega_0(z) \neq \phi$ and $\varepsilon > 0$, $\mu > 0$, there exists a $\rho > 0$ such that for all $z' \in B(z,\rho)$,

$$Df(z,h) \leq \tilde{D}f_\varepsilon(z',h) + \frac{\mu}{2} \qquad \forall h \in S \tag{A11}$$

By Assumption 5, since $B(z,\rho)$ is compact, there exists a $\bar{\tau} > 0$ such that

$$\|\nabla\phi(z',\omega) - \bar{\nabla}\phi(z',\omega;\tau)\| \leq \frac{\mu}{2\sqrt{n}} \qquad \forall \omega \in \Omega, \quad \forall \tau \in [0,\bar{\tau}], \quad \forall z' \in B(z,\rho) \tag{A12}$$

Hence,

$$|\langle \nabla\phi(z',\omega),h\rangle - \langle \bar{\nabla}\phi(z',\omega;\tau),h\rangle| \leq \frac{\mu}{2} \qquad \forall \omega \in \Omega, \quad \forall \tau \in [0,\bar{\tau}], \quad \forall z' \in B(z,\rho) \tag{A13}$$

and it follows that for all $\varepsilon \geq 0$, $\tau \in [0,\bar{\tau}]$,

$$\tilde{D}f_\varepsilon(z',h) \leq \bar{D}f_\varepsilon(z',h;\tau) + \frac{\mu}{2} \qquad \forall h \in S, \quad \forall z' \in B(z,\rho) \tag{A14}$$

Combining (A11) and (A14) we obtain

$$Df(z,h) \leq \bar{D}f_\varepsilon(z',h;\tau) + \mu \qquad \forall h \in S, \quad \forall \tau \in [0,\bar{\tau}], \quad \forall z' \in B(z,\rho) \tag{A15}$$

□

Proposition 4. For any $z \in \mathbb{R}^n$ such that $\Omega_o(z) \neq \phi$, and for any $\varepsilon > 0$, $\mu > 0$, there exist a $\rho > 0$ and a $\bar{\tau} > 0$ such that for all z', $z'' \in B(z,\rho)$

$$Df(z'',h) \leq \bar{D}f_\varepsilon(z',h;\tau) + \mu \qquad \forall h \in S \quad \forall \tau \in [0,\bar{\tau}]. \tag{A16}$$

Proof: By Corollary 2.1, given $z \in \mathbb{R}^n$ such that $\Omega_o(z) \neq \phi$ and $\varepsilon > 0$, $\mu > 0$, there exists a $\rho > 0$ such that for all z', $z'' \in B(z,\rho)$

$$Df(z'',h) \leq \tilde{D}f_\varepsilon(z',h) + \frac{\mu}{2} \qquad \forall h \in S \tag{A17}$$

As a consequence of Assumption 5 there exists a $\bar{\tau} > 0$ such that for all $\varepsilon \geq 0$,

$$\tilde{D}f_\varepsilon(z',h) \leq \bar{D}f_\varepsilon(z',h;\tau) + \frac{\mu}{2} \qquad \forall h \in S \quad \forall z' \in B(z,\rho) \quad \forall \tau \in [0,\bar{\tau}] \tag{A18}$$

Combining (A17) and (A18) we obtain

$$Df(z'',h) \leq \bar{D}f_\varepsilon(z',h;\tau) + \mu \qquad \forall h \in S \quad \forall \tau \in [0,\bar{\tau}] \quad \forall z', z'' \in B(z,\rho) \tag{A19}$$

¤

Proposition 5. For all $z \in \mathbb{R}^n$ such that $\bar{\theta}_o\pi(z;0) < 0$, there exist a $\rho > 0$, an $\bar{\varepsilon} > 0$, and a $\bar{\tau} > 0$ such that

$$\varepsilon(z';\tau) \geq \bar{\varepsilon} \qquad \forall z' \in B(z,\rho), \qquad \forall \tau \in [0,\bar{\tau}] \tag{A20}$$

where $\varepsilon(z',\tau)$ is the value of ε constructed by steps 1 through 6 of the algorithm with $z_i = z'$ and $\tau \in [0,\bar{\tau}]$.

