ON CONSTRAINT DROPPING SCHEMES AND OPTIMALITY FUNCTIONS FOR A CLASS OF OUTER APPROXIMATIONS ALGORITHMS

by

C. Gonzaga and E. Polak

Memorandum No. UCB/ERL M77/68

17 November 1977

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

ON CONSTRAINT DROPPING SCHEMES AND OPTIMALITY FUNCTIONS FOR A CLASS OF OUTER APPROXIMATIONS ALGORITHMS

C. Gonzaga† 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 new class of outer approximations algorithms which incorporate constaint dropping schemes. The algorithms are based on the use of certain types of optimality functions, which are commonly used in minimization algorithms, for defining approximations to stationary points.

---

†On leave from COPPE-UFRJ, Rio de Janeiro, Brazil.

Research sponsored by the Joint Services Electronics Program Contract F44620-76-C-0100 and the National Science Foundation (RANN) Grant ENV76-04264.

1. Introduction

After their introduction in 1960, by Cheney and Goldstein [1] and Kelley [2], in the form of cutting plane methods, and, in 1966, by Levitin and Polyak [3], who treated them in a more abstract setting, outer approximations algorithms went through a decade of stagnation. The reason for this was simple. These methods were intended to solve problems of the form

$$P\colon \min\{f(x) \mid x \in X\} \tag{1}$$

where $X \subset \mathbb{R}^n$ had a very complicated description, e.g., $X = \{x \mid \phi(x,\omega) \leq 0,\ \omega \in \Omega\}$, with $\Omega \subset \mathbb{R}^m$ a set of infinite cardinality (i.e. X is defined by a continum of inequalities). The approach was to substitute for P a sequence of approximating problems

$$P_k\colon \min\{f(x) \mid x \in X_k\},\ k = 0,1,2,\ldots, \tag{2}$$

where $X \subset X_0 \subset X_1 \subset X_2 \subset \ldots$ and the X_k had relatively simple descriptions, e.g. by a finite set of inequalities, $X_k = \{x \mid \phi(x,\omega) \leq 0,$ $\omega \in \Omega_k \subset \Omega\}$ with Ω_k a discrete set. Under certain rules defining the properties of the X_k, one could then show that the accumulation points of the sequence of solutions $\{\hat{x}_k\}$, of the problems P_k, were solutions of P. Unfortunately, in all the specific schemes, the complexity of the description of the X_k (i.e. the number of inequalities involved) grew rapidly with k and quite quickly the problems P_k became almost as difficult as the original problem P.

The first breakthrough came when Topkis [4,5] and Eaves and Zangwill [6] proposed constraint dropping schemes which broke the monotonic growth of the descriptions of the X_k. The Eaves and Zangwill theory in terms of cut set maps is particularly elegant. An interesting further generalization was given by Hogan [7]. Although from a theoretical point of view the work in [4,6,7] was of great importance, it still had several drawbacks from a practical point of view. These are easiest to explain in the Eaves-Zangwill framework, using a simple problem, e.g.

$$P: \min\{f(x) \mid \phi(x,\omega) \leq 0, \ \omega \in \Omega\} \tag{3}$$

where f and ϕ are both differentiable and $x \in \mathbb{R}^n$, $\Omega \subset \mathbb{R}^m$. The Eaves-Zangwill theory requires that we solve, exactly, two problems at each iteration.

$$P_k: \min\{f(x) \mid \phi(x,\omega), \ \omega \in \Omega_k\} \tag{4}$$

where Ω_k is a discrete subset of Ω, to obtain a solution x_k and then

$$\max\{\phi(x_k,\omega) \mid \omega \in \Omega\} \tag{5}$$

to produce a point ω_k. Now, in the abscence of convexity and since only a finite number of iterations of a program for solving (4) and (5) can be used, the best one can hope to achieve is to find an approximation to a stationary point for P_k (rather than to a solution x_k) and, perhaps, an approximation to ω_k. The Eaves-Zangwill theory does not apply to this situation. Second, the constraint dropping schemes (i.e., the dropping of points from Ω_k) is determined by the rate of growth of the cost sequence $\{f(x_k)\}$ relative to the constraint violation sequence $\{\phi(x_k,\omega_k)\}$. As a result, unless a problem is extremely well-scaled, their constraint dropping scheme may fail to operate. The third objection to the early constraint dropping schemes is that when constraint dropping is in operation, only the subsequence of $\{x_k\}$ at which constraints were dropped can be shown to have accumulation points which are solutions of P.

The first two of the above described drawbacks were overcome by Mayne, Polak and Trahan [7] in the framework of an algorithm for computer aided design. The present paper generalizes the work in [8] and eliminates the last objection to the early constraint dropping schemes. In particular, we show in this paper that to be useable in an outer approximations algorithm incorporating approximate evaluations of stationary points and max type operators, an optimality function[†] must have certain properties. We prove that a number of existing optimality functions have this property. Also, we present a number of constraint dropping schemes which do not depend on the growh of the cost sequence $\{f(x_k)\}$ and which have the novel property that <u>any</u> accumulation point of the sequence $\{x_k\}$ is a stationary point for the original problem.

2. New Classes of Outer Approximations Algorithms

The algorithms which we shall present are intended for the solution of problems of the form

$$P_\Omega: \min\{f(x) \mid g^j(x) \leq 0,\ j = 1,2,\dots,\ell;$$
$$\phi^k(x,\omega^k) \leq 0,\ \omega^k \in \Omega^k,\ k = 1,2,\dots,m\} \tag{7}$$

where the functions $f(\cdot)$, $g^j(\cdot)$ and $\phi^k(\cdot,\cdot)$ are continuously differentiable[††] on $\mathbb{R}^n$ and on $\mathbb{R}^n \times \mathbb{R}^{p_k}$, respectively, and Ω^k is a <u>compact</u> subset of

---

[†]We say that $\theta:\mathbb{R}^n \to \mathbb{R}^1$ is an optimality function is $\theta(x) = 0$ for all x solving P and $\theta(x) \leq 0$ for all $x \in \mathbb{R}^n$.

[††]Differentiability in ω is not required by our proofs, but is stipulated as an assumption which is usually required by algorithms which compute approximate solutions to $\max_{\omega^k}\{\phi^k(x,\omega) \mid \omega^k \in \Omega^k\}$.

$\mathbb{R}^{p_k}$, $k = 1,2,\ldots,m$. The symbol Ω is used to denote $\Omega^1 \times \Omega^2 \times \ldots \times \Omega^m$. The problem form (7) is particularly important because many engineering design problems can be transcribed into it.

We shall approximate the problem P_Ω by a sequence of simpler problems of the form (with $i = 1,2,3,\ldots$)

$$P_{\Omega_i}: \min\{f(x) \mid g^j(x) \leq 0,\ j = 1,2,\ldots,\ell;$$
$$\phi^k(x,\omega^k) \leq 0,\ \omega^k \in \Omega_i^k,\ k = 1,2,\ldots,m\} \tag{8}$$

where $\Omega_i^k \subset \Omega^k$. Our aim is to approximate feasible stationary points of P_Ω, i.e., points $\hat{x} \in \mathbb{R}^n$ such that†

$$g^j(\hat{x}) \leq 0\ ,\ j = 1\text{–}\ell;\ \max_{\omega\in\Omega} \phi^k(\hat{x},\omega) \leq 0,\ k = 1\text{–}m \tag{9}$$

and, with $S \triangleq \{h \in \mathbb{R}^n \mid |h^i| \leq 1,\ i = 1,2,\ldots,n\}$

$$\min_{h\in S} \max\{\langle \nabla f(\hat{x}),h\rangle ;\ g^j(\hat{x}) + \langle \nabla g^j(\hat{x}),h\rangle ,\ j = 1,2,\ldots,\ell;$$
$$\phi^k(\hat{x},\omega) + \langle \nabla_x \phi^k(\hat{x},\omega^k),h\rangle ,\ \omega^k \in \Omega^k,\ k = 1,2,\ldots,m\} = 0 \tag{10}$$

we recognize (10) as the Topkis-Veinott [9] multiplier free form of the F. John condition for P_Ω (see, p.8 and p.182 in [10]).

<u>Definition</u>: We shall say that a point $\hat{x} \in \mathbb{R}^n$ is <u>desirable</u> if (9) and (10) are satisfied at $\hat{x}$. We shall denote the set of all desirable points in $\mathbb{R}^n$ by Δ. ¤

We assume that we can "solve" the problems P_{Ω_i} approximately, to the extent of finding a point x_i for which the value of an appropriate optimality function $\theta^i_{\Omega_i}(x_i)$ is small. The superscript is introduced to allow for the possible use of penalty functions. The theory we are about to present is based on our knowledge of phase I-Phase II type methods of feasible directions [11,12] and penalty functions

†We write $j = 1\text{–}\ell$ to denote $j = 1,2,\ldots,\ell$, etc.

