AN IMPROVED ALGORITHM FOR OPTIMIZATION PROBLEMS WITH

FUNCTIONAL INEQUALITY CONSTRAINTS

by

C. Gonzaga, E. Polak, and R. Trahan

Memorandum No. UCB/ERL M78/56

8 September 1977

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

I. Introduction

Recently, Polak and Mayne [1] presented an algorithm for solving problems of the form

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

where $f^o:\mathbb{R}^n \to \mathbb{R}$ and $g^j:\mathbb{R}^n \to \mathbb{R}$, $j = 1,2,\ldots,p$ are continuously differentiable functions and $f^j:\mathbb{R}^n \to \mathbb{R}$, $j = 1,2,\ldots,m$, are functional constraints of the form

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

where $\phi^j:\mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$. It is assumed that $\phi^j(\cdot,\cdot)$ is continuous and that $\nabla_z\phi^j(\cdot,\cdot)$ is continuous for each $j = 1,2,\ldots,m$. The set Ω is a compact interval of the real line. As noted in [1,2,3,4], an important class of engineering design problems can be formulated in the form of (1).

The method in [1] is a phase I-phase II type feasible directions method (see [5]). Since the value of $f^j(z)$ cannot be computed exactly, the algorithm in [1] uses approximations to $f^j(z)$ given by $\max_{\omega \in \Omega_q} \phi^j(z,\omega)$, where Ω_q is a suitably constructed discrete subset of Ω. Also, in [1], in order to make the linear program, which computes the search direction, finite, the sets $\Omega_\varepsilon^j(z) \triangleq \{\omega \in \Omega \mid \phi^j(z,\omega) - \psi_o(z) \geq \varepsilon\}$, where $\psi_o(z) \triangleq \max\{g^j(z),\ j = 1,2,\ldots,p;\ f^j(z),\ j = 1,2,\ldots,m\}$, are also approximated discretely. Unfortunately, the particular choice of the discrete approximation to the sets $\Omega_\varepsilon^j(z)$, used in [1], forces the insertion of a very costly test into the algorithm. This test involves the computation of many inner products which increases both computer time and storage.

In this paper, we present an algorithm which uses a different discrete approximation to the sets $\Omega^j_\varepsilon(z)$, which has the great advantage over the one in [1] that the costly test in [1] need no longer be used. As a result, the new algorithm is much faster. The new algorithm also avoids a number of smaller shortcomings present in the algorithm in [1]. It uses a more satisfactory discretization rule for the approximations of the $f^j(z)$ as well as a better optimality function for the calculation of search directions. Both of these changes further contribute to its superiority over the algorithm in [1]. Although, from a theoretical point of view, the present algorithm does not appear to be all that different from the one in [1], the collective effect of all the changes results in very substantial practical differences, as can be seen from the experimental results in Appendix B, and hence should be of considerable interest to engineers in the area of computer aided design.

II. Definitions and Assumptions

We formalize our remarks about problem (1) by assuming the following hypothesis is true.

Assumption 1. $f^o(\cdot)$ and $g^j(\cdot)$, $j = 1,2,\dots,p$ are continuously differentiable; $\phi^j(\cdot,\cdot)$ and $\nabla_z\phi^j(\cdot,\cdot)$, $j = 1,2,\dots,m$ are continuous. ¤

Because we are using feasible directions type algorithms we introduce the concept of "ε-active" constraints. Given $z \in \mathbb{R}^n$, we define $\psi:\mathbb{R}^n \to \mathbb{R}$ by

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

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

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

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

We identify the set of points in Ω for which $\phi^j(z,\omega)$, $j \in \underline{m}$, is ε-active by defining

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

We assume the following hypotheses are true.

Assumption 2. For all $z \in \mathbb{R}^n$, for all $j \in \underline{m}$, $\Omega^j_0(z)$ is a finite set.

Assumption 3. For all $z \in \mathbb{R}^n$, $\varepsilon > 0$, and $j \in J^f_\varepsilon(z)$, $\Omega^j_\varepsilon(z)$ is the union of a finite number of disjoint intervals, $I^j_{\varepsilon,k}(z)$, $k = 1,2,\ldots,k^j_\varepsilon(z)$, possibly of zero length, i.e.

$$\Omega^j_\varepsilon(z) = \bigcup_{k \in \mathcal{K}^j_\varepsilon(z)} I^j_{\varepsilon,k}(z) \tag{8}$$

where $\mathcal{K}^j_\varepsilon(z) \triangleq \{1,2,\ldots,k^j_\varepsilon(z)\}$

Assumption 4. For all $z \in \mathbb{R}^n$, $\{\nabla_z\phi^j(z,\omega)$, $\omega \in \Omega^j_0(z)$, $j \in J^f_0(z)$; $\nabla g^j(z)$, $j \in J^g_0(z)\}$ is a set of positive linearly independent vectors.† ¤

---

†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 [6].

These three assumptions are identical to those in [1].

We now define two functions which could, theoretically, be used to generate feasible search directions. The first function defines a linear program (LP)(cf. [1]) and the second function defines a quadratic program (QP)(cf. [2]). For any $z \in \mathbb{R}^n$, $\varepsilon \geq 0$, define $\theta^1_\varepsilon : \mathbb{R}^n \to \mathbb{R}$ and $\theta^2_\varepsilon : \mathbb{R}^n \to \mathbb{R}$ by

$$\theta^1_\varepsilon(z) \triangleq \min_{h \in S} \max\{\langle \nabla f^o(z), h\rangle - \gamma\psi_o(z);$$
$$\langle \nabla g^j(z), h\rangle,\ j \in J^g_\varepsilon(z);\ \langle \nabla_z \phi^j(z,\omega), h\rangle,\ \omega \in \Omega^j_\varepsilon(z),\ j \in J^f_\varepsilon(z)\} \tag{9}$$

$$\theta^2_\varepsilon(z) \triangleq \min_{h \in \mathbb{R}^n} \{\tfrac{1}{2}\|h\|^2 + \max\{\langle \nabla f^o(z), h\rangle - \gamma\psi_o(z);$$
$$\langle \nabla g^j(z), h\rangle,\ j \in J^g_\varepsilon(z);\ \langle \nabla_z \phi^j(z,\omega), h\rangle,\ \omega \in \Omega^j_\varepsilon(z),\ j \in J^f_\varepsilon(z)\} \tag{10}$$

where $S \triangleq \{h \in \mathbb{R}^n \mid \|h\|_\infty \leq 1\}$ and $\gamma \geq 1$ is a constant. In [1] the constant γ was not used. We have found through our computational experience that $\gamma \geq 2$ gives best performance. Note that by increasing γ the effect of the cost function, f^o, is lessened whenever $\psi_o(z) > 0$.

The programs defined by (9) and (10) cannot, in general, be solved by conventional LP and QP computer codes because $\Omega^j_\varepsilon(z)$, $j \in \underline{m}$, may contain an infinite number of points. It is for this reason that we approximate $\Omega^j_\varepsilon(z)$ to make the LP or QP tractable while at the same time ensuring that a suitable feasible search direction is computed. In [1], Polak and Mayne approximates $\Omega^j_\varepsilon(z)$ by a discrete set of points given by the set of midpoints of each subinterval $I^j_{\varepsilon,k}(z)$. This approximation provides a low dimensional LP to be solved (or QP as in [2]). But, the direction vector computed with this approximation is not necessarily a satisfactory search direction and a test is necessary to ensure suitability of the vector. This test, Step 5 of Algorithms I and II in [1], involves the computation of inner products for all the points in $\Omega_\varepsilon^j(z)$. This causes two computational problems: (i) the set $\Omega_\varepsilon^j(z)$ has to be stored and (ii) the computation of inner products for all the points in $\Omega_\varepsilon^j(z)$ (even when Ω is discretized) could be very expensive. It is desirable, therefore, to approximate $\theta_\varepsilon^1(z)$ or $\theta_\varepsilon^2(z)$ in some other manner for which the test over all the points in $\Omega_\varepsilon^j(z)$ can be eliminated.

The new algorithm to be presented here is based on a new method of "approximating" $\theta_\varepsilon^1(z)$ and $\theta_\varepsilon^2(z)$. We define the "approximation" to $\Omega_\varepsilon^j(z)$ by

$$\tilde{\Omega}_\varepsilon^j(z) \triangleq \{\omega \in \Omega_\varepsilon^j(z) \,|\, \omega \text{ is a left local maximizer of } \phi^j(z,\cdot)\} \quad . \tag{11}$$

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

$$\phi^j(z,\omega) < \phi^j(z,\bar{\omega}) \qquad \forall \omega \in (\bar{\omega}-\mu,\bar{\omega}) \cap \Omega \tag{12}$$

$$\phi^j(z,\bar{\omega}) \geq \phi^j(z,\omega) \qquad \forall \omega \in (\bar{\omega},\bar{\omega}+\mu) \cap \Omega \tag{13}$$

We now define the functions $\tilde{\theta}_\varepsilon^1 : \mathbb{R}^n \to \mathbb{R}$ and $\tilde{\theta}_\varepsilon^2 : \mathbb{R}^n \to \mathbb{R}$ by

$$\begin{aligned} \tilde{\theta}_\varepsilon^1(z) \triangleq \min_{h \in S} \quad & \max\{\langle \nabla f^o(z),h\rangle - \gamma\psi_o(z); \\ & \langle \nabla g^j(z),h\rangle,\ j \in J_\varepsilon^g(z);\ \langle \nabla_z \phi^j(z,\omega),h\rangle,\ \omega \in \tilde{\Omega}_\varepsilon^j(z), \\ & j \in J_\varepsilon^f(z)\} \end{aligned} \tag{14}$$

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

$$\langle \nabla g^j(z),h\rangle,\ j \in (J^g_\varepsilon(z));\ \langle \nabla_z\phi^j(z,\omega),h\rangle,\ \omega \in \tilde{\Omega}^j_\varepsilon(z),\ j \in J^f_\varepsilon(z)\}\} \tag{15}$$

In order to ensure that $\tilde{\theta}^1_\varepsilon(z)$ and $\tilde{\theta}^2_\varepsilon(z)$ define a finite LP or QP respectively we must require that for all $j \in J^f_\varepsilon(z)$, $\tilde{\Omega}^j_\varepsilon(z)$ is a finite set. Therefore, we assume the following additional hypothesis to be true.