Proof: Given $z \in \mathbb{R}^n$ such that $\bar{\theta}^1_o(z;0) < 0$, let $\mu > 0$ be such that $\bar{\theta}^1_o(z;o) \leq -\mu < 0$. (Consider $\pi = 1$ only.)

Case 1. $\psi(z) < 0$.

By the continuity of $\psi(\cdot)$ there exists a $\tilde{\rho} > 0$ such that for all $z' \in B(z,\tilde{\rho})$,

$$\psi(z') < \frac{\psi(z)}{2} \tag{A21}$$

Let $\tilde{\varepsilon} \triangleq -\frac{\psi(z)}{2}$, then because $\psi_o(z') = 0$, we obtain

$$g(z') - \psi_o(z') < -\varepsilon \tag{A22}$$

$$f(z') - \psi_o(z') < -\varepsilon \tag{A23}$$

for all $z' \in B(z,\tilde{\rho})$ and for all $\varepsilon \in [0,\tilde{\varepsilon}]$. Consequently

$$\bar{D}g_\varepsilon(z',h;\tau) = \tilde{D}f_\varepsilon(z',h;\tau) = -\infty \qquad \begin{array}{l} \forall h \in S \\ \forall z' \in B(z,\tilde{\rho}) \\ \forall \varepsilon \in [0,\tilde{\varepsilon}] \\ \forall \tau \geq 0 \end{array} \tag{A24}$$

Case 2. $\psi(z) \geq 0$, $\Omega_o(z) = \phi$

By the same argument as in Lemma 2.2 there exist a $\tilde{\rho} > 0$ and an $\tilde{\varepsilon} > 0$ such that

$$Dg_\varepsilon(z',h) \leq Dg_o(z,h) + \frac{\mu}{4} \tag{A25}$$

$$\bar{D}f_\varepsilon(z',h;\tau) = \tilde{D}f_o(z,h) = -\infty$$

$$\forall h \in S, \quad \forall z' \in B(z,\tilde{\rho}), \quad \forall \tau \geq 0, \quad \forall \varepsilon \in [0,\tilde{\varepsilon}] \tag{A26}$$

By Assumption 5 there exists a $\tilde{\tau} > 0$ such that

$$\bar{D}g_{\varepsilon}(z',h;\tau) \le Dg_{\varepsilon}(z',h) + \frac{\mu}{4} \qquad \forall z' \in B(z,\tilde{\rho}),\ \forall h \in S,\ \forall \varepsilon \in [0,\tilde{\varepsilon}],\ \forall \tau \in [0,\tilde{\tau}] \tag{A27}$$

From (A25) and (A27) we obtain

$$\bar{D}g_{\varepsilon}(z',h;\tau) \le Dg_{o}(z,h) + \frac{\mu}{2} \qquad \forall z' \in B(z,\tilde{\rho}),\ \forall h \in S,\ \forall \varepsilon \in [0,\tilde{\varepsilon}],\ \forall \tau \in [0,\tilde{\tau}] \tag{A28}$$

Case 3. $\psi(z) \ge 0$, $\Omega_o(z) \ne 0$.

By Proposition 2 there exist a $\rho_1 > 0$, $\tilde{\tau} > 0$ and an $\varepsilon_1 > 0$ such that

$$\bar{D}f_{\varepsilon}(z',h;\tau) \le \tilde{D}f_{o}(z,h) + \frac{\mu}{2} \qquad \forall h \in S,\ \forall z' \in B(z,\rho_1),\ \forall \varepsilon \in [0,\varepsilon_1],\ \forall \tau \in [0,\tilde{\tau}] \tag{A29}$$

Using similar arguments to the ones used in the previous cases it can be shown that there exist a $\tilde{\rho} \in (0,\rho_1]$ and an $\tilde{\varepsilon} \in (0,\varepsilon_1]$ such that

$$\bar{D}g_{\varepsilon}(z',h;\tau) \le Dg_{o}(z,h) + \frac{\mu}{2} \qquad \forall h \in S,\ \forall z' \in B(z,\tilde{\rho}),\ \forall \varepsilon \in [0,\tilde{\varepsilon}],\ \forall \tau \in [0,\tilde{\tau}] \tag{A30}$$

