

**A LAGRANGIAN PENALTY FUNCTION METHOD  
FOR MONOTONE VARIATIONAL INEQUALITIES**

**L.D. Muu and W. Oettli**

Nr. 100 (1989)

# A LAGRANGIAN PENALTY FUNCTION METHOD FOR MONOTONE VARIATIONAL INEQUALITIES

Le D. Muu<sup>1</sup> and W. Oettli

*Universität Mannheim  
6800 Mannheim, Germany*

**Abstract.** A Lagrange-type penalty function method is proposed for a class of variational inequalities. The penalty function may have both positive and negative values. Each penalized subproblem is required to be solved only approximately. A condition under which a Lagrangian penalty function is exact, and an estimate for the penalty coefficient are given.

**Key words.** Penalty function method, variational inequality, monotonicity, coercivity, exactness, algorithm.

## 1. Introduction.

Penalty function methods are fundamental techniques which can be used to transform a constrained problem into a sequence of simpler, unconstrained problems. These methods in their pure forms however often suffer from inherent numerical instability as the penalty parameter tends to 0 or  $+\infty$ . In this paper we use a penalty function proposed in [5] to transform a constrained variational inequality into simpler problems which have to be solved only approximately. The penalty function used here takes positive values outside the feasible region, but in contrast to customary penalty functions it may assume negative values inside the feasible region. The value of this function at a point therefore can be considered as a quantity of reward or penalty for this point depending on its being feasible or not. We would like to call it a Lagrangian penalty function, since it acts like a single cumulative

---

<sup>1</sup> On leave from Institute of Mathematics, Hanoi, by a grant from Alexander-von-Humboldt-Stiftung.

constraint for the feasible region. The penalty parameter is steered adaptively and in nondegenerate situations avoids tending to either 0 or  $+\infty$ .

More precisely the problem we are going to consider is of the following type:

$$x^* \in C, f(x^*, x) \geq 0 \quad \forall x \in C,$$

where  $C$  is a closed convex subset of some reflexive Banach space  $X$ , and  $f$  is a real-valued function defined on  $C \times C$  ( or a larger product set ) such that  $f(x, x) = 0$  for all  $x$  and a certain monotonicity condition is satisfied. This is a rather flexible formulation which subsumes conveniently several standard problems.

We mention only the following:

- Variational operator inequalities:

$$x^* \in C, \langle T(x^*), x - x^* \rangle \geq 0 \quad \forall x \in C,$$

where  $T : C \rightarrow X^*$  is monotone.

- Complementarity problems:

$$x^* \in C, T(x^*) \in C^+, \langle x^*, T(x^*) \rangle = 0,$$

where  $C \subset X$  is a closed convex cone,  $C^+ \subset X^*$  is the nonnegative polar cone of  $C$ , and  $T : C \rightarrow X^*$  is monotone ( a special case of the previous example ).

- Programming problems:

$$x^* \in C, \varphi(x^*) \leq \varphi(x) \quad \forall x \in C,$$

where  $\varphi : C \rightarrow \mathbb{R}$ .

- Saddle point problems:

$$(x_1^*, x_2^*) \in C_1 \times C_2, \varphi(x_1^*, x_2) \leq \varphi(x_1, x_2^*) \quad \forall (x_1, x_2) \in C_1 \times C_2,$$

where  $\varphi : C_1 \times C_2 \rightarrow \mathbb{R}$ .

The paper is divided into four sections. In Section 2 we shall state and prove some results which constitute the foundation of our method. Section 3 is devoted to the

description of the algorithm and its convergence. In the last section exactness will be discussed.

## 2. Problem Statement and Auxiliary Results.

Throughout this paper let  $X$  be a reflexive real Banach space provided with the weak topology. Let  $C$  and  $M$  be nonempty closed convex subsets of  $X$  such that  $C \subset M$ , and let  $f : M \times M \rightarrow \mathbb{R}$ . We shall deal with the following Problem  $(P)$ , which we call a variational inequality problem :

$$x^* \in C, f(x^*, x) \geq 0 \quad \forall x \in C. \quad (P)$$

Suppose given a function  $p : M \rightarrow \mathbb{R}$  satisfying for all  $x \in M$

$$p(x) \leq 0 \iff x \in C.$$

This function, which may assume negative values on  $C$ , will be referred to as Lagrangian penalty or reward-penalty function. For each real  $t > 0$  we denote by  $(P_t)$  the following penalized ( but less constrained ) problem :

$$x_t \in M, tf(x_t, x) + p(x) - p(x_t) \geq 0 \quad \forall x \in M. \quad (P_t)$$

We shall henceforth make the following *assumptions*:

- (i)  $f(x, y) + f(y, x) \leq 0 \quad \forall x, y \in M$  ( monotonicity of  $f$  );  
 $f(x, x) = 0 \quad \forall x \in M$  ;  
 $f$  is convex and lower semicontinuous in the second argument on  $M$  ;  
the function  $\lambda \mapsto f(\lambda x + (1 - \lambda)y, x)$ ,  $0 \leq \lambda \leq 1$ , is upper semicontinuous at  $\lambda = 0$  for all  $x, y \in M$  ( radial continuity of  $f$  );  
there exists  $a \in C$  such that  $f(x, a) \rightarrow -\infty$  as  $\|x\| \rightarrow \infty$ ,  $x \in M$   
( coercivity of  $f$  ).
- (ii)  $p$  is convex, lower semicontinuous on  $M$ , and bounded from below on  $C$ .

