EXACT PENALTY FUNCTIONS AND LAGRANGE MULTIPLIERS

by

J. B. Lasserre

Memorandum No. UCB/ERL M79/31

12 July 1979

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

EXACT PENALTY FUNCTIONS AND LAGRANGE MULTIPLIERS

J.B. Lasserre

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

ABSTRACT

We give necessary and sufficient conditions for a penalty function to be exact. This is an extension to the general case of the result given by Bertsekas for the convex case. An algorithm to minimize the exact penalty function is given. It is based on the same idea as the one used by Demjanov for minimax problems.

Key Words: Nonlinear programming, penalty functions, exact penalty functions, Lagrange multipliers.

Research Sponsored by the National Science Foundation Grant INT78-09263.

I. Introduction

In this paper we are concerned with exact penalty functions Necessary and sufficient conditions for a penalty function to be exact are given. This is a generalization of the result of Bertsekas for the convex programming problem.

In section II we give the result for the well known exact penalty function used by Pietrzykowski, Conn and Han (see [1], [2], [3]). Extension to a class of penalty functions is made in section III.

In section IV an algorithm to minimize the exact penalty function of section II is presented. It is based on the principle used by Demyanov (see [5]) for minimax problems. The direction of descent is calculated by minimizing the directional derivative (we only need to solve a linear programming problem). An example is given (the Rosen Suzuki problem).

II. Exact Penalty Function and Lagrange Multipliers

Let us consider the problem

$$(P) \begin{cases} \min f(x) \\ h_i(x) = 0 \ , \quad i = 1,\dots,m \\ h_i(x) \leq 0 \ , \quad i = m+1,\dots,p \end{cases}$$

$$f: R^n \longrightarrow R \quad \text{twice differentiable}$$

$$h_i: R^n \longrightarrow R \quad \text{twice differentiable}$$

II.1. Notations and Definitions

For $x \in R^n$ and $c \in R^p_+$ let us define the function

$$p(x,c) = f(x) + \sum_{i=1}^{m} c_i |h_i(x)| + \sum_{i=m+1}^{p} c_i (h_i(x))_+$$

where

$$(h_i(x))_+ = \begin{cases} h_i(x) & \text{if } h_i(x) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$p(x,c)$ is continuous but not differentiable.

x^* is a local minimum of $f(x)$ for the problem (P).

We note $H_i(x^*)$ the Hessian matrix of the function $h_i(x)$, evaluated at x^*.

$I(x) = \{i \mid i \geq m+1,\ h_i(x) = 0\}$

$M = \{u \in R^n \mid \langle \nabla h_i(x^*), u\rangle = 0 \text{ if } i \in [i,\dots,m] \cup I(x^*)\}$

II.2. Theorem 1

If x is a point which satisfies the Kuhn Tucker first order conditions ($\lambda_1,\dots,\lambda_p$ are the Lagrange multipliers) and if $c_i \geq |\lambda_i|$ $\forall i$ then in any direction the directional derivative of $p(x,c)$ at x is positive.

Proof. Let $Dp(x,u,c)$ be the directional derivative in the direction u. We have:

$$Dp(x,u,c) = \langle \nabla f(x), u\rangle + \sum_{i=1}^{m} c_i |\langle \nabla h_i(x), u\rangle| + \sum_{i \in I(x)} c_i (\langle \nabla h_i(x), u\rangle)_+ .$$

Then for $i = 1,\dots,m$ we have

$$\lambda_i \langle \nabla h_i(x), u\rangle \leq |\lambda_i| |\langle \nabla h_i(x), u\rangle| \leq c_i |\langle \nabla h_i(x), u\rangle| .$$

For $i \in I(x)$ we have $\lambda_i \geq 0$ and then

$$\lambda_i \langle \nabla h_i(x), u\rangle \leq \lambda_i (\langle \nabla h_i(x), u\rangle)_+ \leq c_i (\langle \nabla h_i(x), u\rangle)_+$$

and for $i \geq m+1$ and $i \notin I(x)$, $\lambda_i = 0$ then

$$Dp(x,u,c) \geq \langle \nabla f(x) + \sum_{i=1}^{p} \lambda_i \nabla h_i(x), u \rangle = 0$$

since

$$\nabla f(x) + \sum_{i=1}^{p} \lambda_i \nabla h_i(x) = 0 \quad \text{(first order condition)}$$

and the proof is complete.

x^* is now a local minimum of $f(x)$ for the problem (P).

