A strongly convergent Hybrid Proximal method in Banach spaces

Rolando Gárciga Otero* B. F. Svaiter†

July 3, 2003

Abstract

This paper is devoted to the study of strong convergence in inexact proximal like methods for finding zeroes of maximal monotone operators in Banach spaces. Convergence properties of proximal point methods in Banach spaces can be summarized as follows: if the operator have zeroes then the sequence of iterates is bounded and all its weak accumulation points are solutions. Whether or not the whole sequence converges weakly to a solution and which is the relation of the weak limit with the initial iterate are key questions. We present a hybrid proximal Bregman projection method, allowing for inexact solutions of the proximal subproblems, that guarantees strong convergence of the sequence to the closest solution, in the sense of the Bregman distance, to the initial iterate.

1 Introduction

Many problems of applied mathematics reduce to finding zeroes of maximal monotone operators, originated e.g. in optimization, equilibrium or in variational inequalities. The proximal point method [19] is among the main tools for finding zeroes of maximal monotone operators in Hilbert spaces, and is also the departure point for the design and analysis of other algorithms.

In some relevant instances, the operator which zeroes are to be found are defined in Banach spaces. Illustrative examples are elliptic boundary value problems (see e.g. [15]), which have the Sobolev spaces, $W^{m,p}(\Omega)$, as its natural domain of definition. Thus, methods for finding zeros of maximal monotone operators in nonhilbertian spaces are also relevant. Extension of the proximal point method to Banach spaces have received some contributions in the works of [1, 5, 7, 13, 11, 12].

Let $T : B \to \mathcal{P}(B^*)$ denotes a maximal monotone operator from a reflexive real Banach space B to parts of its topological dual B^*. Our main problem is to find zeroes of T:

$$Find\ x \in B\ such\ that\ 0 \in T(x). \qquad (1)$$

*Instituto de Economia da Universidade Federal de Rio de Janeiro, Avenida Pasteur 250, Rio de Janeiro, RJ, Brazil (**rgarciga@ie.ufrj.br**). The work of this author was partially supported by FAPERJ/UFF grant no. E-26/152.130/2001.

†Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Rio de Janeiro, RJ, 22460-320, Brazil (**benar@impa.br**). The work of this author was partially supported by CNPq grant no. 302748/2002-4

The proximal point method, for solving this problem, can be formulated as follows: starting from $x^0 \in B$ it generates a sequences of iterates by taking x^{k+1} as the solution of the *k-th proximal sub-problem*, i.e., the unique $x \in B$ such that

$$0 \in T(x) + \lambda_k[f'(x) - f'(x^k)], \tag{2}$$

where $f : B \to \mathbb{R} \cup \{\infty\}$ is a strictly convex and Gâteaux differentiable function on the interior of its domain satisfying some technical assumptions. f' is the Gâteaux derivative of f and $\{\lambda_k\}_k$ is an exogenous sequence of positive parameters. In [13, 11, 12] the authors present inexact versions of the method. In [13] the error criteria are presented in the spirit of those in [19], which are of absolute type. In [11, 12] the methods allow for a relative error through the use of an additional hybrid step extending the works of [21, 20, 23]. In any case, for the more general situation, convergence results can be resumed to those in [5]. It was proved in [5] that if dom $T \subset$ int(dom f) and T has zeroes then the sequence $\{x^k\}$ is bounded and all its weak cluster points are zeroes of T. Actually, it is also requested in [5] that f' be onto, in order to ensure existence of a solution of (2). Additionally, if f' is sequentially weak-to-weak* continuous, then there exists a unique weak cluster point.

When $f = 1/2\,\|x\|_B^2$ and B is a Hilbert space, (2) reduces to the classical proximal point method in Hilbert spaces, and we have weak convergence to a solution [19]. Note that in this case, f' is the identity function, hence sequentially weak-to-weak continuous.

In a nonhilbertian Banach space, the assumption of f' being sequentially weak-to-weak continuous seems to be too demanding. In fact in [8, 9] there are counterexamples showing that in $B = \ell_p$ or $B = \mathcal{L}^p$ $(1 < p < +\infty)$ the function $f = \|\cdot\|_p^r$ $(r > 1)$ does not satisfy this assumption, excepting in the case just mentioned and also the case $B = \ell_p$, $1 < p < +\infty$, and $f = \|\cdot\|_p^p$ (see, for example, Proposition 8.2 in [3]). Thus, we identify the following main questions concerning proximal like methods in nonhilbertian spaces:

1.) Whether or not the whole sequence converges weakly to a solution ?

2.) Which is the relation of the weak limit with the initial iterate ?

3.) What about strong convergence ?