From assumption (i) it follows that  $(P)$  has a solution, see e.g. [2], [6]. It follows moreover from assumptions (i) and (ii) that Problem  $(P_t)$  has a solution for all

$t > 0$ . This follows from the function  $f_t(y, x) := tf(y, x) + p(x) - p(y)$  satisfying again (i) ( in particular  $f_t$  is coercive, i.e.,  $f_t(x, a) \rightarrow -\infty$  as  $\|x\| \rightarrow \infty$ , since  $p$  is bounded from below on  $M$  ). We denote by  $S(\epsilon, t)$  for all real  $\epsilon \geq 0$ ,  $t > 0$  the set of all  $\epsilon$ - solutions of  $(P_t)$ , i.e.,

$$x_t \in S(\epsilon, t) \text{ iff } x_t \in M, tf(x_t, x) + p(x) - p(x_t) \geq -\epsilon \quad \forall x \in M.$$

$S(\epsilon, t)$  is nonempty for all  $\epsilon \geq 0$ ,  $t > 0$ . Furthermore we define for all real  $\epsilon \geq 0$ ,  $t > 0$  the set

$$B(\epsilon, t) := \{x \in M : f(x, a) \geq 0 \text{ or } tf(x, a) + p(a) - p(x) \geq -\epsilon\}.$$

The following properties are easily verified:

- 1) If  $\epsilon' \leq \epsilon$  and  $t' \geq t$ , then  $B(\epsilon', t') \subset B(\epsilon, t)$ .
- 2)  $S(\epsilon, t) \subset B(\epsilon, t)$ .
- 3)  $B(\epsilon, t)$  is bounded in norm.

Indeed : Properties 1) and 2) follow right from the definitions of  $B(\epsilon, t)$  and  $S(\epsilon, t)$ . Property 3) follows from the fact that both  $f(x, a) \rightarrow -\infty$  and  $tf(x, a) + p(a) - p(x) \rightarrow -\infty$  as  $\|x\| \rightarrow \infty$ .

Let now for all real  $\epsilon \geq 0$  denote

$$C(\epsilon) := \{x \in M : p(x) \leq \epsilon\},$$

$$t(\epsilon) := \sup\{t : t = 0 \text{ or } S(\epsilon, t) \cap C(\epsilon) \neq \emptyset\}.$$

It is clear that  $\epsilon' \leq \epsilon$  implies  $0 \leq t(\epsilon') \leq t(\epsilon) \leq \infty$ .

We collect some lemmata, which we need in the sequel.

**Lemma 2.1.** Let  $S$  be a convex subset of  $M$  and let  $h : M \times M \rightarrow \mathbb{R}$  be such that for  $h$  assumption (i) above is satisfied. Then the following statements are equivalent:

- (a)  $y \in S, h(y, x) \geq 0 \quad \forall x \in S;$
- (b)  $y \in S, h(x, y) \leq 0 \quad \forall x \in S.$

**Proof.** Statement (a) implies (b) because of the monotonicity of  $h$ . Assume now that (b) is true. Let  $x \in S$  be arbitrary and  $x_\lambda := \lambda x + (1 - \lambda)y$ ,  $0 < \lambda \leq 1$ . Then  $h(x_\lambda, y) \leq 0$  since  $x_\lambda \in S$ . Therefore

$$0 = h(x_\lambda, x_\lambda) \leq \lambda h(x_\lambda, x) + (1 - \lambda)h(x_\lambda, y) \leq \lambda h(x_\lambda, x).$$

Dividing by  $\lambda$  and letting  $\lambda \rightarrow 0$  we obtain statement (a). Q.E.D.

**Lemma 2.2.** If  $\epsilon_k \searrow 0$  and  $t(\epsilon_{k_0}) < \infty$  for some  $k_0$ , then  $t(\epsilon_k) \searrow t(0)$ .

**Proof.** Let  $\epsilon_k \searrow 0$ . Let  $t_*$  be the limit of  $\{t(\epsilon_k)\}$ . Since  $\{t(\epsilon_k)\}$  is nonincreasing and  $t(\epsilon_k) \geq t(0)$  and  $t(\epsilon_{k_0}) < \infty$ ,  $t_*$  exists and  $t(0) \leq t_* < \infty$ . If  $t(0) = t_*$  we are done. Assume for contradiction that  $t(0) < t_*$ . Let  $t(0) < t < t_*$ . By the definition of  $t(\epsilon_k)$ , there exists a sequence  $\{t_k\}$  such that  $t < t_k \leq t(\epsilon_k)$ ,  $\lim t_k = t_*$  and  $S(\epsilon_k, t_k) \cap C(\epsilon_k) \neq \emptyset$  for every  $k$ . Let  $u^k$  be a point of this set. Then