Assume

(i) at x^* the gradients of the constraints equal to zero are linearly independent. (Note $\lambda_1^*, \ldots, \lambda_p^*$ the Lagrange multipliers which exist.)

(ii) if all the constraints are not linear we assume that $\forall u \in M$, $\exists i \leq m$ or $i \in I(x^*)$ such that $\langle u, H_i(x^*)u \rangle \neq 0$.

II.3. Theorem 2 (Sufficient condition)

x^* is local minimum for the problem (P) and (i) and (ii) hold. Then if $c_i > |\lambda_i^*|$, x^* is a local minimum of $p(x,c)$.

Proof. When $u \notin M$ the directional derivative is strictly positive. (In the proof of Theorem 1 an inequality (at least) must be replaced by a strict inequality.)

When $u \in M$ we have

$$p(x^*+\lambda u,c) - p(x^*,c) = \frac{\lambda^2}{2}[\langle u, \nabla^2 f(x^*)u \rangle + \sum_{i=1}^{m} c_i |\langle u, H_i(x^*)u \rangle| + \sum_{i \in I(x^*)} c_i(\langle u, H_i(x^*)u \rangle)_+] + o(\lambda) .$$

with $\lim_{\lambda \to 0} o(\lambda)/\lambda^2 = 0$. As before we have

$$\lambda_i \langle u, H_i(x^*)u\rangle \le c_i |\langle u, H_i(x^*)u\rangle| \quad \text{for} \quad i \le m$$

and

$$\lambda_i \langle u, H_i(x^*)u\rangle \le c_i(\langle u, H_i(x^*)u\rangle) \quad \text{for} \quad i \in I(x^*)$$

because of assumption (ii) one inequality (at least) is strict. Then

$$\begin{aligned} &[p(x^*+\lambda u,c)-p(x^*,c)]/\lambda^2 \\ &\quad > \langle u, \nabla^2 f(x^*)u\rangle + \langle u, \sum \lambda_i H_i(x^*)u\rangle \\ &\quad \ge 0 \quad \text{(second order necessary condition)} \end{aligned}$$

and then x^* is a local minimum.

If the constraints are linear, then if $u \in M$, $x^*+\lambda u$ satisfies the constraints for λ small enough and as by hypothesis x^* is a local minimum for the problem (P) it is impossible to improve f in the direction u.

<u>Note</u>. In fact assumption (i) could be replaced by a weaker assumption since we simply need that first and second order necessary conditions hold at x^*. This assumption could be: First and second order constraints qualifications hold at x^*.

II.4. Theorem 3 (Necessary condition)

Assume x^* is a local minimum of the problem (P).

Assume (i) holds (λ_i^* are the Lagrange multipliers). Then if $\exists i_0$ such that $c_{i_0} < |\lambda_{i_0}^*|$ (and $\nabla h_{i_0}(x^*) \neq 0$), x^* cannot be a local minimum of $p(x,c)$.

Proof. If $i_0 \geq m+1$ then we must have $h_{i_0}(x^*) = 0$ since otherwise $\lambda_{i_0}^* = 0$ and we cannot have $c_{i_0} < 0$. Let us call v the orthogonal projection of $\nabla h_{i_0}(x^*)$ on the subspace orthogonal to the subspace spanned by the gradients of the constraints which are equal to zero (except $\nabla h_{i_0}(x^*)$). v is not zero because of assumption (i). Then

$$\text{if } \lambda_{i_0} > 0 \text{ take } u = v$$
$$\text{if } \lambda_{i_0} < 0 \text{ take } u = -v$$

then

$$c_{i_0} |\langle \nabla h_{i_0}(x^*), u\rangle| < \lambda_{i_0} \langle \nabla h_{i_0}(x^*), u\rangle$$

and

$$\begin{aligned}(p(x^*+\lambda u,c)-p(x^*,c))/\lambda &= \langle \nabla f(x^*),u\rangle + c_{i_0} |\langle \nabla h_{i_0}(x^*),u\rangle| + o(\lambda)/\lambda \\ &< \langle \nabla f(x^*),u\rangle + \lambda_{i_0} \langle \nabla h_{i_0}(x^*),u\rangle + o(\lambda)/\lambda \\ &= \langle \nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla h_i(x^*),u\rangle + o(\lambda)/\lambda\end{aligned}$$

$$\Rightarrow \lim_{\lambda \to 0} p(x^*+\lambda u,c) - p(x^*,c)/\lambda < 0$$

$\Rightarrow x^*$ is not a local minimum of $p(x,c)$.

