Fixed Point Theory, Volume 6, No. 1, 2005, 25-32
http://www.math.ubbcluj.ro/~nodeacj/sfptcj.htm

PROXIMAL POINT ALGORITHM WITH GENERALIZED BREGMAN FUNCTIONS

CORINA L. CHIRIAC

"Grigore Moisil" High School, Buzău, Romania
E-mail: corinalchiriac@yahoo.com

Abstract. In this paper we introduce a generalized point method for solving variational inequality problems with monotone operators in a Hilbert space. It differs from the classical proximal point method (as discussed by Rockafellar for the problem of finding zeroes of monotone operators) in the use of generalized distances, called Bregman distances, instead of the Euclidian one. These distances play not only a regularization role but a penalization one, forcing the sequence generated by the method to remain in the interior of the feasible set so that the method becomes an interior point one.

Key Words and Phrases: Bregman distance, generalized Bregman function, generalized proximal point algorithm.

2000 Mathematics Subject Classification: 46T99, 47J25, 49J40.

1. Introduction

The proximal point algorithm (PPA) for the problem of finding zeroes of an operator T in a Hilbert space generates a sequence $\left\{x^k\right\} \subset H$ in the following way: it starts with any $x^0 \in H$ and, given x^k, x^{k+1} is taken so that $0 \in T_k\left(x^{k+1}\right)$ where $T_k(x) = T(x) + \lambda_k(x - x^k)$ and $\lambda_k(0, \lambda)$ for some $\lambda > 0$.

Let H be a Hilbert space with inner product $(\cdot, \cdot)$ and let $T : H \to 2^H$ be maximal monotone operator. Given a closed and convex set $K \subset H$, the variational inequality problem $VIP(T, K)$ consists of finding $\overline{x} \in K$ such that there exists $\overline{u} \in T\overline{x}$ satisfying

---

This paper was presented at International Conference on Nonlinear Operators, Differential Equations and Applications held in Cluj-Napoca (Romania) from August 24 to August 27, 2004.

$$(\overline{u}, x - \overline{x}) \geq 0 \text{ for all } x \in K.$$

It is well known that $VIP(T, K)$ is equivalent to finding a zero of the operator $L = T + N_k$, whereN_k is the normal operator associated to the set K. Therefore, PPA can be used to solve $VIP(T, K)$. In the both iteration, x^{k+1} must be a zero of $T_k + N_k$ with T_k as above. Equivalently, x^{k+1} must be a solution of $VIP(T_k, K)$. So, PPA solves the variational inequality problem by solving a sequence of variational inequality problems with the operators T_k.

Although the operators T_k are better conditioned in principle that T (for example $VIP(T_k, K)$ has a unique solution when $VIP(T, K)$ has solutions), the subproblems are structurally as hard to solve as the original problem. In [5] Burachik proposed a *generalized proximal point algorithm* ($GPPA$) for $VIP(T, K)$ (where K has a nonempty interior) which generates subproblems which are structurally simpler than the original problem.

In $GPPA$, it starts with a strictly convex function $g : K \to R$ whose gradient $\bigtriangledown g$ diverges at the boundary ∂K of K, and with the operators

$$T_k(x) = T(x) + \lambda_k \left(\nabla g(x) - \nabla g \left(x^k \right) \right).$$

Strict convexity of g and divergence of ∇g at ∂K guarantee uniqueness and existence of a zero of T_k, which belongs to the interior $\overset{\circ}{K}$ of K. Such a zero is automatically a solution of $VIP(T_k, K)$, but in that method $GPPA$ it can just take x^{k+1} as the zero of T_k without any explicit consideration of the constraint set K, which has been embedded into g (which is in fact an interior penalty function tailored for the set K). In fact, we end up with genuinely unconstrained subproblems.

Another way of looking at PPA and $GPPA$ is the following: for PPA, the factor $x - x^k$ in the definition of T_k is the gradient of $s_k(x) = \left\| x - x^k \right\|^2$ so that, when $Tx = \partial f(x)$ for some convex f, $T_k x = \partial f_k(x)$ with $f_k(x) = f(x) + \lambda_k \left\| x - x^k \right\|^2$ and $VIP(T, K)$ reduces to $\min f_k(x)$ subject to $x \in K$. In the case of $GPPA$, it considers $D_g : K \times \overset{\circ}{K} \to R$, defined as $D_g(x, y) = g(x) - g(y) - (\nabla g(y), x - y)$. If it defines now $s_k(x) = D_g \left(x, x^k \right)$, it gets, from the definition of D_g, $\partial s_k(x) = \nabla g(x) - \nabla g \left(x^k \right)$ so that we again have $T_k(x) = Tx + \lambda_k \partial s_k(x)$, and in the optimization case ($Tx = \partial f(x)$) a zero of

T_k is just a minimizer of

$$f_k(x) = f(x) + \lambda_k D_g(x, x_k).$$

Such a minimizer is unique and belongs to $\overset{\circ}{K}$ due to the properties of g. In this situation, g is said to be a Bregman function and D_g the Bregman distance associated with g.

