THEORETICAL AND COMPUTATIONAL ASPECTS OF THE OPTIMAL DESIGN CENTERING, TOLERANCING AND TUNING PROBLEM

by

E. Polak and A. Sangiovanni-Vincentelli

Memorandum No. UCB/ERL M79/6

8 January 1979

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

THEORETICAL AND COMPUTATIONAL ASPECTS OF THE OPTIMAL DESIGN CENTERING, TOLERANCING AND TUNING PROBLEM

E. Polak and A. Sangiovanni-Vincentelli

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

ABSTRACT

The optimal design centering, tolerancing and tuning problem is transcribed into a mathematical programming problem of the form $P_g : \min\{f(x) \mid \max_{\omega\in\Omega} \min_{\tau\in T} \max_{j\in J} \zeta^j(x,\omega,\tau) \leq 0,\ x \leq 0\}$, $x,\omega,\tau \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}^1$, $\zeta : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to R^1$, continuously differentiable, Ω and T compact subsets of $\mathbb{R}^n$, $J = \{1,\ldots,p\}$. A simplified form of P_g, $P : \min\{f(x) \mid \psi(x) \triangleq \max_{\omega\in\Omega} \min_{\tau\in T} \zeta(x,\omega,\tau) \leq 0\}$ is discussed. It is shown that $\psi(\cdot)$ is locally Lipschitz continuous but not continuously differentiable. Optimality conditions for P based on the concept of generalized gradients are derived. An algorithm, consisting of a master outer approximations algorithm proposed by Gonzaga and Polak and of a new subalgorithm for nondifferentiable problems of the form $P_i : \min\{f(x) \mid \max_{\omega\in\Omega_i} \min_{\tau\in T} \zeta(x,\omega,\tau) \leq 0\}$, where Ω_i is a discrete set, is presented. The subalgorithm is an extension of Polak's method of feasible directions to nondifferentiable problems. The overall algorithm is shown to converge under suitable assumptions.

Research sponsored by the National Science Foundation Grant ENG73-80214-A02 and the Air Force Office of Scientific Research Contract F44620-76-C-0100 and the National Science Foundation (RANN) Grant ENV76-04264.

I. Introduction

A typical electronic design problem involves the determination of optimal values of the design parameters which are constrained to satisfy performance specifications in the presence of tolerances. An important aspect in the manufacture of high performance integrated circuits is the possibility of tuning. In general, tuning is performed by trimming some of the circuit resistors by specialized technological processes, such as by cutting the resistor bodies by a laser beam. While the possibility of tuning a manufactured circuit enables one to increase yield considerably, it also leads to very considerable mathematical and computational difficulties in the optimal design of electronic circuits.

A sophisticated optimal design of an electronic circuit involves the selection of optimal nominal values for the design parameters (center of design), the optimal assignment of tolerances and the optimal determination of the tuning range in order to obtain a specified yield with minimum manufacturing costs. In addition to requiring algorithm for solving the above design problem, it is also necessary to have algorithms for tuning. These algorithms are used after an electronic circuit sample is manufactured, to compute, on the basis of measurements, the amounts by which the circuit elements have to be trimmed.

In [1], Bandler et al. formulated the design centering, tolerancing and tuning problem as a mathematical programming problem. They also discussed the geometrical structure of the problem and introduced some important special cases. However, they did not propose any general algorithm for the solution of this very difficult mathematical programming problem. Algorithms for post manufacture tuning have been discussed in [2-6]. The

approach proposed by Lopresti [2], in particular, is worth mentioning. He proposed a very ingenious optimal tuning scheme based on a transcription of the trimming problem into a quadratic regulator problem.

In Section 2, we shall present a transcription of the Design Centering, Tolerancing and Tuning (DCTT) problem into a mathematical programming problem. This transcription of the DCTT problem will be shown to be equivalent to the one proposed by Bandler et al. [1]. However, it has substantial advantages over the one in [1] from an algorithmic point of view. Our formulation involves functional constraints of the form $\max_{\omega\in\Omega} \min_{\tau\in T} \max_{j\in J} \zeta^j(x,\omega,\tau) \leq 0$, $x,\omega,\tau \in \mathbb{R}^n$, Ω and T compact subsets of $\mathbb{R}^n$, $J = \{1,\ldots,p\}$. While there is a good number of algorithms for optimization problems with constraints of the form $g(x) \leq 0$ (e.g. feasible direction methods, penalty function methods, multiplier methods), only one or two general algorithms with options [7-9] are available for the problem with functional constraints of the form $\max_{\omega\in\Omega} \chi(x,\omega) \leq 0$. When Ω is a polyhedron and $\chi(x,\cdot)$ is convex or one dimensional convex [10], this functional inequality becomes a finite set of inequalities of the form $g(x) \leq 0$ and hence we can use the previously mentioned efficient algorithms. To our knowledge, no algorithms exist in the optimization literature for problems with functional constraints of more complex forms $\max_{\omega\in\Omega} \min_{\tau\in T} \zeta(x,\omega,\tau) \leq 0$ or $\max_{\omega\in\Omega} \min_{\tau\in T} \max_{j\in J} \zeta^j(x,\omega,\tau) \leq 0$.

In Section 3, we concentrate on the simplified DCTT problem involving a functional constraint of the form $\psi(x) \triangleq \max_{\omega\in\Omega} \min_{\tau\in T} \zeta(x,\omega,\tau) \leq 0$. We discuss the differentiability property of $\psi(\cdot)$ and we derive some necessary optimality conditions for the DCTT problem, based on the concept of generalized gradients introduced by Clarke [14,15] for nondifferentiable

optimization problems.

In Section 4, we propose a new algorithm for solving the simplified form of the DCTT problem. This algorithm is based on an outer approximations scheme proposed by Gonzaga and Polak [9]. It requires as a subroutine an algorithm for the solution of a simpler problem of the form $\min\{f(x) \mid \max_{\omega\in\Omega_i} \min_{\tau\in T} \zeta(x,\omega,\tau) \leq 0\}$ where $\Omega_i \subset \Omega$ is a discrete set. We propose two subroutine algorithms for these simpler problems: a conceptual[†] version and a semi-implementable one. The algorithms are extensions of Polak's method of feasible directions [11] to the nondifferentiable case. The algorithms are shown to converge to points satisfying the optimality condition derived in Section 3.

In Section 5, we discuss the simplified DCTT problem under the hypothesis that $\zeta(\cdot,\cdot,\cdot)$ is convex or one dimensional convex [10].

2. Formulation of the DCTT Problem

Let $\phi_o \in \mathbb{R}^n$ be the nominal design (expressed in terms of n parameters); let $\varepsilon \in \mathbb{R}^n$ be the relative tolerance vector[††] and $\phi \in \mathbb{R}^n$ be the tuning vector. Any outcome $\phi \in \mathbb{R}^n$ of the manufacturing process is required to satisfy:

$$\phi_o - E\phi_o \leq \phi \leq \phi_o + E\phi_o \tag{2.1}$$

where

$$E \triangleq \text{diag}(\varepsilon) \tag{2.2}$$

[†] By conceptual we mean an algorithm with unspecified truncation rules for operations that cannot be easily approximated.

[††] We use Bandler's notation in [1] as much as possible. However for the sake of generality we deal with relative tolerances instead of absolute tolerances as in [1].

and $\varepsilon_i \geq 0$, $i = 1,\ldots,n$.

If we introduce a set of scaling parameters $-1 \leq \omega^i \leq 1$, $i = 1,\ldots,n$, then, any outcome of the manufacturing process can be represented as

$$\phi = \phi_o + EQ\phi_o \tag{2.3}$$

where

$$Q \triangleq \text{diag}(\omega). \tag{2.4}$$

Let $\Omega \subset \mathbb{R}^n$ be the subset defined by

$$\Omega = \{\omega \in \mathbb{R}^n \mid -1 \leq \omega^i \leq 1,\ i = 1,\ldots,n\}. \tag{2.5}$$

The tuning vector $\xi \in \mathbb{R}^n$, $\xi \geq 0$, represents the tuning range of each parameter. Therefore an outcome which has been tuned can be expressed by

$$\phi = (I+EQ)\phi_o + \Xi\tau \tag{2.6}$$

where I is the identity matrix,

$$\Xi \triangleq \text{diag}(\xi) \tag{2.7}$$

and $\tau \in \mathbb{R}^n$ is a scaling vector representing the actual amount of trimming used. In general, according to the particular tuning procedure followed, $\tau \in T \subset \mathbb{R}^n$. If two-way tuning (or reversible tuning) is considered, we have

$$T = \{\tau \in \mathbb{R}^n \mid -1 \leq \tau^i \leq 1,\ i = 1,\ldots,n\}. \tag{2.8a}$$

If one-way (or irreversible) tuning (e.g. laser beam trimming of integrated circuits) is considered, we have

$$T = \{\tau \in \mathbb{R}^n \mid 0 \leq \tau^i \leq 1,\ i = 1,\ldots,n\}. \tag{2.8b}$$

The design specifications can be transcribed into inequalities of the form

$$g(\phi) \leq 0 \tag{2.9}$$

where $g : \mathbb{R}^n \to \mathbb{R}^p$. We can assume $g(\cdot)$ to be continuously differentiable (C^1), since that is usually the case in electronic design problems.

In general, the objective function represents the cost of the manufacturing process. We assume that the cost depends on ϕ_o, ε and ξ. We shall denote by $C : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$, the function expressing the cost of the manufacturing process. In the context of electronic design problems, we can assume $C(\cdot,\cdot,\cdot)$ to be continuously differentiable. In [1], the DCTT problem is transcribed into a mathematical programming problem of the following form, where

$$R_o = \{\phi \in \mathbb{R} \mid g(\phi) \leq 0\}: \tag{2.10}$$

$$\tilde{P}: \begin{cases} \text{minimize } C(\phi_o,\varepsilon,\xi) \\ \text{subject to} \\ \forall \omega \in \Omega, \exists \tau \in T \text{ such that} \\ \phi = \phi_o + EQ\phi_o + \Xi\tau \in R_o \qquad (2.11) \\ \text{and} \\ \phi_o,\varepsilon,\xi \geq 0. \qquad (2.12) \end{cases}$$

The constraints (2.11) are not stated in a form which is compatible with algorithmic evaluations. We propose an alternative transcription of $\tilde{P}$ which is suitable for solution by computation.

We define the problem P by:

$$P: \begin{cases} \text{minimize } C(\phi_o,\varepsilon,\xi) \\ \text{subject to} \\ \max_{\omega\in\Omega} \min_{\tau\in T} \max_{j\in J} g^j(\phi) \leq 0;\ J = \{1,\ldots,p\}. \qquad (2.13) \\ \text{and} \\ \phi_o,\varepsilon,\xi \geq 0. \qquad (2.14) \end{cases}$$

Let us see how constraints of the form (2.13) follow from (2.11), and vice versa. The design center ϕ^o is obviously chosen so that $g(\phi^o) \leq 0$. The production process results in a realization with a tolerance $EM\phi^o$ which may result in $g(\phi^o+EM\phi^o) > 0$. The most manufacture tuning process then chooses a $\Xi\tau$ so that $g(\phi^o+EM\phi^o+\Xi\tau) \leq 0$. Since the production process is the "adversary", it results in the first $\max_{\omega\in\Omega}$. The tuning process then does its very best to rectify matters, which is expressed as $\min_{\tau\in T}$. Now the trimming must work for all the specification constraints $g^j(\cdot)$ and hence we get the last $\max_{j\in J}$. Let us now prove rigorously the equivalence between P and $\tilde{P}$.

Proposition 2.1. P is equivalent to $\tilde{P}$.

Proof. Since the objective function $C(\cdot,\cdot,\cdot)$ is the same, and the constraints (2.14) and (2.12) are identical, we have only to prove that a set of variables $x^T = [\phi_o^T, \varepsilon^T, \xi^T]$ satisfies (2.11) if and only if it satisfies (2.13).

For the sake of contradiction, suppose that there exists an $\bar{x}^T = [\bar{\phi}_o^T, \bar{\varepsilon}^T, \bar{\xi}^T]$ which is feasible for $\tilde{P}$ but not for P. Then, there exists at least one index $\bar{j}$, such that

$$\max_{\omega\in\Omega} \min_{\tau\in T} \max_{j\in J} g^j(\bar{\phi}_o+\bar{E}\bar{Q}\bar{\phi}_o+\bar{\Xi}\tau) = \max_{\omega\in\Omega} \min_{\tau\in T} g^{\bar{j}}(\bar{\phi}_o+\bar{E}\bar{Q}\bar{\phi}_o+\bar{\Xi}\tau) > 0. \qquad (2.15)$$

Let $\bar{\omega}$ be a maximizer of (2.15). Then,

$$\min_{\tau\in T} g^{\bar{j}}(\bar{\phi}_o+\bar{E}\bar{Q}\bar{\phi}_o+\bar{\Xi}\tau) > 0. \qquad (2.16)$$

Therefore, there exists at least one ω, viz. $\bar{\omega}$, such that

$$\forall\tau \in T,\ g^{\bar{j}}(\bar{\phi}_o+\bar{E}\bar{Q}\bar{\phi}_o+\bar{\Xi}\tau) > 0 \qquad (2.17)$$

and $\bar{x}$ cannot be feasible for $\tilde{P}$. Now, suppose that $\bar{x}$ is feasible for P but not for $\tilde{P}$. Then, there exists at least one $\omega \in \Omega$, say $\hat{\omega}$, such that for some $\bar{j} \in \{1,2,\ldots,p\}$,

$$\forall \tau \in T,\ g^{\bar{j}}(\bar{\phi}_o + \bar{E}\hat{Q}\bar{\phi}_o + \bar{\Xi}\tau) > 0. \tag{2.18}$$

Therefore

$$\min_{\tau \in T} g^{\bar{j}}(\bar{\phi}_o + \bar{E}\hat{Q}\bar{\phi}_o + \bar{\Xi}\tau) > 0. \tag{2.19}$$

But then

$$\max_{\omega\in\Omega} \min_{\tau\in T} \max_{j\in J} g^{j}(\bar{\phi}_o + \bar{E}Q\bar{\phi}_o + \bar{\Xi}\tau) \geq \max_{\omega\in\Omega} \min_{\tau\in T} g^{\bar{j}}(\bar{\phi}_o + \bar{E}Q\bar{\phi}_o + \bar{\Xi}\tau) \geq \min_{\tau\in T} g^{\bar{j}}(\bar{\phi}_o + \bar{E}\hat{Q}\bar{\phi}_o + \bar{\Xi}\tau) > 0 \tag{2.20}$$

and $\bar{x}$ is not feasible for P, contradicting our hypothesis. ∎

For the sake of notational conciseness, we aggregate the variables $(\phi_o, \varepsilon, \xi)$ into one vector x and rename all functions to obtain a more general form of the problem P:

$$P_g : \min\{f(x) \,|\, \max_{\omega\in\Omega} \min_{\tau\in T} \max_{j\in J} \zeta^{j}(x,\omega,\tau) \leq 0,\ x \geq 0\} \tag{2.21}$$

where $x^T = [\phi_o^T, \varepsilon^T, \xi^T]$, $\zeta^j(x,\omega,\tau) = g^j(\phi_o + EQ\phi_o + \Xi\tau)$ and $f(x) = C(\phi_o, \varepsilon, \xi)$. In the following sections, we will concentrate on the simplest version of P_g, viz.,

$$P : \min\{f(x) \,|\, \max_{\omega\in\Omega} \min_{\tau\in T} \zeta(x,\omega,\tau) \leq 0\}. \tag{2.22}$$

The understanding of the mathematical properties of P and the construction of an algorithm for its solution are a fundamental step towards the solution of the more complicated DCTT problem, P_g.

3. Continuity, Differentiability and Optimality Conditions for the Design Centering, Tolerancing and Tuning Problem

The difficulty in solving P stems from two sources: the reasonably obvious one being that functions of the form (2.13) are quite difficult to evaluate, and the not so obvious one being that these functions are not differentiable. We now summarize the relevant properties of the function $\psi(x) \triangleq \max_{\omega\in\Omega} \min_{\tau\in T} \zeta(x,\omega,\tau)$ which appears in the constraint of P. We shall make use of the following notations:

$$\chi(x,\omega) \triangleq \min_{\tau\in T} \zeta(x,\omega,\tau) \tag{3.1}$$

and

$$\psi(x) \triangleq \max_{\omega\in\Omega} \min_{\tau\in T} \zeta(x,\omega,\tau) = \max_{\omega\in\Omega} \chi(x,\omega). \tag{3.2}$$

Assumption 3.1. $\zeta : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^1$, is C^1 in x,ω,τ.†

The following result is proved in Appendix 0. ¤

Proposition 3.1. The function $\psi(\cdot)$ is locally Lipschitz, i.e., given any bounded subset D of $\mathbb{R}^n$, there exists an $L > 0$, such that for all $x,y \in D$

$$|\psi(x) - \psi(y)| \leq L\|x-y\|. \qquad ¤ \tag{3.3}$$

Corollary 3.1. [16]. The functions $\psi(\cdot)$ and $\chi(\cdot,\cdot)$ are differentiable almost everywhere. ¤

The following results summarizes what we know about the directional derivatives of $\psi(\cdot,\cdot)$ and $\psi(\cdot)$.