II.5. Other Results

As usual we call $L(x,\lambda)$ the Lagrangian of the problem (P):

$$L(x,\lambda) = f(x) + \sum_{i=1}^{p} \lambda_i h_i(x) .$$

<u>Proposition 1</u>. If x is a local minimum of the Lagrangian $L(x,\lambda)$ then if $c_i \geq |\lambda_i|$, x is a local minimum of $p(x,c)$.

<u>Proof</u>. x local minimum implies there exists ε such that for all $x \in B(x,\varepsilon)$

$$p(x,c) = f(x) = f(x) + \sum_{i=1}^{p} \lambda_i h_i(x) \leq f(x') + \sum_{i=1}^{p} \lambda_i h_i(x')$$

$$\leq f(x') + \sum_{i=1}^{m} c_i |h_i(x')| + \sum_{i=m+1}^{p} c_i (h_i(x'))_+ = p(x',c)$$

$$\Rightarrow p(x,c) \leq p(x',c) \quad \text{for all} \quad x' \in B(x,\varepsilon)$$

$$\Rightarrow x \quad \text{a local minimum of} \quad p(x,c)$$

<u>Proposition 2</u>. If (x^*,λ) is a saddle point of the Lagrangian of the problem (P) and if $c_i > |\lambda_i|$, then any global minimum of $p(x,c)$ is the solution of the problem (P).

<u>Proof</u>. Let $\tilde{x}$ be a global minimum of $p(x,c)$. We have then:

$$p(x^*,c) = f(x^*) \leq f(\tilde{x}) + \sum_{i=1}^{p} \lambda_i h_i(\tilde{x})$$

because (x^*,λ) is a saddle point of $L(x,\lambda)$; and we have also

$$f(\tilde{x}) + \sum_{i=1}^{p} \lambda_i h_i(\tilde{x}) \leq f(\tilde{x}) + \sum_{i=1}^{m} c_i |h_i(\tilde{x})| + \sum_{i=m+1}^{p} c_i (h_i(\tilde{x}))_+$$

because $c_i > |\lambda_i|$ and $\lambda_i \geq 0$ for $i \geq m+1$. Moreover, since $\tilde{x}$ is a global minimum of $p(x,c)$ we have

$$f(\tilde{x}) + \sum_{i=1}^{p} c_i |h_i(\tilde{x})| + \sum_{i=m+1}^{p} c_i (h_i(\tilde{x}))_+ \leq p(x^*,c) = f(x^*)$$

Then we must have

(a) $$p(\tilde{x},c) = f(x^*)$$

(b) $$L(\tilde{x},\lambda) = p(\tilde{x},c)$$

From (b) we can write

$$\sum_{i=1}^{m}\left(c_i|h_i(\tilde{x})|-\lambda_i h_i(\tilde{x})\right) + \sum_{i=m+1}^{p}\left(c_i(h_i(\tilde{x}))_+ - \lambda_i h_i(\tilde{x})\right) = 0$$

As each term in the two sums is positive we must have

(1) $$c_i|h_i(\tilde{x})| = \lambda_i h_i(\tilde{x}) \ , \quad i = 1,\dots,m$$

(2) $$c_i(h_i(\tilde{x}))_+ = \lambda_i h_i(\tilde{x}) \ , \quad i = m+1,\dots,p$$

Since $c_i > \lambda_i$ we must have

$$h_i(\tilde{x}) = 0 \ , \quad i = 1,\dots,m$$
$$(h_i(\tilde{x}))_+ = 0 \ , \quad i = m+1,\dots,p \Rightarrow h_i(\tilde{x}) \le 0 \ , \quad \forall i \ge m+1$$

and the proof is complete.