[10], all of which utilize real valued optimality functions $\theta^i_{\Omega'}(\cdot)$, defined on $\mathbb{R}^n$ for discrete subsets $\Omega^{k'} \subset \Omega^k$, and $k = 1,2,\ldots,m$, and all positive integers i. All of these functions have the property that $\theta^i_{\Omega'}(x) \leq 0$ for all $x \in \mathbb{R}^n$ and that if x' is optimal for $P_{\Omega'}$, then $\lim_{i\to\infty} \theta^i_{\Omega'}(x') = 0$. Some of these optimality functions are continuous while others are not. Early examples of such optimality functions can be found in [10], see p. 182. Not all the existing optimality functions can be used in our outer approximations algorithms. Only the ones satisfying Assumptions 1 and 2 are acceptable. We need the following definition. For any subset $\Omega' \in \Omega$, $\psi_{\Omega'}:\mathbb{R}^n \to \mathbb{R}^1$ is defined by

$$\psi_{\Omega'}(x) \triangleq \max\{0;g^j(x),\ j = 1,2,\ldots,\ell;$$
$$\phi^k(x,\omega^k),\ \omega \in \Omega^k,\ k = 1,2,\ldots,m\} \tag{11}$$

<u>Assumption 1</u>: Consider the family of optimality functions $\{\theta^i_{\Omega'}(\cdot)\}$, with Ω' a discrete subset of Ω and i a positive integer. For all $x \in \mathbb{R}^n$, $x \notin \Delta$, there exist $\mu > 0$, $\rho > 0$, $N > 0$ and $\delta \in (0,1)$ (possibly depending on x), such that for all $x' \in B(x,\rho) \triangleq \{x' \mid \|x-x'\| \leq \rho\}$ and all discrete subsets $\Omega'^k \subset \Omega^k$, $k = 1,2,\ldots,m$, satisfying $\psi_{\Omega'}(x) \geq \delta\psi_{\Omega}(x)$, we have

$$\theta^i_{\Omega'}(x') \leq -\mu \quad \text{for all } i \geq N \tag{12}$$

¤

We shall later devote a separate section to showing that a number of common optimality functions satisfy Assumption 1. In the present section we shall only consider its consequences.

Our outer approximations algorithms are of the form of the model below. They differ from one another only by the manner in which the discrete sets Ω^k_i, $k = 1,2,\ldots,m$, are constructed. They all require that we have an algorithm for solving the problem P_{Ω_i}, with Ω_i a discrete set, and another one for approximating the values of the functions $\bar{\psi}^k_{\Omega',k}:\mathbb{R}^n \to \mathbb{R}^1$ defined by

$$\bar{\psi}^k_{\Omega',k}(x) \triangleq \max_{\omega\in\Omega',k} \phi^k(x,\omega^k),\ k = 1,2,\ldots,m \tag{13}$$

with $\Omega',k \subseteq \Omega^k$. To complete our notation, we define $\bar{\psi}^0:\mathbb{R} \to \mathbb{R}^1$ by

$$\bar{\psi}^0(x) \triangleq \max\{g^1(x),g^2(x),\ldots,g^\ell(x)\} \tag{13a}$$

<u>Algorithm Model 1.</u>

<u>Parameters</u>: An infinite sequence $\{\beta_i\}_{i=1}^{\infty}$, $\beta_i > 0$, $\beta_i \to 0$.† ($\theta^i_{\Omega'}(\cdot)$ is a family of optimality functions).

<u>Data</u>: Discrete sets $\Omega^k_0 \subset \Omega^k$, $k = 1,2,\ldots,m$.

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

<u>Step 1</u>: Construct the discrete sets Ω^k_i, $k = 1,2,\ldots,m$.

<u>Step 2</u>: Compute an x_{i+1} such that

$$\theta^i_{\Omega_i}(x_{i+1}) \geq -\beta_i \tag{13b}$$

<u>Step 3</u>: Set $i = i+1$ and go to step 1. ¤

Although all of our methods are summarized in the one to be treated in Theorem 3, it is easier to understand our methods by considering three progressively more sophisticated schemes for constructing the Ω^k_i. The first scheme which does not drop any constraints will be treated in Theorem 1. However, before we can prove this theorem, we need the following proposition.

<u>Proposition 1</u>: Let $\{x_i\}_{i=1}^{\infty}$ be any converging sequence in $\mathbb{R}^n$ with

---

†For example, $\beta_i = \varepsilon\delta^i$, for $\delta \in (0,1)$, or $\beta_i = \varepsilon/i$.

$x_i \to \hat{x}$ as $i \to \infty$ and let $\Omega_i^k \subset \Omega^k$, $i = 1,2,\ldots$, be any sequence of compact sets contained in Ω^k. Then

$$|\bar{\psi}^k_{\Omega_i^k}(x_i) - \bar{\psi}^k_{\Omega_i^k}(\hat{x})| \to 0 \text{ as } i \to \infty \tag{14}$$

Proof: Suppose that (14) does not hold. There exists a $\hat{\delta} > 0$ and a subsequence indexed by $K \subset \{1,2,\ldots\}$ such that

$$|\bar{\psi}^k_{\Omega_i^k}(x_i) - \bar{\psi}^k_{\Omega_i^k}(\hat{x})| \geq \hat{\delta} \text{ for all } i \in K \tag{15}$$

Now, $\bar{\psi}^k_{\Omega_i^k}(x_i) = \phi(x_i,\omega_i^k)$ and $\bar{\psi}^k_{\Omega_i^k}(\hat{x}) = \phi(\hat{x},\hat{\omega}_i^k)$ for some ω_i^k, $\hat{\omega}_i^k$ in Ω_i^k. Without loss of generality, we may assume, therefore, that

$$\phi^k(x_i,\omega_i^k) \geq \phi^k(\hat{x},\hat{\omega}_i^k) + \hat{\delta} \quad \text{for all } i \in K \tag{16}$$

Now, since $\Omega_i^k \subset \Omega^k$, a compact set, and $x_i \to \hat{x}$, there exists an i_o such that

$$\phi^k(\hat{x},\omega_i^k) \geq \phi^k(x_i,\omega_i^k) - \hat{\delta}/2 \text{ for all } i \in K_1 \quad i \geq i_o \tag{17}$$

But (17) and (16) show that $\hat{\omega}_i^k$ is not a maximizer of $\phi^k(\hat{x},\omega^k)$ over Ω_i^k which is a contradiction. Hence the proposition is true. ¤

Corollary: The functions $\bar{\psi}^k_{\Omega',k}(\cdot)$ are continuous, uniformly in $\Omega' \subset \Omega^k$.

Proof: To obtain a contradiction, suppose that $\hat{x} \in \mathbb{R}^n$ is such that given a $\hat{\delta} > 0$ there is no $\hat{\varepsilon} > 0$ such that

$$|\psi^k_{\Omega',k}(x) - \psi^k_{\Omega',k}(\hat{x})| < \hat{\delta}$$

for all $x \in \{x \mid \|x-\hat{x}\| < \hat{\varepsilon}\}$, for all $\Omega'^k \subset \Omega^k$. But then there must exist a sequence $x_i \to \hat{x}$ and set $\Omega_i^k \subset \Omega^k$, $i = 1,2,\ldots$, such that

$$|\psi^k_{\Omega_i k}(x_i) - \psi^k_{\Omega_i k}(\hat{x})| \geq \hat{\delta} \quad \text{for } i = 1,2,\ldots$$

But this contradicts (14) and hence we are done.

$\square$

Our three schemes define the operations to be performed in step 1 of the algorithm Model.

Constraint Construction Scheme 1

Given x_i, (i) compute $\omega_i^k \in \Omega^k$ by approximately evaluating $\max_{\omega^k \in \Omega^k} \phi^k(x_i,\omega^k))$; (ii) set

$$\psi_\Omega^i(x_i) \triangleq \max\{0,\psi^o(x_i);\ \phi^k(x_i,\omega_i^k),\ k = 1,2,\ldots,m\} \tag{18}$$

(iii) if $\psi_\Omega^i(x_i) > 0$, include ω_i^k in Ω_j^k for all $j > i$, for all $k \in \{1,2,\ldots,m\}$ such that

$$\psi_\Omega^i(x_i) = \phi^k(x_i,\omega_i^k) \tag{19}$$

$\square$

In effect in its most economical form, the constraint Construction Scheme 1 only requires that given the sets Ω_i^k, $k = 1,2,\ldots,m$, $\Omega_{i+1}^k = \{\omega_i^k\} \cup \Omega_i^k$ if $\phi^k(x_i,\omega_i^k) = \psi_\Omega^i(x_i)$ and $\psi_\Omega^i(x_i) > 0$, and $\Omega_{i+1}^k = \Omega_i^k$ otherwise; i.e., the approximating constraint sets Ω_i^k are augmented only for those functional inequalities ($\max_{\omega \in \Omega^k} \phi^k(x,\omega) \le 0$) that have been most violated.