The B-functions that generalize Bregman functions has been introduced by Kiwiel and has been used in [8], where a more general approach was developed. In this paper two further assumptions on the function g (continuity in the boundary of K and differentiability in the interior of K) are removed.

Using the papers [5] and [8] we demonstrate that the results (for the variational inequalities) by [5] can also be obtained with functions g as those considered in [8] (it is a conjecture enunciated by Burachik [5]). Our main converge result is the following: Under appropriate assumptions the general algorithm generates a sequence that converges to $x^* \in S$ if and only if $S \neq \emptyset$ (by S an arbitrary B-function, we obtain first a partial convergence result, meaning that the sequence is bounded, the distance between consecutive iterates goes to zero, and every weak cluster point of the sequence is a solution of the problem, and then a full weak convergence result (weak convergence of the whole sequence to a solution of the problem) under a strong hypothesis on B-function.

2. GB-FUNCTIONS

We first recall some useful concepts. The *indicator* function of K is denoted by i_k ($i_k(x) = 0$ if $x \in K$, ∞ otherwise), and its *support* function by

$$i_k^*(\cdot) = \sup_{x \in K}(\cdot, x).$$

For any proper convex function $h : H \to R$, we consider its *difference functions* (see [8])

$$D_h^b(x, y) = h(x) - h(y) - i_{\partial h(y)}^*(x - y) \text{ for all } x, y \in D(h), \tag{1}$$

$$D_h^\sharp(x, y) = h(x) - h(y) + i_{\partial h(y)}^*(y - x) \text{ for all } x, y \in D(h), \tag{2}$$

where ∂h represents the *subdifferential* of h. Its *derivative* in any direction d at a point x where h is finite is denoted by

$$h'(x;d) = \lim_{t\to 0} \frac{[h(x+td) - h(x)]}{t}$$

In [9] was established that

$$h'(x;d) \geq i^*_{\partial h(y)}(d) = \sup\{(g,d) | g \in \partial h(x)\}.$$

From the definition of the support function and (2), we obtain

$$0 \leq D^b_h(x,y) \leq h(x) - h(y) - (\gamma, x-y) \leq D^\sharp_h(x,y), \tag{3}$$

for all $x, y \in D(h)$, $\gamma \in \partial h(y)$. D^b_h and $D^\sharp_h$ generalize the usual Bregman functions from [2], defined by,

$$D_h(x,y) = h(x) - h(y) - (\nabla h(y), x-y)) \text{ for all } x \in D(h), y \in D(\nabla h), \tag{4}$$

since

$$D_h(x,y) = D^b_h(x,y) = D^\sharp_h(x,y) \text{ for all } x \in D(h), y \in D(\nabla h). \tag{5}$$

$D^b_h(x,y)$ and $D^\sharp_h(x,y)$ are used like distances, because for $x \in D(h), y \in D(\partial h)$,

$$\begin{aligned} 0 &\leq D^b_h(x,y) \leq D^\sharp_h(x,y), \text{ and} \\ D^b_h(x,y) &= 0 \Leftrightarrow D^\sharp_h(x,y) = 0 \Leftrightarrow x = y \end{aligned}$$

by strictly convexity.

We shall introduce a new concept starting from the concept of B-function (see Kiwiel [8]).

Definition 1. A closed proper convex function $h : H \to R$ is called *GB-function* (generalized Bregman function) if it satisfies the following conditions:

(B1) h is strictly convex on $D(h)$;

(B2) h is lower semi-continuous on $D(h)$;

(B3) For every $\alpha \in R$ and $x \in D(h)$, the set

$$L^b_h(x,\alpha) = \left\{y \in D(\partial h) | \, D^b_h(x,y) \leq \alpha\right\}$$

is bounded;

(B4) For every $\alpha \in R$ and $x \in D(h)$, if $\{y^k\} \subset L_h^b(x,\alpha)$ is a bounded sequence such that $y^k \rightharpoonup y$, then $\lim_{k\to\infty} D_h^\sharp(y, y^k) = 0$.

(B5) If $\{x^k\} \subset D(h)$ is bounded sequence, $\{y^k\} \subset D(\partial h)$ is such that $y^k \rightharpoonup y$, and $\lim_{k\to\infty} D_h^b(x^k, y^k) = 0$, then $x^k \rightharpoonup y$.