In a practical point of view, using a vector of coefficients instead of one coefficient only could be interesting for problems where, at the solution, the Lagrange multipliers (if they exist) are very different in absolute value. Then if one coefficient only is used it must be greater than the maximum absolute value of the Lagrange multiplier (cf. Theorem 3). Hence some constraints are too penalized and this could be a trouble for the convergence. So, methods used to minimize exactly penalty function (as the algorithm proposed by Conn) can be easily adapted. An heuristic taking into account the last remark would be useful to update the penalty coefficients.

III. Generalization

III.1 Definitions

Let us consider the class of the following continuous penalty functions:

$$p_i(t) = 0 \quad \text{if} \quad t \leq 0$$
$$p_i(t) > 0 \quad \text{otherwise}$$

We define $Dp_i^+(0)$ as

$$\lim_{t \to 0^+} (p(t)/t) \qquad (\text{supposed} < +\infty) .$$

The new exact penalty function is now

$$p(x,c) = f(x) + \sum_{i=1}^{m} p_i[h_i(x)] + p_i[-h_i(x)] + \sum_{i=m+1}^{p} p_i[h_i(x)] .$$

The directional derivative in the direction u at a point z which satisfies the constraints of the original problem is now

$$Dp(z) = \langle \nabla f(z),u \rangle + \sum_{i=1}^{m} Dp_i^+(0) |\langle \nabla h_i(z),u \rangle| + \sum_{i=m+1}^{p} Dp_i^+(0) (\langle \nabla h_i(z),u \rangle)_+ .$$

III.2 Results

We can show that theorems 1, 2, and 3 of Section II are still valid. We have only to replace c_i by $Dp_i^+(0)$ and the proofs are made in the same way. This is the generalization of the result given by Bertsekas for the convex case (see [4]).

IV. The Algorithm

In this section we present an algorithm to minimize the exact penalty function we presented in section II. At each iteration of the algorithm we calculate the descent direction which minimizes the directional derivative by solving a simple linear programming problem.

In fact to avoid jamming at certain points (it is possible since the directional derivative is not continuous) we consider $|h_i(x)| \leq \varepsilon$ as $h_i(x) = 0$; ε is not fixed and is modified according to some rule. (The same principle has been used by Demjanov, see [5].) We show that any accumulation point of a sequence generated by this algorithm has its directional derivatives positive.

First we introduce some notation. $Dp_{\varepsilon,z}$ is a function of two variables α, β.

$$Dp_{\varepsilon,z}(\alpha,\beta) = \langle\nabla f(\alpha),\beta\rangle + \sum_{i\in I_\varepsilon^1(z)} |\langle\nabla h_i(\alpha),\beta\rangle| c_i + \sum_{i\in I_\varepsilon^2(z)} c_i\delta_i\langle\nabla h_i(\alpha),\beta\rangle$$
$$+ \sum_{i\in I_\varepsilon^3(z)} c_i(\langle\nabla h_i(\alpha),\beta\rangle)_+ + \sum_{i\in I_\varepsilon^4(z)} c_i\langle\nabla h_i(\alpha),\beta\rangle$$

where

$$I_\varepsilon^1(z) = \{i \mid 1\leq i\leq m;\ |h_i(z)|\leq\varepsilon\}$$
$$I_\varepsilon^2(z) = \{i \mid 1\leq i\leq m;\ |h_i(z)|>\varepsilon\}$$
$$I_\varepsilon^3(z) = \{i \mid i\geq m+1;\ |h_i(z)|\leq\varepsilon\}$$
$$I_\varepsilon^4(z) = \{i \mid i\geq m+1;\ h_i(z)>\varepsilon\}$$
$$\delta_i = \begin{matrix} +1 & \text{if} & h_i(z) > 0 \\ -1 & \text{if} & h_i(z) < 0 \end{matrix} .$$

$Dp_{0,z}(z,u)$ is the directional derivative of $p(z,c)$ at the point z in the direction u.

We shall call $u_\varepsilon(z)$ the vector such that

$$Dp_{\varepsilon,z}(z,u_\varepsilon(z)) = \min\{Dp_{\varepsilon,z}(z,u):\ \text{subject to } \|u\|_\infty \leq 1\}.$$

Note that

$$Dp_{\varepsilon,z}(\alpha,\beta) \geq Dp_{\varepsilon',z}(\alpha,\beta) \quad \forall\varepsilon\geq\varepsilon'\geq 0 .$$

We define

$$\Delta = \{z \in R^n | Dp_{0,z}(z, u_0(z)) = 0\} .$$

IV.1 The Algorithm

Step 0: Set $i = 0$, $z_i = z_0$ (ε_0 given).