$$t_k f(u^k, x) + p(x) - p(u^k) \geq -\epsilon_k \quad \forall x \in M.$$

This and the monotonicity of  $f$  imply

$$t_k f(x, u^k) - p(x) + p(u^k) \leq \epsilon_k \quad \forall x \in M.$$

Since  $\epsilon_k \leq \epsilon_1$  and  $t_k > t$  for all  $k$  we have  $S(\epsilon_k, t_k) \subset B(\epsilon_k, t_k) \subset B(\epsilon_1, t)$  for every  $k$ . Hence  $\{u^k\}$  is bounded, and we therefore may assume, taking a subsequence if necessary, that  $u^k \rightarrow u^*$ . Letting  $k \rightarrow \infty$  we obtain, from the lower semicontinuity of  $f(x, \cdot)$  and  $p(\cdot)$ ,

$$t_* f(x, u^*) - p(x) + p(u^*) \leq 0 \quad \forall x \in M,$$

from which, by Lemma 2.1, follows  $u^* \in S(0, t_*)$ . This and  $p(u^*) \leq 0$  (since  $p(u^k) \leq \epsilon_k$  for all  $k$ ) imply  $S(0, t_*) \cap C(0) \neq \emptyset$ , hence  $t_* \leq t(0)$ . We arrive at a contradiction. Q.E.D.

**Lemma 2.3.** Let  $0 < t < t'$ ,  $0 \leq \epsilon, \epsilon'$ . Then for each  $x \in S(\epsilon, t)$ ,  $x' \in S(\epsilon', t')$  we have

$$p(x') - p(x) \geq -(\epsilon t' + \epsilon' t)/(t' - t).$$

**Proof.** From the definitions of  $S(\epsilon, t)$  and  $S(\epsilon', t')$  follows

$$tf(x, x') + p(x') - p(x) \geq -\epsilon,$$

$$t'f(x', x) + p(x) - p(x') \geq -\epsilon'.$$

Multiplying the first inequality by  $t'$  and the second by  $t$ , adding and using the monotonicity of  $f$  we obtain  $(t' - t)(p(x') - p(x)) \geq -\epsilon t' - \epsilon' t$ . <sup>2</sup> Q.E.D.

**Lemma 2.4.** For each  $x \in S(0, t)$  one has

$$p(x) > 0 \text{ if } t > t(0),$$

$$p(x) \leq 0 \text{ if } t < t(0).$$

**Proof.** The first assertion is obvious from the definition of  $t(0)$ . To prove the second one we chose  $t'$  such that  $t < t' \leq t(0)$  and  $S(0, t') \cap C \neq \emptyset$ . Let  $x'$  be a point of this set, then  $p(x') \leq 0$ . Applying Lemma 2.3 with  $\epsilon = \epsilon' = 0$  we obtain  $p(x) \leq p(x') \leq 0$ . Q.E.D.

**Lemma 2.5.** Let  $0 \leq \epsilon$ ,  $0 < \xi < 1$  and  $0 < t$ . Then for each  $x \in S(\epsilon, t)$  one has

$$t(\epsilon) \geq t \quad \text{if } p(x) \leq \epsilon, \tag{2.1}$$

$$t(\epsilon) < t/(1 - \xi) \text{ if } p(x) > 2\epsilon/\xi. \tag{2.2}$$

**Proof.** (2.1) follows immediately from the definition of  $t(\epsilon)$ . We have only to prove (2.2). Assume contrary that there is  $x \in S(\epsilon, t)$  with  $p(x) > 2\epsilon/\xi$  and  $t(\epsilon) \geq t/(1 - \xi)$ . Then  $t(\epsilon)/(t(\epsilon) - t) \leq 1/\xi$  and  $t < t(\epsilon)$ . By the definition of  $t(\epsilon)$  there exists a sequence  $\{t_j\}$  such that  $t < t_j \leq t(\epsilon)$ ,  $t_j \rightarrow t(\epsilon)$  and  $S(\epsilon, t_j) \cap C(\epsilon) \neq \emptyset$  for every  $j$ . Letting  $x^j \in S(\epsilon, t_j) \cap C(\epsilon)$  and applying Lemma 2.3 for  $x^j$  and  $x$  we obtain

$$p(x^j) - p(x) \geq -\epsilon(t_j + t)/(t_j - t).$$

<sup>2</sup> We note that this inequality may be considered as an extension of Le Châtelier's principle to  $\epsilon$ -solutions of monotone variational inequalities; we refer to W. Eichhorn and W. Oettli, *Econometrica* 40 (1972), 711 - 717.

This and  $p(x^j) \leq \epsilon$  imply  $p(x) \leq 2ct_j/(t_j - t)$  for every  $j$ . Letting  $j \rightarrow \infty$  it follows

$$p(x) \leq 2\epsilon \leq 2\epsilon/\xi \quad \text{if } t(\epsilon) = \infty,$$

$$p(x) \leq 2ct(\epsilon)/(t(\epsilon) - t) \leq 2\epsilon/\xi \quad \text{if } t(\epsilon) < \infty.$$

Both cases contradict  $p(x) > 2\epsilon/\xi$ .