<u>Assumption 5</u>. For all $z \in \mathbb{R}^n$, $\varepsilon > 0$, and $j \in J^f_\varepsilon(z)$, $\tilde{\Omega}^j_\varepsilon(z)$ is a finite set. ¤

Note that this is a slightly stronger assumption than Assumption 3. In most practical cases, however, Assumption 5 should be satisfied.

The solutions of the programs defined by (14) and (15) will be denoted by $h^1_\varepsilon(z)$ and $h^2_\varepsilon(z)$ respectively. Although $\tilde{\theta}^2_\varepsilon(z)$ defines a QP it may be more efficient to solve its dual [7,8]

$$\tilde{\theta}^2_\varepsilon(z) = \max_{\mu \geqq 0} \{-\tfrac{1}{2}\|\mu^o\nabla f^o(z) + \sum_{j \in J^g_\varepsilon(z)} \mu^j\nabla g^j(z)$$

$$\sum_{j \in J^f_\varepsilon(z)} \sum_{\omega \in \tilde{\Omega}^j_\varepsilon(z)} \mu^{j,\omega}\nabla_z\phi^j(z,\omega)\|^2 - \gamma\mu^o\psi_o(z) \Big|$$

$$\mu^0 + \sum_{j \in J^g_\varepsilon(z)} \mu^j + \sum_{j \in J^f_\varepsilon(z)} \sum_{\omega \in \tilde{\Omega}^j_\varepsilon(z)} \mu^{j,\omega} = 1\} \tag{16}$$

which also defines a QP. The solution of the program defined by (16) will be denoted by $\mu_\varepsilon(z)$. This solution is related to $h^2_\varepsilon(z)$ by

$$h^2_\varepsilon(z) = -[\mu^o_\varepsilon(z)\nabla f^o(z) + \sum_{j \in J^g_\varepsilon(z)} \mu^j_\varepsilon(z)\nabla g^j(z)$$

$$+ \sum_{j \in J^f_\varepsilon(z)} \sum_{\omega \in \tilde{\Omega}^j_\varepsilon(z)]} \mu^{j,\omega}(z)\nabla_z\phi^j(z,\omega)) \tag{17}$$

III. An Algorithm Model

The conceptual algorithm to be presented here is based on an algorithm model which was given in [1] and later in [5]. This model is related to the abstract problem of finding a point in a subset, $\Delta \subset \mathbb{R}^n$, using a search function $A:\mathbb{R}^n \to 2^{\mathbb{R}^n}$. The set of points, Δ, are referred to as desirable points.

Algorithm Model

Data: $z_0 \in \mathbb{R}^n$.

Step 0: Set $i = 0$.

Step 1: If $z_i \in \Delta$, stop; else, compute a $z_{i+1} \in A(z_i)$.

Step 2: Set $i = i + 1$ and go to step 1. ⌑

In the following general convergence theorem we list the assumed properties of the map A. It will be shown later that our conceptual algorithm, for solving problem (1), possesses these properties.

Theorem 1. [1] Suppose the map $A:\mathbb{R}^n \to 2^{\mathbb{R}^n}$, with associated cost function $c:\mathbb{R}^n \to \mathbb{R}$, has the following properties.

1) There exists an $F \subset \mathbb{R}^n$ such that $c = c^1$ on F and $c = c^2$ on F^c, where $c^1, c^2:\mathbb{R}^n \to \mathbb{R}$ are continuous.

2) $A(F) \subset F$.

3) For all $z \in \mathbb{R}^n$ such that $z \notin \Delta$, there exist a $\rho > 0$ and a $\mu > 0$ such that

$$\text{a)}\quad c^1(z'') - c^1(z') \leq -\mu \qquad \forall z' \in B(z,\rho) \cap F,\ \forall z'' \in A(z') \tag{18}$$

$$\text{b)}\quad c^2(z'') - c^2(z') \leq -\mu \qquad \forall z' \in B(z,\rho) \cap F^c,\ \forall z'' \in A(z') \tag{19}$$

where $B(z,\rho) \triangleq \{z' \in \mathbb{R}^n | \|z'-z\| \leq \rho\}$. If the algorithm model constructs an infinite sequence $\{z_i\}_{i=0}^{\infty}$ then any accumulation point $\hat{z}$ of this sequence is desirable, i.e. $\hat{z} \in \Delta$.

The proof of this theorem is given in [1].

As discussed in [1], and to a greater extent in [5], the use of two cost functions allows the "Phase I" process of computing a feasible point to be combined with the "Phase II" process of computing a point in the desirable set, Δ.

IV. The Conceptual Algorithm

We now present a new conceptual algorithm.† This algorithm is not, in general, implementable, since it may not be possible to compute exactly, in a finite time, quantities such as $\max_{\omega \in \Omega} \phi^j(z,\omega)$ or $\tilde{\Omega}^j_\varepsilon(z)$, for $j \in \underline{m}$. This algorithm is, however, used as a prototype for the implementable algorithm to be presented in the next section.

Algorithm I

Data: $\alpha \in (0,1)$, $\beta \in (0,1)$, $\delta \in (0,1]$, $\gamma \geq 1$, $\varepsilon_0 \in (0,\infty)$, $0 < \varepsilon_1 << \varepsilon_0$, $z_0 \in \mathbb{R}^n$, $\pi \in \{1,2\}$, $M > 0$.

Step 0: Set $i = 0$.

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

Step 2: Compute $\tilde{\Omega}^j_\varepsilon(z_i)$.

Step 3: Compute $h^\pi_\varepsilon(z_i)$ and $\tilde{\theta}^\pi_\varepsilon(z_i)$.

†See 1.3 of [9] for a discussion of conceptual vs. implementable algorithms.

Step 4: If $\tilde{\theta}^{\pi}_{\varepsilon}(z_i) \leq -\delta\varepsilon$ (set $\varepsilon(z_i) = \varepsilon$)† go to step 6; else, set $\varepsilon = \varepsilon/2$ and go to step 5.

Step 5: If $\varepsilon \leq \varepsilon_1$ and $\theta^{\pi}_0(z_i) = 0$, stop; else, go to step 2.

Step 6: If $\psi_o(z_i) = 0$ compute the largest step size $\sigma_i = \beta^{\ell_i} \in (0,M]$ $(\ell_i \in \mathbb{Z}_+)^{\dagger\dagger}$ such that

$$f^o(z_i + \sigma_i h^{\pi}_{\varepsilon}(z_i)) - f^o(z_i) \leq -\alpha\sigma_i\delta\varepsilon \tag{20}$$

$$g^j(z_i + \sigma_i h^{\pi}_{\varepsilon}(z_i)) \leq 0 \quad \forall j \in \underline{p} \tag{21}$$

$$f^j(z_i + \sigma_i h^{\pi}_{\varepsilon}(z_i)) \leq 0 \quad \forall j \in \underline{m}. \tag{22}$$

If $\psi_o(z_i) > 0$ compute the largest step size $\sigma_i = \beta^{\ell_i} \in (0,M]$ $(\ell_i \in \mathbb{Z}_+)$ such that

$$\psi(z_i + \sigma_i h^{\pi}_{\varepsilon}(z_i)) - \psi(z_i) \leq -\alpha\sigma_i\delta\varepsilon \tag{23}$$

Step 7: Set $z_{i+1} = z_i + \sigma_i h^{\pi}_{\varepsilon}(z_i)$, $i = i+1$, and go to step 1 (2).††† ⌑

Note that the parameter π must be selected as a data parameter. If $\pi = 1$ is selected then an LP must be solved in step 3; otherwise, with $\pi = 2$, a QP must be solved.

We define F to be the feasible set for (1); i.e.

$$F \triangleq \{z \in \mathbb{R}^n | f^j(z) \leq 0,\ j \in \underline{m};\ g^j(z) \leq 0,\ j \in \underline{p}\} \tag{24}$$

Steps 1 through 7 define a map $A:\mathbb{R}^n \to 2^{\mathbb{R}^n}$ and step 6 provides the property that $A(F) \subset F$. We define $c^1 \triangleq f^o$ and $c^2 \triangleq \psi$ on F and F^c respectively.

---

†The construction of $\varepsilon(z_i)$ is for use in the proofs only.

†† $\mathbb{Z}_+$ denotes the non-negative integers.

†††This algorithm will also work if "go to step 2" is used in step 7. The proof of convergence is substantially more complicated so that we will consider only the case when ε is reset at each iteration.

In order to show that Algorithm I satisfies the assumptions of the convergence theorem of the previous section we require the following three lemmas. No proof is given for Lemma 1 since the proof requires only a slight modification to the proof of Proposition 1 in [1] to be valid here. The proofs for Lemmas 2 and 3 are contained in Appendix A.

Lemma 1. If $z \in F$ is optimal for (1) then $\theta_0^\pi(z) = \tilde{\theta}_0^\pi(z) = 0$ for $\pi \in \{1,2\}$. Furthermore, for all $z \in F^c$, $\theta_0^\pi(z) < 0$ for $\pi \in \{1,2\}$. ¤

Lemma 2. For all $z \in \mathbb{R}^n$ such that $\theta_0^\pi(z) < 0$, $\pi \in \{1,2\}$, there exist a $\rho > 0$, and an $\bar{\varepsilon} > 0$ such that

$$\varepsilon(z') \geq \bar{\varepsilon} \quad \forall z' \in \beta(z,\rho) \tag{25}$$

where $\varepsilon(z')$ is the value of ε constructed by steps 2 through 5 Algorithm I, with $z_i = z'$.