Even in Hilbert spaces, the sequence generated by the exact proximal point method does not converge strongly in general [10]. This problem has been addressed in [22], where it is presented a hybrid proximal-projection method in Hilbert spaces that guarantees strong convergence of the sequence of iterates to the closest solution to the initial iterate. On Banach spaces we mention the recent works in [2, 14]. In [2] the authors present a general approach for convergence of exact proximal like methods. In [14] there is an analysis allowing for approximations of the operator T, but the regularizing parameters are taking converging to zero. The method in [22], joined with the work in [11], is the starting point for this work, which has as its main objective to answer, at least partially, the questions above. We present a hybrid proximal point Bregman projection method that guarantees convergence of the whole sequence to the closest solution, in the sense of the Bregman distance, to the initial iterate. The convergence is always strong. Moreover, the method allows for ϵ-enlarged inexact solutions satisfying an error criterion of relative type.

2 Preliminaries

From now on, B is a reflexive real Banach space. We will use the notation $\langle v, x\rangle$ for the duality product $v(x)$ of $x \in B$ and $v \in B^*$. Convergence in the weak (respectively strong) topology of a sequence will be indicated by $\stackrel{w}{\rightharpoonup}$ (respectively $\stackrel{s}{\longrightarrow}$).

Let $\mathcal{F}$ be the family of functions $f : B \to \mathbb{R}$, which are strictly convex, lower semi-continuous and G-differentiable. The *Bregman distance* associated to $f \in \mathcal{F}$ is defined as $D_f : B \times B \to \mathbb{R}$,

$$D_f(y, x) = f(y) - f(x) + \langle f'(x), y - x\rangle. \tag{3}$$

From this definition, it is straightforward to verify that D_f satisfies the *three-point equality* (see e.g. [23] or [8]):

$$D_f(y, x) = D_f(z, x) + D_f(y, z) + \langle f'(x) - f'(z), z - y\rangle, \tag{4}$$

for any $x, y, z \in B$. As f is strictly convex, the function $D_f(\cdot, x)$ is non negative, strictly convex and $D_f(y, x) = 0$ if and only if $x = y$ (see e.g. 1.1.9 of [8]).

Given a nonempty closed and convex set $C \subset B$ and any $x \in B$, the *Bregman projection*, associated to $f \in \mathcal{F}$, of x over C, usually denoted by $\Pi^f_C(x)$, is defined as the solution of the convex optimization problem $\min_{y\in C} D_f(y, x)$, i.e.,

$$\Pi^f_C(x) = \arg\min_{y\in C} D_f(y, x). \tag{5}$$

The *modulus of total convexity* of $f \in \mathcal{F}$ is the function $\nu_f : B \times \mathbb{R}_+ \to \mathbb{R}$, defined as

$$\nu_f(x, t) = \inf_{\{y\in B:\|y-x\|=t\}} D_f(y, x). \tag{6}$$

A function $f \in \mathcal{F}$ is *totally convex* if $\nu_f(x, t) > 0$ for all $x \in B$ and $t > 0$. Additionally, if $\inf_{x\in E} \nu_f(x, t) > 0$ for each bounded subset $E \subset B$ then f is called *uniformly totally convex.* If $f \in \mathcal{F}$ is totally convex, then

$$\nu_f(x, st) \geq s\nu_t(x, t), \quad \forall s \geq 1, \quad t \geq 0, \quad x \in B, \tag{7}$$

and the Bregman projection associated to f is well defined (see e.g. [8, 1.2.2]).

The assumptions on $f \in \mathcal{F}$ to be considered in the sequel are the following:

H1: The level sets of $D_f(x, \cdot)$ are bounded for all $x \in B$.

H2: Uniform total convexity of f, or equivalently (see [8], 2.1.2), *sequential consistency*[1]: For all $\{x^k\}$, $\{y^k\} \subset B$ such that $\{x^k\}$ is bounded and $\lim_{k\to\infty} D(y^k, x^k) = 0$, it holds that $x^k - y^k \xrightarrow{s} 0$.

H3: The G-derivative of f, f', is uniformly continuous on bounded subsets of B.

Examples of functions in $\mathcal{F}$ satisfying assumptions H1-H3, and also surjectivity of f', are the powers of the norm, $f_r = \frac{1}{r}||\cdot||^r$, $r > 1$, in any uniformly smooth and uniformly convex Banach space B (see e.g. [11]).

We recall that a point-to-set operator $T : B \to \mathcal{P}(B^*)$ is *monotone* if $\langle w - w', x - x' \rangle \geq 0$ for all $x,\ x' \in B$ and all $w \in T(x),\ w' \in T(x')$. A monotone operator is called *maximal monotone* if its graph $G(T) = \{(x, v) \in B \times B^* \mid v \in T(x)\}$ is not properly contained in the graph of any other monotone operator. Given $\epsilon \geq 0$, the ϵ-*enlargement* of a maximal monotone operator T, introduced in [4, 6], is defined by

$$T^\epsilon(x) = \{u \in B^* \mid \langle v - u, y - x \rangle \geq -\epsilon,\ \forall y \in B,\ u \in T(y)\}, \tag{8}$$

for any $x \in B$. Thus, $T \subset T^\epsilon$, in particular, $T^0 = T$. The graph of $T^\cdot$ is demiclosed [6]: if $v^k \in T^{\epsilon_k}(x^k)$ for all k, with ϵ_k converging to $\bar{\epsilon}$, and v^k converges in the weak* (respectively strong) topology of B^* to $\bar{v}$ and x^k converges in the strong (respectively weak) topology of B to $\bar{x}$, then $v \in T^{\bar{\epsilon}}(\bar{x})$.