Step 1: Set $I = \{\ell \,|\, |h_\ell(z_i)| \le \varepsilon_0\}$

$I' = \emptyset$.

Step 2: $\varepsilon(i) = \max_{\ell \in I-I'} \{|h_\ell(z_i)|\}$; (if $I-I' = \emptyset$, $\varepsilon(i) = 0$). If $A = Dp_{\varepsilon(i),z_i}(z_i, u_{\varepsilon(i)}(z_i)) \le -\varepsilon(i)$ go to step 3. Else if $A = 0$, stop. Else set $I' = \{\ell \in I \,|\, |h_\ell(z_i)| = \varepsilon(i)\}$ and go to step 2.

Step 3: $\lambda(z_i) = \operatorname{argmin}_{\lambda \ge 0} p(z_i + \lambda u_{\varepsilon(i)}(z_i), c)$. Set $z_{i+1} = z_i + \lambda(z_i) u_{\varepsilon(i)}(z_i)$, $i = i+1$. Go to step 1.

IV.2 Convergence

Proposition. Any accumulation point of a sequence generated by this algorithm $\in \Delta$.

Proof. Let z be an accumulation point. Suppose the proposition is false. Then

$$z \notin \Delta \Rightarrow Dp_{0,z}(z, u_0(z)) = -\gamma < 0 .$$

First we show that there exists a subsequence $u_i \to \hat{u}$ such that $Dp_{0,z}(z, \hat{u}) < 0$ and then we shall exhibit a contradiction. Let us call $\{\tilde{z}_i\}$ a subsequence which converges towards z, $\tilde{z}_i \to z$. There are two possibilities:

(a) $\exists i_0$ such that $\forall i \ge i_0$, $\varepsilon(i) \ge \delta > 0$.

(b) There exists a subsequence $z_i' \to z$ such that $\varepsilon(i) \to 0$ if $i \to \infty$.

Consider case (a). By continuity of $h_\ell(\cdot)$ we know that there exists $\varepsilon(z)$ such that for all $z' \in B(z,\varepsilon(z))$

$$|h_\ell(z')-h_\ell(z)| < \delta \quad \forall \ell .$$

Then

$$|h_\ell(z)| = 0 \Rightarrow |h_\ell(z')| < \delta \leq \varepsilon(i) , \quad \forall i \geq i_0 .$$

Then as soon as $\tilde{z}_u \in B(z,\varepsilon(z))$ (for $i \geq i_0'$) we have

$$Dp_{\varepsilon(z),\tilde{z}_i}(\tilde{z}_i,u) \geq Dp_{0,z}(\tilde{z}_i,u) , \quad \forall u$$

and

$$Dp_{0,z}(\tilde{z}_i,u_{\varepsilon(i)}(\tilde{z})) \leq Dp_{\varepsilon(i),\tilde{z}_i}(\tilde{z}_i,u_{\varepsilon(i)}(\tilde{z}_i)) \leq -\delta$$

as

$$u_{\varepsilon(i)}(\tilde{z}_i) \in B = \{u \in R^n | \|u\|_\infty \leq 1\} \quad \text{(compact set)}.$$

There exists a subsequence $\{z_i''\}$ such that $z_i'' \to z$, $u_{\varepsilon(i)}(z_i'') \to \hat{u} \in B$ and by continuity of the function $Dp_{0,z}(\cdot,\cdot)$ we must have

$$Dp_{0,z}(z,\hat{u}) \leq -\delta < 0 .$$

Consider now case (b).

$$\varepsilon_1 = \min_{h\ell(z)>0} [|h_\ell(z)|/2]$$

if $h^\ell(z) = 0 \;\; \forall \ell$ set $\varepsilon_i = 1$

$$\varepsilon = \min [\varepsilon_1, |\gamma|/2] .$$

(i) By continuity of $Dp_{0,z}(\cdot,\cdot)$ there exists $\varepsilon(z)$ such that $\forall z' \in B(z,\varepsilon(z))$

$$\min\{Dp_{0,z}(z,u) \,|\, \|u\|_\infty \leq 1\} \leq \frac{1}{2} \min\{Dp_{0,z}(z',u) \,|\, \|u\|_\infty \leq 1\} .$$