Q.E.D.

### 3. Description of the Algorithm.

The following algorithm allows us to find a solution of Problem (P) as a limit point of  $x^k \in S(\epsilon_k, t_k)$  with  $\{\epsilon_k\}$  being a fixed sequence tending decreasingly to zero, and  $t_k$  being determined iteratively. The algorithm is described as follows (some comments are inserted in brackets).

#### Algorithm.

Let  $M, C, p$  be as required before. Take two sequences of positive numbers  $\{\epsilon_k\}$  and  $\{\xi_k\}$  such that  $\epsilon_k \searrow 0$ ,  $\xi_k \searrow 0$ ,  $\epsilon_k/\xi_k \searrow 0$ ,  $\xi_k < 1/2$  for all  $k$ , and take  $c \geq 4$ .

Step 1. [ Initialization ]

Set  $B_1 := \infty$ , choose  $t_1 > 0$ , set  $k = 1$ .

Step 2.

Select  $x^k \in S(\epsilon_k, t_k)$ .

2a) If  $p(x^k) \leq 2\epsilon_k/\xi_k$ , then set  $t_{k+1} := 2t_k$ ,  $B_{k+1} := \infty$ , increase  $k$  by 1 and go to Step 2.

2b) If  $p(x^k) > 2\epsilon_k/\xi_k$ , then set  $B_{k+1} := t_k/(1 - \xi_k)$ , increase  $k$  by 1 and go to Step 3.

Step 3. [ Given  $k$  and  $B_k < \infty$  such that  $t(\epsilon_k) < B_k$  ].

Set  $A_{k,1} := 0$ ,  $B_{k,1} := B_k$ . Set  $j = 1$ .

Step 4. [ Given  $k, j$ ,  $A_{k,j}, B_{k,j}$  such that  $A_{k,j} \leq t(\epsilon_k) < B_{k,j}$  ].

Set  $t_{k,j} := (A_{k,j} + B_{k,j})/2$ ,  $\epsilon_{k,j} := \min\{\epsilon_k, t_{k,j}\epsilon_k\}$ . Select  $x^{k,j} \in S(\epsilon_{k,j}, t_{k,j})$ .

4a) If  $B_{k,j} - A_{k,j} \leq c\xi_k B_k$  ( Clause 1 ) or if  $\epsilon_k < p(x^{k,j}) \leq 2\epsilon_k/\xi_k$  ( Clause 2 ), then set  $x^k := x^{k,j}$ ,  $t_k := t_{k,j}$ ,  $B_{k+1} := B_{k,j}$ , increase  $k$  by 1 and go to Step 3.

- 4b) If  $p(x^{k,j}) \leq \epsilon_k$ , then let  $A_{k,j+1} := t_{k,j}$ ,  $B_{k,j+1} := B_{k,j}$ , increase  $j$  by 1 and go to Step 4 [ lower bound of  $t(\epsilon_k)$  is improved to  $t_{k,j}$ , upper bound is unchanged ].
- 4c) If  $p(x^{k,j}) > 2\epsilon_k/\xi_k$ , then let  $A_{k,j+1} = A_{k,j}$ ,  $B_{k,j+1} = t_{k,j}/(1 - \xi_k)$ , increase  $j$  by 1 and go to Step 4 [ upper bound of  $t(\epsilon_k)$  is improved to  $t_{k,j}/(1 - \xi_k)$ , lower bound is unchanged ].

This completes the description of the algorithm.

**Remark.** The algorithm is a two phases algorithm in the following sense. As long as  $B_k = \infty$  it repeats only Step 2 ( first phase ). As soon as  $B_k < \infty$  it uses only Step 3 and 4 ( second phase ). It is possible without harm that the first phase never terminates. However if the algorithm enters the second phase, then it will be demonstrated that for each fixed  $k$ , Step 4 will be performed only finitely often. Then it comes back to Step 3, thereby increasing  $k$ . Hence for all  $k \geq 1$  the iterates  $t_k$ ,  $B_k$ ,  $x^k$  are well defined. If case 2b) occurs, then it follows from Lemma 2.5 that  $t(\epsilon_k) < B_{k+1}$ , and since  $\epsilon_{k+1} \leq \epsilon_k$  implies  $t(\epsilon_{k+1}) \leq t(\epsilon_k)$ , we have also  $t(\epsilon_{k+1}) < B_{k+1}$ . Case 2b) can occur only once; if it occurs for  $k := k_0$ , say, then we have  $t(\epsilon_k) \leq t(\epsilon_{k_0}) < \infty$  for all  $k \geq k_0$ . In cases 4b) and 4c) it follows from Lemma 2.5 and  $\epsilon_{k,j} \leq \epsilon_k$ , that  $A_{k,j+1} \leq t(\epsilon_k) < B_{k,j+1}$ . In case 4a) the algorithm returns to step 3 and we have then  $t(\epsilon_k) < B_{k,j} = B_{k+1}$ , and since  $t(\epsilon_{k+1}) \leq t(\epsilon_k)$  this implies  $t(\epsilon_{k+1}) < B_{k+1}$ . It will be demonstrated - see (3.1) below - that  $B_{k,j+1} \leq B_{k,j}$ . Therefore we have also  $B_{k+1} \leq B_k$ . When the algorithm exits from the finite loop generated by Step 4 and returns to Step 3 for  $j = j(k)$ , say, then it has produced  $x^k \in S(\epsilon'_k, t_k)$  with  $\epsilon'_k := \min\{\epsilon_k, t_k \epsilon_k\}$  and  $A'_k := A_{k,j(k)}$ ,  $B'_k := B_{k,j(k)} = B_{k+1}$  such that  $A'_k \leq t(\epsilon_k) < B'_k$ . These facts will be used freely in the proof of convergence below.