Apart from this specified restriction, the construction of the Ω_i^k is arbitrary in the sense that any other points $\omega^k \subset \Omega$, not specifically covered by the scheme can be added (or subtracted) from the sets Ω_i^k. The relevance of this is that the most economical form of the Constraint Construction Scheme 1 is not always the best computationally.

Theorem 1: Consider a sequence $\{x_i\}_{i=1}^{\infty}$ constructed by the Algorithm Model 1, using the Constraint Construction Scheme 1. Suppose that

(i) $|\phi^k(x_i,\omega_i^k) - \bar{\psi}^k_{\Omega^k}(x_i)| \to 0$ as $i \to \infty$, for $k = 1,2,\ldots,m$ (i.e. that $|\psi_{\Omega^i}^i(x_i) - \psi_\Omega(x_i)| \to 0$ as $i \to \infty$) (ii) the optimality functions $\theta^i_{\Omega'}(\cdot)$ used in the Algorithm Model 1 satisfy Assumption 1.

Then any accumulation point of the sequence $\{x_i\}_{i=1}^\infty$ is in Δ.

Proof: To obtain a contradiction, suppose that $x_i \overset{K}{\to} \hat{x}$, where $K \subset \{1,2,3,\ldots\}$, and $\hat{x} \notin \Delta$. We consider the various possibilities.

(i) Suppose that $\psi_\Omega(\hat{x}) = 0$, i.e. $\hat{x}$ is feasible for P_Ω. Let $\hat{\delta} \in (0,1)$, $\hat{\rho} > 0$, $\hat{N} > 0$ and $\hat{\mu} > 0$ be as specified in Assumption 1 for $\hat{x}$. Then, since $\psi_{\Omega_i}(x) \geq 0$ for all x and any $\Omega_i \subset \Omega$,

$$\psi_{\Omega_i}(\hat{x}) \geq \hat{\delta}\psi_\Omega(\hat{x}) \text{ for all } i \tag{20}$$

Consequently, there exists an integer $i_o \geq \hat{N}$ such that $x_i \in B(\hat{x},\hat{\rho})$ for $i \geq i_o$, $i \in K$ and

$$\theta^i_{\Omega_i}(x_i) \leq -\hat{\mu} \leq -\beta_i \quad \text{for all } i \geq i_o,\ i \in K \tag{21}$$

which contradicts the construction in Step 2 of Algorithm Model 1.

(ii) Suppose that $\psi_\Omega(\hat{x}) > 0$ and there exists an infinite subsequence $\{x_i\}_{i\in K'}$, with $K' \subset K$, such that $\psi^i_\Omega(x_i) = \bar{\psi}^o(x_i)$ for all $i \in K'$. Then, since $\psi_\Omega(\cdot)$ is continuous and $|\psi^i_\Omega(x_i) - \psi_\Omega(x_i)| \to 0$, as $i \to \infty$, by assumption, we have that $\bar{\psi}^o(x_i) \overset{K'}{\to} \psi_\Omega(\hat{x})$ as $i \to \infty$. But since $\Omega_i \subset \Omega$

$$\psi_\Omega(x_i) \geq \psi_{\Omega_i}(x_i) \geq \psi^o(x_i) \text{ for all } i \tag{22}$$

and hence $\psi_{\Omega_i}(x_i) \overset{K'}{\to} \psi_\Omega(\hat{x})$, as $i \to \infty$. Making use of Proposition 1, we now conclude that $\psi_{\Omega_i}(\hat{x}) \overset{K'}{\to} \psi_\Omega(\hat{x})$, as $i \to \infty$. Let $\hat{\delta} \in (0,1)$, $\hat{N} > 0$, $\hat{\mu} > 0$ be as specified in Assumption 1, then there exists an $i_1 \geq \hat{N}$ such that

$$\psi_{\Omega_i}(\hat{x}) \geq \hat{\delta}\psi_\Omega(\hat{x}) \text{ for all } i \in K',\ i \geq i_1 \tag{23}$$

and hence

$$\theta^{i}_{\Omega_i}(x_i) \leq -\hat{\mu} \leq -\beta_i \text{ for all } i \in K', \; i \geq i_1 \tag{24}$$

which contradicts the construction in step 2 of the Algorithm Model.

(iii) Suppose that $\psi_\Omega(\hat{x}) > 0$ and there exists a $\bar{k} \in \{1,2,\ldots,m\}$ and an infinite subsequence $\{x_i\}_{i\in K''}$, $K'' \subset K$, such that for all $i \in K''$, $0 < \psi^{i}_{\Omega}(x_i) = \phi^{\bar{k}}(x_i,\omega^{\bar{k}}_i)$. Then, because $\psi_\Omega(\cdot)$ is continuous and $|\psi^{i}_{\Omega}(x_i) - \psi_\Omega(x_i)| \to 0$ by assumption, we find that

$$\phi^{\bar{k}}(x_i,\omega^{\bar{k}}_i) \overset{K''}{\to} \psi_\Omega(\hat{x}) \text{ as } i \to \infty \tag{25}$$

Since Ω is compact, we must have

$$|\phi^{\bar{k}}(x_j,\omega^{\bar{k}}_i) - \phi^{\bar{k}}(x_i,\omega^{\bar{k}}_i)| \to 0 \text{ as } i \to \infty, \; i,j \in K'', \; j > i \tag{26}$$

Now, $\omega^{\bar{k}}_i \in \Omega^{\bar{k}}_j$ for all $i \in K''$, $j > i$ by construction. Hence we obtain that

$$\psi_\Omega(x_j) \geq \bar{\psi}^{\bar{k}}_{\Omega^{\bar{k}}_j}(x_j) \geq \phi^{\bar{k}}(x_j,\omega^{\bar{k}}_i) \tag{27}$$

for all $i,j \in K''$, $j > i$. Taking (25) and (26) into account, we conclude that

$$\bar{\psi}^{\bar{k}}_{\Omega^{\bar{k}}_i}(x_i) \overset{K''}{\to} \psi_\Omega(\hat{x}) \text{ as } j \to \infty \tag{28}$$

Hence, since $\psi_\Omega(x_i) \geq \psi_{\Omega_i}(x_i) \geq \bar{\psi}^{\bar{k}}_{\Omega^{\bar{k}}_i}(x_i)$, for all i, we conclude that $\psi_{\Omega_i}(x_i) \to \psi_\Omega(\hat{x})$. The rest of the proof is exactly as for case (ii) and can therefore be omitted.

Since (i), (ii) and (iii) are the only possibilities, we conclude that the theorem holds. ◻

The next approximating constraint construction scheme is similar to the first, in the sense that once a point ω^k_i is added to the set Ω^k_{i+1}

it has to be retained in all subsequent Ω^k_{i+j}, $j = 1,2,\ldots$. However, it uses a considerably milder test and hence augments the Ω^k_i considerably less frequently. In fact, a skillful choice of the parameters $\{\varepsilon_i\}$ in this scheme can result in very few augmentations, indeed.

Constraint Construction Scheme 2.

(a) Specify a decreasing sequence $\{\varepsilon_i\}_{i=1}^{\infty}$, with $\varepsilon_i > 0$ and $\varepsilon_i \to 0$ as $i \to \infty$ (e.g. $\varepsilon_i = \varepsilon_o/i$, or $\varepsilon_i = \varepsilon_o\beta^i$, $\beta \in (0,1)$, etc). (b) Given x_i,
(i) Compute $\omega^k_i \in \Omega^k$, $k = 1,2,\ldots,m$, by approximately evaluating $\max\limits_{\omega^k \in \Omega^k} \phi^k(x_i,\omega^k)$, (ii) Set $\psi^i_\Omega(x_i)$ as in (18), (iii) If $\psi^i_\Omega(x_i) > \varepsilon_i$, include ω^k_i in Ω^k_j for all $j > i$, for all $k \in \{1,2,\ldots,m\}$ satisfying (19).
$\square$

Theorem 2: Consider a sequence $\{x_i\}_{i=1}^{\infty}$ constructed by the Algorithm Model 1, using the Constraint Construction Scheme 2. Suppose

(i) $|\phi^k(x_i,\omega^k_i) - \bar{\psi}^k_{\Omega^k}(x_i)| \to 0$ as $i \to \infty$ for $k = 1,2,\ldots,m$,

(i.e. $|\psi^i_\Omega(x_i) - \psi_\Omega(x_i)| \to 0$ as $i \to \infty$).