Since we shall try proximal like methods for solving problem (1), we then make some comments concerning existence of solutions for the regularized subproblems (2). Denoting by J the normalized duality mapping, then an operator $T : B \to \mathcal{P}(B^*)$ is maximal monotone if, and only if, $T + \lambda J$ is onto for any $\lambda > 0$ (see e.g. [16]). When J is replaced by the G-derivative of a regularizing function $f \in \mathcal{F}$ then $T + \lambda f'$ is onto provided maximal monotonicity of T and surjectivity of $f' : B \to B^*$ (see [5]). We show next that the assumption of surjectivity of f' can be avoided provided existence of solutions for (1).

Lemma 2.1. *Let $T : B \to \mathcal{P}(B^*)$ be maximal monotone. If $T^{-1}(0) \neq \emptyset$ and $f \in \mathcal{F}$ satisfies H1, then for any $\lambda > 0$ and $x \in B$, problem*

$$0 \in T + \lambda[f' - f'(x)]$$

has a (unique) solution.

Proof. Fix $\lambda > 0$ and $x \in B$. To simplify the notation, define $\hat{T} : B \to \mathcal{P}(B^*)$,

$$\hat{T} = T + \lambda[f' - f'(x)].$$

Since f is a proper lower semi-continuous convex function, its subdifferential is maximal monotone [17]. Moreover, $\operatorname{dom} f' = B$. Thus, $\operatorname{dom} T \cap \operatorname{int} \operatorname{dom}(\lambda[f' - f(x)]) = \operatorname{dom} T \neq \emptyset$ and from [18] we conclude that $\hat{T}$ is maximal monotone.

[1]H2 also implies sequential consistency with boundedness of the sequence $\{x^k\}$ replaced by boundedness of $\{y^k\}$. See e.g. [11], Proposition 5.

Recall that B is reflexive. So we can assume, choosing an equivalent norm, that B and its dual B^* are locally uniformly convex (see [25]). Thus we assume that the duality map J is single valued. For any positive integer k, the inclusion $0 \in \hat{T} + (1/k)J$ has solution (see e.g. Theorem 2.11 on page 123 of [16]), which we call x^k. Therefore, for each k there exist a $\hat{v}^k \in B^*$ such that

$$0 = \hat{v}^k + (1/k)Jx^k, \quad \hat{v}^k \in \hat{T}x^k.$$

Hence, there also exist $v^k \in B^*$ satisfying

$$\hat{v}^k = v^k + \lambda[f'(x^k) - f'(x)], \quad v^k \in Tx^k.$$

Note that $v^k + (1/k)Jx^k = \lambda[f'(x) - f'(x^k)]$, for all k. Take now any $\bar{x} \in T^{-1}(0)$. Using the three-point equality (4), monotonicity of T and nonnegativity of D_f we get

$$\begin{aligned} D_f(\bar{x}, x^k) &= D_f(\bar{x}, x) - D_f(x^k, x) + \langle f'(x) - f'(x^k), \bar{x} - x^k \rangle \\ &= D_f(\bar{x}, x) - D_f(x^k, x) + \lambda^{-1} \langle v^k + (1/k)Jx^k, \bar{x} - x^k \rangle \\ &\leq D_f(\bar{x}, x) + (\lambda k)^{-1} \langle Jx^k, \bar{x} - x^k \rangle \\ &= D_f(\bar{x}, x) + (\lambda k)^{-1} [\langle Jx^k, \bar{x} \rangle - \langle Jx^k, x^k \rangle] \end{aligned}$$

Since J is the duality map, $\langle Jp, q \rangle \leq (1/2)||p||^2 + (1/2)||q||^2$ and $\langle Jp, p \rangle = ||p||^2$. Thus,

$$D_f(\bar{x}, x^k) \leq D_f(\bar{x}, x) + (\lambda k)^{-1} \left[(1/2)||\bar{x}||^2 - (1/2)||x^k||^2\right] \leq D_f(\bar{x}, x) + \frac{1}{\lambda}||\bar{x}||^2.$$

Now, in view of H1, the sequence $\{x^k\}_k$ is bounded and so is $\{Jx^k\}_k$. Hence,

$$\lim_{k \to \infty} \left\| \hat{v}^k \right\| = \lim_{k \to \infty} \frac{1}{k} \left\| Jx^k \right\| = 0.$$

As B is reflexive and $\{x^k\}_k$ is bounded, there exist a subsequence $\{x^{k_j}\}_j$ which converges weakly to some $x^\infty \in B$. Since $\hat{v}^{k_j} \in \hat{T}(x^{k_j})$, for all j, and the graph of $\hat{T}$ is demiclosed, it follows $0 \in \hat{T}(x^\infty)$.

Unicity follows from monotonicity of T and strict monotonicity of f'.