Proposition 3.2. [23]. For any $x,h \in \mathbb{R}^n$ and $\omega \in \Omega$, the directional

†Due to the particular structure of $\zeta(\cdot,\cdot,\cdot)$, if $\zeta(\cdot,\cdot,\cdot)$ is C^1 in x, it is also C^1 in ω and τ.

derivative of $\chi(\cdot,\omega)$ exists and is given by

$$d_x\chi(x,\omega;h) \triangleq \lim_{\lambda\downarrow 0} \frac{\chi(x+\lambda h,\omega)-\chi(x,\omega)}{\lambda} = \min_{\tau\in T(x,\omega)} \langle \nabla_x \zeta(x,\omega,\tau),\ h\rangle \qquad (3.4)$$

where, for any $x \in \mathbb{R}^n$, $\omega \in \Omega$,

$$T(x,\omega) \triangleq \{\tau \in T \,|\, \chi(x,\omega) = \zeta(x,\omega,\tau)\}. \qquad (3.5)$$

¤

Proposition 3.3. [23]. Let

$$\Omega(x) \triangleq \{\omega \in \Omega \,|\, \psi(x) = \chi(x,\omega)\}. \qquad (3.6)$$

If for all $x \in \mathbb{R}^n$, $\omega \in \Omega(x)$, $T(x,\omega)$ is a singleton, then for any $x,h \in \mathbb{R}^n$, the directional derivative of $\psi(\cdot)$ exists and is given by

$$d\psi(x;h) = \lim_{\lambda\downarrow 0} \frac{\psi(x+\lambda h)-\psi(x)}{\lambda} = \max_{\omega\in\Omega(x)} \min_{\tau\in T(x,\omega)} \langle \nabla_x \zeta(x,\omega,\tau),h\rangle$$

$$= \max\{\langle \nabla_x \zeta(x,\omega,\tau),h\rangle \,|\, \omega \in \Omega(x),\ \tau \in T(x,\omega)\} \qquad (3.7)$$

¤

Remark 3.1. The above condition on $T(x,\omega)$ is very strict. An obvious case in which it is indeed verified, is when $\zeta(\cdot,\cdot,\cdot)$ is strictly convex in τ and T is convex.

Proposition 3.4. Suppose that Ω is discrete, then for any $x,h \in \mathbb{R}^n$, the directional derivative of $\psi(\cdot)$ exists and is given by

$$d\psi(x;h) \triangleq \lim_{\lambda\downarrow 0} \frac{\psi(x+\lambda h)-\psi(x)}{\lambda} = \max_{\omega\in\Omega(x)} \min_{\tau\in T(x,\omega)} \langle \nabla_x \zeta(x,\omega,\tau),h\rangle . \qquad (3.8)$$

Proof. Since Ω is discrete, $\Omega(x)$ is discrete for all $x \in \mathbb{R}^n$. Suppose $\Omega(x) = \{\bar{\omega}_1,\ldots,\bar{\omega}_i,\ldots,\bar{\omega}_{k(x)}\}$, $k(x) \geq 1$. Then, by (3.6) and by continuity of $\chi(\cdot,\cdot)$, there exists $\bar{\lambda} > 0$ such that for all $\lambda \in [0,\bar{\lambda}]$,

$$\chi(x+\lambda h,\bar{\omega}_i) > \chi(x+\lambda h,\omega),\ i = 1,\ldots,k(x);\ \omega \in \Omega \sim \Omega(x). \tag{3.9}$$

Therefore, for all $\lambda \in [0,\bar{\lambda}]$,

$$\psi(x+\lambda h) = \max_{i=1,\ldots,k(x)} \{\chi(x+\lambda h,\bar{\omega}_i)\}. \tag{3.10}$$

By Proposition 3.2, for all $\lambda \in [0,\bar{\lambda}]$,

$$\begin{aligned}\psi(x+\lambda h) &= \max_{i=1,\ldots,k(x)} \{\chi(x,\bar{\omega}_i) + \lambda d_x\chi(x,\bar{\omega}_i;h) + o^i(\lambda)\} \\ &= \max_{i=1,\ldots,k(x)} \{\psi(x)+\lambda \min_{\tau\in T(x,\bar{\omega}_i)} \langle \nabla_x\zeta(x,\bar{\omega}_i,\tau),h\rangle + o^i(\lambda)\}\end{aligned} \tag{3.11}$$

By (3.8) and (3.11),

$$\begin{aligned}d\psi(x;h) = \lim_{\lambda\downarrow 0} \frac{\psi(x+\lambda h)-\psi(x)}{\lambda} &= \max_{i=1,\ldots,k(x)} \min_{\tau\in T(x,\bar{\omega}_i)} \langle \nabla_x\zeta(x,\bar{\omega}_i,\tau),h\rangle \\ &= \max_{\omega\in\Omega(x)} \min_{\tau\in T(x,\omega)} \langle \nabla_x\zeta(x,\omega,\tau),h\rangle .\end{aligned} \tag{3.12}$$

□

It is reasonable to conjecture that (3.8) is valid in the general case, i.e., when Ω is infinite and $T(x,\omega)$ is not a singleton. The following counter example due to R. T. Rockafellar [27], shows that (3.8) is incorrect in the general case.

Example 3.1. Consider

$$\psi(x) = \max_{\omega\in\Omega} \min_{\tau\in T} \tau(x-\omega) \tag{3.13}$$

with

$$\Omega = [-1,+1],\ T = [-1,+1]. \tag{3.14}$$

It is easy to see that the graph of $\psi(\cdot)$ is as shown in Figure 3.1.

Then,

$$\Omega(x) = \begin{cases} \{+1\}, & x > +1 \\ \{x\}, & -1 \leq x \leq +1 \\ \{-1\}, & x < -1 \end{cases} \tag{3.15}$$

and

$$T(x,\omega) = \begin{cases} \{-1\}, & x > \omega \\ [-1,+1], & x = \omega \\ \{+1\}, & x < \omega \end{cases} \tag{3.16}$$

If (3.8) were valid, we should have

$$d\psi(0;+1) = \max_{\omega\in\Omega(0)} \min_{\tau\in T(0,\omega)} \tau = -1 \tag{3.16a}$$

which is obviously wrong, since from Fig. 3.1, $d\psi(0;+1) = 0$. Consider now

$$\psi'(x) = \max_{\omega\in\Omega'} \min_{\tau\in T} \tau(x-\omega) \tag{3.17}$$

where $\Omega' = \{-1,+1\}$ is a <u>discrete</u> set. The graph of $\psi'(\cdot)$ is shown in Figure 3.2.

In this case

$$\Omega'(x) = \begin{cases} \{+1\}, & x > 0 \\ \{-1,+1\}, & x = 0 \\ \{-1\}, & x < 0 \end{cases} \tag{3.17a}$$

and $T(x,\omega)$ is as in (3.16).

According to (3.8), we have now

$$d\psi'(0;+1) = \max_{\omega\in\Omega(0)} \min_{\tau\in T(0,\omega)} \tau = +1, \tag{3.17b}$$

$$d\psi'(-1;+1) = \max_{\omega\in\Omega(-1)} \min_{\tau\in T(-1,\omega)} \tau = -1 \tag{3.17c}$$

and both results agree with the graph of $\psi'(\cdot)$. ¤

In the general case, it is not even known whether $\psi(\cdot)$ has directional derivatives everywhere. Thus, the function $\psi(\cdot)$ necessitates the introduction of some additional concepts of differentiability, which have been developed by F. Clarke [14,15].

Definition 3.1. The generalized gradient of a locally Lipschitz function $\psi : \mathbb{R}^n \to \mathbb{R}^1$, at x is denoted by $\partial\psi(x)$ and is defined by the convex hull of the set of all the limits of the form

$$\lim_{i\to\infty} \nabla\psi(x+v_i)$$

where $v_i \to 0$ as $i \to \infty$ and the v_i are such that $\nabla\psi(x+v_i)$ is well defined.

Example 3.2. Consider $\psi(\cdot)$ as defined by (3.13). From Fig. 3.1, it is easy to see that $\psi(\cdot)$ is differentiable almost everywhere on $\mathbb{R}^1$. When $x = +1$ and $x = -1$, $\psi(\cdot)$ is not differentiable. By Definition 3.1, we have

$$\partial\psi(x) = \begin{cases} \{-1\}, & x > 1 \\ \text{co}\{-1,0\} = [-1,0], & x = 1 \\ \{0\}, & -1 < x < 1 \\ \text{co}\{-1,0\} = [-1,0], & x = -1 \\ \{-1\}, & x < -1. \end{cases} \tag{3.18}$$

where co denotes the convex hull of the set. ¤

The generalized gradients have certain important properties which

we summarize in Proposition 3.5.

Proposition 3.5. [14,15]. Let $\psi : \mathbb{R}^n \to \mathbb{R}^1$ be a locally Lipschitz function. Let D be any open bounded subset of $\mathbb{R}^n$. Then, for all $x \in D$

(a) $\partial\psi(x)$ is a well defined nonempty convex compact subset of $\mathbb{R}^n$;

(b) the point-to-set map $\partial\psi : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is bounded on D and is upper semicontinuous on D i.e., if $\{x_i\} \subset D$ converges to $\hat{x}$ and $g_i \in \partial\psi(x_i)$ for each i, then each accumulation point $\hat{g}$ of $\{g_i\}$ satisfies $\hat{g} \in \partial\psi(x)$.† ¤

Lebourg [28] proved a very useful mean value theorem for locally Lipschitz functions based on generalized gradients.

Theorem 3.1. [28]. Let D be any bounded subset of $\mathbb{R}^n$ and $\hat{D}$ a convex subset of D. Let $\psi : \mathbb{R}^n \to \mathbb{R}^1$ be a locally Lipschitz function. For each $x, y \in \hat{D}$, there exists $\lambda \in (0,1)$ and $g \in \partial\psi(x+\lambda(y-x))$ such that

$$\psi(x) - \psi(y) = \langle g, x-y \rangle . \tag{3.19}$$

¤

The generalized gradients can be used to extend the F. John [11] optimality condition to P.

Proposition 3.6. [13]. If $\hat{x}$ is optimal for P, then

$$\text{(i)} \quad \psi(\hat{x}) \leq 0 \tag{3.20}$$

---

†This definition of upper semicontinuity due to Clarke [14,15] is different from the one given by Berge [17]: a point-to-set map A: $D \to 2^D \sim \phi$ is upper semicontinuous at D if (i) for each $x \in D$, given a $\delta > 0$, there exists $\rho > 0$ such that for all $y \in B(x,\rho) (\triangleq \{y \mid \|y-x\| \leq \rho\})$

$$N_\delta(A(x)) \triangleq A(x) + B(0,\delta) \supset A(y) \quad \text{(ii) } A(x) \text{ is compact for each } x \in D.$$

However, because $\partial\psi(\cdot)$ is compact, it is easy to see that the two definitions are equivalent as applied to $\partial\psi(\cdot)$.

and

$$\text{(ii)}\quad 0 \in \tilde{M}(\hat{x}) \tag{3.21}$$

where

$$\tilde{M}(x) = \begin{cases} \{\nabla f(x)\} & \text{if } \psi(x) < 0 \\ \text{co}\{\{\nabla f(x)\} \cup \partial\psi(x)\} & \text{if } \psi(x) \geq 0 \end{cases} \tag{3.22}$$

Remark 3.2. The set $\tilde{M}(x)$ can be defined arbitrarily for $\psi(x) > 0$, since the optimality condition involves $\psi(\hat{x}) \leq 0$. Mifflin [13] defines $\tilde{M}(x) = \partial\psi(x)$ when $\psi(x) > 0$. ¤

Proposition 3.7. [13]. The point-to-set map $\tilde{M}(\cdot)$ defined in (3.22) is bounded on bounded subsets of $\mathbb{R}^n$, upper semicontinuous on $\mathbb{R}^n$ and, for each $x \in \mathbb{R}^n$, $\tilde{M}(x)$ is convex. ¤

Now, the problem with (3.21) is that we may not have expressions for $\partial\psi(\hat{x})$ and, hence, no means for verifying (3.21). Here is what we do know.

Proposition 3.8. For any $x \in \mathbb{R}^n$ let the point-to-set map $\Gamma : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ be defined by

$$\Gamma(x) \triangleq \text{co}\{\nabla_x \zeta(x,\omega,\tau),\ \omega \in \Omega(x),\ \tau \in T(x,\omega) \tag{3.23}$$

Then, $\Gamma(\cdot)$ is upper semicontinuous. ¤

To prove this proposition, we need the following lemma.

Lemma 3.1. The point-to-set maps $\Omega(\cdot)$ and $T(\cdot,\cdot)$ defined in (3.11) and (3.9) are upper semicontinuous.

Proof. Let $\{x_i\}$ be a converging sequence and $\hat{x}$ its limit point. Let $\omega_i \in \Omega(x_i)$. Consider an infinite subset $I \subset \{1,2,\ldots.\}$ such that $\omega_i \overset{I}{\to} \hat{\omega}$. (It is always possible to find such a subset since Ω is a compact set.) For the sake of contradiction, suppose that

$$\hat{\omega} \notin \Omega(\hat{x}). \tag{3.24}$$

Let $\tilde{\omega}$ be any element of $\Omega(\hat{x})$. Then,

$$\forall i \in I,\ \chi(x_i,\omega_i) \geq \chi(x_i,\tilde{\omega}). \tag{3.25}$$

By continuity of $\psi(\cdot)$ and of $\chi(\cdot,\cdot)$,

$$\chi(\hat{x},\hat{\omega}) = \lim_{\substack{I \\ i\to\infty}} \chi(x_i,\omega_i) \geq \lim_{\substack{I \\ i\to\infty}} \chi(x_i,\tilde{\omega}) = \chi(\hat{x},\tilde{\omega}) \tag{3.26}$$

in contradiction with (3.24). The same procedure can be followed to prove that $T(\cdot,\cdot)$ is upper semicontinuous. ¤

Proof of Proposition 3.8. Let

$$\tilde{\Gamma}(x) = \{\nabla_x \zeta(x,\omega,\tau) \,|\, \omega \in \Omega(x),\ \tau \in T(x,\omega)\}. \tag{3.27}$$

By the definition of a convex hull, if $\tilde{\Gamma}(\cdot)$ is upper semicontinuous, so is $\Gamma(\cdot) = \text{co}\ \tilde{\Gamma}(\cdot)$.

Let $\{x_i\}$ be a convergent sequence and $\hat{x}$ its limit point. Let I be any infinite subset such that $z_i \overset{I}{\to} \hat{z}$, where $z_i \in \tilde{\Gamma}(x_i)$. We want to show that $\hat{z} \in \tilde{\Gamma}(\hat{x})$. Let ω_i and τ_i be such that

$$z_i = \nabla_x \zeta(x_i,\omega_i,\tau_i). \tag{3.28}$$

Hence $\omega_i \in \Omega(x_i)$ and $\tau_i \in T(x_i,\omega_i)$. Let $I' \subset I$ be any infinite subset such that $\omega_i \overset{I'}{\to} \hat{\omega}$, $\tau_i \overset{I'}{\to} \hat{\tau}$. By Lemma 3.1, $\hat{\omega} \in \Omega(\hat{x})$, $\hat{\tau} \in T(\hat{x},\hat{\omega})$. Hence, by continuity of $\nabla_x \zeta(\cdot,\cdot,\cdot)$,

$$\hat{z} = \lim_{\substack{I \\ i\to\infty}} z_i = \lim_{\substack{I' \\ i\to\infty}} z_i = \lim_{\substack{I' \\ i\to\infty}} \nabla_x \zeta(x_i,\omega_i,\tau_i) = \nabla_x \zeta(\hat{x},\hat{\omega},\hat{\tau}) \tag{3.29}$$

and $\hat{z} \in \tilde{\Gamma}(\hat{x})$. ∎

Proposition 3.9. For any $x \in \mathbb{R}^n$,

$$\partial\psi(x) \subseteq \Gamma(x). \tag{3.30}$$

Proof. Let $x \in \mathbb{R}^n$ be arbitrary and consider any vector z such that there exists a sequence $\{v_i\}$, $v_i \to 0$, and $z = \lim_{i\to\infty} \nabla\psi(x+v_i)$. By Definition 3.1, $z \in \partial\psi(x)$. Since $\psi(\cdot)$ is differentiable at all the points $x + v_i$, for any $h \in \mathbb{R}^n$,

$$d\psi(x+v_i;h) = \lim_{\lambda\downarrow 0} \frac{\psi(x+v_i+\lambda h)-\psi(x+v_i)}{\lambda} = \langle \nabla\psi(x+v_i), h\rangle . \tag{3.31}$$

Moreover, by assumption,

$$\lim_{i\to\infty} \lim_{\lambda\downarrow 0} \frac{\psi(x+v_i+\lambda h)-\psi(x+v_i)}{\lambda} = \langle z, h\rangle . \tag{3.32}$$

By definition of $\psi(\cdot)$, and because $\Omega(x+v_i) \subset \Omega$,

$$\psi(x+v_i+\lambda h) \geq \max_{\omega\in\Omega(x+v_i)} \chi(x+v_i+\lambda h,\omega)$$

$$= \max_{\omega\in\Omega(x+v_i)} [\chi(x+v_i,\omega) + \lambda \frac{\chi(x+v_i+\lambda h,\omega)-\chi(x+v_i,\omega)}{\lambda}]$$

$$= \psi(x+v_i) + \lambda \max_{\omega\in\Omega(x+v_i)} \frac{\chi(x+v_i+\lambda h,\omega)-\chi(x+v_i,\omega)}{\lambda} . \tag{3.33}$$