Lemma 3. For all $z \in \mathbb{R}^n$ such that $\theta_0^\pi(z) < 0$, $\pi \in \{1,2\}$, there exist a $\mu > 0$ and a $\rho > 0$ such that

$$f^o(z'') - f^o(z') \leq -\mu \qquad \begin{array}{l}\forall z' \in B(z,\rho) \cap F \\ \forall z'' \in A(z')\end{array} \tag{26}$$

$$\psi(z'') - \psi(z') \leq -\mu \qquad \begin{array}{l}\forall z' \in B(z,\rho) \cap F^c \\ \forall z'' \in A(z')\end{array} \tag{27}$$

¤

As a consequence of Lemma 1, we define the set of desirable points, Δ, as follows.

$$\Delta \triangleq \{z \in \mathbb{R}^n | \theta_0^\pi(z) = 0,\ \pi \in \{1,2\}\}^\dagger \tag{28}$$

---

†The zeros of $\theta_0^1(\cdot)$ and $\theta_0^2(\cdot)$ coincide; i.e., $\theta_0^1(z) = 0$ if and only if $\theta_0^2(z) = 0$. See [5] for a further discussion of other optimality functions and their use in feasible directions algorithms.

Because of Lemma 2 the algorithm cannot cycle between steps 2 and 4 indefinitely while halving ε. Hence, the map A is well-defined.

Theorem 2. If Algorithm I 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.

$\square$

Proof: We have defined $c^1 \triangleq f^o$ and $c^2 \triangleq \psi$ so that by Assumption 1, c^1 and c^2 are continuous. As previously stated the map A satisfies $A(F) \subset F$. Hence, assumptions 1 and 2 of Theorem 1 are satisfied. From Lemma 3 and the definition of the set of desirable points Δ, it is clear that assumption 3 of Theorem 1 is satisfied. Hence, we can apply Theorem 1 and we obtain the desired conclusion.

$\square$

V. The Implementable Algorithm

Since it is impossible to evaluate in finite time, $\max_{\omega \in \Omega} \phi^j(z,\omega)$ or $\tilde{\Omega}^j_\varepsilon(z)$, $j \in \underline{m}$, exactly, we have developed an implementable version of Algorithm I in which a piecewise linear approximation of each $\phi^j(z,\omega)$, for $j \in \underline{m}$, is used. The method is similar to the one used in [1].

Let $\Omega \triangleq [\omega_o,\omega_c]$ and $\Delta q \triangleq \frac{\omega_c-\omega_o}{q}$ for $q \in \mathbb{Z}_+$. Then let

$$\Omega_q \triangleq \{\omega \in \Omega \,|\, \omega = \omega_o + k\Delta q,\ k = 0,1,2,\ldots,q\} \tag{29}$$

The points in Ω_q will be referred to as mesh points. Given $z \in \mathbb{R}^n$, $j \in \underline{m}$, and $\omega = \lambda\bar{\omega} + (1-\lambda)(\omega+\Delta q)$ with $\lambda \in [0,1]$ and $\bar{\omega} \in \Omega_q$, define

$$\phi^j_q(z,\omega) \triangleq \lambda\phi^j(z,\bar{\omega}) + (1-\lambda)\phi^j(z,\bar{\omega}+\Delta q) \tag{30}$$

and

$$f^j_q(z) \triangleq \max_{\omega \in \Omega_q} \phi^j_q(z,\omega) \tag{31}$$

For any $z \in \mathbb{R}^n$ and $q \in \mathbb{Z}_+$, define $\psi_q : \mathbb{R}^n \to \mathbb{R}$ by

$$\psi_q(z) \triangleq \max\{g^j(z),\ j \in \underline{p};\ f_q^j(z),\ j \in \underline{m}\} \tag{32}$$

and then define $\psi_{q,o} : \mathbb{R}^n \to \mathbb{R}$ by

$$\psi_{q,o}(z) \triangleq \max\{0, \psi_q(z)\} \tag{33}$$

For any $z \in \mathbb{R}^n$, $q \in \mathbb{Z}_+$, and $\varepsilon \geq 0$, we define

$$\Omega_{q,\varepsilon}^j(z) \triangleq \{\omega \in \Omega_q \mid \phi_q^j(z,\omega) - \psi_{q,o}(z) \geq -\varepsilon\} \tag{34}$$

$$\tilde{\Omega}_{q,\varepsilon}^j(z) \triangleq \{\omega \in \Omega_{q,\varepsilon}^j(z) \mid \omega \text{ is a left local maximizer of } \phi_q^j(z,\cdot)\} \tag{35}$$

$$J_{q,\varepsilon}^f(z) \triangleq \{j \in \underline{m} \mid f_q^j(z) - \psi_{q,o}(z) \geq -\varepsilon\} \tag{36}$$

$$J_{q,\varepsilon}^g(z) \triangleq \{j \in \underline{p} \mid g^j(z) - \psi_{q,o}(z) \geq -\varepsilon\} \tag{37}$$

$$\begin{aligned}\theta_{q,\varepsilon}^1(z) \triangleq \min_{h \in S} \max\{&\langle \nabla f^o(z), h\rangle - \gamma\psi_o(z);\\ &\langle \nabla g^j(z), h\rangle,\ j \in J_{q,\varepsilon}^g(z);\\ &\langle \nabla_z \phi^j(z,\omega), h\rangle,\ \omega \in \Omega_{q,\varepsilon}^j(z),\ j \in J_{q,\varepsilon}^f(z)\}\end{aligned} \tag{38}$$

$$\begin{aligned}\tilde{\theta}_{q,\varepsilon}^1(z) \triangleq \min_{h \in S} \max\{&\langle \nabla f^o(z), h\rangle - \gamma\psi_o(z);\\ &\langle \nabla g^j(z), h\rangle,\ j \in J_{q,\varepsilon}^g(z);\ \langle \nabla_z \phi^j(z,\omega), h\rangle,\\ &\omega \in \tilde{\Omega}_{q,\varepsilon}^j(z),\ j \in J_{q,\varepsilon}^f(z)\}\end{aligned} \tag{39}$$

Let $\theta_{q,\varepsilon}^2(z)$ and $\tilde{\theta}_{q,\varepsilon}^2(z)$ denote the obvious modification to (10) and (15) respectively.

In the implementable algorithm to be presented, the conceptual algorithm will be applied to the "approximate" problem

$$\min\{f^o(z) \mid g^j(z) \leq 0,\ j \in \underline{p};\ f_q^j(z) \leq 0,\ j \in \underline{m}\} \tag{40}$$

We must, therefore, ensure that the quantities in (40) satisfy the assumptions for the conceptual algorithm. The only assumption which may be satisfied by the original problem (1), but not for (40), is Assumption 2. This is true since, for some $j \in \underline{m}$, $\Omega_0^j(z)$ may have two points which are adjacent mesh points in Ω_q. If this happens, then $\phi_q^j(z,\cdot)$ would be constant between these two points and a "zero-active flat" would occur. In this case, the approximate problem would not satisfy Assumption 2, since $\Omega_0^j(z)$ being finite implies that $\phi^j(z,\cdot)$ cannot have a "zero-active flat."

While we cannot ensure that Assumption 2 is satisfied by (40) for all $q \in \mathbb{Z}_+$, we have added a subloop in the implementable algorithm to eliminate any "zero-active flats" for any given $z_i \in \mathbb{R}^n$. This subloop decreases the spacing of the points in Ω_q until there are no two adjacent mesh points in $\Omega_{q,0}^j(z)$.

We can now state the implementable algorithm.

<u>Algorithm II</u>

<u>Data</u>: $\alpha \in (0,1)$, $\beta \in (0,1)$, $\delta > 0$, $\gamma \geq 1$, $\varepsilon_0 \in (0,\infty)$, $\mu_1 > 0$, $\mu_2 > 0$, $z_0 \in \mathbb{R}^n$, $q_0 \in \mathbb{Z}_+$, $M > 0$, $\pi \in \{1,2\}$.

<u>Step 0</u>: Set $i = 0$, $q = q_0$, $k = 0$.

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

<u>Step 2</u>: Compute $\tilde{\Omega}_{q,\varepsilon}^j(z_i)$ and $\tilde{\Omega}_{q,0}^j(z_i)$, $j \in \underline{m}$.

<u>Step 2'</u>: If $\tilde{\Omega}_{q,0}^j(z_i)$, $j \in \underline{m}$, contains two adjacent mesh points, set $q = q + 1$ and go to step 2; else, go to step 3.

<u>Step 3</u>: Compute $h_{q,\varepsilon}^{\pi}(z_i)$ and $\tilde{\theta}_{q,\varepsilon}^{\pi}(z_i)$.

Step 4: If $\tilde{\theta}^{\pi}_{q,\varepsilon}(z_i) \le -\delta\varepsilon$ go to step 6; else, go to step 5.

Step 5: If $\varepsilon \le \frac{\mu_1}{2^q}$ and $\psi_{q,o}(z_i) \le \frac{\mu_2}{2^q}$, set $q = q + 1$, $y_k = z_i$, $k = k + 1$, and go to step 1; else, set $\varepsilon = \varepsilon/2$ and go to step 2.