3 The algorithm

In this subsection we present the method under consideration. It accepts inexact solutions of the subproblems, with a criterion which allows for $\epsilon-$enlarged solutions satisfying a relative error measure bounded from above. The algorithm requires an exogenous bounded sequence $\{\lambda_k\} \subset \mathbb{R}_{++}$ and an auxiliary totally convex function $f \in \mathcal{F}$. It is defined as follows:

Algorithm I:

1. Choose $x^0 \in B$.

2. Given x^k, choose $\lambda_k > 0$ and find $\epsilon_k \geq 0$, $\tilde{x}^k$ and v^k satisfying

$$\begin{array}{l} v^k \in T^{\epsilon_k}(\tilde{x}^k) \\ e^k = v^k + \lambda_k[f'(\tilde{x}^k) - f'(x^k)], \end{array} \tag{9}$$

and such that

$$\langle e^k, \tilde{x}^k - x^k\rangle + \epsilon_k \leq \lambda_k D_f(\tilde{x}^k, x^k). \tag{10}$$

3. Let

$$x^{k+1} = \arg \min_{x \in H_k \cap W_k} D_f(x, x^0), \tag{11}$$

where

$$H_k = \{x \in B : \langle v^k, x - \tilde{x}^k\rangle \leq \epsilon_k\}, \quad \text{and} \tag{12}$$

$$W_k = \{x \in B : \langle f'(x^0) - f'(x^k), x - x^k\rangle \leq 0\}. \tag{13}$$

Observe that at iteration k, with x^k and λ_k be given, we are trying to solve the $k-th$ proximal sub-problem (2). But in a relaxed form (9)-(10), which allows for a pair $(\tilde{x}^k, v^k)$ in the graph of T^{ϵ_k} (an enlargement of T) and also an error e^k for the inclusion. Anyway, if $\tilde{x}^k$ is the *exact* solution of problem (2), then there exists $v^k \in B^*$ satisfying

$$0 = v^k + \lambda_k[f'(\tilde{x}^k) - f'(x^k)], \quad v^k \in T(\tilde{x}^k).$$

Hence, $\tilde{x}^k$ and v^k satisfies (9)-(10) with $\epsilon_k = 0$ and $e^k = 0$. Thus, in order to show good definition of the algorithm we just need to ensure existence of solutions for the proximal subproblems and nonemptyness of the closed and convex set $H_k \cap W_k$. In fact, as discussed in the previous section, total convexity of f guarantees good definition of the Bregman projection over $H_k \cap W_k$.

Let S denote the set of solutions of the main problem (1), i.e. the zeroes of the maximal monotone operator T, $S = T^{-1}(0)$. Since the case $S \neq \emptyset$ is the interesting one we separate the analysis. We start by settling the issue of good definition of the algorithm.

Proposition 3.1. *Let $f \in \mathcal{F}$ be a totally convex function and assume that at least one of the following conditions holds:*

a) $S \neq \emptyset$ and f satisfies H1, or

b) $f' : B \to B^*$ is surjective.

Then the algorithm is well defined. Moreover, for all k, $S \subset H_k \cap W_k$.

Proof. Observe first that the proximal subproblems (9)–(10) always has exact solution. We mean, for any $x^k \in B$ and $\lambda_k > 0$ there are $(v^k, \tilde{x}^k) \in G(T)$ with $e^k = 0$ ($\epsilon_k = 0$), which obviously satisfy (10). In fact, apply Lemma 2.1 under assumption a) and Lemma 2.10 and Corollary 3.1 of [5], under assumption b) to get solution for the $k-th$ proximal sub-problem (2). Thus, good definition of the algorithm is reduced to the existence of the Bregman projection in (11). Which, in turn is reduced to nonemptyness of the set $H_k \cap W_k$, since f is totally convex (see e.g. [8]). We separate the proof in two cases corresponding to $S \neq \emptyset$ and $S = \emptyset$.

Assume first that $S \neq \emptyset$. Since $v^k \in T^{\epsilon_k}(\tilde{x}^k)$ we have, for any $\bar{x} \in S$, that $\langle v^k, \bar{x} - \tilde{x}^k \rangle \leq \epsilon_k$. Hence, in view of (12), $S \subset H_k$ always. It is enough to prove that $S \subset W_k$, with W_k given by (13). We proceed by induction in k. If $k = 0$ then $W_0 = B$, which obviously contains S. Assume now that $S \subset W_k$ for a given k. Then $S \subset H_k \cap W_k$. It implies that $H_k \cap W_k \neq \emptyset$, hence there exists a unique x^{k+1} defined by (11). Thus, x^{k+1} satisfies the necessary condition

$$0 \in f'(x^{k+1}) - f'(x^0) + N_{H_k \cap W_k}(x^{k+1}),$$

obtaining,

$$\langle f'(x^0) - f'(x^{k+1}), x - x^{k+1} \rangle \leq 0, \qquad \forall x \in H_k \cap W_k. \tag{14}$$

In particular, (14) holds for any $x \in S$. It follows, from (13), that $S \subset W_{k+1}$.