Hence, by Proposition 3.2,

$$d\psi(x+v_i;h) = \lim_{\lambda\downarrow 0} \frac{\psi(x+v_i+\lambda h)-\psi(x+v_i)}{\lambda} \geq \max_{\omega\in\Omega(x+v_i)} d_x\chi(x+v_i,\omega;h)$$

$$= \max_{\omega\in\Omega(x+v_i)} \min_{\tau\in T(x+v_i,\omega)} \langle \nabla_x \zeta(x+v_i,\omega,\tau), h\rangle \ . \tag{3.34}$$

Let y_i be an element of $\Gamma(x+v_i)$ such that

$$\langle y_i, h\rangle = \max_{\omega\in\Omega(x+v_i)} \min_{\tau\in T(x+v_i,\omega)} \langle \nabla_x \zeta(x+v_i,\omega,\tau), h\rangle \ , \tag{3.35}$$

then we have, for $i = 1,2,\ldots$

$$\langle \nabla\psi(x+v_i), h\rangle \geq \langle y_i, h\rangle \ . \tag{3.36}$$

Now let $K \subset \{0,1,\ldots\}$ be an infinite subset such that $y_i \overset{K}{\to} \hat{y}$. Then $\hat{y} \in \Gamma(x)$ by upper semicontinuity of $\Gamma(\cdot)$ and, since $\nabla\psi(x+v_i) \to z$,

$$\langle z, h\rangle \geq \langle \hat{y}, h\rangle \tag{3.37}$$

i.e., given any $h \in \mathbb{R}^n$ and $z \in \partial\psi(x)$, there exists $\hat{y} \in \Gamma(x)$ such that (3.37) holds. Now, suppose that for some $z \in \partial\psi(x)$, $z \notin \Gamma(x)$ holds. Then, because $\Gamma(x)$ is convex and compact, $\{z\}$ and $\Gamma(x)$ can be separated strictly, i.e., there exists $h \in \mathbb{R}^n$ such that for all $y \in \Gamma(x)$,

$$\langle z, h\rangle < 0 \text{ and } \langle y, h\rangle \geq 0. \tag{3.38}$$

But (3.38) contradicts (3.37) for some $\hat{y} \in \Gamma(x)$ and the proof is completed. ¤

Example 3.3. Consider the function $\psi(\cdot)$ defined by (3.13). The generalized gradient at $x = +1$, $\partial\psi(1)$ is equal to $co\{0,+1\}$. The set $\Gamma(1)$ is equal to $co\{[-1,+1]\} = [-1,+1]$. Then, $\Gamma(1) \supset \partial\psi(1)$, which agrees with Proposition 3.8. ¤

Corollary 3.2. If for any $x \in \mathbb{R}^n$, $\omega \in \Omega(x)$, $T(x,\omega)$ is a singleton, then

$$\Gamma(x) = \partial\psi(x). \tag{3.39}$$

¤

The proof of Corollary 3.2 follows from a result in [23] and from a characterization of the generalized gradient given by Mifflin [19].

Proposition 3.9 leads to a weaker but verifiable optimality condition.

Corollary 3.3. If $\hat{x}$ is optimal for P, then

$$\text{(i)}\quad \psi(\hat{x}) \leq 0 \tag{3.40a}$$

$$\text{(ii)}\quad 0 \in M(\hat{x}) \tag{3.40b}$$

where

$$M(x) = \begin{cases} \{\nabla f(x)\} & \text{if } \psi(x) < 0 \\ \text{co}\{\{\nabla f(x)\} \cup \Gamma(x) & \text{if } \psi(x) \geq 0 \end{cases} \tag{3.41}$$

¤

The proof of Corollary 3.3 follows immediately from Proposition 3.6 and Proposition 3.9.

Corollary 3.4. The point-to-set map $M(\cdot)$ defined in (3.41) is upper semicontinuous. ¤

The proof follows immediately from Proposition 3.8 and Assumption 3.1.

Since our approach to solving P is to replace it with a sequence of approximating problems

$$P_i : \min\{f(x) \mid \psi_i(x) \triangleq \max_{\omega\in\Omega_i} \min_{\tau\in T} \zeta(x,\omega,\tau) \geq 0\} \tag{3.42}$$

where, for all i, $\Omega_i \subset \Omega$ is a discrete set, it is important to investigate further the differentiability property of $\psi_i(\cdot)$. We already know by Proposition 3.4 that $\psi_i(\cdot)$ is directionally differentiable. We are about

to show that $\psi_i(\cdot)$ is <u>continuously</u> directionally differentiable, i.e., that for any $x,h \in \mathbb{R}^n$,

$$\lim_{\lambda \downarrow 0} d\psi_i(x+\lambda h;h) = d\psi_i(x;h). \tag{3.43}$$

To do so, we need the following definition due to Mifflin [19].

<u>Definition 3.2</u>. A locally Lipschitz function $\psi : \mathbb{R}^n \to \mathbb{R}^1$ is <u>semismooth</u> if for any $x,h \in \mathbb{R}^n$ and any sequences $\{\lambda_k\} \subset \mathbb{R}_+$, $\{z_k\}, \{v_k\} \subset \mathbb{R}^n$ such that $\lambda_k \to 0$, $(1/\lambda_k)v_k \to 0$ and $z_k \in \partial\psi(x+\lambda_k h+v_k)$, the sequence $\langle z_k,h \rangle$ converges. ¤

<u>Proposition 3.10</u>. [19]: Let $\psi : \mathbb{R}^n \to \mathbb{R}^1$ be semismooth. Then $\psi(\cdot)$ is directionally differentiable and for any $h \in \mathbb{R}^n$, the directional derivative is given by

$$d\psi(x;h) = \lim \langle z_k,h \rangle \tag{3.44}$$

where $\{z_k\}$ is any sequence as in Definition 3.2. ¤

<u>Corollary 3.5</u>. Let $\psi : \mathbb{R}^n \to \mathbb{R}^1$ be semismooth, then $\psi(\cdot)$ is continuously directionally differentiable, i.e. (3.43) holds at any $x,h \in \mathbb{R}^n$. ¤

The proof follows immediately from Definition 3.2 and Proposition 3.10.

<u>Proposition 3.12</u>. [19]. The function $\chi(\cdot,\cdot)$ is semismooth. ¤

<u>Proposition 3.13</u>. If Ω is a discrete set, the function $\psi(\cdot)$ is semismooth. ¤

The proof of Proposition 3.13 follows immediately from Proposition 3.12 and a result obtained by Mifflin (Theorem 4, [19]).

The most important consequence for us of Proposition 3.13 is the

fact that $\psi_i(\cdot)$ is continuously directionally differentiable. In Section 4, we show that to obtain a feasible descent direction in a convergent algorithm, one needs to compute the set $\Gamma(x)$ at an infinite number of points. Because $\psi_i(\cdot)$ is continuously directionally differentiable, we can construct an adequate approximation to this feasible descent direction by means of a scanning process which results in only one set $\Gamma(x')$ at an appropriate point x'.

4. The Algorithms

Our approach to solving P is to replace it with a sequence of approximating problems

$$P_i : \min\{f(x) \mid \psi_i(x) \triangleq \max_{\omega\in\Omega_i} \min_{\tau\in T} \zeta(x,\omega,\tau) \leq 0\} \tag{4.1}$$

where, for all i, $\Omega_i \subset \Omega$ is a discrete set. We shall show that, under certain conditions, the accumulation points of the sequence of stationary points $\{x_i\}_{i=0}^{\infty}$ of P_i are stationary points of P. The approximating problems are generated by a Master Outer Approximations Algorithm proposed by Gonzaga and Polak [9]. Obviously, one still needs an algorithm for the computation of stationary points of the problems P_i. These problems are very difficult to solve because the $\psi_i(\cdot)$ are not differentiable everywhere. We shall describe a new extension of Polak's method of feasible directions [11] which is capable of solving P_i. Before we present this new algorithm, we introduce the ideas upon which the algorithm is based by discussing how one can solve unconstrained problems of the form

$$\hat{P} : \min\{\psi(x) \mid x \in \mathbb{R}^n\}, \tag{4.2}$$

where $\psi(\cdot)$ is locally Lipschitz. Clarke [14] has shown that a first order necessary optimality condition for $\hat{P}$ is as follows: if $\hat{x}$ is optimal for $\hat{P}$, then

$$0 \in \partial\psi(\hat{x}). \tag{4.3}$$

We will describe an algorithm which, by means of a descent process, generates a sequence of points $\{x_i\}$ whose accumulation points satisfy (4.3). In general, an optimization algorithm consists of two parts: a direction finding subprocedure and a line search-step size finding subprocedure. If $\psi(x)$ were continuously differentiable, then

$h(x) \triangleq -\nabla\psi(x)$ would be the steepest descent direction. In the non-differentiable case, when ψ is convex,

$$h(x) \triangleq \text{argmin}\{\|h\| \,|\, h \in \partial\psi(x)\} \triangleq \text{Nr}\partial\psi(x). \tag{4.4}$$

has been shown to be the steepest descent direction by a number of authors (e.g., Lemarechal [20], Wolfe [21]). In the nonconvex case, Mifflin [13] shows that (4.3) is the steepest descent direction in the sense that the generalized directional derivative at x defined by Clarke is minimized when $h = h(x)$. We now show that the direction given by (4.3) is indeed a descent direction. The derivation of the descent property of $h(x)$ will also be useful to obtain a convergent algorithm based on (4.4). For $\delta > 0$, let

$$N_\delta(\partial\psi(x)) \triangleq \partial\psi(x) + B(0,\delta) \tag{4.5}$$

where $B(\hat{y},\delta) \triangleq \{y \,|\, \|y-\hat{y}\| \leq \delta\}$, be a δ neighborhood of $\partial\psi(x)$. Obviously, there exists $\hat{\delta} > 0$ such that

$$-1/2\ h(x) = \text{Nr}\ N_{\hat{\delta}}(\partial\psi(x)). \tag{4.6}$$

By upper semicontinuity of $\partial\psi(\cdot)$, there exists a $\rho > 0$ such that, for all $y \in B(x,\rho)$,

$$\partial\psi(y) \subset N_{\hat{\delta}}(\partial\psi(x)). \tag{4.7}$$

Now, let $\bar{\lambda} = \rho/\|h(x)\|$. Then by the mean value theorem 3.1 [23] there exists a $\xi \in \text{co}\{x, x+\lambda h(x)\} \subset B(x,\rho)$ and a $g \in \partial\psi(\xi) \subset N_{\hat{\delta}}(\partial\psi(x))$ such that

$$\psi(x+\lambda h(x)) - \psi(x) = \lambda\langle g, h(x)\rangle. \tag{4.8}$$

By (4.6) (see Fig. 4.1), we have that, for any $g \in N_{\hat{\delta}}(\partial\psi(x))$,

$$\langle g+1/2h(x), h(x)\rangle \leq 0. \tag{4.9}$$

Therefore, for all $\lambda \in [0,\bar{\lambda}]$,

$$\psi(x+\lambda h(x)) - \psi(x) \leq -1/2\ \lambda \|h(x)\|^2 \tag{4.10}$$

and $h(x)$ is indeed a descent direction. Therefore, $h(x)$ is a natural choice for a "steepest" descent algorithm. As for the step-size finding subprocedure, our experience with differentiable problems has shown that the Armijo step-size rule [26] works well. Given $h(x)$, the Armijo step-size subprocedure computes the smallest integer $\hat{k}$ such that

$$\psi(x+\beta^{\hat{k}}h(x)) - \psi(x) \leq -\alpha\beta^{\hat{k}}\|h(x)\|^2 \tag{4.11}$$

where α and β are two parameters whose value is between zero and one. In Fig. 4.2, the geometrical interpretation of the Armijo step-size rule is given.

Now, we have the necessary ingredients to describe a natural extension of the steepest descent algorithm for $\hat{P}$.

Algorithm 1: Extension of Steepest Descent Algorithm for Solving $\hat{P}$.

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

Parameters: $\alpha,\beta \in (0,1)$.

Step 0. Set $i = 0$.

Step 1. Compute $\partial\psi(x_i)$ and $h(x_i) \triangleq - Nr\partial\psi(x_i)$.

Stop if $h(x_i) = 0$.

Step 2. Compute the smallest integer $k_i \geq 0$ such that

$$\psi(x_i)+\beta^{k_i}h(x_i)) - \psi(x_i) \leq -\alpha\beta^{k_i}\|h(x_i)\|^2.$$

Step 3. Set $x_{i+1} = x_i + \beta^{k_i}h(x_i)$, $i = i+1$ and go to Step 1. ¤

Unfortunately the algorithm may fail to converge even if we substitute the Armijo step size rule with an exact minimization along the line. The following example is due to Wolfe [21]. Let $x \in \mathbb{R}^2$, $x \triangleq [y,z]^T$, $y,z \in \mathbb{R}^1$,

$$\psi(x) \triangleq \begin{cases} 5(9y^2+16z^2)^{1/2} & y > z \\ 9y + 16z & y \leq z \end{cases} \tag{4.12}$$

As shown in Fig. 4.3, if we start at an initial point x_o such that $y_o > |z_o| > (9/16)^2|y_o|$, the steepest descent algorithm converges to $\hat{x} = [0,0]^T$ which does not satisfy the optimality condition (4.3), since $\partial\psi(0,0) \subset \mathbb{R}^2 = \text{co}\{[u,v]^T | (u/15)^2 + (v/20)^2 = 1, |v| \leq 16\}$ (see Fig. 4.4), and obviously $0 \notin \partial\psi(0,0)$. The convergence of the algorithm to a nonstationary point is due mainly to the lack of continuity of $\partial\psi(\cdot)$. In fact, we cannot find a lower bound on ρ such that (4.7) holds for all x in any neighborhood of a nonstationary point $\hat{x}$. Therefore, $h(x)$ may become feebler and feebler descent direction and may not lead to a uniform decrease of $\psi(\cdot)$, in a neighborhood of $\hat{x}$. In the example, as the points selected by the algorithm converge to the origin, ρ converges to zero. In fact, at all the points x_i generated by the algorithm, the gradient exists and $\partial\psi(x_i)$ is a singleton, while at the origin the gradient is not defined and $\partial\psi(0)$ is a set. To overcome the lack of continuity of $\partial\psi(\cdot)$, we can "smear" $\partial\psi(x)$, by defining the <u>smeared generalized gradient</u>, $\partial_\varepsilon\psi(\cdot)$, as follows

$$\partial_\varepsilon\psi(x) \triangleq \text{co} \bigcup_{y \in B(x,\varepsilon)} \partial\psi(y) \tag{4.13}$$

Smearing has been introduced by Demyanov to overcome the lack of continuity of the directional derivative for min max problems [22]. It has been also used by Bertsekas and Mitter [29] in their ε-subgradient algorithm for nondifferentiable convex problems and by Goldstein in [30]. For a fixed ε, it is easy to see that the smeared generalized gradient $\partial_\varepsilon\psi(x)$ has the same properties as the generalized gradient, i.e., it is convex, compact, bounded on bounded subsets of $\mathbb{R}^n$, and upper semicontinuous. By smearing

$\partial\psi(x)$, we will be able to find a direction, $h_\varepsilon(x) \triangleq -Nr\partial_\varepsilon\psi(x)$, which guarantees a uniform decrease in cost, and, hence the convergence of the descent algorithm.

To demonstrate the effect of smearing, we first describe an algorithm which uses a fixed ε, and which will be shown to converge to points $\hat{x}$ satisfying $0 \in \partial_\varepsilon\psi(\hat{x})$. Then, we will provide a mechanism to drive ε to zero while the algorithm constructs an infinite sequence in order to achieve convergence to points $\hat{x}$ satisfying $0 \in \partial\psi(\hat{x})$.

Algorithm 2: Smeared Steepest Descent Algorithm for Solving $\hat{P}$ (fixed ε).

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

Parameters: $\varepsilon > 0$; $\alpha,\beta \in (0,1)$.

Step 0. Set $i = 0$.

Step 1. Compute $\partial_\varepsilon\psi(x_i)$ and $h_\varepsilon(x_i) \triangleq -N_r\partial_\varepsilon\psi(x_i)$. Stop if $h_\varepsilon(x_i) = 0$.

Step 2. Compute the smallest integer $k_i \geq 0$ such that

$$\psi(x_i+\beta^{k_i}h_\varepsilon(x_i)) - \psi(x_i) \leq -\alpha\beta^{k_i}\|h_\varepsilon(x_i)\|^2 \tag{4.14}$$

Step 3. Set $x_{i+1} = x_i + \beta^{k_i}h_\varepsilon(x_i)$, $i = i + 1$ and go to Step 1. ¤

To prove the convergence properties of the algorithm, it is useful to introduce an optimality function, $\Theta_\varepsilon : \mathbb{R}^n \to \mathbb{R}^1$, defined by

$\Theta_\varepsilon(x) \triangleq \min\{\|h\|^2 \mid h \in \partial_\varepsilon\psi(x)\} = \|h_\varepsilon(x_i)\|^2$.