**Claim 1.** For each fixed  $k$ , Step 4 is performed only finitely often.

**Proof.** We first show that

$$B_{k,j+1} \leq B_{k,j} \leq B_k \quad \forall j. \quad (3.1)$$

To prove (3.1) it is sufficient, since  $B_{k,1} = B_k$ , to show that

$$B_{k,j} \leq B_k \Rightarrow B_{k,j+1} \leq B_{k,j} \quad \forall j.$$

Since in Case 4b)  $B_{k,j+1} = B_{k,j}$  we have only to prove for Case 4c). Then  $B_{k,j+1} = (A_{k,j} + B_{k,j})/(2(1 - \xi_k))$ . Using  $B_{k,j} - A_{k,j} > c\xi_k B_k$  and  $B_{k,j} \leq B_k$  we have  $B_{k,j} - A_{k,j} > c\xi_k B_k \geq c\xi_k B_{k,j}$ . From this and  $c > 2$  follows

$$A_{k,j} + B_{k,j} = 2B_{k,j} + (A_{k,j} - B_{k,j}) \leq 2B_{k,j} - c\xi_k B_{k,j} < 2(1 - \xi_k)B_{k,j}.$$

Hence  $B_{k,j+1} = (A_{k,j} + B_{k,j})/(2(1 - \xi_k)) \leq B_{k,j}$  as required. Now let  $d_j := B_{k,j} - A_{k,j}$ . Then we claim

$$d_j \leq B_k/2^{j-1} + 2B_k(1 - 1/2^{j-1})\xi_k/(1 - \xi_k) \quad \forall j. \quad (3.2)$$

Indeed, (3.2) is true for  $j = 1$ , since  $d_1 = B_{k,1} - A_{k,1} = B_k$ . Assume for induction that it is true for  $j$ , then we prove it for  $j + 1$ . In Case 4b) we have  $d_{j+1} = d_j/2$ , whereas in Case 4c)  $d_{j+1} > d_j/2$ . Hence it is sufficient to prove for Case 4c). In Case 4c) we have

$$B_{k,j+1} = (A_{k,j} + B_{k,j})/(2(1 - \xi_k)), \quad A_{k,j+1} = A_{k,j}.$$

This and (3.1) and  $t_{k,j} = (A_{k,j} + B_{k,j})/2 < B_k$  imply

$$d_{j+1} = d_j/2 - t_{k,j} + t_{k,j}/(1 - \xi_k) = d_j/2 + t_{k,j}\xi_k/(1 - \xi_k) \leq d_j/2 + B_k\xi_k/(1 - \xi_k).$$

Using the induction assumption we then obtain

$$\begin{aligned} d_{j+1} &\leq B_k/2^j + 2B_k(1/2 - 1/2^j)\xi_k/(1 - \xi_k) + B_k\xi_k/(1 - \xi_k) \\ &= B_k/2^j + 2B_k(1 - 1/2^j)\xi_k/(1 - \xi_k). \end{aligned}$$

Hence (3.2) is true for  $j + 1$ , and therefore for all  $j$ . From (3.2) follows

$\lim_{j \rightarrow \infty} d_j \leq 2B_k\xi_k/(1 - \xi_k) < 4B_k\xi_k$ , since  $\xi_k < 1/2$ . From  $c \geq 4$  we deduce that there must exist an index  $j(k)$  for which Clause 1 of Step 4 occurs. Q.E.D.

Let  $K_1 := \{k \in \mathbb{N} : \text{Clause 1 occurs at iteration } k\}$  and let  $K_2 := \{k \in \mathbb{N} : \text{Clause 2 occurs at iteration } k\}$ . Assume that  $K_1$  is not finite.