Now, by the continuity of $\nabla f^{o}(\cdot)$ and $\psi_o(\cdot)$ and by Assumption 5 there exist a $\rho \in (0,\tilde{\rho}]$ and a $\bar{\tau} \in (0,\tilde{\tau}]$ such that

$$\langle \bar{\nabla} f^{o}(z';\tau),h\rangle - \psi_o(z') \le \langle \nabla f^{o}(z),h\rangle - \psi_o(z) + \frac{\mu}{2} \qquad \forall h \in S,\ \forall \varepsilon \in [0,\tilde{\varepsilon}],\ \forall z' \in B(z,\rho),\ \forall \tau \in [0,\bar{\tau}] \tag{A31}$$

Thus,

$$\max\{\langle \bar{\nabla} f^{o}(z';\tau),h\rangle - \psi_o(z');\ \bar{D}g_{\varepsilon}(z',h;\tau);\ \bar{D}f_{\varepsilon}(z',h;\tau)\}$$

$$\leq \max\{\langle \nabla f^{o}(z),h\rangle - \psi_o(z);\ Dg_o(z,h);\ \tilde{D}f_o(z,h)\} + \frac{\mu}{2} \qquad \forall h \in S,\ \forall z' \in B(z,\rho)$$
$$\forall \varepsilon \in [0,\tilde{\varepsilon}],\ \forall \tau \in [0,\bar{\tau}] \tag{A32}$$

Because (A32) holds for all $h \in S$, it follows that

$$\bar{\theta}^{1}_{\varepsilon}(z';\tau) \leq \bar{\theta}^{1}_{o}(z;0) + \frac{\mu}{2} \leq -\frac{\mu}{2} \qquad \forall z' \in B(z,\rho),\ \forall \varepsilon \in [0,\tilde{\varepsilon}],\ \forall \tau \in [0,\bar{\tau}] \tag{A33}$$

Let $\hat{j}(z) \in \mathbb{Z}_+$ be such that $\hat{\varepsilon} \triangleq \varepsilon_o 2^{-\hat{j}(z)} \leq \min\{\tilde{\varepsilon}, \frac{\mu}{2\delta}\}$. Then, from (A33) we obtain

$$\bar{\theta}^{1}_{\hat{\varepsilon}}(z';\tau) \leq -\delta\hat{\varepsilon} \qquad \forall z' \in B(z,\rho),\ \forall \tau \in [0,\bar{\tau}] \tag{A34}$$

In steps 1 through 6 of the algorithm, the quantity $\varepsilon(z';\tau) = \varepsilon_o 2^{-j(z';\tau)}$ is computed where $j(z';\tau) \in \mathbb{Z}_+$ is the smallest integer such that

$$\bar{\theta}^{1}_{\varepsilon(z';\tau)}(z';\tau) \leq -\delta\varepsilon(z';\tau) \tag{A35}$$

for any $z' \in B(z,\rho)$ and $\tau \in [0,\bar{\tau}]$. Comparing (A34) and (A35) we conclude that $j(z';\tau) \geq \hat{j}(z)$, and hence, $\varepsilon(z';\tau) \geq \hat{\varepsilon}$ for all $z' \in B(z,\rho)$ and for all $\tau \in [0,\bar{\tau}]$. ¤

<u>Proposition 6</u>. For all $z \in \mathbb{R}^n$ such that $\bar{\theta}^{\pi}_{o}(z;0) < 0$ there exist a $\rho > 0$, and a $\bar{\tau} > 0$ such that for all $\tau \in [0,\bar{\tau}]$,

$$\text{(a)}\quad Df(z'',h') \leq -\alpha\delta\varepsilon(z';\tau) \qquad \text{if } f(z) = \psi_o(z),\ \forall z'', z' \in B(z,\rho),\ \forall h' \in S(z',\tau) \tag{A36}$$

(b) $\langle \nabla g(z''),h' \rangle \leq -\alpha\delta\varepsilon(z';\tau)$ if $g(z) = \psi_o(z)$

$\forall z'',z' \in B(z,\rho)$

$\forall h' \in S(z',\tau)$ (A37)

(c) $\langle \nabla f^o(z''),h' \rangle \leq -\alpha\delta\varepsilon(z';\tau)$ $\forall z' \in B(z,\rho) \cap F$

$\forall z'' \in B(z,\rho)$, $\forall h' \in S(z',\tau)$ (A38)

where $S(z',\tau)$ is the set of all direction vectors which are solutions to the program defined by $\bar{\theta}^{\pi}_{\varepsilon(z';\tau)}(z';\tau)$.

