ALGORITHMS FOR OPTIMIZATION PROBLEMS

WITH EXCLUSION CONSTRAINTS

by

D. Q. Mayne and E. Polak

Memorandum No. UCB/ERL M85/33

26 April 1985

(cover)

ALGORITHMS FOR OPTIMIZATION PROBLEMS WITH EXCLUSION CONSTRAINTS

by

D. Q. Mayne and E. Polak

Memorandum No. UCB/ERL M85/33

26 April 1985

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

ALGORITHMS FOR OPTIMIZATION PROBLEMS WITH EXCLUSION CONSTRAINTS [1]

D Q Mayne [2] and E Polak [3]

Abstract

This paper proposes algorithms for minimizing a continuously differentiable function $f(x)\colon \mathbb{R}^n \to \mathbb{R}$ subject to the constraint that x does not lie in specified bounded subsets of $\mathbb{R}^n$. Such problems arise in a variety of applications such as tolerance design of electronic circuits and obstacle avoidance in the selection of trajectories for robot arms. Such constraints have the form $\psi(x) \triangleq \min\{g^j(x) \mid j \in J\} \leq 0$. The function ψ is not continuously differentiable. Algorithms based on the use of generalised gradients have considerable disadvantages because of the local concavity of ψ at points where the set $\{j \mid g^j(x) = \psi(x)\}$ has more than one element. Algorithms which avoid these disadvantages are presented and their convergence established.

---

1. Research was sponsored in part by the National Science Foundation under Grant ECS-8121149, The Air Force Office of Scientific Research (AFSC), United States Air Force Under Contract F49620-79-C-0178, The Office of Naval Research Under Grant N00014-83-K-0602, The Air Force Office of Scientific Research Under Grant AFOSR-83-0361, and the Semiconductor Research Consortium under Grant SRC-82-11-008.

2. Department of Electrical Engineering, Imperial College, London SW7 2BT, UK.

3. Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720, USA.

1. INTRODUCTION

We consider optimization problems of the form:

$$\min\{f(x)\,|\,x \notin R_j,\ j \in I;\ g^j(x) \leq 0,\ j \in J\} \tag{1}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable (the cost function), R_j, for each j in I, is a subset of $\mathbb{R}^n$ (an exclusion region) and g_j, for each j in J, is continuously differentiable ($g^j(x) \leq 0$ is a conventional constraint). For each j in I, R_j is defined by:

$$R_j \triangleq \{x \in \mathbb{R}^n\,|\,\psi^j(x) > 0\}. \tag{2}$$

An example of R_j is the set $\{x \in \mathbb{R}^n \,\big|\, \|x\|_\infty < 1\}$ (i.e. $\psi^j(x) = 1 - \|x\|_\infty$); in this case the constraint $x \notin R_j$ is equivalent to the constraint $\psi^j(x) \leq 0$ where $\psi^j(x) \triangleq \min\{1 - x_i, 1 + x_i\,|\,i = 1,\ldots,n\}$. We assume that, in general, $\psi^j\colon \mathbb{R}^n \to \mathbb{R}$ is defined by:

$$\psi^j(x) \triangleq \min\{\phi^{j,k}(x)\,|\,k \in I_j\} \tag{3}$$

where the functions $\phi^{j,k}\colon \mathbb{R}^n \to \mathbb{R}$, $j \in J$, $k \in I_j$, are assumed to be continuously differentiable.

Suppose, for each j in J, that $I_j = \{1,2,\ldots s_j\}$. Then the exclusion contraint $x \notin R_j$ is equivalent to the constraint

$$\psi^j(x) \leq 0 \tag{4}$$

and this in turn is equivalent to the constraint

$$(\phi^{j,1}(x) \leq 0) \text{ or } (\phi^{j,2}(x) \leq 0) \text{ or } \ldots\ \ldots \text{ or } (\phi^{j,s_j}(x) \leq 0) \tag{5}$$

Note the appearance of "or" in (5) compared with the "and" associated with max functions in constraints; the "or" arises from the fact that ψ^j

is defined to be a min function.

The optimization problem can be re-expressed as:

$$\min\{f(x) \mid x \in X\} \tag{6}$$

where

$$X \triangleq \{x \in \mathbb{R}^n \mid \psi^j(x) \leq 0,\ j \in I; g^j(x) \leq 0,\ j \in J\}. \tag{7}$$

The constraint $x \notin \mathbb{R}_j$ (equivalently constraint (5)) has not been extensively studied in the literature; an exception is [1]. Such constraints arise in robotics where each R_j specifies an obstacle to be avoided. They also arise in a sub-problem when outer approximation algorithms are employed (for example to solve problems with infinite dimensional constraints [2]); the sub-problem has the form of problem (6), with X specified by (7), where the sets R_j, $j \in I$ are neighbourhoods of infeasible points previously generated by the main algorithm. A similar problem arises implicitly in the tolerancing problem [2]; in this case a subproblem of the form $\min_x \min_k\{\phi^k(x)\}$ arises, which causes difficulties similar to those arising in problem (1).

The essential feature of the difficulty is the local non-convexity of the level sets of the functions $\psi^j(x)$ at points where the active constraints set $I_j'(x) \triangleq \{k \in I_j \mid \phi^{j,k}(x) = \psi^j(x)\}$ contains more than one element. This is illustrated in <u>Fig. 1(a)</u>. The Clarke generalised gradient [6] is shown in Fig. 1(b); a descent direction, computed using the generalised gradient, will lie in the shaded region. Clearly many permissible search directions are excluded, because a search direction generated using the generalised gradient is a descent direction for <u>each</u> of the active constraints ($\phi^{j,1}$ and $\phi^{j,2}$ in the example of Fig. 1) whereas what is required is a descent direction for <u>any</u> of the active constraints (because $\psi^j(x) = \min_k\{\phi^{j,k}(x) \mid k \in I_j\} = \min_k\{\phi^{j,k}(x) \mid k \in I_j'(x)\}$, a direction h which is a descent direction for $\phi^{j,k}(x)$, where k is any element of $I_j'(x)$, is also a descent direction for $\psi^j(x)$).

2. PRELIMINARIES

Since a conventional constraint may be regarded as a (degenerate) exclusion constraint, our optimization problem may be restated as:

$$P: \quad \min\{f(x) \mid \psi(x) \leq 0\} \tag{7}$$

where $\psi: \mathbb{R}^n \to \mathbb{R}$ is defined by

$$\psi(x) \triangleq \max_{j \in I} \{\psi^j(x)\} \tag{8}$$

and

$$\psi^j(x) \triangleq \min_{k \in I_j} \{\phi^{j,k}(x)\} \tag{9}$$

i.e.

$$\psi(x) = \max_{j \in I} \min_{k \in I_j} \{\phi^{j,k}(x)\}. \tag{10}$$

If ψ^j is a conventional constraint, the cardinality of I_j is unity.