**Claim 2.**  $t_k \rightarrow t(0)$  as  $k \rightarrow \infty$ ,  $k \in K_1$ .

**Proof.** We have  $t(\epsilon_k) < \infty$  for all  $k \in K_1$ . Then Lemma 2.2 can be applied yielding  $t(\epsilon_k) \rightarrow t(0)$  as  $k \rightarrow \infty$ ,  $k \in K_1$ . Let  $j(k)$  be the index for which Clause 1 occurs. Then from the definitions of  $A_{k,j}$  and  $B_{k,j}$  together with Lemma 2.5 and  $\epsilon_{k,j} \leq \epsilon_k$  follows that

$$A_{k,j(k)} \leq t(\epsilon_k) \leq B_{k,j(k)} \quad \forall k \in K_1. \quad (3.3)$$

Furthermore under Clause 1 we have

$$B_{k,j(k)} - A_{k,j(k)} \leq c\xi_k B_k \quad \forall k \in K_1,$$

and since  $B_{k+1} \leq B_k$  for all  $k$  and  $\xi_k \searrow 0$  this implies

$$B_{k,j(k)} - A_{k,j(k)} \rightarrow 0 \quad \text{as } k \rightarrow \infty, k \in K_1.$$

Since  $t_k = t_{k,j(k)} := (A_{k,j(k)} + B_{k,j(k)})/2$  we obtain from (3.3) that  $\{t(\epsilon_k)\}$  and  $\{t_k\}$  have the same limit. Hence  $t_k \rightarrow t(0)$  as  $k \rightarrow \infty$ ,  $k \in K_1$ . Q.E.D.

We now are in the position to prove the convergence of the algorithm.

**Theorem 3.1.** Let assumptions (i) and (ii) hold. Then the sequence  $\{x^k\}$  is bounded, and if in addition

$$p(x) = 0 \quad \forall x \in S(0, t(0)) \quad \text{whenever } 0 < t(0) < \infty, \quad (3.4)$$

then any cluster point of  $\{x^k\}$  solves (P).

**Proof.**

1. We consider first the case  $B_k = \infty$  for all  $k$ , i.e., the first phase does not terminate. In this case  $t_k \nearrow \infty$ . This and  $\epsilon_k \searrow 0$  implies  $B(\epsilon_k, t_k) \subset B(\epsilon_1, t_1)$  for all  $k$ . Since  $x^k \in S(\epsilon_k, t_k) \subset B(\epsilon_k, t_k)$ , the sequence  $\{x^k\}$  is bounded. Let  $x^*$  be

any limit point of  $\{x^k\}$ . We may assume, taking a subsequence if necessary, that  $x^k \rightarrow x^*$ . From 2a) we have  $p(x^k) \leq 2\epsilon_k/\xi_k$ . Thus, by the lower semicontinuity of  $p$ ,  $p(x^*) \leq 0$ , i.e.,  $x^* \in C$ . Furthermore, since  $p$  is bounded from below on  $M$ , there is a constant  $K$  such that  $p(x^k) \geq -K$  for all  $k$ . From this and  $x^k \in S(\epsilon_k, t_k)$  follows by the monotonicity of  $f$  that  $f(x, x^k) \leq (p(x) + K + \epsilon_k)/t_k$  for all  $x \in M$ . Since  $t_k \rightarrow \infty$ , in the limit we obtain  $f(x, x^*) \leq 0$  for all  $x \in M$ . Hence, by Lemma 2.1,  $x^*$  solves (P).

2. Next we consider the situation when the first phase is finite. We may disregard the finitely many iterations of the first phase. Obviously it is sufficient to prove the claimed results for each of the two subsequences  $\{x^k\}$  with  $k \in K_1$  and  $k \in K_2$  respectively. We consider first the case  $k \in K_1$ . In this case Claim 2 implies  $t_k \rightarrow t(0) < \infty$ .

Assume first  $t(0) > 0$ . Then one can choose a real number  $\tau > 0$  and an index  $l$  such that  $t_k > \tau$  for all  $k \in K_1$ ,  $k \geq l$ . This implies  $B(\epsilon_k, t_k) \subset B(\epsilon_l, \tau)$  for all  $k \in K_1$ ,  $k \geq l$ . Hence  $\{x^k\}_{k \in K_1}$  is bounded. Let  $x^*$  be any limit point of this sequence. Taking a subsequence if necessary we may write  $x^k \rightarrow x^*$  as  $k \rightarrow \infty$ ,  $k \in K_1$ . From  $x^k \in S(\epsilon_k, t_k)$  and the monotonicity of  $f$  follows

$$t_k f(x, x^k) - p(x) + p(x^k) \leq \epsilon_k \quad \forall x \in M,$$

from which we obtain in the limit

$$t(0) f(x, x^*) - p(x) + p(x^*) \leq 0 \quad \forall x \in M.$$

Hence, by Lemma 2.1,  $x^* \in S(0, t(0))$ . Then from Assumption (3.4)  $p(x^*) = 0$ , i.e.,  $x^*$  is feasible. Furthermore

$$t(0) f(x^*, x) + p(x) - p(x^*) \geq 0 \quad \forall x \in M.$$

This and  $p(x) \leq 0$  for all  $x \in C$  and  $p(x^*) = 0$  imply that  $x^*$  solves (P).