Step 6: If $\psi_{q,o}(z_i) = 0$, compute the largest step size $\sigma_i = \beta^{\ell_i} \in (0,M]$ $(\ell_i \in \mathbb{Z}_+)$ such that

$$f^o(z_i + \sigma_i h^{\pi}_{q,\varepsilon}(z_i)) - f^o(z_i) \le -\alpha\sigma_i\delta\varepsilon \tag{41}$$

$$g^j(z_i + \sigma_i h^{\pi}_{q,\varepsilon}(z_i)) \le 0 \quad \forall j \in \underline{p} \tag{42}$$

$$f^j_q(z_i + \sigma_i h^{\pi}_{q,\varepsilon}(z_i)) \le 0 \quad \forall j \in \underline{m} \tag{43}$$

If $\psi_{q,o}(z_i) > 0$, compute the largest step size $\sigma_i = \beta^{\ell_i} \in (0,M]$ $(\ell_i \in \mathbb{Z}_+)$ such that

$$\psi_q(z_i + \sigma_i h^{\pi}_{q,\varepsilon}(z_i)) - \psi_q(z_i) \le -\alpha\sigma_i\delta\varepsilon \tag{44}$$

Step 7: Set $z_{i+1} = z_i + \sigma_i h^{\pi}_{q,\varepsilon}(z_i)$, $i = i + 1$, and go to step 1 (2).

¤

In addition to Assumption 2, we require the following hypothesis to be true.

Assumption 6. There exists a $\hat{q} \in \mathbb{Z}_+$ such that for all $z \in \mathbb{R}^n$, and for all $q \ge \hat{q}$, each $\Omega^j_{q,0}(z)$, $j \in \underline{m}$, does not contain adjacent mesh points.

¤

This assumption ensures that there is a uniform minimum distance between the points of $\Omega^j_0(z)$, $j \in \underline{m}$, for all $z \in \mathbb{R}^n$. We now state a result which is an immediate consequence of Assumption 6.

Lemma 4. Algorithm II cannot cycle indefinitely between steps 2 and 2'.

¤

Because of Lemma 4, the only way that Algorithm II can jam up is to cycle between steps 2 and 5, while halving ε. Suppose the algorithm jams up at a point z_i. The cycling can occur if $\tilde{\theta}^{\pi}_{q_i,\varepsilon}(z_i) = 0$ for all $\varepsilon > 0$ and $\psi_{q_i,o}(z_i) > \frac{\mu_2}{2^{q_i}}$, where q_i is the value of q at iteration i. Since $\phi^j_{q_i}(z_i,\cdot)$, $j \in \underline{m}$, has no "zero-active flats," we can apply Lemma 2. If $\theta^{\pi}_{q_i,0}(z_i) < 0$ then by Lemma 2 there exists an $\bar{\varepsilon} > 0$ such that $\tilde{\theta}^{\pi}_{q_i,\varepsilon}(z_i) \leq -\bar{\varepsilon}$. Hence, the only way in which the algorithm can jam up is if $\theta^{\pi}_{q,0}(z_i) = 0$ <u>and</u> $\psi_{q,o}(z_i) > \frac{\mu_2}{2^{q_i}}$. Because $\Omega^j_{q,0}(z_i) \subset \Omega^j_0(z_i)$, $j \in \underline{m}$, we have, by Assumption 4, that if $\psi_{q,o}(z_i) > 0$ then $\theta^{\pi}_{q,0}(z_i) < 0$ which is a contradiction. We conclude that Algorithm II cannot jam up and, hence, it is well-defined.

We now state the main convergence result for Algorithm II. Note that if Algorithm II generates a sequence $\{z_i\}_{i=0}^{\infty}$ then Step 5 sifts out a subsequence $\{y_k\}$.

<u>Theorem 3</u>. If Algorithm II, with $\pi = 1$ or 2, constructs an infinite sequence $\{z_i\}_{i=0}^{\infty}$, then any accumulation point y^* of the sequence $\{y_k\}$ must be desirable; i.e. $y^* \in F$ and $\theta^{\pi}_0(y^*) = 0$.

<u>Proof</u>: We will first show that either $\{y_k\}$ is infinite or else $\{z_i\}_{i=0}^{\infty}$ has no accumulation points (which implies that $\{y_k\}$ can have no accumulation points). Suppose that $\{y_k\}$ is finite; i.e. there exist an $N_1 \in \mathbb{Z}_+$ and a $q \in \mathbb{Z}_+$ such that $\varepsilon(z_i) > \frac{\mu_k}{2^q}$ or $\psi_{q,o}(z_i) > \frac{\mu_2}{2^q}$ for all $i \geq N_1$ where $\varepsilon(z_i)$ is the value of ε constructed by steps 2 through 5 of the algorithm. Such a value of q exists because of Assumption 6.

Let $\hat{z}$ be an accumulation point of $\{z_i\}$; i.e. $z_i \xrightarrow{K} \hat{z}$ for some $K \subset \mathbb{Z}_+$. Because $\psi_{q,o}(\cdot)$ is continuous and, by Theorem 2, $\psi_{q,o}(\hat{z}) = 0$. there exists an $N_2 \geq N_1$ such that $\psi_{q,o}(z_i) \leq \frac{\mu_2}{2^q}$ for all $i \geq N_2$, $i \in K$. Because $\{\varepsilon(z_i)\}_{i\in K}$ is a sequence bounded away from zero there exists a subsequence $\{\varepsilon(z_i)\}_{i\in K'}$, $K' \subset K$, such that $\varepsilon(z_i) \xrightarrow{K'} \hat{\varepsilon} > 0$. Hence, there exists an $N_3 \geq N_1$ such that

$$\tilde{\theta}_{q,\varepsilon(z_i)}(z_i) \leq -\delta\varepsilon(z_i) \leq -\delta\frac{\hat{\varepsilon}}{2} \quad \forall i \geq N_3,\ i \in K' \tag{45}$$

By arguments similar to those used in proving the convergence of the conceptual algorithm it can be shown that there exists a $\sigma > 0$ and an $N_4 \geq N_3$ such that

$$f^o(z_{i+1}) - f^o(z_i) \leq -\sigma \qquad \forall z_i \in F_q,\ i \in K'\ i \geq N_4 \tag{46}$$

$$\psi_q(z_{i+1}) - \psi_q(z_i) \leq -\sigma \qquad \forall z_i \in F_q^c,\ i \in K'\ i \geq N_4 \tag{47}$$

where $F_q \triangleq \{z \in \mathbb{R}^n | g^j(z) \leq 0,\ j \in \underline{p};\ f_q^j(z) \leq 0,\ j \in \underline{m}\}$. Now consider two cases: (i) $z_i \in F_q^c$ for all $i \geq N_4$. For any two successive indices $i, j \in K'$, $j > i \geq N_4$ we have

$$\begin{aligned}\psi_q(z_j) - \psi_q(z_i) = \psi_q(z_j) - \psi_q(z_{j-1}) + \psi_q(z_{j-1}) - \\ \cdots + \psi_q(z_{i+1}) - \psi_q(z_i) \leq -\sigma\end{aligned} \tag{48}$$

Hence, $\{\psi_q(z_i)\}_{i\in K'}$ is not Cauchy which is a contradiction since, by continuity, $\psi_q(z_i) \xrightarrow{K'} \psi_q(\hat{z})$. (ii) If $z_{\hat{i}} \in F_q$ for some $\hat{i} \geq N_4$, then $z_i \in F_q$ for all $i \geq \hat{i}$. For any $i, j \in K'$, $j > i \geq \hat{i}$, we have

$$\begin{aligned}f^o(z_j) - f^o(z_i) = f^o(z_j) - f^o(z_{j-1}) + f^o(z_{j-1}) - \\ \cdots + f^o(z_{i+1}) - f^o(z_i) \leq -\sigma\end{aligned} \tag{49}$$

We again have a contradiction since $f^o(z_i) \overset{K'}{\to} f^o(\hat{z})$. Thus, if $\{y_k\}$ is finite, then $\{z_i\}$ can have no accumulation points.

Let y^* be any accumulation point of $\{y_k\}_{k=0}^{\infty}$; i.e. $y_k \overset{\overline{K}}{\to} y^*$ for some $\overline{K} \subset \mathbb{Z}_+$. Let $\varepsilon(y_k)$ and q_k be the corresponding values of ε and q in the algorithm when y_k is constructed. We then obtain

$$
\begin{aligned}
-\delta \frac{\mu_1}{2^{q_k}} &\leq -\delta\varepsilon(y_k) \\
&< \tilde{\theta}^{\pi}_{q_k,\varepsilon(y_k)}(y_k) \\
&= \min_{h\in S} \max\{\langle \nabla f^o(y_k), h\rangle - \gamma\psi_{q_k,o}(y_k); \\
&\quad \langle \nabla g^j(y_k), h\rangle,\ j \in J^g_{q,\varepsilon(y_k)}(y_k); \\
&\quad \langle \nabla_z \phi^j(z,\omega), h\rangle,\ \omega \in \tilde{\Omega}^j_{q_k,\varepsilon(y_k)}(y_k),\ j \in J^f_{q,\varepsilon(y_k)}(y_k)\} \\
&\leq \min_{h\in S} \max\{\langle \nabla f^o(y_k), h\rangle - \gamma\psi_{q_k,o}(y_k);
\end{aligned}
$$

$$
\gamma[\psi_o(y_k) - \psi_{q_k,o}(y_k)] + \langle \nabla g^j(y_k), h\rangle,\ j \in J^g_{q,\varepsilon(y_k)}(y_k);
$$

$$
\gamma[\psi_o(y_k) - \psi_{q_k,o}(y_k)] + \langle \nabla_z \phi^j(z,\omega), h\rangle,\ \omega \in \Omega^j_{\varepsilon(y_k)}(y_k),\ j \in J^f_{\varepsilon(y_k)}(y_k)\}
$$

$$
= \theta^{\pi}_{\varepsilon(y_k)}(y_k) + \gamma[\psi_o(y_k) - \psi_{q_k,o}(y_k)]^{+} \tag{50}
$$

where we have used the facts that $\tilde{\Omega}^j_{q_k,\varepsilon(y_k)}(y_k) \subset \Omega^j_{\varepsilon(y_k)}(y_k)$, $j \in \underline{m}$, and $\psi_{q_k,o}(y_k) \leq \psi_o(y_k)$, for all k. Since $\psi_{q_k,o}(y_k) \to 0$ as $k \to \infty$ and $|\psi_{q_k,o}(y_k) - \psi_o(y_k)| \to 0$ as $k \to \infty$, $k \in \overline{K}$, it follows from (50) that

$$
\overline{\lim_{k\in\overline{K}}}\ \theta^{\pi}_{\varepsilon(y_k)}(y_k) = 0 \tag{51}
$$

+We have used the expansion for π=1 only. The corresponding result for π=2 should be obvious.