As shown in the Appendix, (10) may be rearranged as:

$$\psi(x) = \min_{k \in K} \max_{j \in J_k} \{\bar{\eta}^{k,j}(x)\} \tag{11}$$

where, for each k in K and j in J_k there exists a j' in I and k' in I_j such that

$$\bar{\eta}^{k,j}(x) = \phi^{j',k'}(x) \tag{12}$$

i.e. (12) is merely a (standard) re-arrangement of (10). The cardinality of K many be high. For each k in K, let $\eta^k: \mathbb{R}^n \to \mathbb{R}$ be defined by

$$\eta^k(x) \triangleq \max_{j \in J_k} \{\bar{\eta}^{k,j}(x)\}\ . \tag{13}$$

For any real valued function ϕ let $\phi(x)_+$ denote $\max\{0,\phi(x)\}$. Then

$$\psi(x) = \min_{k\in K} \{\eta^k(x)\}, \quad \psi(x)_+ = \min_{k\in K}\{\eta^k(x)_+\}, \tag{14}$$

and our optimization problem acquires the form

$$P: \quad \min\{f(x) \mid \min_{k\in K} \{\eta^k(x)\} \leq 0\}. \tag{15}$$

The special feature of the problem arises from the appearance of the min operator in place of the normal max operator in the constraint specification.

The functions $\eta^k: \mathbb{R}^n \to \mathbb{R}$, obtained by maximimizing a finite number of continuously differentiable functions, are not themselves continuously differentiable; however there exist algorithms for minimizing such functions [5].

We propose to solve (15) using an exact penalty function. For each $c > 0$ let $\gamma_c: \mathbb{R}^n \to \mathbb{R}$ be defined by:

$$\gamma_c(x) \triangleq f(x) + c\psi(x)_+ = f(x) + c \min_{k\in K} \{\eta^k(x)_+\} \tag{16}$$

We will show later that, under mild assumptions, for c sufficiently large the constrained optimization problem (15) is equivalent to the unconstrained problem:

$$P_c: \min\{\gamma_c(x) \mid x \in \mathbb{R}^n\}. \tag{17}$$

For each k in K, let $\gamma_c^k: \mathbb{R}^n \to \mathbb{R}$ be defined by:

$$\gamma_c^k(x) \triangleq f(x) + c\eta^k(x)_+ \tag{18a}$$

so that

$$\gamma_c(x) = \min_{k\in K} \{\gamma_c^k(x)\}. \tag{18b}$$

The functions $\gamma_c^k : \mathbb{R}^n \to \mathbb{R}$ are continuous.

Then problem (17) may be expressed as

$$P_c: \quad \min_{x \in \mathbb{R}^n} \{ \min_{k \in K} \{\gamma_c^k (x)\}\} \tag{19}$$

i.e. as the minimization of the <u>minimum</u> of a finite number of functions.

For each k in K let $\hat{\eta}^k : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be defined by

$$\hat{\eta}^k(x, h) \triangleq \max_{j \in J_k} \{\bar{\eta}^{k,j}(x) + \bar{\eta}_x^{k,j}(x)h\}. \tag{20}$$

Similarly let $\hat{\gamma}_c^k (x, h)$ be defined by:

$$\hat{\gamma}_c^k (x, h) \triangleq f(x) + f_x(x)h + c\hat{\eta}^k(x, h)_+. \tag{21}$$

Thus $\hat{\gamma}_c^k (x, h)$ is a first order approximation to $\gamma_c^k (x + h)$ in a sense made precise in Proposition 2. For each k in K let $h_c^k : \mathbb{R}^n \to \mathbb{R}$ be defined by:

$$h_c^k (x) \triangleq \arg \min\{ (1/2) \| h \|^2 + \hat{\gamma}_c^k (x, h)\} \tag{22}$$

and $\theta_c^k : \mathbb{R}^n \to \mathbb{R}$ by:

$$\theta_c^k (x) \triangleq \hat{\gamma}_c^k (x, h_c(x)) - \gamma_c^k (x) \tag{23}$$

Since $h \longmapsto \hat{\gamma}_c^k (x, h)$ is convex, h_c^k is well defined. Clearly $\theta_c^k (x)$ is non-positive for all x. It is shown in the Appendix that $\theta_c^k (x) = 0$ is a necessary condition of optimality for the problem:

$$P_c^k: \quad \min\{\gamma_c^k (x) \,|\, x \in \mathbb{R}^n \}. \tag{24}$$

Indeed, if $\theta_c^k (x) < 0$, then $h_c^k (x)$ is a descent direction for $\gamma_c^k (x)$.

For all $\varepsilon \geq 0$ let the ε-active set $K_\varepsilon(x)$ be defined by:

$$K_\varepsilon(x) \triangleq \{k \in K \mid \eta^k(x) \leq \psi(x) + \varepsilon\}. \tag{25}$$

Note that

$$K_\theta(x) = \{k \in K \mid \eta^k(x) = \psi(x)\}. \tag{26}$$

Let $\theta_c\colon \mathbb{R}^n \to \mathbb{R}$ be defined by

$$\theta_c(x) \triangleq \min\{\theta_c^k(x) \mid k \in K_0(x)\}. \tag{27}$$

Clearly $\theta_c(x) \leq 0$ for all x in $\mathbb{R}^n$, all $c > 0$. We shall show that $\theta_c(x) = 0$ is a necessary condition of optimality for P_c.

Proposition 1

(a) For all $c > 0$, if x is a local solution to P_c (P_c^k) then $\theta_c(x) = 0$ ($\theta_c^k(x) = 0$).

(b) The functions θ_c^k, $h_c^k\colon \mathbb{R}^n \to \mathbb{R}$, $k \in K$, are continuous.

Proof

(a) That $\theta_c^k(x) = 0$ is a necessary condition of optimality for P_c^k is proven in the Appendix.

Suppose that x is optimal for P_c but that $\theta_c(x) < 0$. From the definition of θ_c there exists a $k \in K_0(x)$ such that $\theta_c^k(x) < 0$. Hence x is not locally optimal for P_c^k so that there exists an x' arbitrarily near x such that $\gamma_c^k(x') < \gamma_c^k(x)$. Since $\gamma_c(x') \leq \gamma_c^k(x)$ and $\gamma_c(x) = \gamma_c^k(x)$ (since k is in $K_0(x)$) it follows that $\gamma_c(x') < \gamma_c(x)$, i.e. x' is not locally optimal for P_c, a contradiction. Hence $\theta_c(x) = 0$.

(b) That the functions θ_c^k, h_c^k, $k \in K$, are continuous is proven in the Appendix. □

Comment: Note that we do not claim that θ_c is continuous. Since the set $K_0(x)$ can suddenly decrease, $\theta_c(x)$ can suddenly increase.

3. Algorithms for P_c

We present in this section two algorithms for solving P_c; the first is a simple extension of an algorithm due to Tits [1] for the case when the functions γ_c^k, $k \in K$, are differentiable. The second is a further extension to improve efficiency.

Algorithm 1

Data: $c > 0$, $\varepsilon_0 > 0$, $\alpha \in (0, 1)$, $\beta \in (0, 1)$, $x_0 \in \mathbb{R}^n$.

Step 0: Set $i = 0$.

Step 1: (Determination of search direction)

For all k in $K_{\varepsilon_0}(x_i)$ compute $\theta_c^k(x_i)$ and the corresponding search direction $h_c^k(x_i)$ solving (22) so that

$$\theta_c^k(x_i) = \hat{\gamma}_c^k(x_i, h_c^k(x_i)) - \gamma_c^k(x).$$

Step 2: (Determination of step length)

For all k in $K_{\varepsilon_0}(x_i)$ compute $\lambda_c^k(x_i)$, the largest λ in $S \triangleq \{1, \beta, \beta^2, \ldots\}$ satisfying the Armijo condition

$$\gamma_c^k(x_i + \lambda h_c^k(x_i)) - \gamma_c^k(x_i) \leq -\lambda\alpha\theta_c^k(x_i).$$

Step 3: (Determination of next x)

Determine $k_c(x_i)$, that k in $K_{\varepsilon_0}(x_i)$ which minimizes $\gamma_c^k(x_i + \lambda_c^k(x_i)h_c^k(x_i))$.