Let now  $t(0) = 0$ . Observing that  $x^k \in S(t_k \epsilon_k, t_k)$  we obtain

$$t_k f(x^k, x) + p(x) - p(x^k) \geq -\epsilon_k t_k \quad \forall x \in M. \quad (3.5)$$

Choose  $t' = t_k$ ,  $t = t_k/2$ ,  $\epsilon' = \epsilon_k t_k$ ,  $\epsilon = 0$ ,  $y \in S(0, t)$  and apply Lemma 2.3. We obtain

$$p(x^k) - p(y) \geq -\epsilon_k t_k. \quad (3.6)$$

Adding (3.5) and (3.6) it follows

$$t_k f(x^k, x) + p(x) - p(y) \geq -2\epsilon_k t_k \quad \forall x \in M.$$

Since  $p(y) > 0$  ( Lemma 2.4 and  $t(0) = 0 < t$  ) and  $p(x) \leq 0$  for all  $x \in C$  we have

$$f(x^k, x) \geq -2\epsilon_k \quad \forall x \in C.$$

Choosing  $x = a$  we see, from the coercivity of  $f$ , that  $\{x^k\}_{k \in K_1}$  is bounded. Also, using the monotonicity of  $f$ , it follows

$$f(x, x^k) \leq 2\epsilon_k \quad \forall x \in C. \quad (3.7)$$

Let  $x^*$  be any limit point of  $\{x^k\}_{k \in K_1}$ . We may write, by taking a subsequence if necessary,  $x^k \rightarrow x^*$  as  $k \rightarrow \infty$ ,  $k \in K_1$ . From (3.5) and the monotonicity of  $f$  follows

$$t_k f(x, x^k) - p(x) + p(x^k) \leq \epsilon_k t_k \quad \forall x \in M.$$

Since  $t_k \rightarrow t(0) = 0$  we have  $p(x^*) \leq p(x)$  for all  $x \in M$ . Hence  $x^* \in C$ . From (3.7) follows, in the limit, that  $f(x, x^*) \leq 0$  for all  $x \in C$ . Hence by Lemma 2.1  $x^*$  solves (P).

3. Finally we consider the case  $k \in K_2$ . From Clause 2 we have  $\epsilon_k < p(x^k) \leq 2\epsilon_k/\xi_k$  for all  $k \in K_2$ . From  $x^k \in S(\epsilon_k, t_k)$  and  $\epsilon_k < p(x^k)$  follows  $t_k f(x^k, x) + p(x) \geq 0 \quad \forall x \in M$ , and therefore

$$f(x^k, x) \geq 0 \quad \forall x \in C. \quad (3.8)$$

Choosing  $x = a$  we see, from the coercivity of  $f$ , that  $\{x^k\}_{k \in K_2}$  is bounded. Let  $x^*$  be any limit point of this sequence. Taking again a subsequence if necessary, we may write  $x^k \rightarrow x^*$  as  $k \rightarrow \infty$ ,  $k \in K_2$ . From  $p(x^k) \leq 2\epsilon_k/\xi_k$  follows, in the limit,  $p(x^*) \leq 0$ . Hence  $x^*$  is feasible. From (3.8) and the monotonicity of  $f$  follows

$$f(x, x^k) \leq 0 \quad \forall x \in C.$$

By the lower semicontinuity of  $f(x, \cdot)$  we obtain in the limit

$$f(x, x^*) \leq 0 \quad \forall x \in C,$$

which by Lemma 2.1 implies that  $x^*$  solves  $(P)$ .

Q.E.D.

We give a condition under which Assumption (3.4) of Theorem 3.1 is satisfied.

**Proposition 3.1.** Assumption (3.4) is fulfilled if, in addition,  $p$  is upper semicontinuous and Problem  $(P_{t(0)})$  has a unique solution  $x^0$  whenever  $0 < t(0) < \infty$ .

**Proof.** Let  $\{t_k\}$  be a sequence such that  $t_k \nearrow t(0)$ , and let  $x^k \in S(0, t_k)$ . Then we know as in the proof of Theorem 3.1 that the sequence  $\{x^k\}$  is bounded. Hence a suitable subsequence of  $\{x^k\}$  converges to some  $x'$ . Using in turn the monotonicity of  $f$ , the lower semicontinuity of  $f(x, \cdot)$  and  $p$ , and Lemma 2.1, we obtain  $x' \in S(0, t(0))$ , which means  $x' = x^0$ . Since  $t_k \nearrow t(0)$ , by Lemma 2.4  $p(x^k) \leq 0$ . Hence  $p(x^0) \leq 0$ . Likewise applying Lemma 2.4 to a sequence  $\{t_k\}$  with  $t_k \searrow t(0)$  and using upper semicontinuity of  $p$  we obtain  $p(x^0) \geq 0$ . Hence  $p(x^0) = 0$ .

Q.E.D.

**Remark.** If either  $f$  is strictly monotone on  $M$ , i.e.,  $f(x, y) + f(y, x) < 0$  for every  $x, y \in M$ ,  $x \neq y$ , or  $t(0)f(x, \cdot) + p(\cdot)$  is strictly convex on  $M$ , then the solution of  $(P_{t(0)})$  is unique ( see e.g. [3] ).