This function is zero if and only if $0 \in \partial_\varepsilon\psi(x)$. The Clarke's optimality condition [4.3] can then be restated as follows: if $\hat{x}$ is optimal, then $\Theta_o(\hat{x}) = 0$.

Lemma 4.1. For all $\varepsilon \geq 0$, the optimality function $\theta_\varepsilon(\cdot)$ is lower semicontinuous, i.e., for all $x \in \mathbb{R}^n$, given an $\eta > 0$, there exists $\rho > 0$ such that for all $y \in B(x,\rho)$,

$$\theta_\varepsilon(x) - \theta_\varepsilon(y) \leq \eta. \tag{4.15}$$

Proof. Given $\eta > 0$, there exists $\delta > 0$ such that

$$\|\mathrm{Nr}N_\delta(\partial_\varepsilon\psi(x))\|^2 + \eta \geq \|\mathrm{Nr}\partial_\varepsilon\psi(x)\|^2 = \theta_\varepsilon(x) \tag{4.16}$$

By upper semicontinuity of $\partial_\varepsilon\psi(\cdot)$, there exists $\rho > 0$, such that for all $y \in B(x,\rho)$, $\partial_\varepsilon\psi(y) \subset N_\delta(\partial_\varepsilon\psi(x))$. Hence, for all $y \in B(x,\rho)$,

$$\theta_\varepsilon(y) \triangleq \|\mathrm{Nr}\partial_\varepsilon\psi(y)\|^2 \geq \|\mathrm{Nr}N_\delta(\partial_\varepsilon\psi(x))\| \geq e_\varepsilon(x) - \eta \tag{4.17}$$

and the Lemma is proven. ¤

Theorem 4.1. Let $\{x_i\}$ be an infinite sequence generated by Algorithm 2. Then every accumulation point $\hat{x}$ of $\{x_i\}$ satisfies $0 \in \partial_\varepsilon\psi(\hat{x})$.

Proof. Let $\hat{x}$ be an accumulation point of $\{x_i\}$ and $I \subset \{0,1,\ldots\}$ an infinite subset such that $x_i \overset{I}{\to} \hat{x}$. For the sake of contradiction, suppose $0 \notin \partial_\varepsilon\psi(\hat{x})$. Hence, $\theta_\varepsilon(\hat{x}) > 0$. Since $\partial_\varepsilon\psi(\cdot)$ is bounded on bounded subsets of $\mathbf{R}^n$, there exists a constant b, such that for all $i \in I$,

$$\|h_\varepsilon(x_i)\| \leq b. \tag{4.18}$$

Let $\hat{\lambda} \triangleq \varepsilon/b$. Then, by (4.18) and the mean value Theorem 3.1 [23], for all $\lambda \in [0,\hat{\lambda}]$, for all $i \in I$, there exists $g_i \in \partial_\varepsilon\psi(x_i)$ such that

$$\psi(x_i+\lambda h_\varepsilon(x_i)) - \psi(x_i) = \lambda \langle g_i, h_\varepsilon(x_i)\rangle \tag{4.19}$$

By definition of $h_\varepsilon(x_i)$, we obtain from (4.19), that

$$\psi(x_i+\lambda h_\varepsilon(x_i)) - \psi(x_i) \leq -\lambda\|h_\varepsilon(x_i)\|^2 = -\lambda\theta_\varepsilon(x_i).$$

By lower semicontinuity of $\theta_\varepsilon(\cdot)$, there exists $\hat{\rho}$ such that for all $x \in B(\hat{x},\hat{\rho})$,

$$\theta_\varepsilon(x) \geq \theta_\varepsilon(\hat{x})/2. \tag{4.20}$$

Since $x_i \overset{I}{\to} \hat{x}$, by hypothesis, there exists $\hat{i} \in I$, such that for all $i \in I$, $i \geq \hat{i}$, $x_i \in B(\hat{x},\hat{\rho})$. Therefore, for all $\lambda \in [0,\hat{\lambda}]$

$$\psi(x_i+\lambda h_\varepsilon(x_i)) - \psi(x_i) + \alpha\lambda\,\theta_\varepsilon(x_i) \leq -\lambda\theta_\varepsilon(x_i) + \alpha\lambda\,\theta_\varepsilon(x_i) \leq -(1-\alpha)\lambda\,\theta_\varepsilon(\hat{x})/2. \tag{4.21}$$

Let $\hat{k}$ be such that

$$\beta^{\hat{k}} \leq \hat{\lambda} < \beta^{\hat{k}-1}. \tag{4.22}$$

Then by (4.21), for all $i \in I$, $i \geq \hat{i}$, Algorithms 2 will select a $k_i \leq \hat{k}$. Therefore, for all $i_1 \geq i$, $i_1 \in I$, if $i_2 \in I$ is the index following i_1 in I, then

$$\psi(x_{i_2}) - \psi(x_{i_1}) \leq \psi(x_{i_1} + \beta^{k_{i_1}} h_\varepsilon(x_{i_1})) - \psi(x_{i_1}) \leq -(1-\alpha)\beta^{\hat{k}}\, \theta_\varepsilon(\hat{x})/2 < 0 \tag{4.24}$$

and since $\{\psi(x_i)\}_{i\in I}$ is monotonically decreasing, it is not Cauchy. From the continuity of $\psi(\cdot)$ we now conclude that $\{x_i\}_{i\in I}$ is not Cauchy, contradicting the hypothesis that $\hat{x}$ is an accumulation point of $\{x_i\}_{i\in I}$. ¤

Algorithm 2 is an intermediate step towards a descent algorithm which solves $\hat{P}$. The next step requires the development of a mechanism to drive ε to zero, so that the accumulation points of the sequence generated by the algorithm, $\hat{x}$, will satisfy Clarke's optimality condition (4.3). The mechanism is based upon a check on the value of $\theta_\varepsilon(x_i)$. If the value of $\theta_\varepsilon(x_i)$ is less or equal than ε, then ε is reduced and $\theta_\varepsilon(x_i)$ recomputed. This process will go on until $\theta_\varepsilon(x_i)$ is larger than ε. At this stage, the Armijo step-size rule is applied to find the next point x_{i+1}.

Algorithm 3: Smeared Steepest Descent Algorithm for Solving $\hat{P}$.

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

Parameters: $\varepsilon_o > 0$; $\alpha, \beta \in (0,1)$.

Step 0. Set $i = 0$.

Step 1. Set $\varepsilon = \varepsilon_o$.

Step 2. Compute $\partial_\varepsilon \psi(x_i)$, $h_\varepsilon(x_i)$ and $\theta_\varepsilon(x_i)$.

Step 3. If $\theta_\varepsilon(x_i) \geq \varepsilon$, set $\varepsilon(x_i) = \varepsilon$ and continue, else set $\varepsilon = \varepsilon/2$ and go to Step 2.

Step 4. Compute the smallest integer k_i such that

$$\psi(x_i + \beta^{k_i} h_\varepsilon(x_i)) - \psi(x_i) \leq -\alpha\beta^{k_i}\, \theta_\varepsilon(x_i)$$

Step 5. Set $x_{i+1} = x_i + \beta^{k_i} h_\varepsilon(x_i)$, $i = i + 1$ and go to step 1. ¤

To prove that Algorithm 3 cannot converge to a point $\hat{x}$ such that $\theta_o(\hat{x}) > 0$, we show that in a neighborhood of any point $\hat{x}$ satisfying $\theta_o(\hat{x}) > 0$, the algorithm does not drive ε below a certain lower bound $\hat{\varepsilon} > 0$. Therefore, by Steps 2 and 3, $\theta_{\hat{\varepsilon}}(x) \geq \hat{\varepsilon} > 0$ in a neighborhood of $\hat{x}$. But, then, the behavior of Algorithm 3 is similar to the behavior of Algorithm 2 with $\varepsilon = \hat{\varepsilon}$ and we know by Theorem 4.1 that Algorithm 2 cannot converge to $\hat{x}$ such that $\theta_{\hat{\varepsilon}}(\hat{x}) > 0$.

Lemma 4.2. Let $\varepsilon(x) \triangleq \max\{\varepsilon \mid \varepsilon = 0 \text{ or } \varepsilon_o 2^{-k}, k \in \mathbb{N}, \theta_\varepsilon(x) \geq \varepsilon\}$. If $\theta_o(\hat{x}) > 0$, then, there exist $\hat{\rho} \geq \hat{\varepsilon} > 0$ such that for all $x \in B(\hat{x},\hat{\rho})$

$$\varepsilon(x) \geq \hat{\varepsilon} > 0. \tag{4.25}$$

Proof. Let $\delta > 0$ be such that

$$\|\mathrm{Nr}N_\delta\ \partial\psi(\hat{x})\|^2 \geq \theta_o(\hat{x})/2 \triangleq \varepsilon > 0 \tag{4.26}$$

By upper semicontinuity of $\partial\psi(\cdot)$, there exists $\hat{\rho} > 0$ such that for all $x \in B(\hat{x},2\hat{\rho})$

$$\partial\psi(x) \subset N_\delta\ \partial\psi(\hat{x})). \tag{4.27}$$

Therefore, all $x \in B(\hat{x},\hat{\rho})$

$$\partial_{\hat{\rho}}\psi(x) \subset N_\delta(\partial\psi(\hat{x})). \tag{4.28}$$

Consequently,

$$\theta_{\hat{\rho}}(x) \geq \|\mathrm{Nr}N_\delta(\partial\psi(\hat{x}))\|^2 \geq \varepsilon \tag{4.29}$$

Now, let $\hat{\varepsilon} = \min\{\varepsilon,\hat{\rho}\}$. Then, for all $x \in B(\hat{x},\hat{\rho})$

$$\theta_{\hat{\varepsilon}}(x) \geq \hat{\varepsilon} \tag{4.30}$$

and, by definition of $\varepsilon(x)$, (4.24) holds. ¤

Theorem 4.2. Let $\{x_i\}$ be an infinite sequence generated by Algorithm 3. Then every accumulation point $\hat{x}$ of $\{x_i\}$, is such that $0 \in \partial\psi(\hat{x})$ (or $\theta_o(\hat{x}) = 0$).

Proof. For the sake of contradiction, suppose that there exists an accumulation point of $\hat{x}$ of $\{x_i\}$, such that $\theta_o(\hat{x}) > 0$. Then, by Lemma 4.2, there exist $\hat{\rho},\hat{\varepsilon} > 0$ such that for all $x \in B(\hat{x},\hat{\rho})$, (4.25) holds. Then, for all $x \in B(\hat{x},\hat{\rho})$,

$$\partial_{\varepsilon(x)}\psi(x) \supset \partial_{\hat{\varepsilon}}\psi(x) \tag{4.31}$$

Let $I \subset \{0,1,2,\ldots\}$ be an infinite subset such that $x_i \overset{I}{\to} \hat{x}$. Then, there exists $\hat{i} \in I$, such that for all $i \geq \hat{i}$, $i \in I$,

$$x_i \in B(\hat{x},\hat{\rho}), \tag{4.32}$$

and, therefore $\varepsilon(x_i)$ computed in Step 3, satisfies

$$\varepsilon(x_i) \geq \hat{\varepsilon} \tag{4.33}$$

Let $b > 0$ be such that for all $i \in I$,

$$\|h_{\varepsilon(x_i)}(x_i)\| \leq b; \tag{4.34}$$

and $\hat{\lambda} = \hat{\varepsilon}/b$. Then, by the mean value Theorem 3.1 [23], by (4.31), (4.32) (4.33) and (4.34), for all $\lambda \in [0,\hat{\lambda}]$ for all $i \geq \hat{i}$, $i \in I$ there exists $g_i \in \partial_{\hat{\varepsilon}}\psi(x_i)$ such that

$$\begin{aligned}\psi(x_i + h_{\varepsilon(x_i)}(x_i)) - \psi(x_i) &= \lambda\langle g_i, h_{\varepsilon(x_i)}(x_i)\rangle \leq -\lambda\|h_{\varepsilon(x_i)}(x_i)\|^2 \\ &= -\lambda\,\theta_{\varepsilon(x_i)}(x_i) \leq -\lambda\varepsilon(x_i) \leq -\lambda\hat{\varepsilon}.\end{aligned} \tag{4.35}$$

The remainder of the proof follows the same lines of the proof of Theorem 4.1 and is, henceforth, omitted. ∎

We are now ready to present our algorithm which "solves" the non-differentiable constrained problem P_i, in the sense of that it finds approximations to points $\hat{x} \in \mathbb{R}^n$ which satisfy the optimality condition

$$\psi_i(\hat{x}) \leq 0 \tag{4.36a}$$

$$0 \in M_i(\hat{x}) \tag{4.36b}$$

where

$$M_i(x) \triangleq \begin{cases} \{\nabla f(x)\} & \text{if } \psi_i(x) < 0 \\ co\{\{\nabla f(x)\} \cup \Gamma_i(x)\} & \text{if } \psi_i(x) \geq 0 \end{cases} \tag{4.37}$$

and

$$\Gamma_i(x) \triangleq co\{\nabla_x \zeta(x,\omega,\tau) \,|\, \omega \in \Omega_i(x),\ \tau \in T(x,\omega)\}, \tag{4.38}$$

The algorithm is based on Polak's method of feasible directions [11]. For the differentiable optimization problem

$$P^*: \min\{f(x) \,|\, \psi^*(x) \leq 0\} \tag{4.39}$$

$f: \mathbb{R}^n \to \mathbb{R}^1$, $\psi^*: \mathbb{R}^n \to \mathbb{R}^1$ continuously differentiable, the algorithm is as follows:

Algorithm 4: Polak's Method of Feasible Directions for P*.

Data: $x_o \in F^* \triangleq \{x \,|\, \psi^*(x) \leq 0\}$.

Parameters: $\eta_o > 0$; $\alpha, \beta \in (0,1)$.

Step 0. Set $j = 0$.

Step 1. Set $\eta = \eta_o$.

Step 2. Compute

$$h^*_\eta(x_j) \triangleq -\mathrm{Nr} M^*_\eta(x_j)$$

where

$$M^*_\eta(x_j) \triangleq \begin{cases} \{\nabla f(x_j)\} & \text{if } \psi^*(x_j) < -\eta \\ co\{\{\nabla f(x_j)\} \cup \{\nabla \psi^*(x_j)\}\} & \text{if } \psi^*(x_j) \geq -\eta \end{cases}$$

and

$$\theta_\eta^*(x_j) = \|h_\eta(x_j)\|^2$$

Step 3. If $\theta_\eta^*(x_j) < \eta$, set $\eta = \eta/2$ and go to Step 2.

Else continue.

Step 4. Compute the smallest integer k_j such that

$$f(x_j+\beta^{k_j}h_\eta^*(x_j)) - f(x_j) \leq -\alpha\beta^{k_j}\theta_\eta^*(x_j)$$

and

$$\psi(x_j+\beta^{k_j}h_\eta^*(x_j)) \leq 0.$$

Step 5. Set $x_{j+1} = x_j + \beta^{k_j}h_\eta^*(x_j)$, $j = j + 1$ and go to Step 1. ¤

It can be shown that if the sequence $\{x_j\}$ produced by Algorithm 4 has accumulation points, then every accumulation point $\hat{x}$ produced satisfies the F. John optimality condition, i.e., $\hat{x}$ is is such that $\theta_o(\hat{x}) = 0$ [11].

In Algorithm 4, η plays a role similar to the one played by ε in Algorithm 3. It prevents the step size β^{k_j} from becoming excessively small when x_j is not in a small ball about a point $\hat{x}$ satisfying $\theta_o(\hat{x}) = 0$ [11].

A natural, but simple minded, extension of Algorithm 4 to P_i, results from the substitution of $\Gamma_i(x)$ for $\{\nabla\psi^*(x)\}$ in the expression for the optimality function and in the computation of the descent direction. Unfortunately, the resulting algorithm may converge to a point not satisfying the optimality condition because of the lack of continuity of $\Gamma_i(x)$. In the unconstrained case we overcame the problem of lack of continuity of $\partial\psi(\cdot)$ by smearing it. In the constrained case, we have to smear $\Gamma_i(x)$ to obtain a convergent algorithm. Let

$$\Gamma_i^\varepsilon(x) \triangleq \text{co} \bigcup_{y\in B(x,\varepsilon)} \Gamma_i(y), \tag{4.40}$$

We substitute $\Gamma_i^\varepsilon(x)$ for $\{\nabla\psi^*(x)\}$ in Algorithm 4, and, since we also need a mechanism for reducing ε, as η in Algorithm 4, for the sake of simplicity, we combine these two mechanism into one, controlled by a single parameter, ε. Thus, for any $\varepsilon \geq 0$, we define the set $M_i^\varepsilon(x)$

$$M_i^\varepsilon(x) \triangleq \begin{cases} \{\nabla f(x)\} & \text{if } \psi(x) < -\varepsilon \\ \text{co}\{\{\nabla f(x)\} \cup \Gamma_i^\varepsilon(x)\} & \text{if } \psi(x) \geq -\varepsilon, \end{cases} \tag{4.41}$$

the optimality functions $\theta_i^\varepsilon : \mathbb{R}^n \to \mathbb{R}^1$ by

$$\theta_i^\varepsilon(x) \triangleq \min\{\|h\|^2 \,|\, h \in M_i^\varepsilon(x)\} \tag{4.42}$$

and the descent direction $h_i^\varepsilon(x)$ by

$$h_i^\varepsilon(x) \triangleq -Nr\, M_i^\varepsilon(x). \tag{4.43}$$

It is rather obvious that (4.38) holds at $\hat{x}$ if and only if $\theta_i^o(\hat{x}) = 0$.