It now follows from the upper semicontinuity of the function $\theta_{\varepsilon}^{\pi}(y)$ in both ε and y (at $\varepsilon=0$), that

$$\overline{\lim_{k\in K}}\, \theta^{\pi}_{\varepsilon(y_k)}(y_k) \leq \theta^{\pi}_{0}(y^*) \leq 0 \tag{52}$$

From (51) and (52) we conclude that $\theta_0^{\pi}(y^*) = 0$. From the discussion above, it is clear that $\psi_0(y^*) = 0$ and, hence, $y^* \in F$. ¤

VI. Conclusions

We have presented here a new algorithm that is a substantial improvement over the algorithm in [1] for solving problems with functional inequality constraints. This new algorithm uses a different method in approximating the functional constraints. The use of ε-active local maxima points in Ω, for computing search directions, allows the deletion of a test, which is costly in terms of computation time and storage, from the algorithm in [1]. Also, a new discretization rule is given since the one in [1] will not always be satisfactory.

We have also included a new optimality function which gives rise to a different method of computing search directions. This optimality function gives rise to a better scaled search direction vector and it provides a faster speed of convergence when compared to the method used in [1].

It is hoped that this improved algorithm will provide better computational results when used to solve computer aided design problems. Our computational results have thus far supported this optimism.

ACKNOWLEDGMENT

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

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, 1976, pp. 65-71.
4. V. Zakian and U. Al-Naib, "Design of Dynamical and Centrol Systems by the Method of Inequalities," Proc. IEE, vol. 120, no. 11, Nov. 1973, pp. 1421-1427.
5. E. Polak, R. Trahan, and D.Q. Mayne, "Combined Phase I, Phase II Methods of Feasible Directions," Electronics Research Laboratory Memo No. UCB/ERL M77/39, University of California, Berkeley, May 31, 1977.
6. M. Canon, C. Cullum, E. Polak, "Theory of Optimal Control and Mathematical Programming," New York: McGraw-Hill, 1970.
7. O.L. Mangasarian, "Nonlinear Programming" New York: McGraw-Hill, 1969.
8. O. Pironneau and E. Polak, "Rate of Convergence of a Class of Methods of Feasible Directions," SIAM J. Numer. Anal. vol. 10, March 1973, pp. 161-174.
9. E. Polak, "Computational Methods in Optimization: A Unified Approach," Academic Press, New York, 1971.
10. J. Danskin, "The Theory of Max-Min," Springer-Verlag, Berlin, 1967.

11. C. T. Chen, "Analysis and Synthesis of Linear Control Systems," Holt, Rinehart and Winston, New York, 1975.

Appendix A. Proofs for §§IV.

For the sake of conciseness it will be assumed that m=1 and p=1. The superscripts on f^j and g^j, etc. will be deleted. Of course, all the results hold for the more general case. The proofs will also be given only for $\pi=1$, but they require only slight modifications to hold for $\pi=2$.

Before we prove Lemma 2, we shall prove several preliminary results. For notational convenience, we define

$$Df(z,h) \triangleq \max_{\omega \in \Omega(z)} \langle \nabla_z \phi(z,\omega), h \rangle \tag{A1}$$

where $\Omega(z) \triangleq \{\omega \in \Omega | f(z) = \phi(z,\omega)\}$. It was shown by Danskin [10] that $Df(z,h)$ is the directional derivative of $f(z)$ in the direction h. We also define

$$\tilde{D}f_\varepsilon(z,h) \triangleq \begin{cases} \max\limits_{\omega \in \tilde{\Omega}_\varepsilon(z)} \langle \nabla_z \phi(z,\omega), h \rangle & \text{if } f(z) - \psi_o(z) \geq -\varepsilon \\ -\infty & \text{otherwise} \end{cases} \tag{A2}$$

$$Dg_\varepsilon(z,h) \triangleq \begin{cases} \langle \nabla g(z), h \rangle & \text{if } g(z) - \psi_o(z) \geq -\varepsilon \\ -\infty & \text{otherwise} \end{cases} \tag{A3}$$

Proposition 1. For any $z \in \mathbb{R}^n$ such that $\Omega_0(z) \neq \phi$, and for any $\mu > 0$, there exist a $\rho > 0$, and an $\bar{\varepsilon} > 0$, such that for all $z' \in B(z,\rho)$, and $\varepsilon \in [0,\bar{\varepsilon}]$,

(a) $$\Omega_\varepsilon(z') \subset N_\mu(\Omega_0(z))^+ \tag{A4}$$

+We define the neighborhood of a subset $\tilde{\Omega} \subset \Omega$ of radius $\mu > 0$ by

$$N_\mu(\tilde{\Omega}) \triangleq \{\omega \in \Omega | \; |\omega - \bar{\omega}| \leq \mu \text{ for some } \bar{\omega} \in \tilde{\Omega}\}$$

(b) $\tilde{D}f_{\varepsilon}(z',h) \leq Df(z,h) + \mu, \ \forall h \in S$ (A5)

<u>Proof</u>: (a) Given $z \in \mathbb{R}^n$ such that $\Omega_0(z) \neq \phi$ and $\mu > 0$, suppose the conclusion is not true. Then there exist sequences $\{z_i\}$ and $\{\varepsilon_i\}$ such that $z_i \to z$ and $\varepsilon_i \to 0$ with $\Omega_{\varepsilon_i}(z_i) \not\subset N_\mu(\Omega_0(z))$. This implies that there exist points, $\omega_i \in \Omega_{\varepsilon_i}(z_i)$, $i = 1,2,\cdots$, such that $\omega_i \notin N_\mu(\Omega_0(z))$ for all $i = 1,2,\cdots$. Since Ω is compact, there exists an $\bar{\omega} \in \Omega$ such that $\omega_i \xrightarrow{K} \bar{\omega}$ where $K \subset Z_+$. By the definition of ω_i, $\phi(z_i,\omega_i) \geq \psi_o(z_i) - \varepsilon_i$. Because $\phi(\cdot,\cdot)$ and $\psi_o(\cdot)$ are continuous, $\phi(z,\bar{\omega}) \geq \psi_o(z)$ and hence $\bar{\omega} \in \Omega_0(z)$. Consequently, $\omega_i \in N_\mu(\Omega_0(z))$ for sufficiently large $i \in K$, which is a contradiction. Thus, part (a) must be true. (b) By the continuity of $\nabla_z\phi(\cdot,\cdot)$ and compactness of $\Omega_0(z)$ and S, there exist a $\rho_1 > 0$ and a $\mu_1 > 0$ such that

$$\max_{\omega \in N_{\mu_1}(\Omega_0(z))} \langle \nabla_z\phi(z',\omega),h \rangle - Df(z,h) \leq \mu$$

$$\forall h \in S, \quad \forall z' \in B(z,\rho_1) \tag{A6}$$

From part (a) there exist a $\rho \in (0,\rho_1]$ and an $\bar{\varepsilon} > 0$ such that $\tilde{\Omega}_\varepsilon(z') \subset N_{\mu_1}(\Omega_0(z))$, for all $\varepsilon \in [0,\bar{\varepsilon}]$ and for all $z' \in B(z,\rho)$. From (A6) we obtain

$$\max_{\omega \in \tilde{\Omega}_\varepsilon(z')} \langle \nabla_z\phi(z',\omega),h \rangle \leq Df(z,h) + \mu$$

$$\forall h \in S, \quad \forall z' \in B(z,\rho) \tag{A7}$$

From the definition of $\tilde{D}f_\varepsilon(z',h)$ and (A7), we obtain (A5), the desired result. ◻

<u>Proposition 2</u>. For any $z \in \mathbb{R}^n$ such that $\Omega_0(z) \neq \phi$, and for any $\varepsilon > 0$, $\mu > 0$, there exists a $\rho > 0$ such that for all $z' \in B(z,\rho)$

(a) $$\Omega_0(z) \subset N_\mu(\tilde{\Omega}_\varepsilon(z')) \tag{A8}$$

(b) $$Df(z,h) \leq \tilde{D}f_\varepsilon\,(z',h) + \mu, \quad \forall h \in S. \tag{A9}$$

Proof: (a) Let $\omega^* \in \Omega_0(z)$. Since $\Omega_0(z)$ is finite, there exists a $\mu^* \in (0,\mu]$ such that ω^* is the unique maximizer of $\phi(z,\cdot)$ on the interval $N_{\mu^*}(\omega^*)$. For all $z' \in \mathbb{R}^n$, define

$$\psi_{\omega^*}(z') \triangleq \max_{\omega \in N_{\mu^*}(\omega^*)} \phi(z',\omega) \tag{A10}$$

and

$$\Omega_{\omega^*}(z') \triangleq \{\omega \in N_{\mu^*}(\omega^*) \,|\, \phi(z',\omega) = \psi_{\omega^*}(z')\} \tag{A11}$$

By Proposition 1, there exist a $\rho_1 > 0$ and an $\bar{\varepsilon} \in (0,\varepsilon]$ such that

$$\Omega_{\bar{\varepsilon}}(z') \subset N_{\frac{\mu^*}{2}}(\Omega_0(z)), \quad \forall z' \in B(z,\rho_1) \tag{A12}$$