In the second case we have $S = \emptyset$ and we proceed by induction on k also. For $k = 0$ we know that $W_0 = B$ and H_0 contains e.g. $\tilde{x}^0$, thus $W_0 \cap H_0 \neq \emptyset$. Suppose by induction that $H_n \cap W_n \neq \emptyset$ for $n = 0, 1, ..., k$. Choose $z \in D(T)$, $r = \{\max \|\tilde{x}^n - z\| \mid n = 0, 1, ..., k\} + 1$. Define the function $h : B \to \mathbb{R} \cup \{+\infty\}$ putting $h(x) = 0$ for any $x \in B[z, r] = \{x \in B \mid ||x - z|| \leq r\}$ and $h(x) = +\infty$ for any x out of $B[z, r]$. Since h is a lower semi-continuous proper and convex function its subdifferential ∂h is a maximal monotone operator [17]. Since $z \in int(dom\ \partial h)$ we also have maximal monotonicity of the sum $T' = T + \partial h$ [18]. Note that $T'(x) = T(x)$ for any $x \in B(z, r)$ and, using also [24, Corollary 7.3], we get $T^\epsilon + \partial h \subseteq (T')^\epsilon$. Hence

$$T^{\epsilon_n}(\tilde{x}^n) \subseteq (T')^{\epsilon_n}(\tilde{x}^n) \qquad \text{and} \qquad v^n \in (T')^{\epsilon_n}(\tilde{x}^n) \qquad n = 0, 1, \ldots, k.$$

Consequently, x^n, $\tilde{x}^n$, v^n also satisfy the conditions of the algorithm applied to the problem of finding zeroes of the maximal monotone operator T'. Calling S' the set of solutions of this problem we get that $S' \neq \emptyset$. In fact, dom (T') is contained in $B[z, r]$, thus bounded. It follows that T' has zeroes (see e.g.[3]). Then, the discussed case ensures that x^{k+1} is well defined and $S' \subset H_{k+1} \cap W_{k+1}$.

4 Convergence analysis

We establish next some general properties of the iterates generated by the algorithm, which hold regardless of whether or not the solution set of problem (1), S, is empty. We recall that Proposition 3.1 gives sufficient conditions for the existence of such iterates.

Lemma 4.1. *Let $f \in \mathcal{F}$, $x^0 \in B$ and W_k be defined as in* (13). *Suppose that the algorithm, starting from x^0, reaches iteration k, then* *a) For any $w \in W_k$ it holds*

$$D_f(w, x^0) \geq D_f(w, x^k) + D_f(x^k, x^0). \tag{15}$$

b) x^k is the Bregman projection, associated to f, of x^0 over W_k, i.e.

$$x^k = \Pi^f_{W_k}(x^0) = \arg\min_{x \in W_k} D_f(x, x^0). \tag{16}$$

c) If the algorithm reaches iteration $k+1$ also, then

$$D_f(x^{k+1}, x^k) + \langle \lambda_k^{-1} e_k, x^k - x^{k+1} \rangle \geq D_f(x^{k+1}, \tilde{x}^k). \tag{17}$$

Proof. To prove item a) take any $w \in W_k$. From (13),

$$\langle f'(x^0) - f'(x^k), w - x^k \rangle \leq 0.$$

Using also the three-point property, (4), it follows that,

$$\begin{aligned} D_f(w, x^0) &= D_f(w, x^k) + D_f(x^k, x^0) + \langle f'(x^0) - f'(x^k), x^k - w \rangle \\ &\geq D_f(w, x^k) + D_f(x^k, x^0), \end{aligned}$$

which proves item a).

Item b) follows from *a)* and nonnegativity and strict convexity of $D_f(\cdot, x^k)$. Just note that, in view of (13), $x^k \in W_k$.

Assume now that x^{k+1} is well defined by equation (11). By the three-point property, (4), and (9) we have

$$\begin{aligned} D_f(x^{k+1}, x^k) - D_f(x^{k+1}, \tilde{x}^k) &= D_f(\tilde{x}^k, x^k) + \langle f'(x^k) - f'(\tilde{x}^k), \tilde{x}^k - x^{k+1} \rangle \\ &= D_f(\tilde{x}^k, x^k) + \lambda_k^{-1} [\langle v^k, \tilde{x}^k - x^{k+1} \rangle - \langle e^k, \tilde{x}^k - x^{k+1} \rangle] \\ &\geq D_f(\tilde{x}^k, x^k) + \lambda_k^{-1} \left[-\epsilon_k - \langle e^k, \tilde{x}^k - x^k \rangle + \langle e^k, x^{k+1} - x^k \rangle \right] \\ &\geq \lambda_k^{-1} \langle e^k, x^{k+1} - x^k \rangle. \end{aligned}$$

Here we used, in the first inequality, that $x^{k+1} \in H_k$ and in the last inequality the error criterion (10).

The next proposition resumes the global behavior of the algorithm.