(ii) The optimality functions $\theta^i_{\Omega'}(\cdot)$ used in the Algorithm Model 1 satisfy Assumption 1.

Then any accumulation point of the sequence $\{x_i\}_{i=1}^{\infty}$ is in Δ.

Proof: To obtain a contradiction, suppose that $x_i \xrightarrow{K} \hat{x}$, where $K \subset \{1,2,3,\ldots\}$ and $\hat{x} \notin \Delta$.

As in the proof of Theorem 1, we consider three possibilities.

(i) Suppose that $\psi_\Omega(\hat{x}) = 0$. Then we get a contradiction exactly as in the corresponding case of Theorem 1.

(ii) Suppose that $\psi_\Omega(\hat{x}) > 0$ and there exists an infinite subsequence $\{x_i\}_{i\in K'}$, $K' \subset K$, such that $\psi_\Omega^i(x_i) = \bar{\psi}^o(x_i)$ for all $i \in K'$. Then, again, we get a contradiction exactly as the corresponding case of Theorem 1.

(iii) Suppose that $\psi_\Omega(\hat{x}) > 0$ and there exists an infinite subsequence $\{x_i\}_{i\in K''}$, $K'' \subset K$, and a $\bar{k} \in \{1,2,\ldots,m\}$ such that

$$\psi_\Omega^i(x_i) = \phi^{\bar{k}}(x_i,\omega_i^k) > \varepsilon_i \text{ for all } i \in K'' \tag{29}$$

Then (28) can be established as in case (iii) of Theorem 1 and the rest of the proof by contradiction is exactly the same as for Theorem 1.

¤

Our third approximating constraint construction scheme is a generalization of the ones proposed by Eaves and Zangwill [6] and by Mayne, Trahan and Polak [8]. Like those schemes, it will retain a particular constraint for a certain number of approximating problems, and then drop it. However, it stores more information than the schemes in [6] and [8] and therefore leads both to better computational behavior and to a more interesting convergence theorem.

<u>Constraint Construction Scheme 3</u>.

(a) Specify a double indexed sequence $\{\varepsilon_{ij}\}_{\substack{i=0\\ j\le i}}^{\infty}$ such that (i) $\varepsilon_{ij} > 0$ for all $i,j \geq i$; (ii) $\varepsilon_{ij} \to \bar{\varepsilon}_j$ as $i \to \infty$, uniformly in j; and (iii) $\bar{\varepsilon}_j > \varepsilon_{ij}$ for $i \geq j$, and $\bar{\varepsilon}_j \to 0$ as $j \to \infty$. (For example, $\varepsilon_{ij} = \delta^j - \delta^i$, with $\delta \in (0,1)$, or $\varepsilon_{ij} = \bar{\varepsilon}_j - \bar{\varepsilon}_i$, where $\bar{\varepsilon}_i \downarrow 0$).

(b) Given x_i; (i) compute $\omega_i^k \in \Omega^k$, $k = 1,2,\ldots,m$, by approximately evaluating $\max\limits_{\omega^k\in\Omega^k} \phi^k(x_i,\omega^k)$ and store it. (ii) Set $\psi_\Omega^i(x_i)$ as in (18) and store it. (iii) For all $j \in \{1,2,\ldots,i\}$ such that

$\psi^j_\Omega(x_j) > \varepsilon_{ij}$, include ω^k_j in Ω^k_{i+1}, for all $k \in \{1,2,\dots,m\}$ satisfying (19). ◻

For a comparison with the Eaves-Zangwill scheme [6], we set $\varepsilon_{ij} \triangleq f(x_i) - f(x_j)$, $i \geq j$. Their rule is to store only the last $\psi_\ell(x_j)$, ω^k_j and $f(x_j)$ at which constraints were dropped and to include all ω^k_ℓ in Ω^k_i, $j \leq \ell < i$ for all $k \in \{1,2,\dots,m\}$ satisfying (19) (at ℓ), whenever $\psi_\Omega(x_j) > \varepsilon_{ij}$. The Mayne-Polak-Trahan scheme [8] is similar to the Eaves-Zangwill one, except that it sets $\varepsilon_{ij} = \dfrac{f(x_i)-f(x_j)+\mu\beta^i}{\tau(1-\beta^i)}$ where $\beta \in (0,1)$ and $\tau > 0$, $\mu > 0$. Thus, the schemes in [6] and in [8], slowly accumulate constraints (i.e. ω^k_j), then drop them in mass then accumulate them again. This type of oscillatory behavior results in poor computational properties. Also, since only one $\psi_\Omega(x_j)$ is stored at any time, convergence properties in [6], [8] can only be established for the subsequence at which constraints were dropped, rather than for the whole sequence. Our Constraint Construction Scheme 3 does not lead to the type of oscillatory behavior mentioned above and does permit to establish convergence properties for the entire sequence $\{x_i\}$. It does share, with the schemes in [6;8] the property that it retains a certain ω^k_j in Ω^k_i until $i-j$ has become sufficiently large for $\psi^j_\Omega(x_j) \leq \varepsilon_{ij}$ to take place, and then drop it.

<u>Theorem 3</u>: Consider a sequence $\{x_i\}_{i=1}^\infty$ constructed by the Algorithm Model 1, using the Constraint Construction Scheme 2. Suppose that

(i) $|\phi^k(x_i,\omega^k_i) - \bar{\psi}_{\Omega^k}(x_i)| \to 0$ as $i \to \infty$, for $k = 2,\dots,m$,

(i.e. $|\psi^i_{\Omega_i}(x_i) - \psi_\Omega(x_i)| \to 0$ as $i \to \infty$) (ii) The optimality functions $\theta^i_{\Omega'}(\cdot)$ used in the Algorithm Model 1 satisfy Assumption 1.

Then any accumulation point of the sequence $\{x_i\}_{i=1}^{\infty}$ is in Δ.

<u>Proof</u>: We note that since $\varepsilon_{ij} > \bar{\varepsilon}_j$ for all $i \geq j$ and all j, when Scheme 3 is used, a point ω_j^k satisfying (19) is always included in all Ω_i^k, $i > j$, whenever $\psi_\Omega^i(x_j) > \bar{\varepsilon}_j$. Since the $\bar{\varepsilon}_j$ satisfy the properties of the $\{\varepsilon_i\}$ specified in Scheme 2, theorem 3 follows directly from Theorem 2.

¤

It may sometimes be difficult to show that an optimality function $\theta_{\Omega'}^i(\cdot)$ satisfies Assumption 1. In that case one can make use of Assumption 2, below. It is satisfied by the optimality functions used in [8].

<u>Assumption 2</u>: Consider the family of optimality functions $\{\theta_{\Omega'}^i(\cdot)\}$, where the Ω_i are discrete subsets of Ω. If $\{x_i\}_{i=1}^{\infty}$ is a converging sequence in $\mathbb{R}^n$, with $x_i \to \hat{x}$, with $\psi_\Omega(\hat{x}) = 0$, and $\theta_{\Omega_i}^i(x_i) \to 0$ as $i \to \infty$, then $\hat{x} \in \Delta$.

¤

When Assumption 2 is in force, we must use a different algorithm model.

<u>Algorithm Model 2</u>

<u>Parameters</u>: An infinite sequence $\{\beta_i\}_{i=0}^{\infty}$, $\beta_i > 0$, $\beta_i \to 0$.

<u>Data</u>: Discrete sets Ω_0^k, $k = 1,2,\ldots,m$ contained in Ω.

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

<u>Step 1</u>: Construct the discrete sets Ω_i^k, $k = 1,2,\ldots,m$.

<u>Step 2</u>: Compute an x_i such that

$$0 \geq \theta_{\Omega_i}^i(x_i) \geq -\beta_i \text{ and } \psi_{\Omega_i}(x_i) \leq \beta_i \tag{30}$$

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

Theorem 4: Consider a sequence $\{x_i\}_{i=0}^{\infty}$ constructed by Algorithm Model 2, using the Constrain Construction Scheme 1, 2, or 3. Suppose that

(i) $|\phi^k(x_i,\omega_i^k) - \bar{\psi}_{\Omega^k}(x_i)| \to 0$ as $i \to \infty$ for $k = 1,2,\ldots,m$, (i.e., that $|\psi_\Omega^i(x_i) - \psi_\Omega(x_i)| \to 0$ as $i \to \infty$);

(ii) the optimality functions $\theta_{\Omega^i}^i(\cdot)$ used in the Algorithm Model 2 satisfy Assumption 2.

Then any accumulation point of the sequence $\{x_i\}_{i=1}^{\infty}$ is in Δ.

Proof: We only need to prove this theorem for the case where the Constraint Construction Scheme 2 is used, since both Scheme 1 and Scheme 3 can be seen to be special cases of it.