Set $x_{i+1} = x_i + \lambda_c^{k_c(x_i)}(x_i) h_c^{k_c(x_i)}(x_i)$.

Set $i = i + 1$. Go to Step 1. □

Theorem 1

Any accumulation point x^* of an infinite sequence $\{x_i\}$ generated by Algorithm 1 satisfies $\theta_c(x^*) = 0$.

Proof

If can be shown (see the proof of Theorem 2 for a similar result) that if $\theta_c^k(x) < 0$ (for any x in $\mathbb{R}^n$, any k in K) then there exist a $\varepsilon > 0$ and a $\delta > 0$ such that

$$\gamma_c^k(A_c^k(x')) - \gamma_c^k(x') \le -\delta \, , \; A_c^k(x') \triangleq x' + \lambda_c^k(x') h_c^k(x')$$

for all x' in $B_\delta(x) \triangleq \{y \mid \|y-x\| \le \delta\}$. This establishes the hypothesis employed by Tits [1]; hence it is easily shown that any accumulation point x^* of an infinite sequence generated by the algorithm satisfies $\theta_c^k(x^*) = 0$ for all k in $K_0(x^*)$, i.e. $\theta_c(x^*) = 0$. □

Computational expense arises in Algorithm 1 in two ways, firstly the computation of $\theta_c^k(x_i)$ for all k in $K_{\varepsilon_0}(x_i)$ (this can be reduced by reducing ε_0) and, secondly, the computation of $A_c^k(x_i)$ for all k in $K_{\varepsilon_0}(x_i)$. The second algorithm reduces the latter computation by employing a single Armijo type computation rather than one for each element of $K_{\varepsilon_0}(x_i)$. In order to specify the algorithm we introduce the function $\hat{\gamma}_c : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by:

$$\hat{\gamma}_c(x, h) \triangleq \min_{k \in K_{\varepsilon_0}(x)} \{\hat{\gamma}_c^k(x, h)\}. \tag{29}$$

Because $\gamma_c^k(x) > \gamma_c(x) + \varepsilon_0$ for all k not in $K_{\varepsilon_0}(x)$ it follows that $\hat{\gamma}_c$ is a first order approximation to γ_c in the sense that $|\gamma_c(x + \lambda h) - \hat{\gamma}_c(x,\lambda h)| = o(\lambda)$.

To specify the second algorithm we need to introduce the following definition. In step 3 of Algorithm 2, $h_c(x, \lambda)$ denotes (with some abuse of notation) $\lambda h_c^{k^0}(x)$ where k^0 is <u>any</u> k in $\arg\min_{k \in K_{\varepsilon_0}(x)} \{\hat{\gamma}_c^k(x, \lambda h_c^k(x))\}$.

The term $h_c(x, \lambda)$ replaces the term λh in more conventional algorithms. In the algorithm to be proposed both the step length (λ) <u>and</u> the search direction ($h_c(x, \lambda)$) change as λ varies. In conventional algorithms the search direction does <u>not</u> vary. It follows from this definition that

$$\hat{\gamma}_c(x, h_c(x, \lambda)) = \min_{k \in K_{\varepsilon_0}(x)} \{\hat{\gamma}_c^k(x, \lambda h_c^k(x))\}.$$

(independently of the k^0 employed in constructing $h_c(x, \lambda)$).

A typical plot of $\hat{\gamma}_c(x, h_c(x, \lambda))$ as λ varies from 0 to 1 is shown in Fig. 2. Note that the function is the minimum of a set of convex functions and is neither convex nor concave.

Algorithm 2

Data: $c > 0$; $\varepsilon_0 > 0$; $\beta \in (0, 1)$; $x_0 \in \mathbb{R}^n$.

Step 0: Set $i = 0$.

Step 1: (Determination of search direction)

For all k in $K_{\varepsilon_0}(x_i)$ compute $\theta_c^k(x_i)$ and the corresponding search direction $h_c^k(x_i)$.

Step 2: (Determination of step length)

Compute $\lambda_c(x_i)$, the largest λ in S such that:

$$[\gamma_c(x_i+h_c(x_i,\lambda))-\gamma_c(x_i)]\leq(1/2)[\hat{\gamma}_c(x_i,h_c(x_i,\lambda))-\gamma_c(x_i)] .$$

Step 3: Set $x_{i+1} = x_i + h_c(x_i, \lambda_c(x_i))$.

Set $i = i+1$.

Go to step 1. □

To analyse this algorithm we require the following result:

Proposition 2

For all x in $\mathbb{R}^n$, all $\eta > 0$, there exists a $\delta > 0$ such that

$$|\eta^k(x' + h) - \hat{\eta}^k(x', h)| \leq \eta\|h\|$$

for all x' in $B_\delta(x)$, all h in $B_\delta(0)$ and all k in K.

Proof: See Appendix. □

We can now state our main result.

Theorem 2

Suppose Algorithm 2 generates an infinite sequence $\{x_i\}$. Then any accumulation point x^* of $\{x_i\}$ satisfies $\theta_c(x^*) = 0$.

Proof

Suppose $x_i \xrightarrow{I} x^*$ where I is some subsequence of $\{0,1,2,...\}$ and that, contrary to what is to be proven, $\theta_c(x^*) < 0$. By construction, there exists a k^* in $K_0(x^*)$ such that

$$\theta_c^{k^*}(x^*) = \theta_c(x^*) .$$

Since $K_{\varepsilon_0}(x) \triangleq \{k \in K \mid \eta^k(x) \leq \psi(x) + \varepsilon_0\}$ and h_c^k is continuous for all k in K, (see Prop. A1) there exists an $\varepsilon_1 > 0$ and a compact subset H of $\mathbb{R}^n$ such that:

$$k^* \in K_{\varepsilon_0}(x)$$

$$K_{\varepsilon_0}(x) \subset K_{\varepsilon_0}(x^*)$$

$$\gamma_c(x) = \min_{k \in K_{\varepsilon_0}(x)} \gamma_c^k(x) = \min_{k \in K_{\varepsilon_0}(x^*)} \gamma_c^k(x) \tag{30}$$

and

$$h_c(x, \lambda) \in H \tag{31}$$

for all x in $B_{\varepsilon_1}(x^*)$ and all λ in $[0, 1]$. Hence

$$\begin{aligned}\hat{\gamma}_c(x, h_c(x, \lambda)) &\leq \hat{\gamma}_c^{k^*}(x, \lambda h_c^{k^*}(x)) \\ &\leq \gamma_c^{k^*}(x) + \lambda\theta_c^{k^*}(x) \end{aligned} \tag{32}$$

for all λ in $[0, 1]$ and all x in $B_{\varepsilon_1}(x^*)$. Let $L \triangleq \max\{\|h\| \,|\, h \in H\}$.