Proof: By Proposition 5 there exist a $\rho_1 > 0$, an $\hat{\varepsilon} > 0$, and a $\tau_1 > 0$ such that $\varepsilon(z';\tau) \geq \hat{\varepsilon}$ for all $z' \in B(z,\rho_1)$ and for all $\tau \in [0,\tau_1]$.

(a) If $f(z) = \psi_o(z)$, by Proposition 4 there exist a $\rho_2 \in (0,\rho_1]$ and a $\tau_2 \in (0,\tau_1]$ such that for all $z'',z' \in B(z,\rho_2)$

$$Df(z'',h') \leq \bar{D}f_{\varepsilon}(z',h';\tau) + (1-\alpha)\ \delta\hat{\varepsilon}$$

$$\leq -\delta\varepsilon(z';\tau) + (1-\alpha)\ \delta\hat{\varepsilon}$$

$$\leq -\alpha\delta\varepsilon(z';\tau) \qquad \forall h' \in S(z',\tau)$$

$$\forall \tau \in [0,\tau_2] \qquad \text{(A39)}$$

(b) If $g(z) = \psi_o(z)$, there exists a $\rho_3 \in (0,\rho_1]$ such that $g(z') - \psi_o(z') \geq -\hat{\varepsilon}$ for all $z' \in B(z,\rho_3)$. Also, there exist a $\rho_4 \in (0,\rho_3]$ and a $\tau_3 \in (0,\tau_1]$ such that for all $z'',z' \in B(z,\rho_4)$

$$\langle \nabla g(z''),h' \rangle \leq \langle \bar{\nabla} g(z';\tau),h' \rangle + (1-\alpha)\ \delta\hat{\varepsilon} \qquad \forall h' \in S(z';\tau)$$

$$\forall \tau \in [0,\tau_3] \qquad \text{(A40)}$$

Since $g(z') - \psi_o(z') \geq -\hat{\varepsilon} \geq -\varepsilon(z';\tau)$ for all $z' \in B(z,\rho_4)$

$$\langle \nabla g(z''),h' \rangle \leq \bar{D}g_{\varepsilon(z';\tau)}(z',h',\tau) + (1-\alpha)\ \delta\hat{\varepsilon}$$

$$\leq -\delta\varepsilon(z';\tau) + (1-\alpha)\delta\hat{\varepsilon}$$

$$\leq -\alpha\delta\varepsilon(z';\tau)$$

$$\forall z'', z' \in B(z,\rho_4)$$
$$\forall h' \in S(z';\tau),\ \forall \tau \in [0,\tau_3] \qquad \text{(A41)}$$

(c) By continuity and Assumption 5 there exist a $\rho_5 \in (0,\rho_1]$ and a $\tau_4 \in (0,\tau_1]$ such that

$$\langle \nabla f^o(z''),h\rangle \leq \langle \bar{\nabla} f^o(z';\tau),h\rangle + (1-\alpha)\delta\hat{\varepsilon} \quad \forall z'',z' \in B(z,\rho_5)$$
$$\forall h \in S,\ \forall \tau \in [0,\tau_4] \qquad \text{(A42)}$$

If $\psi_o(z') = 0$, then

$$\langle \nabla f^o(z''),h'\rangle \leq -\delta\varepsilon(z';\tau) + (1-\alpha)\ \delta\hat{\varepsilon}$$

$$\leq -\alpha\delta\varepsilon(z';\tau) \qquad \forall z' \in B(z,\rho_5) \cap F$$
$$\forall z'' \in B(z,\rho_5),\ \forall h' \in S(z',\tau)$$
$$\forall \tau \in [0,\tau_4] \qquad \text{(A43)}$$

Let $\rho = \min\{\rho_2,\rho_4,\rho_5\}$ and $\bar{\tau} = \min\{\tau_2,\tau_3,\tau_4\}$ and we are done. ◻