#### 4. Exactness.

The exactness of a penalty function is an interesting question which earned attention of many authors ( see e.g. [1,3,4] ). Exactness means that there exists a parameter  $t$  such that the solution-set of  $(P_t)$  is contained in the solution-set of  $(P)$ . In order to treat this question and to give a lower estimate for  $t(0)$  we require that the feasible set  $C$  is given by

$$C := \{x \in M : g(x) \leq 0\},$$

where  $g = (g_1, \dots, g_m) : M \rightarrow \mathbb{R}^m$  with each  $g_j$  convex on  $M$ . Let  $p$  be defined by  $p(x) := l(g(x))$ , where  $l : \mathbb{R}^m \rightarrow \mathbb{R}$  is convex on  $\mathbb{R}^m$  ( hence continuous ) and

has the property

$$l(y) \leq 0 \iff y \leq 0.$$

Moreover  $l$  should be such that  $p$  is convex. Let  $x^*$  be a solution of Problem (P). Then, since  $f(x^*, x^*) = 0$ ,  $x^*$  is also a solution of the convex programming problem

$$\min\{f(x^*, x) : x \in C\} = \min\{f(x^*, x) : x \in M, g(x) \leq 0\}.$$

Assume in addition that the constraints  $g(x) \leq 0$  satisfy a constraint qualification, e.g. Slater's condition. Then there exists a multiplier vector  $\lambda \in \mathbb{R}^m$ ,  $\lambda \geq 0$  such that

$$0 \leq f(x^*, x) + \lambda^T g(x) \quad \forall x \in M.$$

Assume  $\lambda \neq 0$ . It is then easily seen that the convex programming problem  $\min\{l(y) : \lambda^T y \geq 0\}$  has optimal value 0. For this problem Slater's condition is satisfied, thus there exists a multiplier  $t_* \geq 0$  satisfying

$$0 \leq l(y) - t_* \lambda^T y \quad \forall y \in \mathbb{R}^m.$$

It follows then

$$0 \leq t_* f(x^*, x) + t_* \lambda^T g(x) \leq t_* f(x^*, x) + l(g(x)) = t_* f(x^*, x) + p(x) \quad \forall x \in M.$$

From  $x^* \in C$  follows  $p(x^*) \leq 0$ . Choosing  $x = x^*$  in the above inequality we obtain

$$0 \leq t_* f(x^*, x^*) + p(x^*) = p(x^*) \leq 0.$$

Hence  $p(x^*) = 0$ , and then

$$t_* f(x^*, x) + p(x) - p(x^*) \geq 0 \quad \forall x \in M.$$

This means that  $x^* \in S(0, t_*)$ . Since  $x^* \in C$  we have  $t_* \leq t(0)$ .

Assume now that  $t_* < t < t(0)$  and that  $x_t \in S(0, t)$ . Then from Lemma 2.3 applied to  $x_t$  and  $x^*$  together with  $p(x^*) = 0$  and Lemma 2.4 follows  $p(x_t) = 0$ .

This implies that  $x_t$  solves Problem  $(P)$ . Observing that under Assumption (3.4) every point of  $S(0, t(0))$  is a solution of  $(P)$  we then have obtained the following result:

Let  $x^*$  be a solution of  $(P)$ , let  $\lambda \geq 0$  be a multiplier of the convex program

$$\min\{f(x^*, x) : x \in M, g(x) \leq 0\}.$$

Assume that  $\lambda \neq 0$ . Let  $t_* \geq 0$  be a multiplier of the convex program

$$\min\{l(y) : \lambda^T y \geq 0\}.$$

Then  $t_* \leq t(0)$ , and every solution of  $(P_t)$  with  $t_* < t < t(0)$  solves  $(P)$ . If  $0 < t(0) < \infty$  and Assumption (3.4) is satisfied, then every solution of  $(P_{t(0)})$  solves  $(P)$ .

#### References.

- [1] D. P. Bertsekas, Necessary and sufficient conditions for a penalty method to be exact. *Math. Programming* 9 (1975), 251-269.
- [2] J. Gwinner, On the regularization of monotone variational inequalities. *Operations Research Verfahren* 28 (1978), 374-386.
- [3] J. Gwinner, On the penalty method for constrained variational inequalities. In: *Optimization - Theory and Algorithms*, edited by J.-B. Hiriart-Urruty, W. Oettli and J. Stoer, pp. 197-211. Marcel Dekker, New York, 1981.
- [4] S. P. Han and O. L. Mangasarian, Exact penalty functions in nonlinear programming. *Math. Programming* 17 (1979), 251-269.
- [5] L. D. Muu, An augmented penalty function method for solving a class of variational inequalities. *U.S.S.R. Comput. Math. and Math. Phys.* 26 (1986), no. 6, 117-122.
- [6] W. Oettli, Monoton-konvexe Funktionen - eine Bemerkung zum Satz von Browder-Minty. In: *Advances in Mathematical Optimization*, edited by J. Guddat et al. ( *Mathematical Research*, Vol. 45 ), pp. 130-136. Akademie-Verlag, Berlin, 1988.