In other words there exists $\varepsilon(z)$ such that $\forall z' \in B(z,\varepsilon(z))$

$$Dp_{0,z}(z',u_0(z')) \leq -\frac{1}{2}\gamma .$$

(ii) By continuity of h_ℓ, $\forall\ell$, we can write there exists $\varepsilon'(z)$ such that

$$h_\ell(z) = 0 \Rightarrow |h_\ell(z')| \leq \varepsilon$$
$$h_\ell(z) \neq 0 \Rightarrow |h_\ell(z')| > \varepsilon$$

$\forall z' \in B(z,\varepsilon'(z))$.

(iii) Call $\{z_i'\}$ the subsequence which converges towards z and such that $\varepsilon(i) \to 0$ as $i \to \infty$. Then

$$z_i' \to z \Rightarrow \exists i_0 \text{ such that } i \geq i_0 \Rightarrow z_i' \in B(z,\varepsilon_1(z))$$
$$\varepsilon(i) \to 0 \Rightarrow \exists i_0 \text{ such that } i \geq i_0' \Rightarrow \varepsilon(i) \leq \varepsilon$$

(where $\varepsilon_1(z) = \min(\varepsilon(z),\varepsilon'(z))$) and then because of the definition of $\varepsilon(i)$ at step 2 of the algorithm, we have

$$Dp_{\varepsilon(i),z_i'}(\alpha,\beta) = Dp_{0,z}(\alpha,\beta)$$

$\forall z_i'$; $i \geq \max(i_0,i_0')$ and we have

$$Dp_{\varepsilon(i),z_i'}(z_i',u_{\varepsilon(i)}(z_i')) \leq -\varepsilon(i)$$

(because of (i)). As $u_{\varepsilon(i)}(z_i') \in B$ there exists a subsequence $z_i'' \to z$ such that $u_{\varepsilon(i)}(z_i'') \to \hat{u} \in B$ and by continuity of $Dp_{0,z}(\cdot,\cdot)$ we have

$$Dp_{0,z}(z_i'',u_{\varepsilon(i)}(z_i'')) \longrightarrow Dp_{0,z}(z,\hat{u}) = -\gamma < 0 .$$

Hence we have shown that there exists in any case a subsequence

$$z_i \to z$$
$$u_{\varepsilon(i)}(z_i) \to \hat{u}$$

such that

$$Dp_{0,z}(z,\hat{u}) < 0 .$$

Define, now

$$\hat{\lambda} = \underset{\lambda \geq 0}{\operatorname{argmin}}\; p(z+\lambda\hat{u},c)$$

and

$$\delta(z) = p(z+\hat{\lambda}\hat{u},c) - p(z) < 0 .$$

Since $p(z_i,c)$ is a decreasing sequence which converges towards $p(z,c)$, by continuity of $p(\cdot,c)$ we must have

$$\forall\varepsilon\ \exists i_0 \text{ such that } i \geq i_0 \Rightarrow |p(z_i,c)-p(z_{i+1},c)| < \varepsilon$$

(take $\varepsilon = +|\delta(z)|/2$).

$$\exists\varepsilon(z) \text{ such that } \|u'-\hat{u}\| \leq \varepsilon(z) \text{ and } \|z'-z\| \leq \varepsilon(z)$$
$$\Rightarrow p(z'+\hat{\lambda}\hat{u},c) - p(z') \leq \frac{1}{2}\delta(z) = \frac{1}{2}(p(z+\hat{\lambda}\hat{u},c) - p(x)) .$$

There exists i_0 such that $\|u_i-\hat{u}\| \leq \varepsilon(z)$ and $\|z_i-z\| \leq \varepsilon(z)$ $\forall i \geq i_0'$ (since $u_i \to \hat{u}$ and $z_i \to z$). Then

$$i \geq \max(i_0,i_0')$$
$$\Rightarrow \frac{|\delta(z)|}{2} > |p(z_{i+1},c)-p(z_i,c)| \geq |p(z_i+\lambda(z_i)u_{\varepsilon(i)}(z_i),c)-p(z_i,c)|$$
$$\geq |p(z_i+\hat{\lambda}u_{\varepsilon(i)}(z_i),c)-p(z_i,c)| \geq \frac{1}{2}|\delta(z)|$$
$$\Rightarrow \text{contradiction} \Rightarrow z \in \Delta .$$

Q.E.D.