<u>Lemma 5</u>. For all $z \in \mathbb{R}^n$ such that $\bar{\theta}_o^{-1}(z;0) < 0$ there exist a $\mu > 0$, $\rho > 0$, and a $\bar{\tau} > 0$ such that

$$f^o(z'') - f^o(z') \leq -\mu \qquad \forall z' \in B(z,\rho) \cap F$$
$$\forall z'' \in A(z';\tau)$$
$$\forall \tau \in (0,\bar{\tau}] \qquad \text{(A44)}$$

$$\psi(z'') - \psi(z') \leq -\mu \qquad \forall z' \in B(z,\rho) \cap F^c$$
$$\forall z'' \in A(z';\tau)$$
$$\forall \tau \in (0,\bar{\tau}] \qquad \text{(A45)}$$

Proof: Because the proof of this lemma is so similar to that of Lemma 2.3 we shall include only the case when $\psi(z) < 0$.

By Proposition 6 there exist a $\rho_o > 0$, $\hat{\varepsilon} > 0$, and a $\tau_o > 0$ such that $\varepsilon(z';\tau) \geq \hat{\varepsilon}$ for all $z' \in B(z,\rho_o)$ and for all $\tau \in (0,\tau_o]$.

Case 1. $\psi(z) < 0$.

By continuity there exists a $\rho_1 \in (0,\rho_o]$ such that $\psi(z') < 0$ for all $z' \in B(z,\rho_1)$; i.e. $B(z,\rho_1) \subset F$. By Proposition 6 there exist a $\bar{\rho} \in (0,\rho_1]$ and a $\bar{\tau} \in (0,\tau_o]$ such that $\langle \nabla f^o(z''),h' \rangle \leq -\alpha\delta\varepsilon(z';\tau)$, for all $h' \in S(z',\tau)$, for all $\tau \in (0,\bar{\tau}]$, and for all $z'',z' \in B(z,\bar{\rho})$. It follows from the definition of the directional derivative that for any $z',h \in \mathbb{R}^n$ and $\lambda \geq 0$,

$$f^o(z'+\lambda h) - f^o(z') = \int_0^\lambda \langle \nabla f^o(z'+sh),h \rangle \, ds \tag{A46}$$

For $z' \in B(z,\bar{\rho}/2)$, and for any $h' \in S(z',\tau)$, $\tau \in (0,\bar{\tau}]$, it follows that $z' + \lambda h' \in B(z,\bar{\rho})$ for all $\lambda \in [0, \frac{\bar{\rho}}{2\sqrt{n}}]$ because $\|h'\| \leq \sqrt{n}$. From (A46) and our choices of $\bar{\rho}$ and $\bar{\tau}$ we obtain

$$f^o(z'+\lambda h) - f^o(z') \leq -\lambda\alpha\delta\varepsilon(z';\tau) \qquad \forall z' \in B(z,\bar{\rho}/2), \; \forall h' \in S(z',\tau), \; \forall \lambda \in [0,\bar{\rho}/2\sqrt{n}] \tag{A47}$$

Because $z' + \lambda h' \in B(z,\bar{\rho})$ for all $\lambda \in [0,\bar{\rho}/2\sqrt{n}]$, $z' \in B(z,\bar{\rho}/2)$, and $h' \in S(z',\tau)$, $\tau \in (0,\bar{\tau}]$,

$$\psi(z'+\lambda h') \leq 0 \qquad \forall h' \in S(z',\tau), \quad \forall \lambda \in (0,\bar{\rho}/2\sqrt{n}], \quad \forall \tau \in (0,\bar{\tau}]. \tag{A48}$$

Let $\hat{\lambda} = \min\{\lambda_o,\bar{\rho}/2\sqrt{n}\}$; then let $\hat{k}(z) \in \mathbb{Z}_+$ be such that $\beta^{\hat{k}(z)} \leq \hat{\lambda} \leq \beta^{\hat{k}(z)-1}$. In steps 8 through 12 (assuming $\lambda \geq \tau\cdot\lambda_{min}$) of the algorithm the smallest integer $k(z') \in \mathbb{Z}_+$ is chosen such that $\lambda(z') \triangleq \lambda_o\beta^{k(z')}$ satisfies

$$f^o(z'+\lambda(z')h') - f^o(z') \leq -\lambda(z')\ \alpha\delta\varepsilon(z',\tau) \tag{A49}$$