From Proposition 2 (with $\eta = -\theta_c^{k^*}(x^*)/8L$) and the continuity of $\theta_c^{k^*}$, there exists an ε_2 in $(0, \varepsilon_1/2]$ such that

$$|\gamma_c^k(x + \lambda h) - \hat{\gamma}_c^k(x, \lambda h)| \leq -(\lambda/8)\theta_c^{k^*}(x^*) \tag{33}$$

$$\theta_c^{k^*}(x) \in [(3/2)\theta_c^{k^*}(x^*), (3/4)\theta_c^{k^*}(x^*)]$$

$$x + \lambda h \in B_{\varepsilon_1}(x^*)$$

for all x in $B_{\varepsilon_2}(x^*)$, all h in H, all k in K and all λ in $[0, \lambda_1]$ where λ_1 is the largest λ in S satisfying $\|\lambda_1 h\| \leq \varepsilon_2$ for all $h \in H$. It follows (see Appendix) from (29), (30) and (33) that

$$|\gamma_c(x + \lambda h) - \hat{\gamma}_c(x, \lambda h)| \leq -(\lambda/8)\theta_c^{k^*}(x^*)$$

for all x in $B_{\varepsilon_2}(x^*)$, all h in H and all λ in $[0, \lambda_1]$.

Hence

$$\begin{aligned} \gamma_c(x + h_c(x,\lambda)) &\leq \hat{\gamma}_c(x, h_c(x,\lambda)) - (\lambda/8)\theta_c^{k^*}(x^*) \\ &\leq \hat{\gamma}_c(x, h_c(x, \lambda)) - \lambda(3/16)\theta_c^{k^*}(x) \end{aligned} \tag{34}$$

for all x in $B_{\varepsilon_2}(x^*)$ and all λ in $[0, \lambda_1]$. From (32)

$$\begin{aligned} -\lambda\theta_c^{k^*}(x) &\leq |\hat{\gamma}_c(x, h_c(x, \lambda) - \gamma_c^{k^*}(x)| \\ &\leq |\hat{\gamma}_c(x, h_c(x, \lambda) - \gamma_c(x)| \\ &\quad + |\gamma_c(x) - \gamma_c^{k^*}(x)| \end{aligned} \tag{35}$$

for all x in $B_{\varepsilon_1}(x^*)$, all λ in $[0, \lambda_1]$. Since $\gamma_c^{k^*}(x^*) = \gamma_c(x^*)$, we can choose an ε_3 in $(0, \varepsilon_2]$ such that

$$|\gamma_c(x) - \gamma_c^{k*}(x)| \leq -(3\lambda_1/32)\theta_c^{k*}(x^*)$$

$$\leq -(\lambda_1/8)\theta_c^{k*}(x)$$

so that, from (35),

$$-\lambda_1\theta_c^{k*}(x) \leq (8/7)\,|\hat{\gamma}_c(x,\ h_c(x,\ \lambda_1)) - \gamma_c(x)| \tag{36}$$

for all x in $B_{\varepsilon_3}(x^*)$. From (34) and (36)

$$\gamma_c(x + h_c(x,\ \lambda_1)) - \gamma_c(x) \leq (11/14)[\,\hat{\gamma}_c(x,h_c(x,\ \lambda_1) - \gamma_c(x)]$$

for all x in $B_{\varepsilon_3}(x^*)$. Hence the Armijo step length $\lambda_c(x)$ is not less than λ_1 for all x in $B_{\varepsilon_3}(x^*)$.

Using the above bounds:

$$\gamma_c(x + h_c(x,\lambda_1)) - \gamma_c(x) \leq \hat{\gamma}_c^{k*}(x,\lambda_1 h_c^{k*}(x)) - \gamma_c^{k*}(x)$$

$$-(\lambda_1/3)\theta_c^{k*}(x^*) + |\gamma_c(x) - \gamma_c^{k*}(x)|$$

$$\leq (7\lambda_1/8 - \lambda_1/12)\,\theta_c^{k*}(x^*)$$

for all x in $B_{\varepsilon_3}(x^*)$. Since the Armijo step length is not less than λ_1 it follows that

$$\gamma_c(A_c(x)) - \gamma_c(x) \leq (19/24)\lambda_1\theta_c^{k*}(x^*)$$

for all x in $B_{\varepsilon_3}(x^*)$, where

$$A_c(x) \triangleq x + h_c(x,\ \lambda_c(x))$$

is the successor point to x generated by the algorithm. The desired result follows from a standard algorithm model[4]. □

4. CHOICE OF THE PENALTY PARAMETER

We have presented in Section 3 two algorithms for solving P_c. An algorithm for solving P requires the addition of a rule for adaptively adjusting c. Our rule will be based on the algorithm model in [4] and requires a test function $x \mapsto t_c(x)$. If the test $t_c(x) \leq 0$ is satisfied the penalty parameter c is left unaltered; otherwise c is increased. Before proposing the test function we establish a few properties of P and P_c. The first result is obvious.

Proposition 3

Suppose $\hat{x}$ is a local minimizer for P_c and is feasible for P. Then $\hat{x}$ is a local minimizer for P. □

Before proceeding it is helpful to define a few terms. For all k in K let

$$J_0^k(x) \triangleq \{j \in J_k | \bar{\eta}^{k,j}(x) = \eta^k(x)_+\}$$

denote the active constraint set for the constraint $\eta^k(x) \leq 0$; note that $J_0^k(x)$ is empty if $\eta^k(x) < 0$. For all k in K let $\partial\eta^k$ and $\partial\eta^k(\cdot)_+$ denote the Clarke generalised gradient [6] of, respectively, η^k and $\eta^k(\cdot)_+$. These satisfy:

$$\partial\eta^k(x) = \mathrm{co}\{\nabla\bar{\eta}^{k,j}(x),\ j \in J_0^k(x)\}$$

and

$$\partial\eta^k(x)_+ = \mathrm{co}\{0;\ \nabla\bar{\eta}^{k,j}(x),\ j \in J_0^k(x)\} \text{ if } \eta^k(x) = 0$$
$$= 0 \quad \text{if } \eta^k(x) < 0.$$

We can express the condition of optimality $(\theta_c^k(x) = 0)$ for problem $P_c^k : \min \gamma_c^k$ in terms of generalised gradients.

Proposition 4

Suppose x is such that $\eta^k(x) \leq 0$ and $c > 0$. Then

(i) $\theta_c^k(x) = 0$

if and only if

(ii) $0 \in \nabla f(x) + c\partial\eta^k(x)_+$.

Proof

Suppose $\eta^k(x) = 0$. If follows from the definition of θ_c^k that $\theta_c^k(x) = 0$ if and only if

$$<\nabla f(x),h> + c \max\{0; <\nabla\bar{\eta}^{k,j}(x),h>, j \in J_0^k(x)\} \geq 0$$

for all h in $\mathbb{R}^n$. This, in turn, is easily seen to be equivalent to

$$<\nabla f(x), h> + c \max\{<\xi, h>|\xi \in \partial\eta^k(x)_+\} \geq 0$$

or

$$\max\{<\xi, h>|\xi \in \nabla f(x) + c\partial\eta^k(x)_+\} \geq 0 \tag{38}$$

for all h in $\mathbb{R}^n$.

Suppose then that $\theta_c^k(x) = 0$ but that 0 does not lie in the convex set $\nabla f(x) + c\partial\eta^k(x)_+$; let g denote the closest point in this set to the origin. Then $h = -g$ satisfies $<h, \xi> < 0$ for all ξ in $\nabla f(x) + c\partial\eta^k(x)_+$, contradicting (38).

If, on the other hand, 0 lies in the convex set $\nabla f(x) + c\partial\eta^k(x)_+$ then (38) is true so that $\theta_c^k(x) = 0$.