Thus, suppose that $x_i \xrightarrow{K} \hat{x}$, with $K \subset \{1,2,3,\ldots\}$ and that $\hat{x} \notin \Delta$. First, suppose that $\psi_\Omega(\hat{x}) = 0$. Since $\theta_{\Omega_i}^i(x_i) \to 0$ as $i \to \infty$ by construction, it follows from Assumption 2 that $\hat{x} \in \Delta$ and we get a contradiction.

Therefore, suppose that $\psi_\Omega(\hat{x}) > 0$. Then, since $\psi_\Omega(x_i) \xrightarrow{K} \psi_\Omega(\hat{x})$ by continuity and $|\psi_\Omega^i(x_i) - \psi_\Omega(x_i)| \to 0$ as $i \to \infty$, by assumption, we must have $\psi_\Omega^i(x_i) \xrightarrow{K} \psi_\Omega(\hat{x})$. Therefore, since $\varepsilon_i \to 0$ as $i \to \infty$, there exists an integer $i_o > 0$ such that $\psi_\Omega^i(x_i) > \varepsilon_i$ for all $i \geq i_o$, $i \in K$, and consequently, $\omega_i^k \in \Omega_j^k$ for all $j > i$, $i \in K$ and $k \in \{1,2,\ldots,m\}$ such that (19) holds. We now distinguish between two possibilities.

(i) there is an infinite subsequence, $\{x_i\}_{i\in K'}$, with $K' \subset K$, such that $\psi_\Omega^i(x_i) = \bar{\psi}^o(x_i)$ for all $i \in K'$.

In this case, we get that

$$\psi_{\Omega_i}(x_i) \geq \bar{\psi}^o(x_i) = \psi_\Omega^i(x_i) \text{ for all } i \in K' \tag{31}$$

But $\psi_\Omega^i(x_i) \xrightarrow{K} \psi_\Omega(\hat{x})$ and hence there exists an $i_1 \geq 0$ such that $\psi_{\Omega_i}(x_i) \geq \psi_\Omega(\hat{x})/2 > \beta_i$ for all $i \geq i_1$, $i \in K$. But this contradicts (30). Hence, consider the second alternative. (ii) There exists an infinite subsequence $\{x_i\}_{i \in K''}$, with $K'' \subset K$ and a $\bar{k} \in \{1,2,\ldots,m\}$ such that $\psi_\Omega^i(x_i) = \phi^{\bar{k}}(x_i,\omega_i^{\bar{k}}) > \epsilon_i$ for all $i \in K''$. Then, because we are using Constraint Construction Scheme 2, we obtain that

$$\psi_{\Omega_i}(x_i) \geq \phi^{\bar{k}}(x_i,\omega_j^{\bar{k}}) \text{ for all } i > j \geq i_o,\ j \in K'' \tag{32}$$

Since $\omega_j^{\bar{k}} \in \Omega_i^{\bar{k}}$, by construction, in this case. Now, because Ω is compact,

$$|\phi^{\bar{k}}(x_i,\omega_j^{\bar{k}}) - \phi^{\bar{k}}(x_j,\omega_j^{k})| \to 0 \tag{33}$$

as $i,j \to \infty$, $i,j \in K''$, $i > j$. Since $\psi_\Omega^j(x_j) = \phi^{\bar{k}}(x_j,\omega_j^{\bar{k}}) \xrightarrow{K''} \psi_\Omega(\hat{x})$ as $j \to \infty$, we obtain from (33) that

$$\phi^{\bar{k}}(x_i,\omega_j^{\bar{k}}) \to \psi_\Omega(\hat{x}) \tag{34}$$

as $i,j \to \infty$, $i,j \in K''$, $i > j$. Hence, there exists an $i_2 \geq i_o$ such that, by (32),

$$\psi_{\Omega_i}(x_i) \geq \psi_\Omega(\hat{x})/2 > \beta_i \text{ for all } i \geq i_2,\ i \in K'' \tag{35}$$

But this contradicts (30) and hence we are done. ◻

3. Optimality Functions for Outer Approximations Algorithms

We shall now present a few optimality functions which satisfy Assumptions 1 and 2. First we show that any family of optimality functions satisfying Assumption 1 must also satisfy Assumption 2.

Proposition 2: Suppose $\{\theta_{\Omega_i}^i(\cdot)\}$, $\Omega_i \subset \Omega$, is a sequence of optimality functions satisfying Assumption 1. Then it also satisfies Assumption 2.

Proof: Suppose that $x_i \to \hat{x}$ as $i \to \infty$, with $\psi(\hat{x}) = 0$, and that $\theta^i_{\Omega_i}(x_i) \to 0$ as $i \to \infty$. Then $\hat{x} \in \Delta$, for otherwise, by Assumption 1 (since $\psi_{\Omega_i}(\hat{x}) \geq \hat{\delta}\psi_{\Omega}(\hat{x})$ for any $\hat{\delta} \in (0,1)$ and all i) there exists a $\hat{\mu} > 0$ and an i_o such that $\theta^i_{\Omega_i}(x_i) \leq -\hat{\mu}$ for all $i \geq i_o$ which contradicts $\theta^i_{\Omega}(x_i) \to 0$ as $i \to \infty$.

¤

The first two optimality functions that we consider are independent of the superscript i and hence we shall drop it for these cases. These optimality conditions are normally used in methods of feasible directions (see [10],[11],[12]) for computing descent directions. Since, as we should show this satisfy Assumption 1, we conclude that methods of feasible directions based on these functions are suitable for solving problems $P_{\Omega'}$.

Consider the functions, with $\Omega'^k \subset \Omega^k$, introduced in [13] by Pironneau and Polak,

$$\theta_{\Omega}(x) \triangleq \min_h\{\tfrac{1}{2}\|h\|^2 + \max\{\langle \nabla f(x),h\rangle;\ g^j(x) + \langle \nabla g^j(x),h\rangle,\ j = 1,2,\ldots,\ell;$$
$$\phi^k(x,\omega^k) + \langle \nabla_x\phi^k(x,\omega^k),h\rangle,\ \omega^k \in \Omega'^k,\ k = 1,2,\ldots,m\} - \psi_{\Omega'}(x) \tag{37}$$

Since (37) is an extremely messy expression, we shall show (without loss of generality) that it satisfies Assumption 1 by considering only the special case where $m = 1$ and $\ell = 0$, i.e., no $g^j(\cdot)$ constraints. For this case superscripts can be dropped, $P_{\Omega'}$ becomes $\min\{f(x)\,|\,\phi(x,\omega) \leq 0,\ \omega \in \Omega'\}$ and (37) simplifies to

$$\theta_{\Omega'}(x) = \min_n\{\tfrac{1}{2}\|h\|^2 + \max\{\langle \nabla f(x),h\rangle;\ \phi(x,\omega) + \langle \nabla_x\phi(x,\omega),h\rangle,$$
$$\omega \in \Omega\} - \psi_{\Omega'}(x) \tag{38}$$

Assumption 3: For every $x \in \mathbb{R}^n$, $0 \notin \underset{\omega \in \Omega_0(x)}{\text{co}} \nabla_x \phi(x,\omega)$, where

$$\Omega_0(x) \triangleq \{\omega \in \Omega \,|\, \phi(x,\omega) = \psi_\Omega(x)\} \tag{39}$$

and co denotes the convex hull of the set in question. ☐

Theorem 5: Suppose that Assumption 3 is satisfied. Then family of optimality functions defined by (37) satisfies Assumption 1.