Proposition 4.2. *Let $f \in \mathcal{F}$ satisfying assumptions H2 and H3. Suppose that $\lambda_k \leq \overline{\lambda}$ for all k and some $\overline{\lambda}$ and assume that the algorithm generates an infinite sequence $\{x^k\}$ with ϵ_k converging to zero and $\lambda_k^{-1} e^k \xrightarrow{s} 0$. Then either $\{D_f(x^k, x^0)\}$ converges, $\{x^k\}$ is bounded and each of its weak accumulation points belongs to $S \neq \emptyset$, or $S = \emptyset$, $\{x^k\}$ is unbounded and $\lim_k D_f(x^k, x^0) = +\infty$.*

Proof. ¿From (11) we know that for any k, $x^{k+1} \subset H_k \cap W_k \subset W_k$. Hence, Lemma 4.1-(a) gives us

$$D_f(x^{k+1}, x^0) \geq D_f(x^{k+1}, x^k) + D_f(x^k, x^0) \geq D_f(x^k, x^0).$$

Thus, the sequence $\{D_f(x^k, x^0)\}_k$ is nondecreasing and

$$\sum_{k=0}^{n} D_f(x^{k+1}, x^k) \leq D_f(x^{n+1}, x^0) - D_f(x^0, x^0) = D_f(x^{n+1}, x^0). \tag{18}$$

Assume first that $\{Df(x^k, x^0)\}$ is bounded, thus convergent. Then the sum in (18) converges. Consequently $\lim_k D_f(x^{k+1}, x^k) = 0$, which in turn implies that $x^{k+1} - x^k \xrightarrow{s} 0$ (see H2). Since $\{\lambda_k^{-1} e^k\}$ is bounded, Lemma 4.1-(c) ensures that $\lim_k D_f(x^{k+1}, \tilde{x}^k) = 0$ also. Then, $x^{k+1} - \tilde{x}^k \xrightarrow{s} 0$ and $\tilde{x}^k - x^k \xrightarrow{s} 0$. Observe also that in this case $\{x^k\}$ is necessarily bounded. In fact, if the sequence $\{x^k\}_k$ is unbounded then, there is some subsequence $\{x^{j_k}\}$ such that $\lim_{k\to\infty} ||x^{j_k} - x^0|| = +\infty$, and

$$D_f(x^{j_k}, x^0) \geq \nu_f(x^0, ||x^{j_k} - x^0||) \geq ||x^{j_k} - x^0|| \nu_f(x^0, 1) \quad \text{for k big enough.} \tag{19}$$

Here we used the property, described in equation (7), of totally convex functions . From (19) we get that $\lim_{k\to\infty} D_f(x^{j_k}, x^0) = +\infty$. A contradiction.

Combining now (9), $e^k \xrightarrow{s} 0$ and property H3 we get that $v^k \xrightarrow{s} 0$. Taking any weak limit x^∞ of the bounded sequence $\{\tilde{x}^k\}$ we find: $\tilde{x}^{j_k} \xrightarrow{w} x^\infty$, $v^{j_k} \in T^{\epsilon_{j_k}}(\tilde{x}^{j_k})$, $v^{j_k} \xrightarrow{s} 0$ and $\lim_{k\to\infty} \epsilon_{j_k} = 0$. Then, $0 \in T^0(x^\infty) = T(x^\infty)$ in view of demiclosedness of $T^{()}$. In particular $S \neq \emptyset$.

Let us suppose now that $S = \emptyset$. Then, by the preceding assertion, $\lim_{k\to\infty} D_f(x^k, x^0) = +\infty$. Since f has full domain then $\lim_{k\to\infty} D_f(x^k, x^0) = +\infty$ also implies $\{x^k\}$ unbounded in view of H3, because in such situation $D_f(\cdot, x^0)$ is bounded on bounded subsets of B (see [11], Proposition 4). Thus boundedness of $\{D_f(x^k, x^0)\}$ and $\{x^k\}$ are equivalent.

4.1 Strong Convergence

We are now in conditions to resume the main properties of the algorithm for the case of interest: when the operator has zeroes. Essentially, the algorithm generates a strongly convergent sequence to the solution of (1), which is closest to the initial iterate in the Bregman distance sense.

Theorem 4.3. *Assume that $S \neq \emptyset$. Let $f \in \mathcal{F}$ be a regularizing function satisfying assumptions H1, H2 and H3 and suppose that $\lambda_k \leq \overline{\lambda}$ for all k and some $\overline{\lambda}$. Then, the algorithm starting from any $x^0 \in B$ generates an infinite sequence $\{x^k\}$. Moreover, if $(\lambda_k^{-1} e^k, \epsilon_k) \xrightarrow{s} 0$ then $\{D_f(x^k, x^0)\}$ converges to $\inf_{z\in S} D_f(z, x^0)$ and $\{x^k\}$ converges strongly to $\hat{x} = \Pi_S^f(x^0) = \arg\min_{z\in S} D_f(z, x^0)$.*

Proof. Note that the Bregman projection of the initial iterate x^0 over S, $\hat{x} = \Pi_S^f(x^0)$, exists because the solution set is closed, convex and we assumed it to be nonempty and f is totally convex. From