Algorithm 5: Extension of Polak's Method of Feasible Directions to P_i.

Data: $x_o \in F_i \triangleq \{x \,|\, \psi_i(x) \leq 0\}$.

Parameters: $\varepsilon_o > 0$, $\alpha, \beta \in (0,1)$.

Step 0. Set $j = 0$.

Step 1. Set $\varepsilon = \varepsilon_o$.

Step 2. Compute

$$h_i^\varepsilon(x_j) \triangleq -Nr\, M_i^\varepsilon(x_j).$$

and

$$\theta_i^\varepsilon(x_j) \triangleq \|h_i^\varepsilon(x_j)\|^2.$$

Step 3. If $\theta_i^\varepsilon(x_j) \leq \varepsilon$, set $\varepsilon = \varepsilon/2$ and go to Step 2. Else set $\varepsilon(x_j) = \varepsilon$ and continue.

Step 4. Compute the smallest integer k_j such that

$$f(x_j + \beta^{k_j} h_i^\varepsilon(x_j)) - f(x_j) \leq -\alpha\beta^{k_j}\theta_i^\varepsilon(x_j)$$

and

$$\psi(x_j+\beta^{k_j}h_i^{\varepsilon}(x_j)) \leq 0.$$

Step 5. Set $x_{j+1} = x_j + \beta^{k_j}h_i^{\varepsilon}(x_j)$, $j = j + 1$ and go to Step 1. ¤

The following standard assumption for feasible directions methods is needed to ensure that Algorithm 5 does not stop at x_o and that the sequence constructed by Algorithm has accumulation points.

Assumption 4.1. The set $F_i'(x_o)$ defined by

$$F_i'(x_o) = \{x \in \mathbb{R}^n \,|\, f(x) - f(x_o) \leq 0,\ \psi_i(x) \leq 0\}$$

is compact and has an interior. ¤

The proof of the following theorem can be found in Appendix 1.

Theorem 4.3. Let Assumption 4.1 be satisfied. Let $\{x_j\}$ be any infinite sequence generated by Algorithm 5. Then, every accumulation point $\hat{x}$ of $\{x_j\}$, satisfies $\psi_i(\hat{x}) \leq 0$ and $\theta_i(\hat{x}) = 0$. ¤

Algorithm 5 is a conceptual algorithm since it may involve an infinite number of gradient evaluations to compute $\Gamma_i^{\varepsilon}(x)$. Fortunately, the fact that $\psi_i(\cdot)$ is continuously directionally differentiable (see Corollary 3.5 and Proposition 3.13) make it possible to devise an algorithm based on an approximation to the descent direction $h_i^{\varepsilon}(x)$ which does not involve the computation of the entire set $M_i^{\varepsilon}(x)$. The approximation subprocedure is based on the line search introduced in [13,21]. In particular, suppose that $\psi_i(x_j) \geq -\varepsilon$. Then, we start our approximation by computing $\tilde{M}_i^o(x_i,\varepsilon) \triangleq M_i^o(x_j) \triangleq co\{\{\nabla f(x_j)\} \cup \Gamma_i(x_i)\}$, and $\tilde{h} \triangleq -Nr\ \tilde{M}_i^o(x_j,\varepsilon)$. Because of the lack of continuity of $M_i^o(\cdot)$, $\tilde{h}$ may be a "bad" descent direction for $\psi_i(\cdot)$. Therefore, we have to insert a test in the algorithm that detects if $\tilde{h}$ is indeed a bad descent direction. The test consists of checking if the resulting step-size is too small,

i.e., if for the smallest positive integer $\tilde{k}$ such that $\beta^{\tilde{k}}\|\tilde{h}\| \leq \varepsilon$, $x_j + \beta^{\tilde{k}}\tilde{h}$ is not feasible. If, according to this test, $\tilde{h}$ is not acceptable, then we add to $\tilde{M}_i^0(x_j,\varepsilon)$ a set $\Gamma_i(y_1)$ computed at a point y_1 in the ball $B(x_j,\varepsilon)$ and we compute $\tilde{h} = -Nr\ \tilde{M}_i^1(x_j,\varepsilon)$, where $\tilde{M}_i^1(x_j,\varepsilon) \triangleq co\{\tilde{M}_i^0(x_i,\varepsilon) \cup \Gamma_i(y_1)\}$. By doing so, we are sure to improve the approximation to $h_i^\varepsilon(x_j)$. However, the new approximation may still not be sufficiently "good" in the sense specified above and we may need to add a new set $\Gamma_i(y_2)$. In order to obtain a satisfactory approximation to $h_i^\varepsilon(x_j)$, in a finite number of steps, the points y_r have to be chosen so that the sets $\Gamma_i(y_r)$ add enough "new" information about $h_i^\varepsilon(x_j)$ to $\tilde{M}_i^r(x_j,\varepsilon)$. It turns out (see [13,21]) that if y_r is such that $\Gamma_i(y_r)$ contains an element v satisfying $\langle \tilde{h},v\rangle \geq -\gamma\|\tilde{h}\|^2$ where $\gamma \in (0,1)$, then the requirements on $\Gamma_i(y_r)$ are met. To find a point y_r such that $\Gamma_i(y_r)$ satisfies the above condition, we proceed as follows. Since $\psi_i(x_j) \leq 0$ and $\psi_i(x_j+\beta^{\tilde{k}}\tilde{h}) > 0$, there must exist a point $x_j + \tilde{\lambda}\,\tilde{h}$, $\tilde{\lambda} \leq \beta^{\tilde{k}}$ such that $d\psi_i(x_j+\tilde{\lambda}\tilde{h};\tilde{h}) > 0$. By Proposition 3.13 and Corollary 3.5, we know that $\psi_i(\cdot)$ is continuously directionally differentiable. Hence, there exists a $\delta > 0$ such that, for all $\lambda \in [\tilde{\lambda}-\delta,\tilde{\lambda}+\delta]$, $d\psi_i(x_j+\lambda\tilde{h};\tilde{h}) > -\gamma\|\tilde{h}\|^2$. Therefore, if we apply a bisection procedure to the interval $[0,\beta^{\tilde{k}}]$, after a finite number of steps, we obtain $\tilde{\lambda}_r$ such that $d\psi_i(x_j+\tilde{\lambda}_r\tilde{h};\tilde{h}) > -\gamma\|h\|^2$. Since, by Proposiion 3.4, there exists $v \in \Gamma_i(x_j+\tilde{\lambda}_r\tilde{h})$ such that $\langle v,\tilde{h}\rangle = d\psi_i(x_j+\tilde{\lambda}_r\tilde{h};\tilde{h}) > -\gamma\|\tilde{h}\|^2$, we add $\Gamma_i(x_j+\tilde{\lambda}_r\tilde{h})$ to $\tilde{M}_i^r(x_j,\varepsilon)$.

Algorithm 6: Extension of Polak's Method of Feasible Directions to P_i (semi-implementable version).

Data: $x_o \in F_i$.

Parameters: $\tilde{\varepsilon}_o > 0$, $\alpha,\beta,\gamma \in (0,1)$.

Step 0. Set $j = 0$.

Step 1. Set $p = 0$, $r = 0$, $D = \{x_j\}$.

Direction Finding Subprocedure.

Step 2. Set $r = r + 1$ and compute

$$\tilde{M}_i^r(x_j,\tilde{\varepsilon}_p) \triangleq \begin{cases} \{\nabla f(x_j)\} & \text{if } \psi_i(x_j) < -\tilde{\varepsilon}_p \\ \text{co}\{\{\nabla f(x_j)\} \cup \text{co}\{\bigcup_{y\in D\cap B(x_j,\tilde{\varepsilon}_p)} \Gamma_i(y)\}\} & \text{if } \psi_i(x_j) \geq -\tilde{\varepsilon}_p. \end{cases}$$

Step 3. Compute

$$\tilde{h} = -Nr\ \tilde{M}_i^r(x_j,\tilde{\varepsilon}_p)$$

and

$$\tilde{\theta}_i^r(x_j,\tilde{\varepsilon}_p) \triangleq \|\tilde{h}\|^2$$

Step 4. If $\tilde{\theta}_i^r(x_j,\tilde{\varepsilon}_p) \leq \tilde{\varepsilon}_p$, set $\tilde{\varepsilon}_{p+1} = \tilde{\varepsilon}_p/2$, $r = 0$, $p = p + 1$ and go to Step 2. Else continue.

Step-size Computation Subprocedure.

Step 5. Set $k = 0$.

Step 6. Set $k = k + 1$. If $\beta^k\|\tilde{h}\| \leq \tilde{\varepsilon}_p$ and $\psi_i(x_j+\beta^k\tilde{h}) > 0$, go to Step 9. Else continue.

Step 7. If

$$f(x_j+\beta^k\tilde{h}) - f(x_j) \leq -\alpha\beta^k\tilde{\theta}_i^r(x_j,\tilde{\varepsilon}_p)$$

and

$$\psi_i(x_j+\beta^k\tilde{h}) \leq 0,$$

continue. Else go to Step 6.

Step 8. Set $k_j = k$, $\tilde{\varepsilon}(x_j) = \tilde{\varepsilon}_p$, $\tilde{h}(x_j) = \tilde{h}$, $x_{j+1} = x_j + \beta^{k_j}\tilde{h}(x_j)$, $j = j + 1$ and go to Step 1.

Approximation Refinement of $M_i^{\varepsilon}(x_j)$ Subprocedure.

Step 9. Set $m = 0$, $\lambda_o = 0$, $\rho_o = \beta^k$, $\mu_o = \beta^k$.

Comment: λ_m, ρ_m are the left and right boundary points of the interval in $\mathbb{R}^1$ to be searched; μ is the midpoint of this interval (except for μ_o).

Step 10. Compute

$$d\psi_i(x_j+\mu_m\tilde{h};\tilde{h}) = \max_{\omega\in\Omega_i(x_j+\lambda_m\tilde{h})} \min_{\tau\in T(x_j+\lambda_m\tilde{h},\omega)} \langle \nabla_x \zeta(x_j+\mu_m\tilde{h},\omega,\tau),\tilde{h}\rangle$$

Step 11. If

$$d\psi_i(x_j+\mu_m\tilde{h};\tilde{h}) \geq \gamma\tilde{\theta}_i^r(x_j,\tilde{\varepsilon}_p),$$

set $\tilde{\lambda}_r = \mu_m$, $D = D \cup \{x_j+\tilde{\lambda}_r\tilde{h}\}$, and go to Step 2. Else, continue.

Step 12. If $\psi_i(x_j+\mu_m\tilde{h}) > 0$, set $\rho_{m+1} = \mu_m$, $\lambda_{m+1} = \lambda_m$ and continue. Else set $\rho_{m+1} = \rho_m$, $\lambda_{m+1} = \mu_m$ and continue.

Step 13. Set $m = m + 1$, $\mu_m = (\lambda_m+\rho_m)/2$ and go to Step 10. ¤

Algorithm 6 is convergent in the sense of the following theorem, whose proof can be found in Appendix 2.

Theorem 4.4. Suppose that Assumption 4.1 is satisfied. Consider any sequence $\{x_j\}$ generated by Algorithm 6. If $\{x_j\}$ is finite because Algorithm 6 jammed up at x_s, then $\psi_i(x_s) \leq 0$ and $\theta_i^o(x_s) = 0$. If $\{x_j\}$ is infinite, then every accumulation point $\hat{x}$ of $\{x_j\}$ is such that $\psi_i(\hat{x}) \leq 0$ and $\theta_i^o(\hat{x}) = 0$.

Remark 4.1. Algorithm 5 is only semi-implementable because at a point x, it requires the precise evaluation of $\psi_i(x)$ and the computation of $\Gamma_i(x)$, which in turn requires the computation of the sets $\Omega_i(x)$ and $T(x,\omega)$. It is still to be established how to approximate $\psi_i(x)$ and $\Gamma(x)$ in order to make Algorithm 6 implementable without destroying its convergence properties. ¤

Remark 4.2. Algorithm 6 requires a starting point $x_o \in F_i$. If such a point is not available, Algorithm 6 can be easily modified to obtain a semi-implementable algorithm for the unconstrained minimization of

$\psi_i(\cdot)$. The modified algorithm stops when it finds an $\tilde{x}$ such that $\psi_i(\tilde{x}) \leq 0$. By setting $x_o = \tilde{x}$, we can now apply Algorithm 6 to P_i. It is also possible to derive an extension to the nondifferentiable case of the phase I - phase II feasible directions algorithm of Polak, Trahan and Mayne [12]. This approach has two advantages over the previous one: (i) a separate computer program for the computation of x_o is not needed; (ii) while producing a feasible point, the algorithm does not completely ignore the cost function.

We now present the Master Outer Approximations Algorithm.

Algorithm 7: Master Outer Approximations Algorithm [9].

Data: A discrete set $\Omega_o \subset \Omega$.

Parameters: $K \geq 10$, $L \geq 2$, $\eta_o > 0$.

Step 0. Set $i = 0$.

Step 1. Compute a point z_i such that $\theta_i^{\eta_i}(z_i) \leq \eta_i$ and $\psi_i(z_i) \leq \eta_i$,* (i.e. solve P_i approximately).

Step 2. Compute $\psi(z_i)$ and an $\omega_i \in \Omega(z_i)$.†

Step 3. Include in Ω_{i+1} all ω_j, $j \leq i$, such that

$$\psi(z_j) > K\left[\frac{1}{(1+j)^{1/2}} - \frac{1}{(1+i)^{1/2}}\right]. \tag{4.43}$$

Step 4. Set $\eta_{i+1} = \eta_i/2$, set $i = i + 1$ and go to Step 1. ¤

The crucial part of Algorithm 7 is Step 3, the constraint construction scheme. The scheme used in Algorithm 7 is a generalization of the

---

†This requirement can be relaxed to that of requiring a progressively finer approximation to the ω_i in step 2.

*Since we use a feasible direction subalgorithm for all i we produce a point z_i such that $\psi_i(z_i) \leq 0$. Because of the condition $\psi_i(z_i) \leq \eta_i$, we could use as subalgorithms, algorithms based on penalty functions.

constraint construction schemes introduced by Eaves and Zangwill [7] and by Mayne, Trahan and Polak [8]. It will retain a particular constraint for a certain number of approximating problems and drop it when, presumably, it is no longer relevant. The test (4.43) in Step 3 "measures" the relevance of a constraint at a certain iteration i. The test checks if the constraints were "sufficiently" violated when they were added to the constraint set in relation to the number of approximating problems in which they have been taken into account. The test is more and more difficult to pass as i-j increases, so that "old" constraints are likely to be dropped. By choosing a large K, we make the test more difficult to pass and Algorithm 7 is likely to drop a constraint after only a few iterations. We expect Algorithm 7 to perform better if used in an interactive computing facility with a graphic display terminal. In fact, in that environment, the user could select interactively the parameters of the algorithm, add or retain a constraint that Algorithm 7 would drop if operated in batch mode, so as to improve the computational behavior of the outer approximations scheme. Not all the existing optimality functions can be used in Algorithm 7 to ensure its convergence. They have to satisfy certain conditions. Lemma 4.3 (whose proof can be found in Appendix 3) shows that the family of optimality functions $\theta_i^{\varepsilon}(\cdot)$, satisfies the condition specified in [9].

Lemma 4.2. If $\{\eta_i\} \subset \mathbb{R}_+$ is such that $\eta_i \to 0$ and $\{z_i\} \subset \mathbb{R}^n$ is such that $z_i \to \hat{z}$, with $\psi(\hat{z}) = 0$, and, in addition, $\theta_i^{\eta_i}(x_i) \to 0$ as $i \to \infty$, then $\theta(\hat{x}) \triangleq \|\mathrm{Nr}M(\hat{z})\|^2 = 0$ (i.e., $0 \in M(\hat{z})$). ¤

Since Lemma 4.2 holds, it follows from [9] that Algorithm 7 is convergent in the following sense.

Theorem 4.4 [9]. Let $\{z_i\}$ be an infinite sequence constructed by Algorithm 7, then every accumulation point $\hat{z}$ of $\{z_i\}$, satisfies $\psi(\hat{z}) \leq 0$ and $0 \in M(\hat{z})$. ¤

When used as a subroutine of Algorithm 7, Algorithm 6 has to be slightly modified. In particular, we have to add an exist to the Master Algorithm between Steps 3 and 4:

Step 3' (Exit to Master Algorithm). If $\tilde{\varepsilon}_p \leq \eta_i$ and $\tilde{\theta}_i^r(x_j,\tilde{\varepsilon}_p) \leq \eta_i$, set $z_i = x_j$ and go to Step 2 of Master Algorithm. Else continue. ¤

In order to conclude that the overall algorithm formed by Algorithm 7, with Algorithm 6 with the above indicated modification, used as a subroutine, it remains to prove that, when Algorithm 6 terminates to return to the Master Algorithm, $\theta_i^{\eta_i}(x_j) \leq \eta_i$. This is rather obvious because, by definition of $\tilde{\theta}_i^r(x_j,\tilde{\varepsilon}_p)$ and by Step 3', when Algorithm 6 stops, we have

$$\eta_i \geq \tilde{\varepsilon}_p \geq \tilde{\theta}_i^r(x_j,\tilde{\varepsilon}_p) \geq \tilde{\theta}_i^r(x_j,\eta_i) \geq \theta_i^{\eta_i}(x_j). \tag{4.45}$$

This ends our discussion of the general case. In the next section we shall consider the important special cases when $\zeta(\cdot,\cdot,\cdot)$ has certain convexity properties.