By the continuity of $\psi_{\omega^*}(\cdot)$ and $\psi_o(\cdot)$, there exists a $\rho_2 \in (0,\rho_1]$ such that for all $z' \in B(z,\rho_2)$,

$$\psi_{\omega^*}(z') - \psi_{\omega^*}(z) \geq -\frac{\bar{\varepsilon}}{2} \tag{A13}$$

$$-\psi_o(z') + \psi_o(z) \geq -\frac{\bar{\varepsilon}}{2} \tag{A14}$$

Since $\psi_{\omega^*}(z) = \psi_o(z)$ (because $\omega^* \in \Omega_0(z)$) we obtain

$$\psi_{\omega^*}(z') - \psi_o(z') \geq -\bar{\varepsilon}, \quad \forall z' \in B(z,\rho_2) \tag{A15}$$

From (A12) and (A15) we conclude that $\Omega_{\omega^*}(z') \subset \Omega_{\overline{\varepsilon}}(z') \subset N_{\frac{\mu^*}{2}}(\Omega_0(z))$ for all $z' \in B(z,\rho_2)$. But this implies that $\Omega_{\omega^*}(z') \subset N_{\frac{\mu^*}{2}}(\omega^*)$ for all $z' \in B(z,\rho_2)$; i.e., $\Omega_{\omega^*}(z')$ consists of unconstrained local maximizers of $\phi(z',\omega)$ over $N_{\mu^*}(\omega^*)$. Consequently, the points in $\Omega_{\omega^*}(z')$ are local maximizers of $\phi(z',\cdot)$ over Ω. From (A15) we obtain

$$\phi(z',\omega(z')) - \psi_o(z') \geq -\overline{\varepsilon} \geq -\varepsilon \tag{A16}$$

where $\omega(z') \triangleq \min\{\omega \mid \omega \in \Omega_{\omega^*}(z')\}$, (leftmost maximizer). Thus, $\omega(z') \in \tilde{\Omega}_\varepsilon(z')$. Also, $\omega^* \in N_\mu(\tilde{\Omega}_\varepsilon(z'))$ because $|\omega(z') - \omega^*| < \mu^* \leq \mu$.

The above argument can be repeated for each $\omega^* \in \Omega_0(z)$. Let $\rho_2(\omega^*)$ be the associated radius of the ball about z for which the argument holds. Then let $\rho = \min_{\omega^* \in \Omega_0(z)} \{\rho_2(\omega^*)\}$ and we are done with part (a).

(b) Because $\nabla\phi(\cdot,\cdot)$ is continuous, and $\Omega_0(z)$ is a finite set, and S is compact, there exist a $\mu_1 > 0$ and a $\rho_1 > 0$ such that for any $\omega \in \Omega_0(z)$

$$\begin{aligned} &\langle \nabla_z\phi(z,\omega),h \rangle - \langle \nabla_z\phi(z',\omega'),h \rangle \leq \mu \\ &\qquad \forall h \in S, \quad \forall |\omega-\omega'| \leq \mu_1, \quad \forall z' \in B(z,\rho_1) \end{aligned} \tag{A17}$$

We then obtain

$$\begin{aligned} Df(z,h) &= \max_{\omega \in \Omega_0(z)} \langle \nabla_z\phi(z,\omega),h \rangle \\ &\leq \max_{\omega' \in \tilde{\Omega}} \langle \nabla_z\phi(z',\omega'),h \rangle + \mu \\ &\qquad \forall h \in S, \quad \forall z' \in B(z,\rho_1) \end{aligned} \tag{A18}$$

where $\tilde{\Omega}$ is any set such that $\Omega_0(z) \subset N_{\mu_1}(\tilde{\Omega})$. By part (a), there exists a $\rho \in (0,\rho_1]$ such that $\Omega_0(z) \subset N_{\mu_1}(\tilde{\Omega}_\varepsilon(z'))$ for all $z' \in B(z,\rho)$. Therefore,

$$Df(z,h) \leq \max_{\omega' \in \tilde{\Omega}_\varepsilon(z')} \langle \nabla_z \phi(z',\omega'),h \rangle + \mu$$

$$= \tilde{D}f_\varepsilon(z',h) + \mu$$

$$\forall h \in S, \quad \forall z' \in B(z,\rho) \tag{A19}$$

□

The following result is a corollary to Propositions 1 and 2.

Corollary 1. For any $z \in \mathbb{R}^n$ such that $\Omega_0(z) \neq \phi$, and for any $\varepsilon > 0$, $\mu > 0$, there exists a $\rho > 0$ such that for all z', $z'' \in B(z,\rho)$,

(a) $$\Omega(z'') \subset N_\mu(\tilde{\Omega}_\varepsilon(z')) \tag{A20}$$

(b) $$Df(z'',h) \leq \tilde{D}f_\varepsilon(z',h) + \mu, \quad \forall h \in S \tag{A21}$$

Proof: Let $\varepsilon > 0$ and $\mu > 0$ be arbitrary. Then by Proposition 1, there exist a $\rho_1 > 0$, and an $\bar{\varepsilon} \in (0,\varepsilon]$, such that for all $z'' \in B(z,\rho_1)$ and $\varepsilon' \in [0,\bar{\varepsilon}]$

$$\Omega_{\varepsilon'}(z'') \subset N_{\frac{\mu}{2}}(\Omega_0(z)) \tag{A22}$$

and

$$\tilde{D}f_{\varepsilon'}(z'',h) \leq Df(z,h) + \frac{\mu}{2}, \quad \forall h \in S \tag{A23}$$

By Proposition 2, there exists a $\rho_2 \in (0,\rho_1]$ such that for all $z' \in B(z,\rho_2)$

$$\Omega_0(z) \subset N_{\frac{\mu}{2}}(\tilde{\Omega}_\varepsilon(z')) \tag{A24}$$

and

$$Df(z,h) \leq \tilde{D}f_{\varepsilon}(z',h) + \frac{\mu}{2}, \quad \forall h \in S \tag{A25}$$

Since $f(\cdot)$ and $\psi_0(\cdot)$ are continuous and $f(z) = \psi_0(z)$ there exists a $\rho_3 \in (0,\rho_2]$ such that $f(z'') - \psi_0(z'') \geq -\bar{\varepsilon}$ for all $z'' \in B(z,\rho_3)$. Thus, $\Omega(z'') \subset \Omega_{\bar{\varepsilon}}(z'')$ and we obtain part (a) by combining (A22) and (A24). Similarly, because $\Omega(z'') \subset \tilde{\Omega}_{\bar{\varepsilon}}(z'')$, we have that $Df(z'',h) \leq \tilde{D}f_{\bar{\varepsilon}}(z'',h)$ and therefore part (b) follows from (A23) and (A25). ▫

Lemma 2. For all $z \in \mathbb{R}^n$ such that $\theta_0^{\pi}(z) < 0$, $\pi \in \{1,2\}$, there exist a $\rho > 0$, and an $\bar{\varepsilon} > 0$ such that

$$\varepsilon(z') \geq \bar{\varepsilon} \qquad \forall z' \in B(z,\rho) \tag{A26}$$

where $\varepsilon(z')$ is the value of ε constructed by steps 2 through 5 of Algorithm I with $z_i = z'$.

Proof: Let $\mu > 0$ be such that $\theta_0(z) \leq -\mu < 0$.+ Because $\Omega_0(z) = \tilde{\Omega}_0(z)$, it follows that

$$\theta_0(z) = \tilde{\theta}_0(z) = \min_{h \in S} \max\{\langle \nabla f^o(z),h \rangle - \gamma\psi_0(z); \; Dg_0(z,h); \; \tilde{D}f_0(z,h)\} \tag{A27}$$

We now consider 3 cases.

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

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

+The proof will be valid for π=1 only. We drop the superscript on $\theta_{\varepsilon}(z)$, etc.

$z' \in B(z,\tilde{\rho})$,

$$\psi(z') < \frac{\psi(z)}{2} \tag{A28}$$

Let $\tilde{\varepsilon} \triangleq -\frac{\psi(z)}{2}$; then we obtain

$$g(z') - \psi_0(z') < -\varepsilon$$

$$f(z') - \psi_0(z') < -\varepsilon \qquad \forall z' \in B(z,\tilde{\rho}), \quad \forall\, \varepsilon \in [0,\tilde{\varepsilon}] \tag{A29}$$

Consequently,

$$Dg_\varepsilon(z',h) = \tilde{D}f_\varepsilon(z',h) = -\infty$$

$$\forall h \in S, \quad \forall z' \in B(z,\tilde{\rho}), \quad \forall \varepsilon \in [0,\tilde{\varepsilon}]. \tag{A30}$$

Case 2. $\psi(z) \geq 0$, $\Omega_0(z) = \phi$.

Because $\Omega_0(z) = \phi$, there exists an $\tilde{\varepsilon} > 0$ defined by $\tilde{\varepsilon} \triangleq -\frac{1}{2}$ $(f(z) - \psi_o(z))$. By the continuity of $f(\cdot)$ and $\psi_o(\cdot)$, there exists a $\rho_1 > 0$ such that $f(z') - \psi_o(z') < -\varepsilon$ for all $z' \in B(z,\rho_1)$ and for all $\varepsilon \in [0,\tilde{\varepsilon}]$. By the continuity of $\nabla g(\cdot)$, there exists a $\tilde{\rho} \in (0,\rho_1]$ such that

$$\langle \nabla g(z'),h \rangle \leq \langle \nabla g(z),h \rangle + \frac{\mu}{2}$$

$$\forall h \in S, \quad \forall z' \in B(z,\tilde{\rho}) \tag{A31}$$

Consequently,

$$Dg_\varepsilon(z',h) \leq Dg_0(z,h) + \frac{\mu}{2}$$

$$\tilde{D}f_\varepsilon(z',h) = \tilde{D}f_0(z,h) = -\infty$$

$$\forall h \in S, \quad \forall z' \in B(z,\tilde{\rho}), \quad \forall \varepsilon \in [0,\tilde{\varepsilon}] \tag{A32}$$

Case 3. $\psi(z) \geq 0$, $\Omega_0(z) \neq \phi$.