The case when $\eta^k(x) < 0$ is trivial; both $\theta_c^k(x) = 0$ and $0 \in \nabla f(x) + c\partial\eta^k(x)_+$ if and only if $\nabla f(x) = 0$. □

Proposition 5

(a) Suppose that x is feasible for P and for some $c > 0$ satisfies $\theta_c(x) = 0$. Then:

$$\left.\begin{array}{ll} \text{(i)} & \text{if } \psi(x) = 0,\ 0 \in \operatorname{co}\{\nabla f(x),\ \partial\eta^k(x)\} \text{ for all } k \text{ in } K_0(x) \\ \text{(ii)} & \text{if } \psi(x) < 0,\ \ \nabla f(x) = 0. \end{array}\right\} \quad (39)$$

(b) If $\hat{x}$ is a local minimizer for P, then $\hat{x}$ satisfies (39).

Proof

(a) Since $\theta_c(\hat{x}) = 0$ it follows that $\theta_c^k(\hat{x}) = 0$ for all k in $K_0(\hat{x})$. From Proposition 4, $0 \in \nabla f(\hat{x}) + c\partial\eta^k(\hat{x})_+$ for all k in $K_0(\hat{x})$. Hence there exist non negative multipliers μ^j such that $\nabla f(\hat{x}) + \Sigma c\mu^j\nabla\bar{\eta}^{k,j}(x) = 0$, the summation being over the set $J_0^k(\hat{x})$; dividing by $1 + c\Sigma\mu^j$ yields (if $\psi(x) = 0$)

$$0 \in \operatorname{co}\{\nabla f(\hat{x}),\ \partial\eta^k(\hat{x})\} \text{ for all } k \text{ in } K_0(\hat{x})$$

Clearly $J_0^k(\hat{x})$ is empty if $\psi(x) < 0$ so that in this case $\nabla f(\hat{x}) = 0$.

(b) If $\hat{x}$ is a local minimizer of P, it is also a local minimizer for P^k for all k in $K_0(\hat{x})$. The result follows from [6]. □

We introduce next a constraint qualification assumed to hold in the sequel.

Constraint Qualification

For all x such that $\psi(x) \geq 0$, for all k in $K_0(x)$

$$0 \notin \partial\eta^k(x) = \operatorname{co}\{\nabla\bar{\eta}^{k,j}(x),\ j \in J_0^k(x)\}.$$ □

The constraint qualification ensures that $\psi(x)$ can be reduced at all non-feasible x and thus ensures the existence of finite penalties.

Proposition 6

Suppose $\hat{x}$ is a local minimiser for P so that $\hat{x}$ is feasible for P and satisfies (39). Then there exists a $\hat{c} > 0$ such that $\theta_c(\hat{x}) = 0$ for all $c \geq \hat{c}$.

Proof

Suppose $\psi(\hat{x}) = 0$. It follows from (39) and the constraint qualification that there exist non-negative multipliers $\mu^{k,j}$ satisfying

$$\nabla f(\hat{x}) + \sum_{j \in J_0^k(\hat{x})} \mu^{k,j} \nabla \bar{\eta}^{k,j}(\hat{x}) = \nabla f(\hat{x}) + \sum_{j \in J_0^k(\hat{x})} (\mu^{k,j}/c) c \nabla \bar{\eta}^{k,j}(\hat{x})$$

$$= 0$$

for all k in $K_0(\hat{x})$, $c > 0$. There exists a $\hat{c} > 0$ such that

$$\sum_{j \in J_0^k(\hat{x})} (\lambda^{k,j}/c) \leq 1$$

so that:

$$\sum_{j \in J_0^k(\hat{x})} (\lambda^{k,j}/c) c \nabla \bar{\eta}^{k,j}(\hat{x}) \in c \partial \eta^k(\hat{x})_+$$

for all $c \geq \hat{c}$, all k in $K_0(\hat{x})$.

Hence:

$$0 \in \{\nabla f(\hat{x}) + c \partial \eta^k(\hat{x})_+\}$$

for all k in $K_0(\hat{x})$, all $c \geq \hat{c}$. It follows from Proposition 4 that, for all $c \geq \hat{c}$, $\theta_c^k(\hat{x}) = 0$ if $k \in K_0(\hat{x})$. Hence $\theta_c(\hat{x}) = 0$ for all $c \geq \hat{c}$.

The case $\psi(\hat{x}) < 0$ is trivial. □

We can now define the test function; for all $c > 0$, $t_c : \mathbb{R}^n \to \mathbb{R}$ is defined by:

$$t_c(x) \triangleq \theta_c(x) + \psi(x)_+/c \; . \tag{40}$$

The properties required by the test function are specified in our next result.

<u>Proposition 7</u>

(i) If for some $c > 0$ $\theta_c(x) = 0$ and $t_c(x) \leq 0$, then x satisfies (39) and $\psi(x) \leq 0$ (i.e. x is a stationary point for P).

(ii) For all $c > 0$, if $x_i \to x^*$ as $i \to \infty$ and $t_c(x_i) \leq 0$ for all i, then $t_c(x^*) \leq 0$.

(iii) For all $\hat{x}$ there exist positive $\hat{c}$ and $\hat{\varepsilon}$ such that $t_c(x) \leq 0$ for all $c \geq \hat{c}$ and all x in $B_{\hat{\varepsilon}}(\hat{x})$. □

The proof of this result is deferred to the appendix.

The algorithm for solving P can be presented. In the algorithm $\{c_j\}$ is an infinite sequence of positive numbers tending monotonically to infinity.

Algorithm 3

Data: $\varepsilon_0 > 0$; $\alpha, \beta \in (0,1)$; $\{c_j\}$, an infinite sequence such that $c_0 > 0$ and $c_j \nearrow \infty$ as $j \to \infty$; $x_0 \in \mathbb{R}^n$.

Step 0: Set $i = 0$.

Step 1: If $t_{c_j}(x_i) > 0$, set $z_j = x_i$ and choose j^* so that c_{j^*} is the lowest c in $\{c_j\}$ satisfying $t_{c_j^*}(x_i) \leq 0$. Set $j = j^*$.

Step 2: If $\theta_{c_j}(x_i) = 0$ stop. Else compute x_{i+1} using Steps 1-3 of Algorithm 1 or 2. Set $i=i+1$ and go to Step 1. □

Theorem 3

(i) If the sequence $\{z_j\}$ is finite then either the sequence $\{x_i\}$ is finite with its last element x_k satisfying $\psi(x_k) = 0$ and (39) or it is infinite and any accumulation point $\hat{x}$ satisfies $\phi(\hat{x}) = 0$ and (39).

(ii) If $\{z_j\}$ is infinite then it has no accumulation points.

Proof

This result follows from Theorem 2, Proposition 7 and Theorem 4 of [4], if we note that hypothesis (ii) in this result (the continuity of t_c for all $c > 0$) is merely used to establish hypothesis (ii) in Proposition 7 above.

Corollary

If the sequence $\{x_i\}$ is bounded, then $\{z_j\}$ is finite. □

5. CONCLUSION

Exclusion constraints have the unusual feature that the set of feasible search directions is not necessarily convex. Thus algorithms based on the use of generalised gradients may jam at points which are not local minima of the constrained objective function. Two algorithms which avoid this problem are presented. These algorithms effectively explore a set of possible search directions (rather than one). As pointed out in [1] the cardinality of K is very high if the cardinality of I is high. Hence it is desirable that the number of exclusion regions is not too high.

6. REFERENCES

[1] A Tits, "Comments on the Optimal Design, Centering, Tolerancing and Tuning Problem", JOTA,

[2] D Q Mayne, E Polak, A Voreadis, "A Cut Map Algorithm for Design Problems with Parameter Tolerances", IEEE Trans. Circuits and Systems, Vol CAS-29, pp 35-45, 1982.