5. The Convex Case

When $\zeta(x,\cdot,\cdot)$ is convex, or one dimensional convex [10] and Ω is a hyper rectangle, the DCTT problem becomes much simpler. In particular, we show that if $\zeta(x,\cdot,\cdot)$ is convex or one dimensional convex, then

$$\max_{\omega\in\Omega} \min_{\tau\in T} \zeta(x,\omega,\tau) = \max_{\omega\in\tilde{\Omega}} \min_{\tau\in T} \zeta(x,\omega,\tau) \tag{5.1}$$

where $\tilde{\Omega}$ is a discrete set.

Proposition 5.1. Suppose that for any $x \in \mathbb{R}^n$, $\zeta(x,\cdot,\cdot)$ is convex (one dimensional convex). Then, for any $x \in \mathbb{R}^n$, $\chi(x,\cdot)$ is convex (one dimensional convex).

Proof. If $\zeta(x,\cdot,\cdot)$ is convex, then for any ω_1, ω_2, τ_1, $\tau_2 \in \mathbb{R}^n$, $\lambda \in [0,1]$,

$$\zeta(x,\lambda\omega_i+(1-\lambda)\omega_2,\lambda\tau_1+(1-\lambda)\tau_2) \leq \lambda\zeta(x,\omega_1,\tau_1) + (1-\lambda)\zeta(x,\omega_2,\tau_2) \tag{5.2}$$

By definition of $\chi(x,\cdot)$ and from (5.2), for any x, ω_1, $\omega_2 \in \mathbb{R}^n$, $\lambda \in [0,1]$,

$$\chi(x,\lambda\omega_1+(1-\lambda)\omega_2) = \min_{\tau\in T}\zeta\ (x,\lambda\omega_1+(1-\lambda)\omega_2,\tau)$$

$$\leq \zeta(x,\lambda\omega_1+(1-\lambda)\omega_2),\lambda\hat{\tau}_1+(1-\lambda)\hat{\tau}_2) \tag{5.3}$$

where $\hat{\tau}_1$ is any element of $T(x,\omega_1)$ and $\hat{\tau}_2$ is any element of $T(x,\omega_2)$. Therefore, by (5.2), for any x, ω_1, $\omega_2 \in \mathbb{R}^n$, $\lambda \in [0,1]$

$$\chi(x,\lambda\omega_1+(1-\lambda)\omega_2) \leq \lambda\zeta(x,\omega_1,\hat{\tau}_1) + (1-\lambda)\zeta(x,\omega_2,\hat{\tau}_2)$$

$$= \lambda\chi(x,\omega_1) + (1-\lambda)\chi(x,\omega_2). \tag{5.4}$$

The proof for the one dimensional convex case is a straightforward generalization of the above steps. ∎

By Proposition 5.1,

$$\psi(x) = \max_{\omega \in \Omega} \chi(x,\omega). \tag{5.5}$$

Bandler [10] has proven that, if Ω is a polytope and $\chi(x,\cdot)$ is convex (or one dimensional convex and Ω is a hyper rectangle with sides parallel to the axes), the maximum of $\chi(x,\cdot)$ is achieved at the vertices of the polytope Ω. So we can replace (5.5) by

$$\psi(x) = \max_{\omega \in \tilde{\Omega}} \chi(x,\omega) \tag{5.6}$$

where $\tilde{\Omega}$ is the set of vertices of the polytope. Therefore, we can use Algorithm 6 to solve P when $\zeta(x,\cdot,\cdot)$ is convex or one dimensional convex and Ω is an appropriate polytope.

6. Conclusion

The algorithm which we have presented is, to our knowledge, the only available algorithm for problems with constraints of the form $\psi(x) \stackrel{\Delta}{=} \max_{\omega \in \Omega} \min_{\tau \in T} \zeta(x,\omega,\tau) \leq 0$. Because the functions $\psi(x)$ are so difficult to evaluate and because they are not differentiable, the algorithm is quite complex and computations with it will be costly. This cause of difficulty is not likely to diminish in the future, even if, as we hope, better algorithms will be developed. However, in a design situation, on the basis of the anticipated benefits, one can justify quite considerable computing costs and we expect that algorithms of the type presented in this paper will slowly find their way into the designer's arsenal.

Appendix 0: Proof of Proposition 3.1

We first prove that $\chi(\cdot,\cdot)$ is locally Lipschitz. Let $x_1, x_2 \in D$, a compact subset of $\mathbb{R}^n$, and let $\omega_1, \omega_2 \in \Omega$. For $k = 1,2$, let τ_k be a minimizer of $\zeta(x_k,\omega_k,\cdot)$ over T. Then, $\chi(x_k,\omega_k) = \zeta(x_k,\omega_k,\tau_k)$ and $\chi(x_k,\omega_k) \leq \zeta(x_k,\omega_k,\tau_\ell)$ with $\ell \neq k(\ell=1,2)$. Hence

$$- |\zeta(x_1,\omega_1,\tau_2) - \zeta(x_2,\omega_2,\tau_2)| \leq \zeta(x_1,\omega_1,\tau_1) - \zeta(x_2,\omega_2,\tau_2)$$

$$= \chi(x_1,\omega_1) - \chi(x_2,\omega_2) \leq \zeta(x_1,\omega_1,\tau_2) - \zeta(x_2,\omega_2,\tau_2)$$

$$\leq |\zeta(x_1,\omega_1,\tau_2) - \zeta(x_2,\omega_2,\tau_2)|. \qquad \text{(A.0.1)}$$

Since T is compact and ζ is C^1 (and hence also locally Lipschitz), there exists $L \geq 0$ such that for any $x_1, x_2 \in D$, $\omega_1, \omega_2 \in \Omega$ and any $\tau \in T$,

$$|\zeta(x_1,\omega_1,\tau) - \zeta(x_2,\omega_2,\tau)| \leq L\|y_1 - y_2\| \qquad \text{(A.0.2)}$$

where $y_i^T = [x_i^T, \omega_i^T]$, $i = 1,2$. It now follows from (A.0.1) that $\chi(\cdot,\cdot)$ is locally Lipschitz. By similar means we can easily show that $\psi(\cdot)$ is locally Lipschitz. ¤

Appendix 1: Proof of Theorem 4.3

We shall need the following facts and lemma.

Fact A.1.1. For all $\varepsilon \geq 0$ the point-to-set map $\Gamma_i^{\varepsilon}(\cdot)$ is upper semicontinuous. ¤

Fact A.1.2. For all $\varepsilon \geq 0$, the point-to-set map $M_i^{\varepsilon}(\cdot)$ is upper semicontinuous. ¤

Let

$$\varepsilon(x) \triangleq \max \{\varepsilon | \varepsilon = \varepsilon_o 2^{-k} \text{ or } \varepsilon = 0,\ k \in \mathbb{N},\ \theta_i^{\varepsilon}(x) \geq \varepsilon\}. \tag{A.1.1}$$

We recall that $F_i \triangleq \{x | \max_{\omega \in \Omega_i} \min_{\tau \in T} \zeta(x,\omega,\tau) \leq 0\}$.

Lemma A.1. For all $\hat{x} \in F_i$, if $\theta_i^o(\hat{x}) > 0$, then there exist $\hat{\rho} \geq \hat{\varepsilon} > 0$, such that for all $x \in B(\hat{x},\hat{\rho})$

$$\varepsilon(x) \geq \hat{\varepsilon} > 0. \tag{A.1.2}$$

Proof. Let $\delta > 0$ be such that

$$\|\mathrm{Nr} N_\delta(M_i^o(\hat{x}))\|^2 \geq \theta_i^o(\hat{x})/2 \triangleq \varepsilon > 0. \tag{A.1.3}$$

By upper semicontinuity of $M_i^o(\cdot)$, there exists $\hat{\rho} > 0$ such that for all $x \in B(\hat{x},2\hat{\rho})$

$$M_i^o(x) \subset N_\delta(M_i^o(\hat{x})). \tag{A.1.4}$$

Therefore, for all $x \in B(\hat{x},\hat{\rho})$

$$M_i^{\hat{\rho}}(x) \subset N_\delta(M_i^o(\hat{x})) \tag{A.1.5}$$

and hence

$$\theta_i^{\hat{\rho}}(x) \geq \|\mathrm{Nr}N_\delta(M_i^o(\hat{x}))\| \geq \varepsilon \tag{A.1.6}$$

Now let $\hat{\varepsilon} = \min\{\varepsilon,\hat{\rho}\}$. Then for all $x \in B(\hat{x},\hat{\rho})$

$$\theta_i^{\hat{\varepsilon}}(x) \geq \hat{\varepsilon} \tag{A.1.7}$$

and, by definition of $\varepsilon(x)$, (A.1.2) holds. ¤

Proof of Theorem 4.3. For the sake of contradiction, suppose that there exists an accumulation point $\hat{x}$ of $\{x_j\}$, such that $\theta_i^o(\hat{x}) > 0$. First, note that since $f(x_j)$ decreases monotonically, we must have $f(x_j) \to f(\hat{x})$. Next, by Lemma A.1, there exists $\hat{\rho}$, $\hat{\varepsilon} > 0$ such that for all $x \in B(\hat{x},\hat{\rho})$, (A.1.2) holds. Therefore, for all $x \in B(\hat{x},\hat{\rho})$,

$$M_i^{\varepsilon(x)}(x) \supset M_i^{\hat{\varepsilon}}(x). \tag{A.18}$$

Let J be an infinite subset such that $x_i \overset{J}{\to} \hat{x}$. Then, by the mean value theorem, we have that for all j,

$$\begin{aligned} &f(x_j+\lambda h_i^{\varepsilon(x_j)}(x_j)) - f(x_j) + \alpha\lambda\theta_i^{\varepsilon(x_j)}(x_j) \\ &\leq \lambda[\sup_{r\in(0,1]} \|\nabla f(x_j+r\lambda h_i^{\varepsilon(x_j)}(x_j) - \nabla f(x_j)\| \, \|h_i^{\varepsilon(x_j)}(x_j)\| \\ &+ \langle \nabla f(x_j), h_i^{\varepsilon(x_j)}(x_j)\rangle + \alpha\theta_i^{\varepsilon(x_j)}(x_j)]. \end{aligned} \tag{A.1.9}$$

By definition of $h_i^{\varepsilon(x_j)}(x_j)$, for all j,

$$\langle \nabla f(x_j), h_i^{\varepsilon(x_j)}(x_j)\rangle \leq -\theta_i^{\varepsilon(x_j)}(x_j) \tag{A.1.10}$$

Since $x_j \overset{J}{\to} \hat{x}$, there exists a $\hat{j} \in J$, such that for all $j \geq \hat{j}$, $j \in J$,

$$x_j \in B(\hat{x},\hat{\rho}), \tag{A.1.11}$$

and, therefore, $\varepsilon(x_j)$ computed in Step 3, satisfies

$$\theta_i^{\varepsilon(x_j)}(x_j) > \varepsilon(x_j) \geq \hat{\varepsilon} \tag{A.1.12}$$

by Lemma A.1.1. Now, since $\nabla f(\cdot)$ is continuous and $M_i^o(\cdot)$ is bounded on bounded subsets of $\mathbb{R}^n$, $\|h^{\varepsilon(x_j)}(x_j)\|$ is bounded on the sequence $\{x_j\}$ and hence there exists a $b > 0$ such that for all $j \in J$,

$$\|h_i^{\varepsilon(x_j)}(x_j)\| \leq b \tag{A.1.13}$$

Next because of the uniform continuity of $\nabla f(\cdot)$ on compact subsets of $\mathbb{R}^n$, and (A.1.9), (A.1.10), there exists a $\hat{\lambda} > 0$, such that, for all $j \geq \hat{j}$, $j \in J$, for all $\lambda \in [0,\hat{\lambda}]$

$$f(x_j+\lambda h_i^{\varepsilon(x_j)}(x_j)) - f(x_j) + \alpha\lambda\theta_i^{\varepsilon(x_j)}(x_j) \leq -\lambda/2(1-\alpha)\hat{\varepsilon} < 0. \tag{A.1.14}$$

Now, suppose that there exists an infinite subset $J' \subset J$ such that $\psi_i(x_j) \leq -\varepsilon(x_j)$. Then for all $j \geq \hat{j}$, $j \in J'$,

$$\psi_i(x_j) < -\hat{\varepsilon}. \tag{A.1.15}$$

Therefore by uniform continuity of $\psi_i(\cdot)$ on compact subsets of $\mathbb{R}^n$, there exists ρ' such that for all $j \geq \hat{j}$, $j \in J'$, for all $x \in B(x_j,\rho')$

$$\psi_i(x) \leq 0. \tag{A.1.16}$$

Because of (A.1.13), it now follows that there exists a $\lambda' > 0$, such that for all $\lambda \in [0,\lambda']$, for all $j \geq \hat{j}$, $j \in$,

$$\psi_i(x_j+\lambda h_i^{\varepsilon(x_j)}(x_j)) \leq 0. \tag{A.1.17}$$

Let $\overline{k} \geq 0$ be such that

$$\beta^{\bar{k}} = \bar{\lambda} \leq \min\{\lambda', \hat{\lambda}\}. \tag{A.1.18}$$

Then, for all $j \geq \hat{j}$, $j \in J'$, the algorithm will select, in Step 4, $k_j \leq \bar{k}$, and hence

$$f(x_j + \beta^{k_j} h_i^{\varepsilon(x_j)}(x_j)) - f(x_j) \leq -\alpha\beta^{k_j}\theta_i^{\varepsilon(x_j)}(x_j) \leq -\alpha\beta^{\bar{k}}\hat{\varepsilon}. \tag{A.1.19}$$

Therefore $\{f(x_j)\}$ and hence also $\{f(x_j)\}_{j\in J'}$ is not Cauchy and by continuity of $f(\cdot)$, we conclude that $\{x_j\}_{j\in J'}$, is not Cauchy, contradicting the hypothesis that $\hat{x}$ is an accumulation point of $\{x_j\}_{j\in J}$, $J \supseteq J'$.

If no infinite subset of J satisfies (A.1.15), then there exists $j' \geq \hat{j}$, such that, for all $j \geq j'$, $j \in J$,

$$\psi_i(x_j) \geq -\varepsilon(x_j). \tag{A.1.20}$$

Then, $M_i^{\varepsilon(x_j)}(x_j) \supset \Gamma_i^{\varepsilon(x_j)}(x_j)$ and by Lemma A.1.1, for all $j \geq j'$,

$$M_i^{\varepsilon(x_j)}(x_j) \supset M_i^{\hat{\varepsilon}}(x_j) \supset \Gamma_i^{\hat{\varepsilon}}(x_j). \tag{A.1.21}$$

By the mean value Theorem 3.1 and by Proposition 3.9, we have, for all j

$$\psi_i(x_j + \lambda h_i^{\varepsilon(x_j)}(x_j)) - \psi_i(x_j) = \lambda\langle g, h_i^{\varepsilon(x_j)}(x_j)\rangle \tag{A.1.22}$$

for some $g \in \Gamma_i^{\lambda\|h_i^{\varepsilon(x_j)}(x_j)\|}(x_j)$. By (A.1.13), there exists $\lambda^* > 0$, such that for all $j \geq j'$,

$$\lambda^* \|h_i^{\varepsilon(x_j)}(x_j)\| \leq \hat{\varepsilon}. \tag{A.1.23}$$

Therefore, by (A.21), and by (A.1.2), for all $j \geq j'$, for all $\lambda \in [0,\lambda^*]$,

$$\psi_i(x_j+\lambda h_i^{\varepsilon(x_j)}(x_j)) - \psi(x_j) = \lambda\langle g,h_i^{\varepsilon(x_j)}(x_j)\rangle \leq -\lambda\theta_i^{\varepsilon(x_j)}(x_j)$$

$$\leq -\lambda\hat{\varepsilon} < 0. \tag{A.1.24}$$

Let $\bar{k} \geq 0$ be such that

$$\beta^{\bar{k}} = \bar{\lambda} \leq \min\{\lambda^*,\hat{\lambda}\}. \tag{A.1.25}$$

Then, for all $j \geq j'$, $j \in J$, because of (A.1.9) and (A.1.24), the algorithm will select a $k_j \leq \bar{k}$, and therefore

$$f(x_j+\beta^{k_j}h_i^{\varepsilon(x_j)}(x_j)) - f(x_j) \leq -\alpha\beta^{\bar{k}}\hat{\varepsilon},$$

which shows that the sequence $\{f(x_j)\}$, and hence also $\{f(x_j)\}_{j\in J}$, is not Cauchy, contradicting the hypothesis that $\{x_j\}_{j\in J}$ converges to $\hat{x}$. ¤

Appendix 2. Proof of Theorem 4.4

To prove Theorem 4.4, we need the following lemma and facts.

Lemma A.2.1. The $M_i^{\varepsilon}(x_j)$ Approximation Refinement subprocedure of Algorithm 6 terminates in a finite number of iterations.