Definition 2. A GB-function h is said to be zone coercive if, in addition to B1-B5 the following condition holds:

(B6) For every $y \in H$, there exists an $x \in K$ such that $y \in \partial h(x)$.

These definitions allow us to state and to study an algorithm for solving VIP-(1). Before that we shall enumerate some results which are useful in our analysis.

3. Pseudo-monotonicity and regularity

Our convergence theorems require a condition on the operator T, namely pseudo-monotonicity, which we will discuss next.

Definition 3. (Browder [4]). Let H be a Hilbert space and $T : H \to 2^H$ such that $D(T)$ is closed and convex. T is said to be *pseudo-monotone* if and only if it satisfies the following condition:

For any sequence $\{x^k\} \subset D(T)$, converging weakly to an element $x^0 \in D(T)$, and any sequence $\{w^k\} \subset H$, with $w^k \in Tx^k$ for all k, such that

$$\limsup_{k\to\infty} \left(w^k, x^k - x^0\right) \le 0,$$

there exists an element $w^0 \in Tx^0$, for each $y \in D(T)$, such that

$$\left(w^0, x^0 - y\right) \le \liminf_{k\to\infty} \left(w^k, x^k - y\right). \blacksquare$$

The next proposition lists several conditions which ensure the pseudo-monotonicity.

Definition 4. A maximal monotone operator $T : H \to 2^H$ is said to be regular if and only if for all $u \in R(T)$ and for all $x \in D(T)$ it follows that

$$\sup_{(y,v)\in G(T)} (v - u, x - y) < \infty.$$

4. STATEMENT OF THE ALGORITHM

Let K be a closed and convex subset of a Hilbert space H, $h : K \to R$ a strictly convex lower semi-continuous function and $\{\lambda_k\}$ a sequence of positive real numbers bounded above by some $\lambda > 0$. We define the generalized proximal point algorithm $GPPA$ by:

(I) *Initialization.* $x^0 \in K$ and $\gamma^0 \in \partial h\left(x^0\right)$

(II) *Iterative step.* Given x^k and $\gamma^k \in \partial h\left(x^k\right)$, let $x^{k+1} \in H$ be such that

$$T\left(x^{k+1}\right) + \lambda_k\left(\partial h\left(x^{k+1}\right) - \gamma^k\right) \ni 0,$$

where $\gamma^k \in \partial h\left(x^k\right)$.

As a consequence, we have to solve the problem:

Find $x^{k+1} \in K$ such that there exist $u^k \in Tx^{k+1}$ and $\gamma^{k+1} \in \partial h\left(x^{k+1}\right)$ satisfying (called generalized rule)

(GR) $\qquad \lambda_k \gamma^k = u^k + \lambda_k \gamma^{k+1}.$

Note that the existence of x^{k+1} satisfying (II) is not immediate at all and will be ensured only under some assumption on T and h. This issue is the content of the next section.

Existence result. For proving that the sequence given by (I)-(II) is well defined and contained in K, we will present the following result. We recall that an operator A is *strictly monotone* if and only if

$$(u - v, x - y) > 0$$

for all $x, y \in H$, $x \neq y$ and $u \in Ax$, $v \in Ay$.

We are now in the position to establish the following

Theorem 5. *Let $\left\{x^k\right\}$ be the sequence given by (10)-(11) and $T : H \to 2^H$ a maximal monotone operator. Suppose that the following conditions hold*:

(i) $D(T) \bigcap D(\partial h) \neq \emptyset$,

(ii) *h is zone coercive (i.e., it satisfies* (B6))

Then $\left\{x^k\right\}$ is well defined and contained in K.

We continue with a convergence result of the sequence generated by (I)-(II).

Theorem 6. *Let $T : H \to 2^H$ be a maximal monotone operator. Consider $VIP(T, K)$ where K is a closed convex subset of H. Let $h : K \to R$ be a GB-function, i.e., it satisfies B1-B5, and suppose that the following conditions hold:*

(i) $D(T)\bigcap D(\partial h)\neq\emptyset$,

(ii) $VIP(T,K)$ *has solutions* (S *is the set of solutions of* $VIP(T,K)$),

(iii) T *is pseudo-monotone*,

(iv) $\lambda_k\in(0,\lambda]$ *for some* $\lambda>0$,

(v) H *is zone coercive, i.e., satisfies* (B6).