[3] E Polak, "Computational Methods in Optimization", Academic Press, New York, 1971.

[4] E Polak, "On the Global Stabilization of Locally Convergent Algorithms", Automatica, 12, pp 337-342, 1976.

[5] E Polak, D Q Mayne and Y Wardi, "On the Extension of Constrained Optimization Algorithms from Differentiable to Nondifferentiable Problems", SIAM J Control & Opt., Vol 21, No 2, pp 179-203, 1983.

[6] F H Clarke, "Optimization and Nonsmooth Analysis", Wiley, New York, 1983.

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

APPENDIX

Proof of Proposition 7

(i) Suppose that $c > 0$, $\theta_c(x) = 0$ and $t_c(x) \leq 0$. Hence from (40), if follows that $\psi(x)_+ = 0$ (so that $\psi(x) \leq 0$). Since $\theta_x(x) = 0$ it follows from part (a) of Proposition 5 that (39) holds.

(ii) Suppose that $c > 0$, $x_i \to x^*$ as $i \to \infty$ and $t_c(x_i) \leq 0$ for all i. From (40)

$$\psi(x_i)_+/c \leq -\theta_c(x_i) \tag{A1}$$

for all i. There exists a neighbourhood N of x* in which $K_0(x) \subset K_0(x^*)$. Let $\pi\colon \mathbb{R}^n \to \mathbb{R}$ be defined by:

$$\pi(x) \triangleq \min\{\theta_c^k(x) \,|\, k \in K_0(x^*)\}. \tag{A2}$$

Clearly π is continuous, $\pi(x) \leq \theta_c(x)$ for all x in N and $\pi(x^*) = \theta_c(x^*)$. Hence, for all i sufficiently large, it follows from (A1) that

$$\psi(x_i)_+/c \leq -\theta_c(x_i) \leq -\pi(x_i),$$

so that

$$\psi(x^*)_+/c \leq -\pi(x^*) = -\theta_c(x^*).$$

Hence

$$t_c(x^*) \leq 0.$$

(iii)

(α) Suppose $\psi(\hat{x}) < 0$. There exists an $\hat{\varepsilon} > 0$ such that $\psi(x)_+ = 0$ for all x in $B_{\hat{\varepsilon}}(\hat{x})$. Hence $t_c(x) \leq 0$ for all $c > 0$, and all x in $B_{\hat{\varepsilon}}(\hat{x})$.

(β) Suppose $\psi(\hat{x}) \geq 0$. By virtue of the constraint qualification $0 \notin \mathrm{co}\{\nabla\bar{\eta}^{k,j}(\hat{x}),\ j \in J_0^k(\hat{x})\}$ for all k in $K_0(\hat{x})$. Also there exists a

$\varepsilon_1 > 0$ such that $K_0(x) \subset K_0(\hat{x})$ for all x in $B_{\varepsilon_1}(\hat{x})$. Since $\eta^k(x) = \max\{\bar{\eta}^{k,j}(x) \mid j \in J_k\}$ there exists an $\varepsilon_2 \in (0, \varepsilon_1]$ such that

$$\min\{\bar{\eta}^{k,j}(x) + \langle\nabla\bar{\eta}^{k,j}(x), h\rangle \mid j \in J_0^k(\hat{x})\}$$

$$\geq \max\{\bar{\eta}^{k,j}(x) + \langle\nabla\bar{\eta}^{k,j}(x), h\rangle \mid j \in J_k \setminus J_0^k(\hat{x})\} \quad \text{(A3)}$$

for all x in $B_{\varepsilon_2}(\hat{x})$, all h in $H \triangleq B_{\varepsilon_2}(0)$ and all k in $K_0(\hat{x})$.

For all k, let $\pi^k: \mathbb{R}^n \to \mathbb{R}$ be defined by:

$$\pi^k(x) \triangleq \min_{h\in H} \max_{j\in J_0^k(\hat{x})} \{\langle\nabla\bar{\eta}^{k,j}(x), h\rangle\} \quad \text{(A4)}$$

Clearly π^k is continuous and, by virtue of the constraint qualification, satisfies

$$\pi^k(\hat{x}) \leq -2\hat{\delta}$$

for some $\hat{\delta} > 0$ and all k in $K_0(\hat{x})$. Hence, there exists an $\varepsilon_3 \in (0, \varepsilon_2]$ such that

$$\pi^k(x) \leq -\hat{\delta} \quad \text{(A5)}$$

for all x in $B_{\varepsilon_3}(\hat{x})$ and all k in $K_0(\hat{x})$.

From the definition of θ_c^k:

$$\theta_c^k(x) \leq \min_{h\in H}\{(1/2)\|h\|^2 + \hat{\gamma}_c^k(x,h) - (1/2)\|h_c^k(x)\|^2 - \gamma_c^k(x)\}.$$

Using the definition of $\hat{\gamma}_c^k$ we obtain

$$\theta_c^k(x) \leq \min_{h\in H} \max\{(1/2)\|h\|^2 + c\bar{\eta}^{k,j}(x) - c\eta^k(x)_+ + \langle\nabla f(x) + c\nabla\bar{\eta}^{k,j}(x), h\rangle,$$

$$j \in J_k;$$

$$(1/2)\|h\|^2 - c\eta^k(x)_+ + \langle\nabla f(x), h\rangle\} - (1/2)\|h_c^k(x)\|^2 \quad \text{(A6)}$$

The second term in the braces accounts for the case when $\hat{\eta}^k(x,h)_+ = \max\{\hat{\eta}^k(x,h), 0\} = 0$. Because of inequality (A3), J_k in (A6) can be replaced by $J_0^k(\hat{x})$ if $x \in B_{\varepsilon_3}(\hat{x})$. Let $h^k(x)$ denote the minimizing h in (A4) so that (by (A4) and (A5)):

$$\langle \nabla \bar{\eta}^{k,j}(x), h^k(x) \rangle \leq -\delta \tag{A7}$$

for all $x \in B_{\varepsilon_3}(\hat{x})$, all k in $J_0^k(\hat{x})$.

Let b denote an upper bound for $\langle \nabla f(x), h \rangle$ as x ranges over $B_{\varepsilon_3}(\hat{x})$ and h ranges over H. Substituting $\alpha h^k(x)$ for h and $J_0^k(\hat{x})$ for J_k in (A6) yields:

$$\theta_c^k(x) \leq \min_{\alpha \in [0,1]} \max\{(1/2)\varepsilon_2^2\alpha^2 + b\alpha - c\alpha\delta; \\ (1/2)\varepsilon_2^2\alpha^2 + b\alpha - c\eta^k(x)_+\} \tag{A8}$$

for all x in $B_{\varepsilon_3}(\hat{x})$ and all k in $K_0(\hat{x})$. Since $\theta_c(x) = \min\{\theta_c^k(x) \mid k \in K_0(x)\}$, since $K_0(x) \subset K_0(\hat{x})$ for all x in $B_{\varepsilon_3}(\hat{x})$ and since $\eta^k(x) = \psi(x)$ of $k \in K_0(x)$ it follows

$$\theta_c(x) \leq \min_{\alpha \in [0,1]} \max\{(1/2)\varepsilon_2^2\alpha^2 + b\alpha - c\alpha\delta; (1/2)\varepsilon_2^2\alpha^2 + b\alpha - c\psi(x)_+\} \tag{A9}$$

for all x in $B_{\varepsilon_3}(\hat{x})$. Hence for all x in $B_{\varepsilon_3}(\hat{x})$:

$$t_c(x) \leq \min_{\alpha \in [0,1]} \max\{(1/2)\varepsilon_2^2\alpha^2 - (c\delta - b)\alpha + \psi(x)_+/c; \\ (1/2)\varepsilon_2^2\alpha^2 + b\alpha + \psi(x)_+/c - c\psi(x)_+\} \tag{A10}$$