Proof. For the sake of contradiction, suppose that the subprocedure does not terminate finitely. Then, for all m,

$$d\psi_i(x_j+\mu_m\tilde{h};\tilde{h}) < -\gamma\hat{\theta}_i^r(x_j,\tilde{\varepsilon}_p) < 0. \tag{A.2.1}$$

Moreover, the bisection implemented in the subprocedure produces infinite sequences $\{\lambda_m\}$, $\{\mu_m\}$ and $\{\rho_m\}$, which satisfy

$$\lim_{m\to\infty} \lambda_m = \lim_{m\to\infty} \rho_m = \lim_{m\to\infty} \mu_m = \hat{\mu}. \tag{A.2.2}$$

By Proposition 3.13, Corollary 3.5, and (A.27),

$$\lim_{m\to\infty} d\psi_i(x_j+\mu_m\tilde{h};\tilde{h}) = d\psi_i(x_j+\hat{\mu}\tilde{h};\tilde{h}) \leq -\gamma\tilde{\theta}_i^r(x_j,\tilde{\varepsilon}_p) < 0. \tag{A.2.3}$$

By the test Step 12 in Algorithm 6, for all m,

$$\psi_i(x_j+\lambda_m\tilde{h}) \leq 0 \tag{A.2.4}$$

and

$$\psi_i(x_j+\rho_m\tilde{h}) > 0. \tag{A.2.5}$$

Hence,

$$\psi_i(x_j+\hat{\mu}\tilde{h}) \leq 0. \tag{A.2.6}$$

By subtracting (A.2.6) from (A.2.5) and dividing by $\rho_m - \hat{\mu}$, we obtain

for all m, since $\rho_m - \hat{\mu} > 0$,

$$[\psi_i(x_j+\rho_m\tilde{h}) - \psi_i(x_j+\hat{\mu}\tilde{h})]/(\rho_m-\hat{\mu}) > 0. \tag{A.2.7}$$

By the definition of directional derivative and (A.2.7), we have

$$\lim_{m\to\infty}[\psi_i(x_j+\hat{\mu}\tilde{h}+(\rho_m-\hat{\mu})\tilde{h}) - \psi(x_j+\hat{\lambda}\tilde{h})]/(\rho_m-\hat{\lambda})$$

$$= d\psi_i(x_j+\hat{\mu}\tilde{h};\tilde{h}) \geq 0 \tag{A.2.8}$$

which contradicts (A.2.3). ¤

Fact A.2.1. Given any $\varepsilon \geq 0$, for all r, for all j,

$$\tilde{\theta}_i^r(x_j,\varepsilon) \geq \theta_i^\varepsilon(x_j) \tag{A.2.9}$$

and

$$\tilde{\varepsilon}(x_j) \geq \varepsilon(x_j) \triangleq \max\{\varepsilon \mid \varepsilon = \varepsilon_o 2^{-k} \text{ or } \varepsilon = 0,\ k \in \mathbb{N},\ \theta_i^\varepsilon(x_j) > \varepsilon\} \tag{A.2.10}$$

Proof. Equation (A.2.9) follows from the definition of $\tilde{\theta}_i^r(x_j,\varepsilon)$ and from the fact that for all r, for all j, $\tilde{M}_i^r(x_j,\varepsilon) \subseteq M_i^\varepsilon(x_j)$. Equation (A.2.10) follows from the definition of $\tilde{\varepsilon}(x_j)$ and from (A.2.9). ¤

Fact A.2.2. Let $\tilde{M}$ be a compact, converse subset of a compact convex set $M \subset \mathbb{R}^n$ and let $\gamma \in (0,1)$. Let $\tilde{h} \triangleq -\text{Nr}\tilde{M}$ and let Γ be a subset of M, such that for some $v \in \Gamma$

$$\langle v,\tilde{h}\rangle \geq -\gamma\|\tilde{h}\|^2. \tag{A.2.11}$$

Then $\tilde{h}' \triangleq -\text{Nr co}\{\tilde{M} \cup \Gamma\}$ satisfies

$$\|\tilde{h}'\|^2 \leq \max\{\gamma, 1-(1-\gamma)^2\|\tilde{h}\|^2/4C^2\}\|\tilde{h}\|^2 \tag{A.2.12}$$

where $C \triangleq \max\{\|h\| \,|\, h \in M\}$.

Proof. The fact follows from [21] (Theorem 1, part (i)) and from [13] (Lemma 4.4). ¤

Proof of Theorem 4.4. Suppose that Algorithm 6 constructs a finite sequence $\{x_j\}$ with last element x_s. Since it is obvious that $\psi_i(x_s) \leq 0$, we have to prove that $\theta_i^o(x_s) = 0$. For the sake of contradiction suppose that $\theta_i^o(x_s) > 0$. By Lemma A.1.1, $\varepsilon(x_s) = \varepsilon_o 2^{-\hat{\ell}} > 0$. By Fact A.2.1, Algorithm 4.6 jams up with $p = \hat{p}$, such that

$$\tilde{\varepsilon}_{\hat{p}} \geq \varepsilon(x_s) > 0. \tag{A.2.13}$$

This implies that $\hat{p} \leq \hat{\ell}$ and Algorithm 6 cannot jam up between Step 2 and Step 4. By Lemma A.2.1, it cannot jam up between Step 9 and Step 12. Therefore, jamming of Algorithm 6 may only be caused by infinite cycling between Step 2 and the $M_i^{\varepsilon}(x_j)$ Approximation Refinement Subprocedure, i.e., by $r \to \infty$. By (A.2.13), we have, for all r,

$$\tilde{\theta}_i^r(x_j, \tilde{\varepsilon}_{\hat{p}}) > \tilde{\varepsilon}_{\hat{p}}. \tag{A.2.14}$$

Therefore, for all r,

$$\tilde{M}_i^{r+1}(x_s, \tilde{\varepsilon}_{\hat{p}}) = \text{co}\{\tilde{M}_i^r(x_s, \tilde{\varepsilon}_{\hat{p}}) \cup \Gamma_i(x_s + \tilde{\lambda}_r \tilde{h}) \ ,$$

By Step 6 and by the $M_i^{\varepsilon}(x_j)$ Approximation Refinement Subprocedure, for all r,

$$\Gamma_i(x_s + \tilde{\lambda}_r \tilde{h}) \subseteq M_i^{\tilde{\varepsilon}_{\hat{p}}}(x_s), \tag{A.2.15}$$

and

$$\tilde{M}_i^r(x_s, \tilde{\varepsilon}_{\hat{p}}) \subseteq M_i^{\tilde{\varepsilon}_{\hat{p}}}(x_s). \tag{A.2.16}$$

$M_i^{\tilde{\varepsilon}_{\hat{p}}}(x_s)$ is convex and compact. By Step 11 in Algorithm 4.6, there exists $v_r \in \Gamma_i(x_s+\tilde{\lambda}_r h)$ such that

$$\langle v_r, -Nr\tilde{M}_i^r(x_s,\tilde{\varepsilon}_{\hat{p}})\rangle \geq -\gamma\|Nr\tilde{M}_i^r(x_s,\tilde{\varepsilon}_{\hat{p}})\|^2. \tag{A.2.17}$$

Therefore, by Fact A.2.2, for all r,

$$\theta_i^{r+1}(x_s,\tilde{\varepsilon}_{\hat{p}}) \leq \max\{\gamma, 1-(1-\gamma)^2\frac{\tilde{\theta}_i^r(x_s,\tilde{\varepsilon}_{\hat{p}})}{4C^2}\}\,\tilde{\theta}_i^r(x_s,\tilde{\varepsilon}_{\hat{p}})$$

By Fact A.2.1,

$$\tilde{\theta}_i^{r+i}(x_s,\tilde{\varepsilon}_{\hat{p}}) \leq \max\{\gamma, 1-(1-\gamma)^2\frac{\theta_i^{\tilde{\varepsilon}_{\hat{p}}}(x_s)}{4C^2}\}\tilde{\theta}_i^r(x_s,\tilde{\varepsilon}_{\hat{p}})$$

$$\leq \hat{L}\tilde{\theta}_i^r(x_s,\tilde{\varepsilon}_{\hat{p}}) \tag{A.2.19}$$

where $\hat{L} \triangleq \max\{\gamma, 1-(1-\gamma)^2\theta_i^{\tilde{\varepsilon}_{\hat{p}}}(x_s)/4C^2\} < 1$ does not depend on r. Therefore, for all r

$$\tilde{\theta}_i^r(x_s,\tilde{\varepsilon}_{\hat{p}}) \leq \hat{L}^r\tilde{\theta}_i^o(x_s,\tilde{\varepsilon}_{\hat{p}}). \tag{A.2.20}$$

Hence, there exists $\hat{r}$ such that

$$\tilde{\theta}_i^{\hat{r}}(x_s,\tilde{\varepsilon}_p) \leq \tilde{\varepsilon}_{\hat{p}} \tag{A.2.21}$$

which contradicts (A.2.14).

Now, suppose that Algorithm 6 constructs an infinite sequence $\{x_j\}$. Since it is obvious that by continuity of $\psi_i(\cdot)$, for every accumulation point $\hat{x}$ of $\{x_j\}$, $\psi_i(\hat{x}) \leq 0$, we have to prove that for every accumulation point $\hat{x}$ of $\{x_j\}$, $\theta_i^o(\hat{x}) = 0$. Suppose for the sake of contradiction, that there exists an accumulation point $\hat{x}$ of $\{x_j\}$, such that $\theta_i^o(\hat{x}) > 0$. Let

J be an infinite subset of the integers such that $x_j \overset{J}{\to} \hat{x}$. By the first part of the proof of Theorem 4.3, and by definition of $\tilde{M}_i^r(x_j,\tilde{\varepsilon}_p)$, there exists $\hat{\varepsilon} > 0$, $\lambda > 0$ and $\hat{j} \in J$, such that for all $j \geq \hat{j}$, $j \in J$, for all $\lambda \in [0,\hat{\lambda}]$,

$$f(x_j+\lambda\tilde{h}(x_j))-f(x_j)+\alpha\lambda\tilde{\theta}_i^r(x_j,\tilde{\varepsilon}_p) \leq -\lambda/2(1-\alpha)\hat{\varepsilon} < 0 \qquad \text{(A.2.22)}$$

Since $\nabla f(\cdot)$ is continuous and $\Gamma_i(\cdot)$ is bounded on bounded subsets, there exists $b > 0$, such that for all $j \in J$,

$$\|\tilde{h}(x_j)\| \leq b. \qquad \text{(A.2.23)}$$

Let $\lambda^* \triangleq \min\{\hat{\varepsilon}/b,\hat{\lambda}\}$. Then, by Step 6 of Algorithm 6 and by Fact A.2.1, for all r, for all $\lambda \in [0,\lambda^*]$, for all $j \geq \hat{j}$, $j \in J$,

$$f(x_j+\lambda\tilde{h}(x_j)) - f(x_j) + \alpha\lambda\tilde{\theta}_i^r(x_j,\tilde{\varepsilon}_p) \leq -\lambda/2(1-\alpha)\hat{\varepsilon} < 0 \qquad \text{(A.2.24)}$$

and

$$\psi_i(x_j+\lambda\tilde{h}(x_j)) \leq 0. \qquad \text{(A.2.25)}$$

Let $\hat{k}$ be such that $\beta^{\hat{k}} \leq \lambda^* < \beta^{\hat{k}-1}$. By (A.2.24) and (A.2.25), Algorithm 6 selects $k_j \leq \hat{k}$. Therefore, we have for all $j \geq \hat{j}$, $j \in J$,

$$f(x_j+\beta^{k_j}\tilde{h}(x_j)) - f(x_j) \leq -\alpha\beta^{\hat{k}}\hat{\varepsilon} \qquad \text{(A.2.26)}$$

and, since $f(x_{j+1}) < f(x_j)$ for all j, the sequence $\{f(x_j)\}_{j\in J}$ is not Cauchy, contradicting the hypothesis that $\{x_j\}_{j\in J}$ converges to $\hat{x}$. ¤

Appendix 3. Proof of Lemma 4.3

For the sake of contradiction suppose that $\theta(\hat{z}) > 0$. Since by hypothesis $\psi(\hat{z}) \leq 0$,

$$\psi_i(\hat{z}) \leq \psi(\hat{z}) \leq 0, \text{ for all } i. \tag{A.3.1}$$

From (A.3.1) and the definition of $M_i^o(\cdot)$, we have

$$M(\hat{z}) \supseteq M_i^o(\hat{z}). \tag{A.3.2}$$

Therefore, for all i,

$$\theta_i^o(\hat{z}) \triangleq \|\text{Nr}M_i^o(z)\|^2 \geq \|\text{Nr}M(\hat{z})\|^2 \triangleq \theta(\hat{z}) > 0. \tag{A.3.3}$$

By (A.3.3), there exists $\delta > 0$ such that

$$\|\text{Nr}N_\delta(M(\hat{z}))\|^2 \geq \theta(\hat{z})/2 > 0. \tag{A.3.4}$$

By (A.3.2) and by upper semicontinuity of $M_i^o(\cdot)$, there exists $\hat{\rho} > 0$ such that, for all $z \in B(\hat{z},2\hat{\rho})$

$$M_i^o(z) \subset N_\delta(M(\hat{z})). \tag{A.3.5}$$

Therefore, for all $z \in B(\hat{z},\hat{\rho})$, from the definition of $M_i^\varepsilon(\cdot)$,

$$M_i^{\hat{\rho}}(z) \subset N_\delta(M(\hat{z})). \tag{A.3.6}$$

Hence, for all $z \in B(\hat{z},\hat{\rho})$, by (A.3.2) and (A.3.6)

$$\theta_i^{\hat{\rho}}(z) \geq \theta(\hat{z})/2 > 0 \tag{A.3.7}$$

Since $z_i \to \hat{z}$ by hypothesis, there exists $\hat{i}$ such that for all $i \geq \hat{i}$, $z_i \in B(\hat{z},\hat{\rho})$, and hence,

$$\theta_i^{\hat{\rho}}(z_i) \geq \theta(\hat{z})/2 > 0. \tag{A.3.8}$$

By Step 6 of Algorithm 7, $\eta_i \to 0$ as $i \to \infty$, and, hence, there exists $i' \geq \hat{i}$, such that for all $i \geq i'$

$$\eta_i \leq \hat{\rho}. \tag{A.3.9}$$

Therefore, for all $i \geq i'$,

$$\theta_i^{\eta_i}(z_i) \geq \theta_i^{\hat{\rho}}(z_i) \geq \theta(\hat{z})/2 > 0, \tag{A.3.10}$$

which contradicts the hypothesis of Lemma 4.3. ¤

REFERENCES

[1] J. W. Bandler, P. C. Liu and H. Tromp, "A Nonlinear Programming Approach to Optimal Design Centering, Tolerancing and Tuning," IEEE Trans. on Circuits and Systems, vol. CAS-23, n. 3, March 1976, pp. 155-165.

[2] P. V. Lopresti, "Optimum Design of Linear Tuning Algorithms," IEEE Trans. on Circuits and Systems, vol. CAS-24, n. 3, March 1977, pp. 144-151.

[3] J. F. Pinel, "Computer-Aided Network Tuning," IEEE Trans. on Circ. Theory, vol. CT-17, n. 1, Jan. 1971, pp. 192-194.

[4] G. Müller, "On Computer-Aided Tuning of Microwave Filters," IEEE Proc. ISCAS 1976, Münich, May 1976, pp. 209-211.

[5] E. Lüder and B. Kaiser, "Precision Tuning of Miniaturized Circuits," IEEE Proc. ISCAS 1976, Münich, May 1976, pp. 722-725.

[6] R. L. Adams and V. K. Manaktala, "An Optimization Algorithm Suitable for Computer-Assisted Network Tuning," IEEE Proc. 1975 ISCAS, Boston 1975, pp. 210-212.

[7] B. C. Eaves and W. I. Zangwill, "Generalized Cutting Plane Algorithms," SIAM Journ. Control, vol. 9, n. 4, pp. 529-542.

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

[9] C. Gonzaga and E. Polak, "On Constraint Dropping Schemes and Optimality Functions for a Class of Outer Approximation Algorithms," Memo. #M77-68, ERL, University of California, Berkeley, November 1977, also SIAM J. on Control and Optimization, to appear.

[10] J. W. Bandler, "Optimization of Design Tolerances Using Nonlinear Programming," J. Optimization Theory and Applications, vol. 14, July 1974, pp. 99-114.

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

[12] E. Polak, R. Trahan and D. Q. Mayne, "Combined Phase I-Phase II Methods of Feasible Directions," Memo. #M77-39, ERL, University of California, Berkeley, May 1977, also Math. Programming, to appear.

[13] R. Mifflin, "An Algorithm for Constrained Optimization with Semismooth Functions," Math. of Oper. Research, vol. 2, n. 2, May 1977, pp. 191-207.

[14] F. H. Clarke, "A New Approach to Lagrange Multipliers," Math. of Oper. Research, vol. 1, 1976, pp. 165-174.

[15] F. H. Clarke, "Generalized Gradients and Applications," Trans. Amer. Math. Soc., vol. 205, 1975, pp. 247-262.

[16] E. M. Stein, "Singular Integrals and Differentiability Properties of Functions," Princeton Math. Ser., n. 30, Princeton University Press, Princeton, N. J., 1970.