By Proposition 1, there exist a $\rho_1 > 0$ and $\varepsilon_1 > 0$ such that

$$\tilde{D}f_\varepsilon(z',h) \leq \tilde{D}f_0(z,h) + \frac{\mu}{2}$$
$$\forall h \in S, \quad \forall z' \in B(z,\rho_1), \quad \forall \varepsilon \in [0,\tilde{\varepsilon}] \tag{A33}$$

where we have used the fact that $Df(z,h) = \tilde{D}f_0(z,h)$. By the same arguments as in the two previous cases, there exist a $\tilde{\rho} \in (0,\rho_1]$ and $\tilde{\varepsilon} \in (0,\varepsilon_1]$ such that

$$Dg_\varepsilon(z',h) \leq Dg_0(z,h) + \frac{\mu}{2}$$
$$\forall h \in S, \quad \forall z' \in B(z,\tilde{\rho}), \quad \forall \varepsilon \in [0,\tilde{\varepsilon}] \tag{A34}$$

This completes the third case.

By the continuity of $\nabla f^o(\cdot)$ and $\psi_o(\cdot)$, there exists a $\rho \in (0,\tilde{\rho}]$, (using the appropriate $\tilde{\rho}$ from case 1, 2, or 3 above), such that

$$\langle \nabla f^o(z'),h \rangle - \gamma\psi_o(z') \leq \langle \nabla f^o(z),h \rangle - \gamma\psi_o(z) + \frac{\mu}{2}$$
$$\forall h \in S, \quad \forall z' \in B(z,\rho) \tag{A35}$$

Using $\tilde{\varepsilon}$ from case 1, 2, or 3, as appropriate, we obtain,

$$\max\{\langle \nabla f^o(z'),h \rangle - \gamma\psi_o(z');\ Dg_\varepsilon(z',h);\ \tilde{D}f_\varepsilon(z,h)\}$$
$$\leq \max\{\langle \nabla f^o(z),h \rangle - \gamma\psi_o(z);\ Dg_0(z,h);\ \tilde{D}f_0(z,h)\}\ \frac{\mu}{2}$$
$$\forall h \in S, \quad \forall z' \in B(z,\rho), \quad \forall \varepsilon \in [0,\tilde{\varepsilon}] \tag{A36}$$

Because (A36) holds for all $h \in S$, it follows that

$$\tilde{\theta}_\varepsilon(z') \leq \theta_0(z) + \frac{\mu}{2} \leq -\frac{\mu}{2}$$
$$\forall z' \in B(z,\rho), \quad \forall \varepsilon \in [0,\tilde{\varepsilon}] \tag{A37}$$

Let $\hat{j}(z) \in \mathbb{Z}_+$ be an integer such that $\hat{\varepsilon} \overset{\Delta}{=} \varepsilon_0 2^{-\hat{j}(z)} \leq \min\{\tilde{\varepsilon}, \frac{\mu}{2\delta}\}$. Then, from (A37) we obtain

$$\tilde{\theta}_{\hat{\varepsilon}}(z') \leq -\delta\hat{\varepsilon}, \quad \forall z' \in B(z,\rho) \tag{A38}$$

The algorithm constructs $\varepsilon(z') = \varepsilon_0 2^{-j(z')}$ where $j(z')$ is the smallest nonnegative integer such that $\tilde{\theta}_{\varepsilon(z')}(z') \leq -\delta\varepsilon(z')$. In view of (A38), we conclude that $\hat{j}(z) \geq j(z')$. Hence, $\varepsilon(z') \geq \hat{\varepsilon}$ for all $z' \in B(z,\rho)$.

¤

<u>Proposition 3</u>. For all $z \in \mathbb{R}^n$ such that $\theta_0(z) < 0$, there exists a $\rho > 0$ such that

(a) $Df(z'',h') \leq -\alpha\delta\varepsilon(z')$+ if $f(z) = \psi_o(z)$

$$\forall z'', z' \in B(z,\rho), \quad \forall h' \in S(z') \tag{A38}$$

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

$$\forall z'', z' \in B(z,\rho), \quad \forall h' \in S(z') \tag{A39}$$

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

$$\forall z'' \in B(z,\rho), \quad \forall h' \in S(z') \tag{A40}$$

where $S(z') \subset S$ is the set of all direction vectors which are solutions to the program defined by (14) with $z = z'$ and $\varepsilon = \varepsilon(z')$.

<u>Proof</u>: From Lemma 2 there exist a $\rho_1 > 0$ and an $\hat{\varepsilon} > 0$ such that $\varepsilon(z') \geq \hat{\varepsilon}$ for all $z' \in B(z,\rho_1)$.

(a) If $f(z) = \psi_o(z)$, then by Corollary 1, there exists a $\rho_2 \in (0,\rho_1]$ such that for all $z'', z' \in B(z,\rho_2)$

+Note that α and δ are data parameters in Algorithm I. δ is used in the ε test in Step 4 and α is used in the step length calculation in Step 6.

$$Df(z'',h') \le \tilde{D}f_{\varepsilon}(z',h') + (1-\alpha)\delta\hat{\varepsilon}$$

$$\le - \delta\varepsilon(z') + (1-\alpha)\delta\hat{\varepsilon}$$

$$\le - \alpha\delta\varepsilon(z') \quad \forall h' \in S(z') \tag{A41}$$

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

$$\langle \nabla g(z''),h' \rangle \le \langle \nabla g(z'),h' \rangle + (1-\alpha)\delta\hat{\varepsilon} \qquad \forall h' \in S(z') \tag{A42}$$

Because $g(z') - \psi_o(z') \ge - \hat{\varepsilon} \ge - \varepsilon(z')$ for all $z' \in B(z,\rho_4)$, we obtain from (A42)

$$\langle \nabla g(z''),h' \rangle \le Dg_{\varepsilon(z')}(z',h') + (1-\alpha)\delta\hat{\varepsilon}$$

$$\le - \delta\varepsilon(z') + (1-\alpha)\delta\hat{\varepsilon}$$

$$\le - \alpha\delta\varepsilon(z') \quad \forall z'',\ z' \in B(z,\rho_4), \quad \forall h' \in S(z') \tag{A43}$$

(c) By uniform continuity there exists a $\rho_5 \in (0,\rho_1]$ such that

$$\langle \nabla f^o(z''),h \rangle \le \langle \nabla f^o(z'),h \rangle + (1-\alpha)\delta\hat{\varepsilon} \qquad \forall h \in S, \quad \forall z'',\ z' \in B(z,\rho_5) \tag{A44}$$

By the definition of $\varepsilon(z')$, if $\psi_o(z') = 0$, we obtain

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

$$\le - \alpha\delta\varepsilon(z') \qquad \forall z' \in B(z,\rho_5) \cap F$$

$$\forall z'' \in B(z,\rho_5), \quad \forall h' \in S(z') \tag{A45}$$

Let $\rho \triangleq \min\{\rho_2, \rho_4, \rho_5\}$ and the proof is complete. ◻

Lemma 3. For all $z \in \mathbb{R}^n$ such that $\theta_0^\pi(z) < 0$, $\pi \in \{1,2\}$, there exist a $\mu > 0$ and a $\rho > 0$ such that

$$f^o(z'') - f^o(z') \leq -\mu \qquad \forall z' \in B(z,\rho) \cap F \qquad \forall z'' \in A(z') \tag{A46}$$

$$\psi(z'') - \psi(z') \leq -\mu \qquad \forall z' \in B(z,\rho) \cap F^c \qquad \forall z'' \in A(z') \tag{A47}$$

Proof: We shall consider the case, $\pi=1$, only. By Lemma 2, there exist a $\rho_0 > 0$ and an $\hat{\varepsilon} > 0$ such that $\varepsilon(z') \geq \hat{\varepsilon}$ for all $z' \in \dot{B}(z,\rho_0)$.

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

By the definition of $\psi(z)$, either $f(z) = \psi(z)$, or $g(z) = f(z)$, or both. Suppose,

(i) $g(z) \leq \psi(z)$. By continuity there exists a $\rho_1 \in (0,\rho\ ]$ such that $g(z') < \psi(z')$ for all $z' \in B(z,\rho_1)$ and hence $\psi(z') = f(z')$ for all $z' \in B(z,\rho_1)$. By Proposition 3, there exists a $\rho \in (0,\rho_1]$ such that $Df(z'',h') \leq -\alpha\delta\varepsilon(z')$ for all z'', $z' \in B(z,\rho)$, and $\langle \nabla f^o(z''),h' \rangle \leq -\alpha\delta\varepsilon(z')$ for all $z' \in B(z,\rho) \cap F$, $z'' \in B(z,\rho)$ and for all $h' \in S(z')$. It follows from the definition of the directional derivative that for any z', $h \in \mathbb{R}^n$ and $\lambda \geq 0$,

$$f(z' + \lambda h) - f(z') = \int_o^\lambda Df(z' + sh,h)ds \tag{A48}$$

For any $z' \in B(z,\rho/2)$, and for all $\lambda \in [0, \frac{\rho}{2\sqrt{n}}]$, it follows that

$z' + \lambda h' \in B(z,\rho)$, for all $h' \in S(z')$, since $\|h'\| \leq \sqrt{n}$. We then obtain from (A48)

$$f(z' + \lambda h') - f(z') \leq - \lambda\alpha\delta\varepsilon(z') \qquad \forall z' \in B(z,\rho/2),\ \forall h' \in S(z'),\ \forall \lambda \in [0,\rho/2\sqrt{n}] \tag{A49}$$