Proof: We shall only give a proof for the special case (38). It is quite easy to see that $\Delta = \{x \,|\, \theta_\Omega(x) = 0,\ \psi_\Omega(x) = 0\}$. Since by assumption $0 \notin \underset{\omega \in \Omega_0(x)}{\text{co}} \nabla_x \phi(x,\omega)$ for all $x \in \mathbb{R}^n$, it is easy to see that $\theta_\Omega(x) < 0$ for all $x \in \mathbb{R}^n$ such that $\psi_\Omega(x) > 0$ and hence

$$\Delta = \{x \,|\, \theta_\Omega(x) = 0\} \tag{40}$$

Now, suppose that $\hat{x} \notin \Delta$; therefore $\theta_\Omega(\hat{x}) < 0$. Then, for any $\Omega' \subset \Omega$, we get, from (38)

$$\theta_{\Omega'}(\hat{x}) = \theta_\Omega(\hat{x}) + [\psi_\Omega(\hat{x}) - \psi_{\Omega'}(\hat{x})] \tag{41}$$

Since $\theta_\Omega(\cdot)$ is continuous and by the Corollary to Proposition 1, $\psi_{\Omega'}(\cdot)$ is continuous uniformly in Ω', there exist $\hat{\delta} \in (0,1)$ and $\hat{\rho} > 0$ such that

$$\hat{\delta}\psi_\Omega(\hat{x}) \geq \psi_\Omega(\hat{x}) + \frac{1}{4}\theta_\Omega(\hat{x}) \tag{42}$$

$$\theta_\Omega(x) \leq \frac{1}{2}\theta_\Omega(\hat{x}) \text{ for all } x \in B(\hat{x},\hat{\rho}) \tag{43}$$

and for any $\Omega' \subset \Omega$ such that $\psi_{\Omega'}(\hat{x}) \geq \hat{\delta}\psi_\Omega(\hat{x})$,

$$\psi_{\Omega'}(x) \geq \psi_\Omega(x) + \frac{1}{2}\theta_\Omega(x) \text{ for all } x \in B(\hat{x},\hat{\rho}) \tag{44}$$

Hence, from (41) and (44), for all $x \in B(\hat{x},\hat{\rho})$,

$$\theta_{\Omega'}(x) \leq \theta_\Omega(x) - \frac{1}{2}\theta_\Omega(x) \leq \frac{1}{4}\theta_\Omega(\hat{x}) \triangleq -\hat{\mu} \tag{45}$$

which completes our proof. ☐

Next, continuing in the simplified framework of the problem $P_\Omega: \min\{f(x) \mid \phi(x,\omega) \leq 0,\ \omega \in \Omega\}$, which results in no loss of generality, we define a new optimality function, which we obtain from the test in Polak's method of feasible directions [10], as follows. For any $\varepsilon \geq 0$, $\Omega' \subseteq \Omega$ and $x \in \mathbb{R}^n$, let

$$\Omega'_\varepsilon(x) \triangleq \{\omega \in \Omega' \mid \phi(x,\omega) \geq \psi_{\Omega'}(x) - \varepsilon\} \tag{46}$$

and let

$$\gamma^\varepsilon_{\Omega'}(x) = \min_{h \in C} \max\{\langle \nabla f(x), h\rangle - \psi_{\Omega'}(x);\ \langle \nabla_x \phi(x,\omega), h\rangle,\ \omega \in \Omega_\varepsilon(x)\} \tag{47}$$

with $C = \{h \in \mathbb{R}^n \mid |h^i| \leq 1,\ i = 1,2,\ldots,n\}$. Let $\beta \in (0,1)$, $\rho > 0$ be given. Then we define

$$\theta_{\Omega'}(x) \triangleq \min_k\{-\beta^k \rho \mid \gamma^\varepsilon_{\Omega'}(x) \leq -\varepsilon,\ \varepsilon = \beta^k \rho,\ k = 0,1,2,3,\ldots\} \tag{48}$$

It is easy to show (see [10]) that

$$\Delta = \{x \in \mathbb{R}^n \mid \gamma^0_\Omega(x) = 0,\ \psi_\Omega(x) = 0\} \tag{49}$$

Since $\gamma^\varepsilon_\Omega(x) \geq \gamma^0_\Omega(x)$ for all $\varepsilon \geq 0$ and $0 \geq \gamma^\varepsilon_\Omega(x)$ always holds, we must have also that

$$\Delta = \{x \in \mathbb{R}^n \mid \theta_\Omega(x) = 0,\ \psi_\Omega(x) = 0\} \tag{50}$$

If we assume that Assumption 3 holds, i.e. that $0 \notin \underset{\omega \in \Omega_0(x)}{\mathrm{co}} \nabla_x \phi(x,\omega)$ for all $x \in \mathbb{R}^n$, then $\psi_\Omega(x) = 0$ can be removed from (50), since, in that case, $\gamma^0_\Omega(x) < 0$ for all x such that $\psi_\Omega(x) > 0$.

<u>Lemma 1</u>: Suppose that Assumption 3 is satisfied and let $\Omega' \subseteq \Omega$ be any compact set. Under these assumptions,

a) If $\hat{x}$ is optimal for $P_{\Omega'}$ then $\theta_{\Omega'}(\hat{x}) = 0$. Furthermore,

$$\Delta = \{x \in \mathbb{R}^n \,|\, \theta_{\Omega}(x) = 0\} \tag{51}$$

b) For any $\hat{x} \in \mathbb{R}^n$ such that $\theta_{\Omega'}(\hat{x}) < 0$, there exist $\hat{\rho} < 0$ and $\hat{\varepsilon} > 0$ such that $\theta_{\Omega'}(x) \leq -\hat{\varepsilon}$ for all $x \in B(\hat{x},\hat{\rho})$.

c) If $\hat{x}$ is such that $\theta_{\Omega'}(\hat{x}) = 0$ and Ω' is finite, then $\theta_{\Omega'}(\cdot)$ is continuous of $\hat{x}$.

<u>Proof</u>: a) Referring to [10], we see that $\gamma^{0}_{\Omega'}(\hat{x}) = 0$ if $\hat{x}$ is optimal for $P_{\Omega'}$ and hence, since $\gamma^{0}_{\Omega'}(\hat{x}) \leq \gamma^{\varepsilon}_{\Omega'}(\hat{x})$ for all $\varepsilon \geq 0$, it follows from (48) that $\theta_{\Omega'}(\hat{x}) = 0$. The fact that (51) holds follows from (50) and Assumption 3 which guarantees that $\gamma^{0}_{\Omega}(x) < 0$ for all x such that $\psi_{\Omega}(x) > 0$.

b) Suppose that $\theta_{\Omega'}(\hat{x}) < 0$ (i.e. $\gamma_{\Omega'}(\hat{x}) < 0$). Our first observation is that the map $(x,\varepsilon) \to \Omega'_{\varepsilon}(x)$ is u.s.c., i.e., given $\hat{x}$ and $\hat{\delta} > 0$, there exist $\hat{\varepsilon}_o > 0$ and $\hat{\rho}_o > 0$ such that

$$\Omega'_{\varepsilon}(x) \subset N_{\hat{\delta}}(\hat{x}) \text{ for all } \varepsilon \in [0,\hat{\varepsilon}_o],\ x \in B(\hat{x},\hat{\rho}_o) \tag{52}$$

where

$$N_{\hat{\delta}}(\hat{x}) \triangleq \{\omega \in \mathbb{R}^p \,|\, \|\omega-\omega'\| \leq \hat{\delta}, \text{ for some } \omega' \in \Omega'_0(\hat{x})\} \tag{53}$$

i.e. $N_{\delta}(\hat{x}) \supset \Omega'_0(\hat{x})$. Let $\hat{\delta} > 0$ be such that, with $N_{\hat{\delta}}(\hat{x})$ defined by (53), and

$$\bar{\gamma}^{\hat{\delta}}_{\Omega'}(x) \triangleq \min_{h\in C} \max\{\langle \nabla f(x),h\rangle - \psi_{\Omega'}(x);\ \langle \nabla_x\phi(x,\omega),h\rangle,\ \omega \in N_{\hat{\delta}}(\hat{x})\} \tag{54}$$

we have

$$\bar{\gamma}^{\hat{\delta}}_{\Omega}(\hat{x}) \leq \gamma^{0}_{\Omega'}(\hat{x})/2 \tag{55}$$

Note that $\bar{\gamma}^{\hat{\delta}}_{\Omega'}(\cdot)$ is a continuous function.

Now, let $\hat{\varepsilon}_o > 0$, $\hat{\rho}_o > 0$ be such that (52) holds and let $\hat{\rho} \in (0,\hat{\rho}_o]$ be such that $\bar{\gamma}^{\hat{\delta}}_{\Omega'}(x) \leq \gamma^{0}_{\Omega'}(\hat{x})/4$ for all $x \in B(\hat{x},\hat{\rho})$. Then, for all

$\varepsilon \in [0,\hat{\varepsilon}_o]$ and for all $x \in B(\hat{x},\hat{\rho})$,

$$\gamma^{\varepsilon}_{\Omega'}(x) \leq \bar{\gamma}^{-\hat{\delta}}_{\Omega'}(x) \leq \gamma^{0}_{\Omega'}(\hat{x})/4$$

where the first inequality holds because $\Omega_{\varepsilon}(x) \subset N_{\hat{\delta}}(\hat{x})$. Let $\hat{k} \geq 0$ be any integer such that $\gamma^{0}_{\Omega'}(\hat{x})/4 \leq -\beta^{\hat{k}}\rho \triangleq -\hat{\varepsilon}$ and $\hat{\varepsilon} \in (0,\hat{\varepsilon}_o]$. Then, for all $x \in B(\hat{x},\hat{\rho})$

$$\gamma^{\hat{\varepsilon}}_{\Omega}(x) \leq -\hat{\varepsilon} \tag{56}$$

and therefore, by definition, $\theta_{\Omega'}(x) \leq -\hat{\varepsilon}$ for all $x \in B(\hat{x},\hat{\rho})$.