Since $t_c(x) = \theta_c(x) \leq 0$ if $\psi(x) \leq 0$ we confine attention to those x such that $\psi(x) = \psi(x)_+ > 0$. The first term in (A10) is negative if

$$\alpha(c\delta - b - \alpha\varepsilon_2^2/2) \leq \psi(x)_+/c$$

i.e., if

$$c\delta - b - \alpha\varepsilon_2^2/2 > 0 \tag{A11}$$

and

$$\alpha \geq \psi(x)_+/[c(c\delta - b - \alpha\varepsilon_2^2/2)] \tag{A12}$$

Since $\alpha \in [0, 1]$, inequalities (A11) and (A12) hold if

$$c > (b + \varepsilon_2^2/2)/\delta \tag{A13}$$

and

$$\alpha \geq \ell(c)\psi(x)_+ \tag{A14}$$

where

$$\ell(c) \triangleq 1/[c(c\delta - b - \varepsilon_2^2/2)] \tag{A15}$$

The second term in (A10) is negative if

$$\alpha[b + \alpha\varepsilon_2^2/2] \leq \psi(x)_+[c - 1/c] \tag{A16}$$

and

$$c > 1 \tag{A17}$$

Clearly (A16) holds if

$$\alpha \leq \psi(x)_+[c - 1/c]/[b + \alpha\varepsilon_2^2/2]$$

which, in turn, is implied by

$$\alpha \leq \psi(x)_+ u(c) \tag{A18}$$

where

$$u(c) \triangleq [c - 1/c]/[b + \varepsilon_2^2/2] \tag{A19}$$

(since $\alpha \in [0, 1]$). Clearly $\hat{c}$ can be chosen so that $c > (b + \varepsilon_2^2/2)/\delta$, $c > 1$ and $\ell(c) < u(c)$ for all $c \geq \hat{c}$ yielding $t_c(x) \leq 0$ for all x in $B_{\varepsilon_3}(\hat{x})$ and all $c \geq \hat{c}$ which is the desired result. □

Proposition A1

(i) Suppose x is a local minimizer for P_c^k with $c > 0$. Then $\theta_c^k(x) = 0$.

(ii) The functions θ_c^k, h_c^k are continuous.

Proof

(i) Suppose $\theta_c^k(x) < 0$. Hence there exists a h in H such that $\hat{\gamma}_c^k(x, h) - \gamma_c(x) = -d < 0$. If follows from Proposition 2 and the differentiability of f that there exists a $\delta > 0$ such that: $|\gamma_c^k(x + \alpha h) - \hat{\gamma}_c^k(x, \alpha h)| \leq \alpha d/2$ for all α in $[0, \delta]$. Because of convexity $\hat{\gamma}_c^k(x, \alpha h) - \gamma_c(x) \leq -\alpha d$ for all α in $[0, 1]$. Hence $\gamma_c^k(x + \alpha h) - \gamma_c^k \leq -\alpha d/2$ for all α in $[0, \delta]$, contradicting local optimality.

(ii) We first prove that h_c^k is continuous. Let $\xi: \mathbb{R}^n \to \mathbb{R}$ be defined by

$$\xi(x) \triangleq \min\{(1/2)\|h\|^2 + \hat{\gamma}_c^k(x,h) \mid h \in \mathbb{R}^n\}.$$

Suppose that $x_i \to \hat{x}$ as $i \to \infty$ and let

$$h_i \triangleq \arg\min\{(1/2)\|h\|^2 + \hat{\gamma}_c^k(x_i,h) \mid h \in \mathbb{R}^n\}$$

for all i (i.e. $h_i \triangleq h_c^k(x_i)$). Clearly

$$\xi(x_i) \leq \hat{\gamma}_c^k(x_i, 0) = \gamma_c^k(x_i)$$

for all i and

$$(1/2)\|h\|^2 + \hat{\gamma}_c^k(x, h) \to \infty \text{ as } \|h\| \to \infty$$

uniformly for all x in a compact set. Hence the infinite sequence $\{\|h_i\|\}$ is bounded.

Now

$$\overline{\lim}\, \xi(x_i) \leq \overline{\lim}\, \{(1/2)\|\hat{h}\|^2 + \hat{\gamma}_c^k(x_i, \hat{h})\}$$

$$= \xi(\hat{x})$$

where $\hat{h} \triangleq h_c^k(\hat{x})$, so that ξ is u.s.c.

Suppose that $\underline{\lim}\, \xi(x_i) < \xi(\hat{x})$. Since:

$$\underline{\lim}\, \xi(x_i) = \underline{\lim}\, \{(1/2)\|h_i\|^2 + \hat{\gamma}_c^k(x_i, h_i)\}$$

$$= \underline{\lim}\, \{(1/2)\|h_i\|^2 + \hat{\gamma}_c^k(\hat{x}, h_i)\}$$

it follows that

$$\underline{\lim}\, \{(1/2)\|h_i\|^2 + \hat{\gamma}_c^k(\hat{x}, h_i)\} < \xi(\hat{x})$$

But this contradicts the optimality of $\hat{h}$. Hence $\underline{\lim}\, \xi(x_i) \geq \xi(\hat{x})$, i.e.

$$\overline{\lim}\, \xi(x_i) \leq \xi(\hat{x}) \leq \underline{\lim}\, \xi(x_i)$$

and therefore $\lim \xi(x_i) = \xi(\hat{x})$, i.e. ξ is continuous.

Since $\{h_i\}$ is bounded, $h_i \xrightarrow{K} h^*$ along some subsequence K. But this implies that h^* is a minimizer for the problem $\min\{(1/2)\|h\|^2 + \hat{\gamma}_c(\hat{x},h)\}$. But the minimizer is unique, so $h^* = \hat{h}$. Since $\{h_i\}$ is bounded and has a unique accumulation point it follows that $h_i \to \hat{h}$, i.e. h_c^k is continuous. Since h_c^k is continuous it follows easily that θ_c^k is continuous. □

Proposition A2

$$|\max\{A,B\} - \max\{C,D\}| \leq \max\{|A-C|,|B-D|\}.$$

Proof

Let $\alpha \triangleq \max\{A,B\}$, $\beta \triangleq \max\{C,D\}$,

$$\phi \triangleq |\alpha - \beta|$$

$$= \max\{\alpha - \beta, \beta - \alpha\}$$

$$\alpha - \beta = \max\{A - \beta, B - \beta\} \leq \max\{A-C, B-D\}$$

$$\beta - \alpha = \max\{C - \alpha, D - \alpha\} \leq \max\{C-A, D-B\}$$

Hence

$$\phi \leq \max\{A-C, B-D, C-A, D-B\}$$

$$= \max\{|A-C|, |B-D|\}. \qquad \square$$

Proposition A3

If $|a_i - b_i| \leq \delta$ for all i in I

then

$$|\min_{i\in I} a_i - \min_{i\in I} b_i| \leq \delta$$

Proof

Let $\alpha \triangleq \min_{i\in I} a_i$,

$$\beta \triangleq \min_{i\in I} b_i$$

and suppose $a_j = \alpha$, $b_k = \beta$,

Then:

$$\alpha - \beta = \alpha - b_k \leq a_k - b_k \leq \delta$$

and

$$\beta - \alpha = \beta - a_j \leq b_j - a_j \leq \delta$$

Hence $|\alpha - \beta| \leq \delta$. □

Finally, since the result is well known in the literature on digital logic, we illustrate the assertion that

$$\max_{j\in I} \min_{k\in I_j} \{\phi^{j,k}(x)\} \leq 0 \text{ can be replaced by } \min_{k\in K} \max_{j\in J_k} \{\bar{\eta}^{k,j}(x)\} \leq 0.$$