Note. Of course in the implementable version, we replace step 3 of the algorithm by an Armijo stepsize rule (as in the steepest descent algorithm). Then step 3 is now

Step 3: Compute the first integer k_i such that

$$p(z_i+\beta^{k_i}u_{\varepsilon(i)}(z_i),c) - p(z_i,c) \leq \alpha\beta^{k_i}Dp_{\varepsilon(i),z_i}(z_i,u_{\varepsilon(i)}(z_i))$$

where α and β are given, $\alpha \in [0,1]$, $\beta \in [0,1]$. $z_{i+1} = z_i + \beta^{k_i} u_{\varepsilon(i)}(z_i)$, $i = i+1$, go to step 1.

IV.3 An Example

To illustrate the conclusion we made in the end of section II, we applied this algorithm for the Rosen Suzuki problem which is

$$\min f(x) = x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4$$

subject to

$$\text{(i)} \quad 2x_1^2 + x_2^2 + x_3^2 + 2x_1 + x_2 - x_4 - 5 \leq 0$$
$$\text{(ii)} \quad x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 - 8 \leq 0$$
$$\text{(iii)} \quad x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 - 10 \leq 0$$

The solution is

$$f(x) = -44 \quad \text{at} \quad x^* = (0,1,2,-1)$$

with constraints (i) and (ii) active. The Lagrange multipliers are (2,1,0).

(a) With $c_1 = 2.001$, $c_2 = 1.001$, $c_3 = 0.001$ the results after <u>25 iterations</u> are (we started at (0,0,0))

$Dp_{\varepsilon,x^*}(x^*,u_\varepsilon(x^*)) \geq -0.0001$

$x^* = (0.00001,\ 1.00000,\ 2.00000,\ -1.00001)$

$f(x^*) = -44.00007$

The constraints (i) and (ii) are 0.00002 and 0.00003.

(b) With $c_1 = 3 = c_2 = c_3$ we have after <u>59 iterations</u>

$Dp_{\varepsilon,x^*}(x^*,u_\varepsilon(x^*)) \geq -0.0001$

$x^* = (0.00001, 1.00001, 2.00000, -0.99998)$

$f(x^*) = -44.00002$

The constraints (i) and (ii) are 0.00002 and 0.00003.

We can see that although the Lagrange multipliers are not very different the convergence is much slower with one coefficient only.

References

[1] PIETRZYKOWSKI, T.

An exact potential method for constrained maxima, SIAM J. of Numerical Analysis, vol. 6, no. 2, June 1969.

[2] CONN, A.

Constrained optimization using a nondifferentiable penalty function, SIAM J. of Numerical Analysis, vol. 10, no. 4, September 1973.

[3] HAN, S.P.

A globally convergent method for nonlinear programming, JOTA, vol. 22, no. 3, July 1977.

[4] BERTSEKAS, D.P.

Necessary and sufficient conditions for a penalty method to be exact, Mathematical Programming, vol. 9, 1975, pp. 87-89.

[5] DEMJANOV, V.F.

Algorithm for some minimax problems, Journal of Computer and Systems Sciences, vol. 2, 1968, pp. 342-380.

[6] ZANGWILL, W.I.

Nonlinear programming via penalty functions, Management Science, vol. 13, 1967, pp. 344-358.

[7] FIACCO, A.V. and McCORMICK, G.P.

Nonlinear Programming: Sequential Unconstrained Minimization Techniques, Wiley, New York, 1968.

[8] LUENBERGER, D.G.

Introduction to Linear and Nonlinear Programming, Addison-Wesley, Reading, Mass., 1972.

[9] POLAK, E.

Computational Methods in Optimization. A Unified Approach, Academic Press, New York, 1971.

Just before we sent this paper to the printer we became aware of a paper by S. P. Han and O. L. Mangasarian, "Exact penalty function in nonlinear programming." The sufficient condition of Section II is also demonstrated, but under stronger assumption. (Second order sufficiency condition is required).