c) Now suppose that $\theta_{\Omega'}(\hat{x}) = 0$, and for the sake of contradiction, suppose that $\theta_{\Omega'}(\hat{x})$ is not continuous at $\hat{x}$. Then there exists a sequence $x_i \to \hat{x}$ as $i \to \infty$ and a $\delta > 0$ such that

$$\theta_{\Omega'}(x) \leq -\delta < 0 \text{ for all } i \tag{57}$$

Since Ω' is discrete, there exists a $\hat{\rho} > 0$ such that $\Omega'_{\delta/2}(x) \supset \Omega'_0(\hat{x})$ for all $x \in B(\hat{x},\hat{\rho})$. Hence, by continuity of $\bar{\gamma}^{0}_{\Omega'}(\cdot)$ ($\hat{\delta} = 0$ in (54)), there exists an $i_o \geq 0$ such that

$$-\frac{\delta}{2} < \bar{\gamma}^{o}_{\Omega'}(x_i) \leq \gamma^{\delta/2}_{\Omega'}(x_i) \text{ for all } i \geq i_o \tag{58}$$

But this implies that $\theta_{\Omega'}(x_i) \geq -\delta/2$ for all $i \geq i_o$ which contradicts (57). This completes our proof.

□

<u>Theorem 5</u>: Suppose that Assumption 3 is satisfied, then the optimality functions $\theta_{\Omega'}(\cdot)$ defined by (48) satisfy Assumption 1.

<u>Proof</u>: Suppose $\hat{x} \notin \Delta$. Then, by Lemma 1, $\theta_{\Omega}(\hat{z}) < 0$ and there exist $\rho_1 > 0$ and $\hat{\varepsilon} > 0$ such that

$$\theta_{\Omega}(x) \leq -\hat{\varepsilon} \text{ for all } x \in B(\hat{x},\rho_1) \tag{59}$$

Let $\hat{\delta} \in (0,1)$ be such that

$$\hat{\delta}\psi_{\Omega}(\hat{x}) \geq \psi_{\Omega}(\hat{x}) - \hat{\varepsilon}/4 \tag{60}$$

Since Ω is compact, $\phi(\cdot,\omega)$ is continuous, uniformly in $\omega \in \Omega$, and hence there exists a $\hat{\rho} \in (0,\rho_1]$ such that for any finite subset $\Omega' \subset \Omega$ and for all $x \in B(\hat{x},\hat{\rho})$

$$\psi_{\Omega'}(x) - \psi_{\Omega'}(\hat{x}) \geq \phi(x,\hat{\omega}) - \phi(\hat{x},\hat{\omega}) \geq -\hat{\varepsilon}/8 \tag{61}$$

where $\hat{\omega} \in \arg\max_{\omega\in\Omega'} \phi(\hat{x},\omega)$, and also (by continuity of $\psi_{\Omega}(\cdot)$)

$$\psi_{\Omega}(x) \leq \psi_{\Omega}(\hat{x}) + \hat{\varepsilon}/8 \tag{62}$$

Now suppose that $\Omega' \subset \Omega$ is a finite set satisfying

$$\psi_{\Omega'}(\hat{x}) \geq \hat{\delta}\psi_{\Omega}(\hat{x}) \geq \psi_{\Omega}(\hat{x}) - \hat{\varepsilon}/4 \tag{63}$$

and suppose that $x \in B(\hat{x},\hat{\rho})$. Then, making use of (61) and (62) and (63), we obtain

$$\psi_{\Omega'}(x) \geq \psi_{\Omega'}(\hat{x}) - \hat{\varepsilon}/8 \geq \psi_{\Omega}(\hat{x}) - \hat{\varepsilon}/8 - \hat{\varepsilon}/4 \geq \psi_{\Omega}(x) - \hat{\varepsilon}/2 \tag{65}$$

Therefore, since $\Omega'_{\hat{\varepsilon}/2}(x) \subset \Omega'_{\hat{\varepsilon}}(x) \subset \Omega_{\hat{\varepsilon}}(x)$ always holds, for all $x \in B(\hat{x},\hat{\rho})$, we obtain

$$\begin{aligned}
\gamma_{\Omega'}^{\hat{\varepsilon}/2}(x) &= \min_{h\in C} \max\{\langle \nabla f(x),h\rangle - \psi_{\Omega'}(x);\ \langle \nabla_x\phi(x,\omega),h\rangle,\ \omega \in \Omega'_{\hat{\varepsilon}/2}(x)\} \\
&\leq \min_{h\in C} \max\{\langle \nabla f(x),h\rangle - \psi_{\Omega}(x) + \hat{\varepsilon}/2;\ \hat{\varepsilon}/2 + \langle \nabla_x\phi(x,\omega),h\rangle, \\
&\qquad \omega \in \Omega_{\hat{\varepsilon}}(x)\} = \gamma_{\Omega}^{\hat{\varepsilon}}(x) + \hat{\varepsilon}/2 \leq -\hat{\varepsilon}/2
\end{aligned} \tag{65}$$

The last inequality follows because, with $\varepsilon(x) \triangleq -\theta_{\Omega}(x)$, by (59) and (48), for all $x \in B(\hat{x},\hat{\rho})$,

$$\gamma_{\Omega}^{\varepsilon(x)}(x) \leq -\varepsilon(x) \leq -\hat{\varepsilon} \tag{66}$$

which implies that $\hat{\varepsilon} \leq \varepsilon(x)$ for all $x \in B(\hat{x},\hat{\rho})$ and therefore,

$$\gamma_{\Omega}^{\hat{\varepsilon}}(x) \leq \gamma_{\Omega}^{\varepsilon(x)}(x) \leq -\hat{\varepsilon} \text{ for all } x \in B(\hat{x},\hat{\rho}) \tag{67}$$

It now follows from (48) and (65) that

$$\theta_{\Omega'}(x) \leq -\hat{\varepsilon}/2 \triangleq -\hat{\mu} < 0, \text{ for all } x \in B(\hat{x},\hat{\rho}) \tag{68}$$

which completes our proof.

$\square$

To conclude this section, we show that penalty functions can also be used as subroutines for solving the problems $P_{\Omega'}$ in our outer approximations methods. We continue to restrict ourselves to the special case where P_{Ω} is $\min\{f(x) \mid \phi(x,\omega) \leq 0,\ \omega \in \Omega\}$, since there is no loss of generality is doing so, but the notational simplification is great.

Let $\{s_i\}_{i=1}^{\infty}$ be an infinite sequence such that $s_i > 0$ and $s_i \to 0$ as $i \to \infty$ (e.g. $s_i = s_o/i$, or $s_i = \beta^i$, with $\beta \in (0,1)$), and let $\Omega' \subset \Omega$ be any finite set, with cardinality $\nu_{\Omega'}$. Then we define, $p_{\Omega'}:\mathbb{R}^n \to \mathbb{R}^1$ and, for $i = 1,2,3,\ldots$, $f^i_{\Omega'}:\mathbb{R}^n \to \mathbb{R}^1$ by

$$p_{\Omega'}(x) \triangleq \frac{1}{\nu'_{\Omega}} \sum_{\omega \in \Omega'} [\max\{0,\phi(x,\omega)\}]^2 \tag{69}$$

and

$$f^i_{\Omega'}(x) \triangleq f(x) + \frac{1}{s_i} p_{\Omega'}(x) \tag{70}$$

We then define the optimality functions $\theta^i_{\Omega'}(\cdot)$ by

$$\theta^i_{\Omega'}(x) \triangleq -\|\nabla f^i_{\Omega'}(x)\|,\ \Omega' \subset \Omega,\ i = 1,2,3,\ldots \tag{71}$$

with Ω' always a finite subset of Ω. A standard assumption in penalty function methods is that for any x such that $\psi_{\Omega'}(x) > 0$, $\nabla p_{\Omega'}(x) \neq 0$ (see [10]) or the somewhat stronger assumption that $0 \notin \underset{\omega \in \Omega'}{\text{co}}\ \nabla\phi(x,\omega)$

where co denotes the convex hull of the set specified. When extended to the problem P_Ω, the latter assumption becomes $0 \notin \underset{\omega\in\Omega}{\mathrm{co}}\ \nabla\phi(x,\omega)$ for all x such that $\psi_\Omega(x) > 0$. This, in turn, leads to the following assumption which we shall need to show that the optimality functions (71) satisfy Assumption 1.