Consider the simple case

$$\max_{j\in\{1,2\}} \min_{k\in\{1,2\}} \{\phi^{j,k}(x)\} \leq 0.$$

This is clearly equivalent to the following

$$[\phi^{1,1}(x) \leq 0 \ \underline{\text{or}} \ \phi^{1,2}(x) \leq 0]$$

$$\underline{\text{and}} \quad [\phi^{2,1}(x) \leq 0 \ \underline{\text{or}} \ \phi^{2,2}(x) \leq 0]$$

This in turn is equivalent to:

$$[\phi^{1,1}(x) \leq 0 \text{ and } \phi^{2,1}(x) \leq 0]$$

$$\underline{\text{or}} \quad [\phi^{1,1}(x) \leq 0 \text{ and } \phi^{2,2}(x) \leq 0]$$

$$\underline{\text{or}} \quad [\phi^{1,2}(x) \leq 0 \text{ and } \phi^{2,2}(x) \leq 0]$$

or $[\phi^{1,2}(x) \leq 0$ and $\phi^{2,2}(x) \leq 0]$.

which is equivalent to

$$\min_{k\in\{1,2,3,4\}} \max_{j\in\{1,2\}} \{\bar{\eta}^{k,j}(x)\} \leq 0$$

where $\bar{\eta}^{1,1} = \phi^{1,1}$, $\bar{\eta}^{1,2} = \phi^{2,1}$, $\bar{\eta}^{2,1} = \phi^{1,1}$, $\bar{\eta}^{2,2} = \phi^{2,2}$, $\bar{\eta}^{3,1} = \phi^{1,2}$, $\bar{\eta}^{3,2} = \phi^{2,2}$, $\bar{\eta}^{4,1} = \phi^{1,1}$, $\bar{\eta}^{4,2} = \phi^{2,2}$.

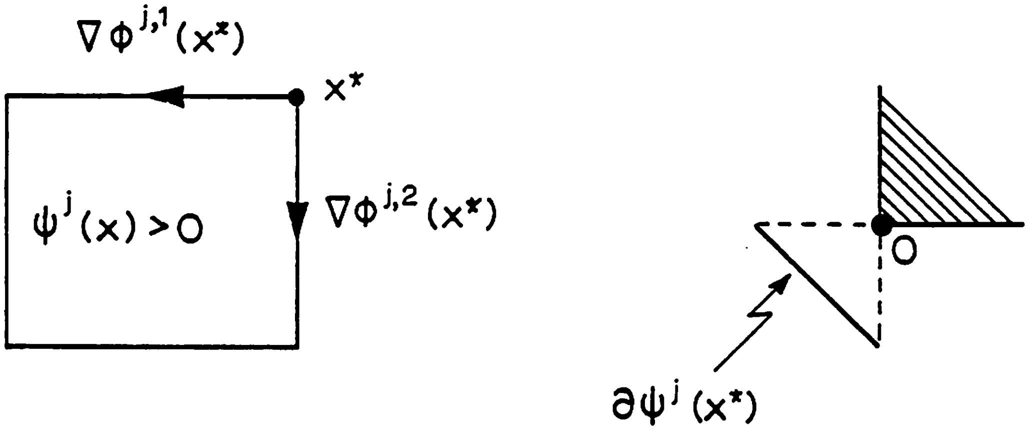
The result is easily generalised.

Captions for Figures

Fig 1: (a) Exclusion Constraint

(b) Generalized Gradient

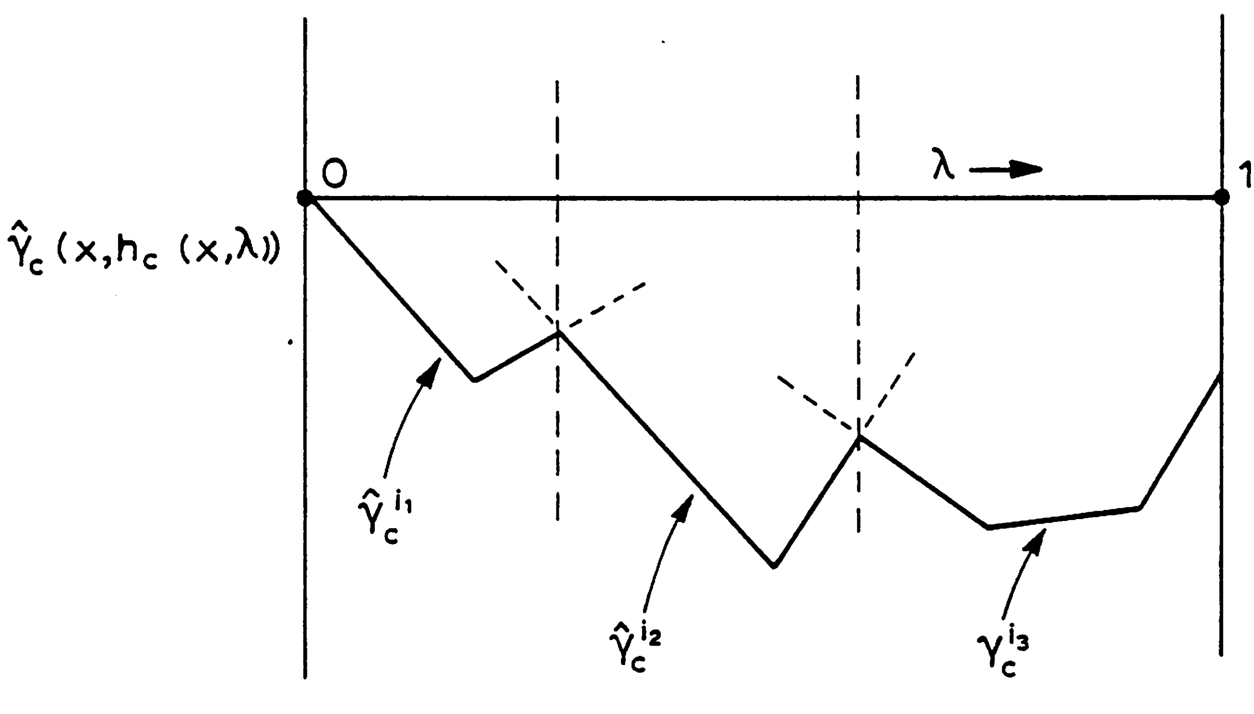
Fig 2: Step Length Determination



$\nabla\phi^{j,1}(x^*)$
x^*
$\psi^j(x) > 0$
$\nabla\phi^{j,2}(x^*)$
0
$\partial\psi^j(x^*)$

(a) (b)

Fig. 1



0
$\lambda \rightarrow$
1
$\hat{\gamma}_c(x, h_c(x,\lambda))$
$\hat{\gamma}_c^{i_1}$
$\hat{\gamma}_c^{i_2}$
$\gamma_c^{i_3}$

Fig. 2