Then the sequence $\{x_k\}$ *generated by (10)-(11) satisfies the following*:

(a) *The sequence* $\left\{D_h^k\left(z,x^k\right)\right\}$, *where*

$$D_h^k\left(z,x^k\right):=h(z)-h\left(x^k\right)-\left(\gamma^k,z-x^k\right),\gamma^k\in\partial h\left(x^k\right)$$

is nonincreasing for all $z\in S$,

(b) $\left\{x^k\right\}$ *is bounded and has weak cluster points,*

(c) $\lim_{k\to\infty}D_h^k\left(x^k,x^{k+1}\right)=0$,

(d) *If* $\overline{x}$ *is a weak cluster point of* $\left\{x^k\right\}$, *then there exists* $\overline{u}\in T\overline{x}$ *such that*

$$(\overline{u},z-x)=0,\text{ where }z\in S.$$

The convergence of $\left\{x^k\right\}$ to a solution of $VIP(T,K)$ is obtained, if we shall add some conditions.

Definition 7. (Burachik [5]). The operator $T:H\to 2^H$ is *paramonotone* in a convex set C if it is monotone and $(z-z',w-w')=0$ with $z,z'\in C,w\in Tz,w'\in Tz'$ implies $w\in Tz',w'\in Tz$.

Corollary 8. *Suppose that all the conditions of Theorem 11 hold. Assume furthermore that* T *is paramonotone in* K. *Then the sequence* $\left\{x^k\right\}$ *generated by (10)-(11) is bounded,* $\lim_{k\to\infty}D_h^k\left(x^{k+1},x^k\right)=0$, *and all weak cluster points of* $\left\{x^k\right\}$ *are solutions of* $VIP(T,K)$.

We close this section showing that the existence of solutions of $VIP(T,K)$ is a necessary condition for convergence of the sequence $\left\{x^k\right\}$ generated by (I)-(II). In fact, we will show that if the solution set S of $VIP(T,K)$ is empty, then $\left\{x^k\right\}$ is unbounded. We need a preparatory result.

Lemma 9. *Under the hypotheses of Corollary 8, let* $\left\{x^k\right\}$ *be the sequence generated by (I)-(II). If* $S=\emptyset$ *then* $\left\{x^k\right\}$ *is unbounded.*

We summarize our results in the following theorem.

Theorem 10. *Assume that the hypotheses of Theorem 11 hold. Let* $\left\{x^k\right\}$ *be the sequence generated by (10)-(11). Then*

(i) *If* $S\neq\emptyset$, *then the weak cluster points of* $\left\{x^k\right\}$ *belongs to* S.

(ii) *If* $S=\emptyset$, *then* $\left\{x^k\right\}$ *is unbounded.*

References

[1] Y. Alber, D. Butnariu, *Convergence of Bregman projection methods for solving consistent convex feasibility problems in reflexive Banach spaces*, J. Optim. Theory Appl., **92** (1997), 33-61.

[2] L. M. Bregman, *The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming*, USSR Computational Mathematics and Mathematical Physics, **7** (1967), 200-217.

[3] H. Brezis, A. Haraux, *Image d'une somme d'operateurs monotone et applications*, Israel J. Math., **23** (1976), 165-186.

[4] F. E. Browder, *Nonlinear operators and nonlinear equation of evolution in Banach space*, Proc. Sympos. Pure Math. **18**, American Mathematical Society, Providence, RI, 1976.

[5] R. S. Burachik, A. N. Iusem, *A generalized proximal point algorithm for the variational inequality problem in a Hilbert space*, SIAM J. Optim., **8** (1998), 197-216.

[6] R. S. Burachik, A. N. Iusem, *A generalized proximal point algorithm for the nonlinear complementary problem*, RAIRO Oper. Res., **33** (1999), 447-479.

[7] C. Chiriac, *A numerical analysis of variational inequalities*, Ph. D. Thesis, "Ovidius" University, Constanta 2002.

[8] K. Kiwiel, *Proximal minimization methods with generalized Bregman functions*, SIAM J. Control Optim., **35** (1997), 1142-1168.

[9] B. Martinet, *Regularisation d'inequations variationelles par approximations succesives*, RAIRO Rech. Oper., **4** (1970), 154-158.

[10] D. Pascali, S. Sburlan, *Nonlinear Mappings on Monotone Type*, Editura Academiei, Bucharest, 1978.

[11] D. Pascali, *Methods of the monotonicity in Hilbert spaces*, in Analiza neliniara şi aplicatii in mecanica, Editura Academiei R.S. Romania, Bucuresti, 1977.

[12] R. T. Rockafellar, *On the maximality of sums of nonlinear monotone operators*, Trans. Amer. Math. Soc., **149** (1970), 75-88.

[13] R. T. Rockafellar, *Monotone operators and proximal point algorithm*, SIAM J. Control Optim., **14** (1976), 877-898.

Received October 27, 2004; Revised December 15, 2004.