Because $f(z') = \psi(z')$ for all $z' \in B(z,\rho)$ we obtain

$$\psi(z' + \lambda h') - \psi(z') \leq - \lambda\alpha\delta\varepsilon(z') \qquad \forall z' \in B(z,\rho/2),\ \forall h' \in S(z'),\ \forall \lambda \in [0,\frac{\rho}{2\sqrt{n}}] \tag{A50}$$

For all $z' \in B(z,\rho/2) \cap F$ we obtain from Proposition 3

$$\begin{aligned} f^o(z' + \lambda h') - f^o(z') &= \int_o^\lambda \langle \nabla f^o(z' + sh'),h' \rangle ds \\ &\leq - \lambda\alpha\delta\varepsilon(z') \qquad \forall h' \in S(z'),\ \forall \lambda \in [0,\frac{\rho}{2\sqrt{n}}] \end{aligned} \tag{A51}$$

From (A50) we have that for all $z' \in B(z,\frac{\rho}{2}) \cap F$,

$$\psi(z' + \lambda h') \leq 0 \quad \forall h' \in S(z'),\ \forall \lambda \in [0,\frac{\rho}{2\sqrt{n}}] \tag{A52}$$

Let $M' = \min\{M, \frac{\rho}{2\sqrt{n}}\}$; then let $\hat{k}(z) \in \mathbb{Z}_+$ be such that $\beta^{\hat{k}(z)} \leq M' \leq \beta^{\hat{k}(z)-1}$. In step 6 of the algorithm the smallest integer $k(z') \in \mathbb{Z}_+$ is chosen such that $\beta^{k(z')} \in (0,M]$ and

$$\psi(z' + \beta^{k(z')}h') - \psi(z') \leq - \beta^{k(z')} \alpha\delta\varepsilon(z') \quad \text{if } z' \in F^c \tag{A53}$$

or

$$f^o(z' + \beta^{k(z')}h') - f^o(z') \leq - \beta^{k(z')} \alpha\delta\varepsilon(z')$$
$$\psi(z' + \beta^{k(z')}h') \leq 0 \qquad \text{if } z' \in F \tag{A54}$$

Comparing (A50), (A51), and (A52) with (A53) and (A54), we conclude that $k(z') \leq \hat{k}(z)$ for all $z' \in B(z,\frac{\rho}{2})$. Hence, $-\beta^{k(z')} \leq - \beta^{\hat{k}(z)}$ and we obtain

$$\psi(z'') - \psi(z') \leq - \beta^{\hat{k}(z)}\alpha\delta\varepsilon(z') \leq - \beta^{\hat{k}(z)}\alpha\delta\hat{\varepsilon}$$
$$\forall z' \in B(z,\rho/2) \cap F^c \tag{A55}$$

$$f^o(z'') - f^o(z') \leq - \beta^{\hat{k}(z)}\alpha\delta\varepsilon(z') \leq - \beta^{\hat{k}(z)}\alpha\delta\hat{\varepsilon}$$
$$\psi(z'') \leq 0 \qquad \forall z \in B(z,\rho/2) \cap F \tag{A56}$$

where $z'' \triangleq z' + \beta^{k(z')}h'$; i.e. $z'' \in A(z')$.

(ii) $f(z) < \psi(z)$. The argument is the same as part (i) except that $f(\cdot)$ and $g(\cdot)$ are interchanged.

(iii) $f(z) = g(z) = \psi(z)$. The argument is similar to part (i) where both $f(\cdot)$ and $g(\cdot)$ are written using integral expansions.

Case 2. $\psi(z) < 0$.

By continuity there exists a $\rho_1 \in (0,\rho_0]$ such that $\psi(z') < 0$ for all $z' \in B(z,\rho_1)$; i.e. $B(z,\rho_1) \subset F$. By Proposition 3 there exists a $\rho \in (0,\rho_1]$ such that $\langle \nabla f^o(z''),h' \rangle \leq - \alpha\delta\varepsilon(z')$ for all $h' \in S(z')$ and for all z'', $z' \in B(z,\rho)$. Using the same argument as in Case 1, we obtain

$$f^o(z'') - f^o(z') \leq - \alpha\beta^{\hat{k}(z)}\delta\hat{\varepsilon}$$
$$\psi(z'') \leq 0, \qquad \forall z' \in B(z,\rho/2) \tag{A57}$$

where $z'' \in A(z')$.

Using the results of Case 1 or Case 2 as appropriate, if we let $\hat{\rho} = \frac{\rho}{2}$ and $\hat{\mu} = \beta^{\hat{k}(z)} \alpha \delta \hat{\varepsilon}$, then we are done. ¤

Appendix B

We now present a design problem in which a PID controller is to be designed for the system in Figure 1. The transfer function is

$$G(s) = \frac{1}{(s+3)(s^2 + 2s + 2)} \tag{B1}$$

and we wish to choose the gains, z, for the PID series compensator $H(z,s) = z^1 + z^2/s + z^3 s$. We wish to minimize the mean square error in the zero-state response to a step input, subject to the constraint that the phase margin is not smaller than 45°. The cost is

$$f^o(z) = \int_0^\infty e^2(z,t)\,dt$$

$$= \frac{z^2(122 + 17z^1 + 6z^3 - 5z^2 + z^1 z^3) + 180z^3 - 36z^1 + 1224}{z^2(408 + 56z^1 - 50z^2 + 60z^3 + 10z^1 z^3 - 2(z^1)^2)} \tag{B2}$$

where we have used Parseval's theorem and the tables in [11]

The phase margin constraint is formulated as the inequality constraint $\phi^1(z,\omega) \leq 0$ for $\omega \in \Omega \triangleq [10^{-6}, 30.0]$ where

$$\phi^1(z,\omega) \triangleq \text{Im } T(z,\omega) - 3.33\,(\text{Re } T(z,\omega))^2 + 1.0 \tag{B3}$$

and $T(z,\omega) = 1 + H(z, j\omega)\,G(j\omega)$. This constraint, when satisfied, defines a region in the complex plane in which the Nyquist plot will be located. The excluded region is a parabolic region as shown in Figure 2. We also impose conventional constraints on the PID controller gains, $0 \leq z^1 \leq 100$, $0.1 \leq z^2 \leq 100$, and $0 \leq z^3 \leq 100$.

The parameters used in the implementable algorithm are

$\alpha = 0.2$, $\beta = 0.3$, $\delta = 10^{-3}$, $\gamma = 2.0$, $\varepsilon_o = 0.2$, $\mu_1 = 10^{-3}$, $\mu_2 = 10^{-2}$, $q_o = 128$, $M = 15.0$, $\pi = 2$.

Our results are tabulated in Table 1. The approximate CPU time was 32 seconds on a CDC 6400 computer.

Table 1.

i	z_i^1	z_i^2	z_i^3	$f^o(z_i)$	$\varepsilon(z_i)$	q
0	1.000	1.000	1.000	3.131	0.20	128
10	21.964	21.952	30.646	0.186	2.0×10^{-4}	128
20	15.400	36.702	35.357	0.176	2.4×10^{-5}	128
30	17.826	40.770	34.017	0.175	6.1×10^{-6}	128
40	16.626	42.552	34.781	0.175	3.1×10^{-6}	128
50	17.746	42.152	34.108	0.175	3.1×10^{-6}	512
60	17.017	42.712	34.555	0.175	1.5×10^{-6}	512
68	16.928	42.974	34.617	0.175	7.6×10^{-7}	512

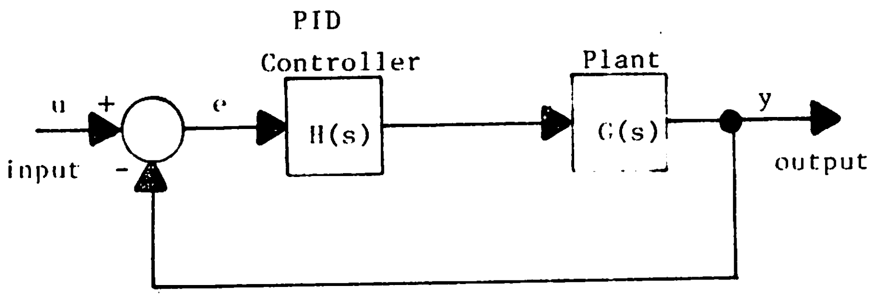

Fig. 1. System block diagram for PID controller design problem.

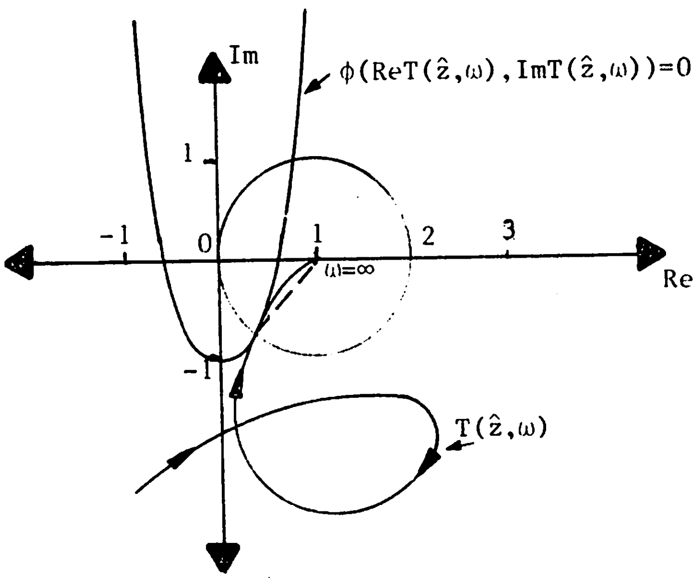

Fig. 2. Nyquist plot for system of Fig. 1 with gains of $\hat{z}^1=16.928$, $\hat{z}^2=42.974$, and $\hat{z}^3=34.617$.