[17] C. Berge, Topological Spaces, MacMillan, New York, 1963.

[18] O. H. Merril, "Applications and Extensions of an Algorithm that Computes Fixed Points of Certain Upper Semicontinuous Point-to-Set Mappings," Ph.D. Dissertation, University of Michigan, Ann Arbor, 1972.

[19] R. Mifflin, "Semismooth and Semiconvex Functions in Constrained Optimization," SIAM J. on Control and Optimization, Nov. 1977, pp. 959-972.

[20] C. Lemarechal, "An Extension of Davison Methods to Nondifferential Problems," in M. L. Balinski and P. Wolfe, eds., Nondifferentiable Optimization, Mathematical Programming Study 3, North Holland, Amsterdam, 1975, pp. 95-109.

[21] P. Wolfe, "A Method of Conjugate Subgradients for Minimizing Nondifferentiable Functions," ibid, pp. 145-173.

[22] V. F. Demyanov, "Algorithm for some Minimax Problems," Journal of Computer and Systems Sciences, vol. 2, 1968, pp. 342-380.

[23] V. F. Demyanov, "Differentiability of a Maximum Function: I," USSR Comp. Math. and Math. Phys., vol. 8, n. 6, 1968, pp. 1-15.

[24] O. Pironneau and E. Polak, "Rate of Convergence of a Class of Methods of Feasible Directions," SIAM J. Numer. Anal., vol. 10, 1973, pp. 161-174.

[25] O. L. Mangasarian, "Dual Feasible Direction Algorithms," in A. V. Balakrishman (ed.) Techniques of Optimization, 1971, pp. 67-88.

[26] L. Armijo, "Minimization of Functions Having Continuous Partial Derivatives," Pacific J. Math, vol. 16, 1966, pp. 1-3.

[27] R. T. Rockafellar, private communications, 1978.

[28] G. Lebourg, "Valeur Moyenne pour gradient généralisé," C. R. Acad. Sci. Paris, vol. 281, 1975, pp. 795-797.

[29] D. P. Bertsekas and S. K. Mitter, "A Descent Numerical Method for Optimization Problems with Nondifferentiable Cost Functionals," SIAM J. on Control, vol. 11, 1973, pp. 637-652.

[30] A. A. Goldstein, "Optimization of Lipschitz Continuous Functions," Math. Programming, vol. 13, 1977, pp. 14-22.

FIGURE CAPTIONS

Fig. 3.1. The graph of $\psi(x) = \max_{\omega\in[-1,+1]} \min_{\tau\in[-1,+1]} \tau(x-\omega)$.

Fig. 3.2. The graph of $\psi'(x) = \max_{\omega\in\{-1,+1\}} \min_{\tau\in[-1,+1]} \tau(x-\omega)$.

Fig. 4.1. Construction of $N_{\hat{\delta}}(\partial\psi(x))$ and property of $\langle g + 1/2\ h(x),\ h(x)\rangle$.

Fig. 4.2. Geometrical interpretation of the Armijo step-size rule when $\psi(\cdot)$ is differentiable at x.

Fig. 4.3. Contours of constant values of $\psi(\cdot)$ and a steepest descent path.

Fig. 4.4. The generalized gradient $\partial\psi(0)$.

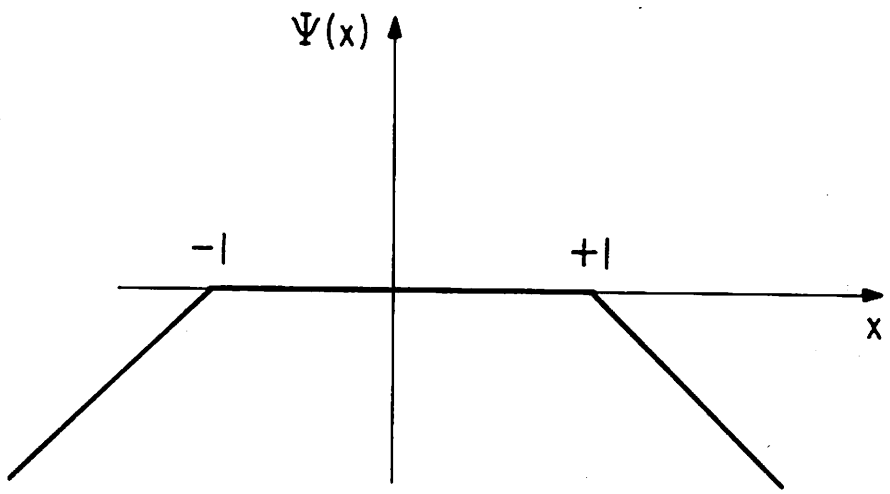

Fig. 3.1. The graph of $\psi(x) = \max_{\omega\in[-1,+1]} \min_{\tau\in[-1,+1]} \tau(x-\omega)$.

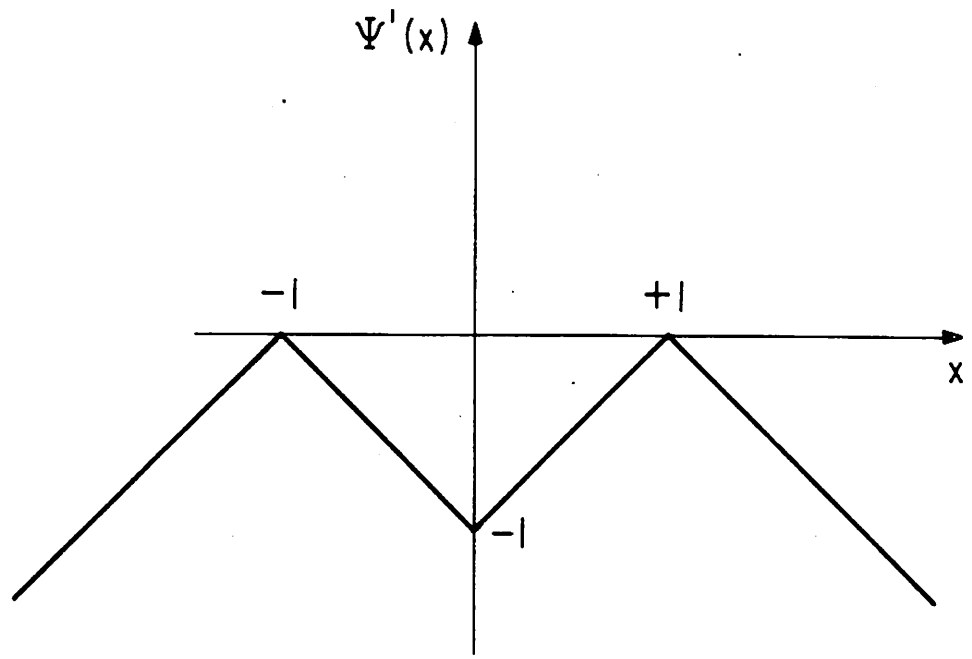

Fig. 3.2. The graph of $\psi'(x) = \max_{\omega\in\{-1,+1\}} \min_{\tau\in[-1,+1]} \tau(x-\omega)$.

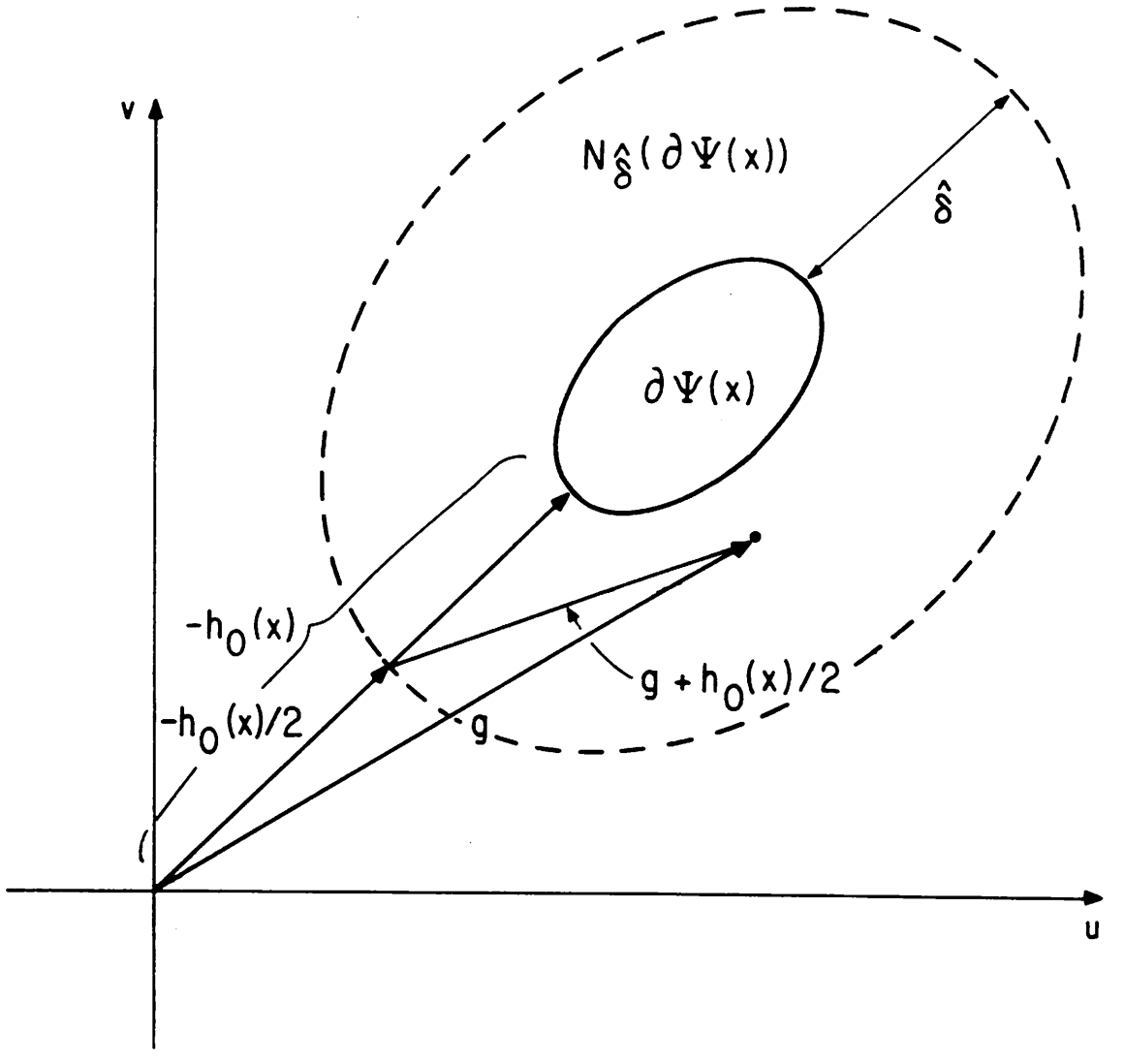

Fig. 4.1. Construction of $N_{\hat{\delta}}(\partial\psi(x))$ and property of $\langle g + 1/2\ h(x),\ h(x)\rangle$.

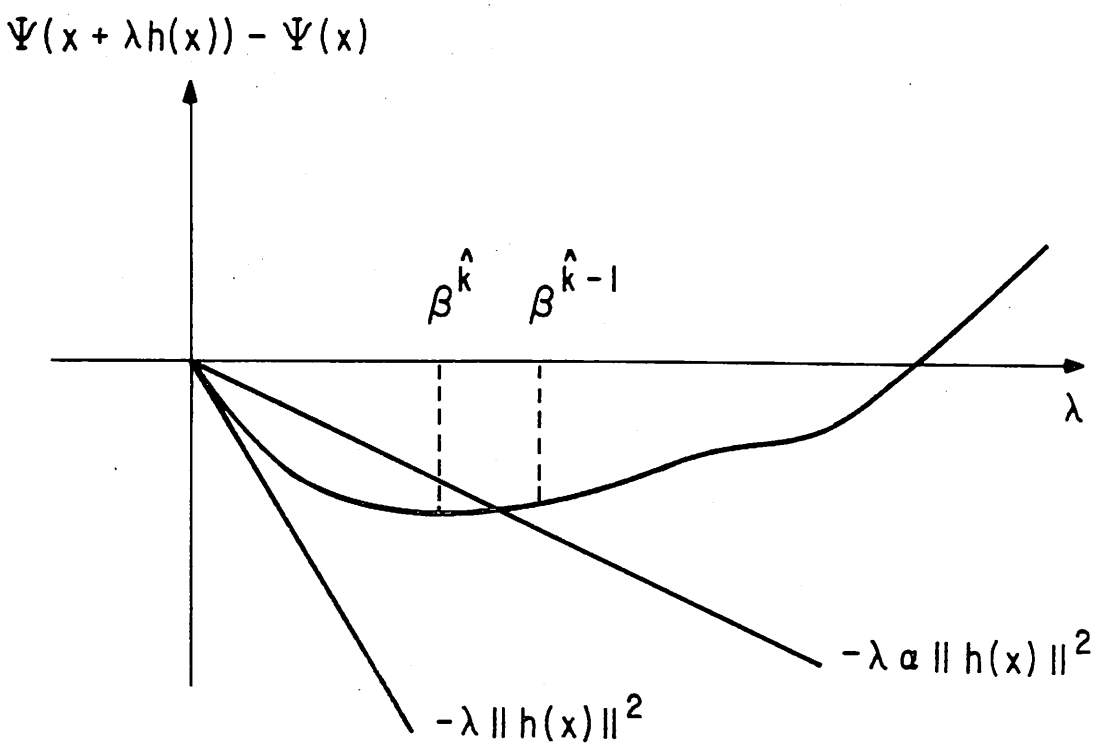

Fig. 4.2. Geometrical interpretation of the Armijo step-size rule when $\psi(\cdot)$ is differentiable at x.

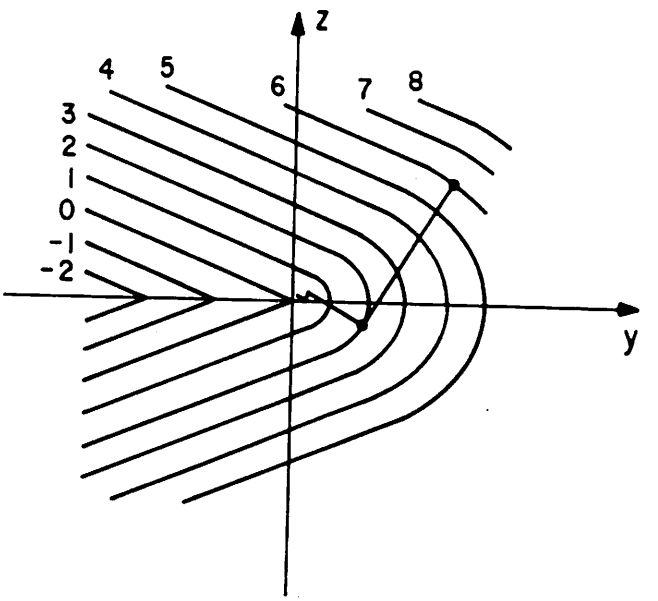

Fig. 4.3. Contours of constant values of $\psi(\cdot)$ and a steepest descent path.

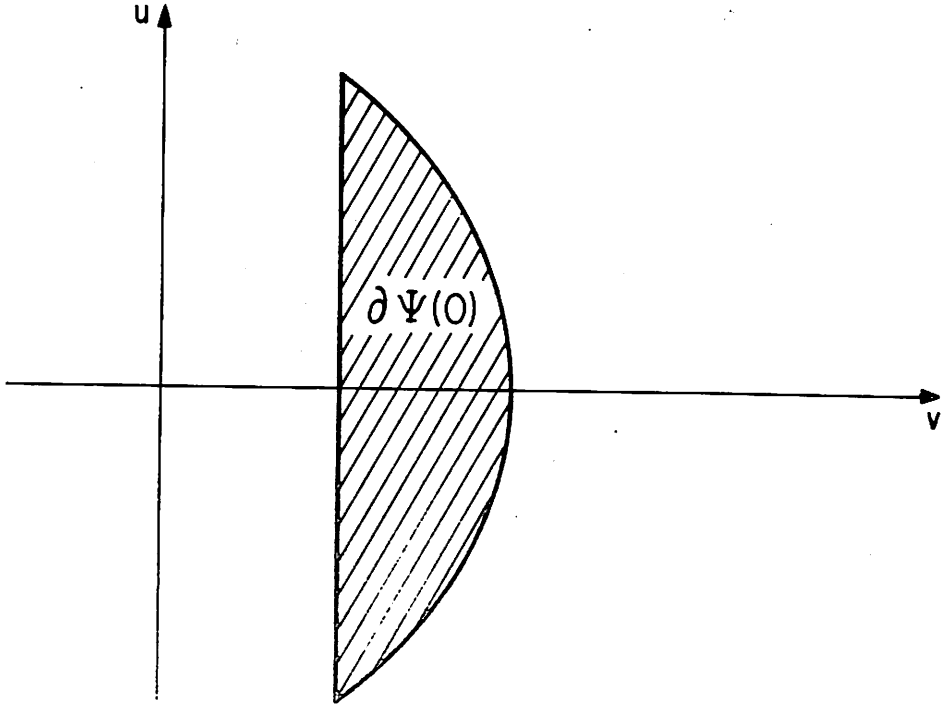

Fig. 4.4. The generalized gradient $\partial\psi(0)$.