(11) we know that $D_f(x^{k+1},x^0) \le D_f(x,x^0)$ for all $x \in S \subset H_k \cap W_k$ and, particularly, for $\hat{x}$. Since $D_f(x^{k+1},x^0) \ge D_f(x^k,x^0)$ (see Lemma 4.1-(a)), it holds

$$D_f(x^k,x^0) \le D_f(\hat{x},x^0). \tag{20}$$

Then, $\{D_f(x^k,x^0)\}$ converges and $\{x^k\}$ is bounded . Let

$$\alpha = \lim_{k\to\infty} D_f(x^k,x^0) = \sup_k D_f(x^k,x^0), \tag{21}$$

and choose any weakly convergent subsequence $x^{j_k} \xrightarrow{w} x^\infty$. Then, from Proposition 4.2, $x^\infty \in S$. Consequently,

$$D_f(\hat{x},x^0) \le D_f(x^\infty,x^0) \le \liminf_k D_f(x^{j_k},x^0) = \alpha, \tag{22}$$

where the last inequality follows from the lower semi-continuity of an $f \in \mathcal{F}$. Thus, of $D_f(\cdot,x^0)$. From equations (20)-(22) we get $\alpha = D_f(\hat{x},x^0) = D_f(x^\infty,x^0)$. Consequently $x^\infty = \hat{x}$, meaning that $\{x^k\}_k$ has a unique weak accumulation point and converges weakly to $\hat{x}$. Moreover, from equation (15) in Lemma 4.1-(a), with $w = \hat{x} \in W_k$, and taking limits, it follows

$$\limsup_k D_f(\hat{x},x^k) \le \limsup_k [D_f(\hat{x},x^0) - D_f(x^k,x^0)] = 0.$$

Thus, $\lim_k D_f(x^k,\hat{x}) = 0$. Then, property H2 ensures $x^k \xrightarrow{s} \hat{x}$, i.e. the convergence is strong.

Corollary 4.4. *Let $f \in \mathcal{F}$ be a regularizing function satisfying assumptions H1, H2 and H3 and suppose that $\lambda_k \le \overline{\lambda}$ for all k and some $\overline{\lambda}$. Assume that $S \neq \emptyset$ and that for all k we choose the error criterion*

$$\|e^k\|_* \|x^k - \tilde{x}^k\| + \epsilon_k \le \lambda_k D_f(\tilde{x}^k,x^k) \tag{23}$$

instead of (10) *with the the additional assumption that $e^k = 0$ when $\tilde{x}^k = x^k$* [2]. *Then the algorithm remains well defined. Moreover, if $\{\lambda_k^{-1}e^k\}$ is bounded then $(\lambda_k^{-1}e^k,\epsilon_k) \xrightarrow{s} 0$ and $\{x^k\}$ converges strongly to $\hat{x} = \Pi_S^f(x^0) = \arg\min_{z\in S} D_f(z,x^0)$ and $\{D_f(x^k,x^0)\}$ converges to $D_f(\hat{x},x^0)$.*

Proof. Good definition of the algorithm follows from Proposition 3.1-(a). Concerning the global behavior of the method just note that this error criterion is more demanding that the error in (10). Since $S \neq \emptyset$ then the argument used in the proof of Theorem (4.3) until (20) ensures that $D_f(x^k,x^0)$ converges and $\{x^k\}$ is bounded. Then the sum in (18) converges and $\lim_k D_f(x^{k+1},x^k) = 0$, which in turn implies that $x^{k+1} - x^k \xrightarrow{s} 0$ (see H2). Since $\{\lambda_k^{-1}e^k\}$ is bounded Lemma 4.1-(c) ensures

[2] Assumption H3 implies $\lim_{y\to x} \frac{D_f(y,x)}{||y-x||} = 0$. Then, fixed x, the function defined by $\varphi_x(y) = \frac{D_f(y,x)}{||y-x||}$, when $y \neq x$ and $\varphi_x(x) = 0$ is continuous at x. Moreover, from (23), we have $\|e^k\|_* \le \lambda_k \varphi_{x^k}(\tilde{x}^k)$. Hence, e^k should be zero when $\tilde{x}^k = x^k$.

that $\lim_k D_f(x^{k+1}, \tilde{x}^k) = 0$ also. Then, $x^{k+1} - \tilde{x}^k \xrightarrow{s} 0$ and $\tilde{x}^k - x^k \xrightarrow{s} 0$. Hence ϵ_k converges to zero. Combining now (23) and property H3 we get

$$\lim_{k\to\infty} \|\lambda_k^{-1} e^k\|_* \leq \lim_{k\to\infty} \frac{D_f(\tilde{x}^k, x^k)}{||x^k - \tilde{x}^k||} = 0.$$

Thus, $(\lambda_k^{-1} e^k, \epsilon_k) \xrightarrow{s} 0$ and we can apply Proposition 4.2 and Theorem 4.3.

4.2 The case of no solutions: $S = \emptyset$

We resume next the properties of the algorithm when the operator has not zeroes.