$$\psi(z'+\lambda(z')h') \leq 0 \tag{A50}$$

if $z' \in F$, and $h' \in S(z',\tau)$. Comparing (A47) and (A48) with (A49) and (A50) we conclude that $k(z') \leq \hat{k}(z)$ for all $z' \in B(z,\bar{\rho}/2)$. Hence $-\rho^{k(z')} \leq -\beta^{k(z)}$ and we obtain

$$\begin{aligned} f^o(z'+\lambda(z')h') - f^o(z') &\leq -\beta^{\hat{k}(z)}\alpha\delta\varepsilon(z';\tau) \\ &\leq -\beta^{\hat{k}(z)}\alpha\delta\hat{\varepsilon} \end{aligned} \tag{A51}$$

$$\psi(z'+\lambda(z')h') \leq 0 \tag{A52}$$

for all $z' \in B(z,\bar{\rho}/2)$, $h' \in S(z',\tau)$, and for all $\tau \in (0,\bar{\tau}]$.

Since Subprocedure 1 maintains feasibility and does not increase the cost we obtain

$$f^o(z'') - f^o(z') \leq -\beta^{\hat{k}(z)}\alpha\delta\hat{\varepsilon} \tag{A53}$$

$$\psi(z'') \leq 0 \tag{A54}$$

for all $z' \in B(z,\bar{\rho}/2)$ where $z'' \triangleq S_1(z'+\lambda(z')h',\tau)$; i.e. $z'' \in A(z';\tau)$, $\tau \in (0,\bar{\tau}]$.

Case 2. $\phi(z) \geq 0$. This case is proved in the same manner as case 1 using the arguments contained in the proof of Lemma 2.3.

If we let $\rho \triangleq \bar{\rho}/2$ and $\mu \triangleq \beta^{\hat{k}(z)}\alpha\delta\hat{\varepsilon}$, we are done. ∎

References

[1] E. Polak and D. Q. Mayne, "An Algorithm for Optimization Problems with Functional Inequality Constraints," IEEE Trans. Auto. Contr. vol. AC-21, pp. 184-193, April 1976.

[2] E. Polak and R. Trahan, "An Algorithm for Computer Aided Design Problems," Proc. of the 1976 IEEE Conf. on Dec. and Contr., Dec. 1976, pp. 537-542.

[3] E. Polak, K. S. Pister, and D. Ray, "Optimal Design of Framed Structures Subjected to Earthquakes," Engineering Optimization, vol. 2, pp. 65-71. 1976.

[4] V. Zakian and U. Al-Naib, "Design of Dynamical and Control Systems by the Method of Inequalities," Proc. IEE, vol. 120, no. 11, pp. 1421-1427, Nov. 1973.

[5] E. Polak, "Computational Methods in Optimization: A Unified Approach," Academic Press, New York, 1971.

[6] G. Zoutendijk, "Methods of Feasible Directions," Elsevier, Amsterdam, 1970.

[7] R. Klessig and E. Polak, "A Method of Feasible Directions Using Function Approximations, with Applications to Min Max Problems," Journal of Mathematical Analysis and Applications, vol. 41, no. 3, pp. 583-602, March 1973.

[8] N. V. Banitchouk, V. M. Petrov, and R. L. Chernousko, "Numerical Solution of Problems with Variational Limits by the Method of Local Variations," Zh. Vychisl. Mat. Mat. Fiz., vol. 6, pp. 947-961, 1966.

[9] R. Hooke, and T. A. Jeeves, "Direct Search Solution of Numerical and Statistical Problems," J. Assoc. Comput. Mach., vol. 8, pp. 212-221, 1961.

[10] H. H. Rosenbrock, "An Automatic Method for Finding the Greatest or Least Value of a Function," Computer Journal, vol. 6, pp. 175-184, 1960.

[11] D. G. Luenberger, "Introduction to Linear and Nonlinear Programming," Addison-Wesley, Reading, Mass., 1973.

[12] M. Canon, C. Cullum, and E. Polak, "Theory of Optimal Control and Mathematical Programming," McGraw-Hill, New York, 1970.

[13] A. R. Curtis and J. K. Reid, "The Choice of Step Lengths When Using Differences to Approximate Jacobian Matrices," J. Inst. Math. and Applic., vol. 13, pp. 121-126, 1974.

[14] D. Ray, K. S. Pister and E. Polak, "Sensitivity Analysis for Hysteretic Dynamic Systems: Theory and Applications," Comp. Meth. in Appl. Mech. and Eng., vol. 14, pp. 179-208, 1978.