<u>Assumption 4</u>: For every $\varepsilon > 0$, there exists an $\eta > 0$ such that for any $x \in \mathbb{R}^n$, any $\Omega' \subset \Omega$ finite, if $\psi_{\Omega'}(x) \geq \varepsilon$, then $\|\nabla p_{\Omega'}(x)\| \geq \eta$.
$\square$

<u>Theorem 7</u>: Suppose Assumption 4 is satisfied. Then the family of optimality functions defined by (72) satisfies Assumption 1.

<u>Proof</u>: Let $\hat{x} \in \mathbb{R}^n$ be such that $\hat{x} \notin \Delta$.

a) Suppose that $\psi_\Omega(\hat{x}) = 0$ and that the $\theta^i_{\Omega'}(\cdot)$ do not satisfy Assumption 1 at $\hat{x}$. Then, since $\psi_{\Omega'}(x) \geq \hat{\delta}\psi_\Omega(\hat{x})$ for any $\hat{\delta} > 0$ and $\Omega' \subset \Omega$, we can construct sequences $\{x_i\}_{i=1}^\infty$, $\{\Omega_i\}_{i=1}^\infty$, $\{\mu_i\}_{i=1}^\infty$, such that $x_i \to \hat{x}$ as $i \to \infty$, Ω_i are discrete subsets of Ω, $\mu_i > 0$ and $\mu_i \to 0$ as $i \to \infty$, such that

$$\theta_{\Omega_i}(x_i) = -\|\nabla f(x_i) + \frac{1}{\nu_\Omega} \sum_{\omega\in\Omega_i,} \frac{2}{s_i} \max\{0,\phi(x,\omega)\}\nabla_x\phi(x_i,\omega)\}\| \geq -\mu_i \tag{72}$$

If there exists an infinte subsequence $\{\Omega_i\}_{i\in K}$, $K \subset \{1,2,\ldots\}$ such that $\Omega_i = \phi$ for all $i \in K$, then (72) implies that $\nabla f(x_i) \overset{K}{\to} 0$ as $i \to \infty$ and hence that $\nabla f(\hat{x}) = 0$. But this is impossible since $\hat{x} \notin \Delta$. Hence no such subsequence satisfying (72) can exist. Now, let $\nu_i = \nu_{\Omega_i}$ and let

$$\pi_i \triangleq \frac{1}{\nu_i} \sum_{\omega\in\Omega_i} \frac{2}{s_i} \max\{0,\phi(x,\omega)\nabla_x\phi(x,\omega) \tag{73}$$

Then, from (72)

$$\|\pi_i\| \leq \|\nabla f(x_i)\| + \mu_i,\ i = 1,2,3,\ldots \tag{74}$$

which shows that the $\|\pi_i\|$ are bounded. Consequently, there exists a bound M, such that, because of Caratheodory's theorem [14],

$$\pi_i = \sum_{k=1}^{p+1} \lambda_i^k \nabla\phi(x_i,\omega_i^k),\ i = 1,2,3,... \tag{75}$$

with $\omega_k^i \in \Omega$, and $0 \leq \lambda_i^k \leq M$. Since Ω is compact, there must exist an infinite subset $K' \subset \{1,2,3,...\}$ such that $\omega_i^k \overset{K'}{\to} \hat{\omega}^k$, $k = 1,2,...,p+1$, as $i \to \infty$, and $\lambda_i^k \to \hat{\lambda}^k \geq 0$, $k = 1,2,...,p+1$, as $i \to \infty$. Substituting into (72) and taking limits, we get that

$$\nabla f(\hat{x}) + \sum_{k=1}^{p+1} \lambda_i^k \nabla_x \phi(\hat{x},\hat{\omega}^k) = 0 \tag{76}$$

But this shows that $\hat{x}$ satisfies the Kuhn-Tucker conditions and hence $\hat{x} \in \Delta$ (which only requires the F. John condition) and hence we get a contradiction. Thus, the $\theta_{\Omega'}^i(\cdot)$ satisfy Assumption 1 at any $\hat{x} \notin \Delta$ such that $\psi_\Omega(\hat{x}) = 0$.

b) Now suppose that $\psi_\Omega(\hat{x}) > 0$. Let $\hat{\delta} > 0$ be arbitrary. Then, by Assumption 4, there exists a $\hat{\mu} > 0$ such that $\|\nabla p_{\Omega'}(x)\| \geq \hat{\mu}$ for all $x \in \mathbb{R}^n$ $\Omega' \subset \Omega$ finite, such that $\psi_{\Omega'}(x) > \frac{\hat{\delta}}{2}\psi_\Omega(\hat{x})$.

Since the $\psi_{\Omega'}(\cdot)$ functions are continuous, uniformly in Ω' (see Corollary to Proposition 1), there exists a $\hat{\rho} > 0$ such that if $\Omega' \subset \Omega$ is finite and if $\psi_{\Omega'}(\hat{x}) \geq \hat{\delta}\psi_\Omega(\hat{x})$, then $\psi_{\Omega'}(x) \geq \frac{1}{2}\hat{\delta}\psi_\Omega(\hat{x})$ for all $x \in B(\hat{x},\hat{\rho})$ and consequently $\|\nabla p_{\Omega'}(x)\| \geq \hat{\mu}$ for all $x \in B(\hat{x},\hat{\rho})$. Hence, if $\psi_{\Omega'}(\hat{x}) \geq \hat{\delta}\psi_\Omega(\hat{x})$ and $x \in B(\hat{x},\hat{\rho})$,

$$\begin{aligned}\theta_{\Omega'}^i(x) &= -\|\nabla f(x) + \frac{1}{s_i}\nabla p_{\Omega'}(x)\| \leq -\frac{1}{s_i}\|\nabla p_{\Omega'}(x)\| + \|\nabla f(x)\| \\ &\leq -\frac{1}{s_i}\hat{\mu} + M\end{aligned} \tag{77}$$

where $M = \max\{\|\nabla f(x)\| \,|\, x \in B(\hat{x},\hat{\rho})\}$. Since $s_i \to 0$ as $i \to \infty$, there exists an N such that $-(\frac{1}{s_i}\hat{\mu}+M) \leq -\hat{\mu}$ for all $i \geq N$ and hence we see that Assumption 1 holds at $\hat{x}$. This completes our proof. ¤

Conclusion

The algorithms described in this paper are directly implementable, since all the required computations can be carried out by a finite number of operations. They differ from the earlier versions of outer approximations algorithms in four respects. They require more storage (scheme 3), but they have better convergence properties, they are implementable, and they have more parameters to be specified by the user. The last property may be considered to be undesirable by some. However, we feel that this freedom to select parameters is very important, since it can be exploited constructively in interactive computing schemes utilizing a graphic display terminal.

REFERENCES

[1] Cheney, E. W, and Goldstein, A. A., "Newton's Method for Convex Programming and Tchebycheff Approximation," Numerische Matematik, I, pp. 253-68, 1959.

[2] Kelley, J. E., "The Cutting-Plane Method for Solving Convex Programs," J. SIAM, VIII, No. 4, pp. 703-12, 1960.

[3] Levitin, E. S. and Polyak, B. T., "Constrained Minimization Methods," Zh. Vychisl. Mat. i Mat Fiz. Vol. 6, No. 5, pp. 781-823, 1966.

[4] Topkis, D. M., "Cutting Plane Methods Without Nested Constraints," Opns. Res. 18, pp. 404-413, 1970.

[5] Topkis, D. M., "A Note on Cutting Plane Methods Without Nested Constraints," Opns. Res. Center, University of California, Report No. 69-36, 1969.

[6] Eaves, B. C. and Zangwill, W. I., "Generalized Cutting Plane Algorithms," SIAM, J. Control, Vol. 9, No. 4, pp. 529-542.

[7] Hogan, W. W., "Applications of a General Convergence Theory for Outer Approximations Algorithms," Math. Progr. Vol. 5, No. 2, pp. 151-168, 1973.

[8] Mayne, D. Q., Polak, E. and Trahan, R., "An Outer Approximations Algorithm for Computer-Aided Design Problems," Memo #M77-10, Electronics Res. Lab., University of California, Berkeley, Feb. 1977.

[9] Topkis, D. M. and Veinott, A., "On the Convergence of Some Feasible Directions Algorithms for Nonlinear Programming," J. SIAM, Contr. Vol. 5, No. 2, pp. 268-79 (1967).

[10] Polak, E., "Computational Methods in Optimization," Academic Press, 1971.

[11] Polak, E., Trahan, R. and Mayne, D. Q., "Combined Phase I-Phase II Methods of Feasible Directions," Memo #M77-39, Electronics Res. Lab., University of California, Berkeley.

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

[13] Pironneau, O. and Polak, E., "Rate of Convergence of a Class of Methods of Feasible Directions," J. SIAM Num. Anal. Vol. 10, No. 1, pp. 161-173, 1973.

[14] Rockafellar, R. T., "Convex Analysis," Princeton Univ. Press, 1970.