Theorem 4.5. *Assume that $S = \emptyset$. Let $f \in \mathcal{F}$ be a regularizing function satisfying assumptions H2 and H3 with surjective derivative. Suppose that $\lambda_k \leq \overline{\lambda}$ for all k and some $\overline{\lambda}$. Then, the algorithm starting from any $x^0 \in B$ generates an infinite sequence $\{x^k\}$. If $(\lambda_k^{-1} e^k, \epsilon_k) \xrightarrow{s} 0$ then $\lim_k D_f(x^k, x^0) = +\infty$ and $\{x^k\}$ is unbounded.*

Proof. Since f' is surjective Proposition 3.1-(b) ensures good definition of the algorithm. The second part of the statement follows from $S = \emptyset$ and Proposition 4.2.

References

[1] Alber, Y. I., Burachik, R. S., and Iusem, A. N. A proximal point method for nonsmooth convex optimization problems in Banach spaces. *Abstr. Appl. Anal. 2*, 1-2 (1997), 97–120.

[2] Baushke, H., and Combettes, P. Construction of best Bregman approximations in reflexive Banach spaces. Accepted for the Proceedings of the AMS.

[3] Browder, F. E. Nonlinear operators and nonlinear equations of evolution in Banach spaces. In *Nonlinear functional analysis (Proc. Sympos. Pure Math., Vol. XVIII, Part 2, Chicago, Ill., 1968)*. Amer. Math. Soc., Providence, R. I., 1976, pp. 1–308.

[4] Burachik, R. S., Iusem, A. N., and Svaiter, B. F. Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Anal. 5*, 2 (1997), 159–180.

[5] Burachik, R. S., and Scheimberg, S. A proximal point algorithm for the variational inequality problem in Banach spaces. *SIAM J. Control Optim. 39*, 5 (2001), 1633–1649.

[6] Burachik, R. S., and Svaiter, B. F. ϵ-enlargements of maximal monotone operators in Banach spaces. *Set-Valued Anal. 7*, 2 (1999), 117–132.

[7] Butnariu, D., and Iusem, A. N. On a proximal point method for convex optimization in Banach spaces. *Numer. Funct. Anal. Optim. 18*, 7-8 (1997), 723–744.

[8] Butnariu, D., and Iusem, A. N. *Totally convex functions for fixed points computation and infinite dimensional optimization.* Kluwer Academic Publishers, Dordrecht, 2000.

[9] Gárciga Otero, R. *Inexact versions of proximal point and cone-constrained augmented Lagrangians in Banach spaces.* PhD thesis, IMPA, Rio de Janeiro, November 2001.

[10] Güler, O. On the convergence of the proximal point algorithm for convex minimization. *SIAM J. Control Optim. 29*, 2 (1991), 403–419.

[11] Iusem, A. N., and Gárciga Otero, R. Inexact versions of proximal point and augmented Lagrangian algorithms in Banach spaces. *Numer. Funct. Anal. Optim. 22*, 5-6 (2001), 609–640.

[12] Iusem, A. N., and Gárciga Otero, R. Augmented Lagrangian methods for cone-constrained convex optimization in Banach spaces. *J. Nonlinear Convex Anal. 3*, 2 (2002).

[13] Kassay, G. The proximal points algorithm for reflexive Banach spaces. *Studia Univ. Babeş-Bolyai Math. 30* (1985), 9–17.

[14] Kassay, G., and Butnariu, D. On a regularization method for ill posed inclusions and variational inequalities. Talk of Dan Butnariu at the IV Brazilian workshop on Continuous Optimization. Rio, 2002.

[15] Mosco, U. Perturbation of variational inequalities. In *Nonlinear Functional Analysis (Proc. Sympos. Pure Math., Vol. XVIII, Part 1, Chicago, Ill., 1968).* Amer. Math. Soc., Providence, R.I., 1970, pp. 182–194.

[16] Pascali, D., and Sburlan, S. *Nonlinear mappings of monotone type.* Martinus Nijhoff Publishers, The Hague, 1978.

[17] Rockafellar, R. T. On the maximal monotonicity of subdifferential mappings. *Pacific J. Math. 33* (1970), 209–216.

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

[19] Rockafellar, R. T. Monotone operators and the proximal point algorithm. *SIAM J. Control Optim. 14*, 5 (1976), 877–898.

[20] Solodov, M. V., and Svaiter, B. F. A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Anal. 7*, 4 (1999), 323–345.

[21] Solodov, M. V., and Svaiter, B. F. A hybrid projection-proximal point algorithm. *J. Convex Anal. 6*, 1 (1999), 59–70.

[22] Solodov, M. V., and Svaiter, B. F. Forcing strong convergence of proximal point iterations in a Hilbert space. *Math. Program. 87*, 1, Ser. A (2000), 189–202.

[23] Solodov, M. V., and Svaiter, B. F. An inexact hybrid generalized proximal point algorithms and some new results on the theory of Bregman functions. *Math. Oper. Res. 51* (2000), 479–494.

[24] Svaiter, B. F. A family of enlargements of maximal monotone operators. *Set-Valued Anal. 8*, 4 (2000), 311–328.

[25] Troyanski, S. L. On locally uniformly convex and differentiable norms in certain non-separable Banach spaces. *Studia Math. 37* (1970/71), 173–180.