[15] C. Gonzaga, E. Polak and R. Trahan, "An Improved Algorithm for Optimization Problems with Functional Inequality Constraints," University of California, Electronics Research Laboratory Memorandum M78/56, September 1977.

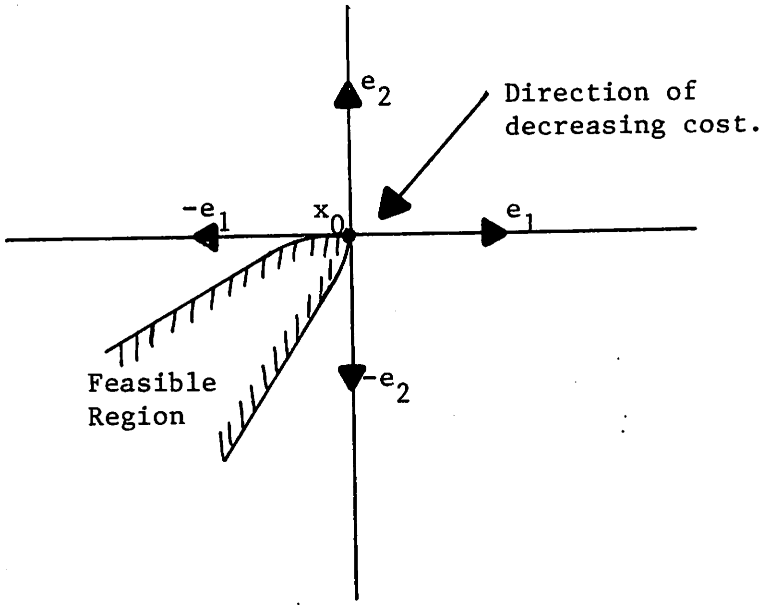

Fig. 1. A nonstationary point at which no feasible step exists along the coordinate directions.

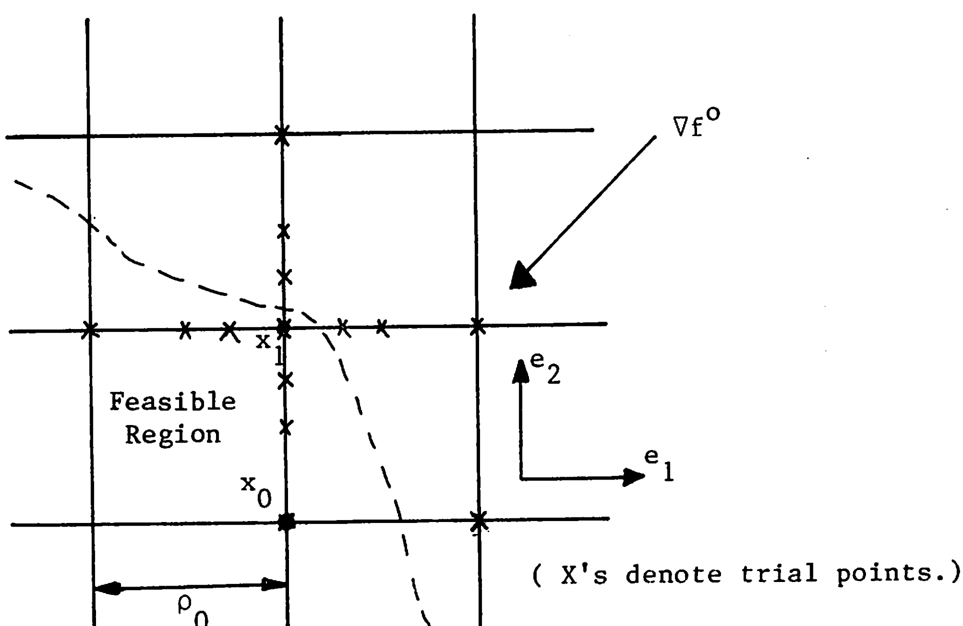

Fig. 2. Calculation of $S_1(x_0,\tau)$ with $\tau=\rho_0/4$. No feasible step of length greater than or equal to τ exists for which $f^o(x_1)$ can be decreased. Hence, $x_1=S_1(x_0,\tau)$.