Journal of Nonlinear and Convex Analysis
Volume 10, Number 3, 2009, 471–485

Yokohama Publishers
ISSN 1880-5221 ONLINE JOURNAL

A STRONG CONVERGENCE THEOREM FOR A PROXIMAL-TYPE ALGORITHM IN REFLEXIVE BANACH SPACES

SIMEON REICH AND SHOHAM SABACH

Abstract. We establish a strong convergence theorem for a proximal-type algorithm which approximates (common) zeroes of maximal monotone operators in reflexive Banach spaces. This algorithm employs a well-chosen convex function. The behavior of the algorithm in the presence of computational errors and in the case of zero free operators is also analyzed. Finally, we mention several corollaries, variations and applications.

1. Introduction

In this paper X denotes a real reflexive Banach space with norm $\|\cdot\|$ and X^* stands for the (topological) dual of X endowed with the induced norm $\|\cdot\|_*$. We denote the value of the functional $\xi \in X^*$ at $x \in X$ by $\langle \xi, x \rangle$. An operator $A : X \to 2^{X^*}$ is said to be *monotone* if for any $x, y \in \operatorname{dom} A$, we have

$$\xi \in Ax \text{ and } \eta \in Ay \quad \Longrightarrow \quad \langle \xi - \eta, x - y \rangle \geq 0.$$

(Recall that the set $\operatorname{dom} A = \{x \in X \ : \ Ax \neq \varnothing\}$ is called the *effective domain* of such an operator A.) A monotone operator A is said to be *maximal* if graph A, the graph of A, is not a proper subset of the graph of any other monotone operator. In this paper $f : X \to (-\infty, +\infty]$ is always a proper, lower semicontinuous and convex function, and $f^* : X^* \to (-\infty, +\infty]$ is the Fenchel conjugate of f. The set of nonnegative integers will be denoted by $\mathbb{N}$.

The problem of finding an element $x \in X$ such that $0^* \in Ax$ is very important in Optimization Theory and related fields. For example, if A is the subdifferential ∂f of f, then A is a maximal monotone operator and the equation $0^* \in \partial f(x)$ is equivalent to the problem of minimizing f over X. One of the methods for solving this problem in Hilbert space is the well-known proximal point algorithm. Let H be a Hilbert space and let I denote the identity operator on H. The proximal point algorithm generates, for any starting point $x_0 = x \in H$, a sequence $\{x_n\}_{n \in \mathbb{N}}$ in H by the rule

$$x_{n+1} = (I + \lambda_n A)^{-1} x_n, \quad n = 0, 1, 2, \ldots, \tag{1.1}$$

2000 *Mathematics Subject Classification.* Primary 47H05; Secondary 47J25.

Key words and phrases. Banach space, Bregman projection, Legendre function, maximal monotone operator, monotone operator, proximal point algorithm, resolvent, totally convex function.

The first author was partially supported by the Israel Science Foundation (Grant 647/07), by the Fund for the Promotion of Research at the Technion, and by the Technion President's Research Fund.

where $\{\lambda_n\}_{n\in\mathbb{N}}$ is a given sequence of positive real numbers. Note that (1.1) is equivalent to

$$0 \in Ax_{n+1} + \frac{1}{\lambda_n}\left(x_{n+1} - x_n\right), \quad n = 0, 1, 2, \ldots.$$

This algorithm was first introduced by Martinet [26] and further developed by Rockafellar [34], who proves that the sequence generated by (1.1) converges weakly to an element of $A^{-1}(0)$ when $A^{-1}(0)$ is nonempty and $\liminf_{n\to+\infty}\lambda_n > 0$. Furthermore, Rockafellar [34] asks if the sequence generated by (1.1) converges strongly. This question was answered in the negative by Güler [22], who presented an example of a subdifferential for which the sequence generated by (1.1) converges weakly but not strongly; see [6] for a more recent and simpler example. More recently, Solodov and Svaiter [37] have modified the proximal point algorithm in order to generate a strongly convergent sequence. They introduce the following algorithm:

$$\begin{cases} x_0 \in H, \\ 0 = v_n + \frac{1}{\lambda_n}\left(y_n - x_n\right), \qquad v_n \in Ay_n, \\ H_n = \left\{z \in H : \langle v_n, z - y_n\rangle \leq 0\right\}, \\ W_n = \left\{z \in H : \langle x_0 - x_n, z - x_n\rangle \leq 0\right\}, \\ x_{n+1} = \mathrm{P}_{H_n\cap W_n}(x_0), \qquad n = 0, 1, 2, \ldots. \end{cases} \tag{1.2}$$

Here, for each $x \in H$ and each nonempty, closed and convex subset C of H, the mapping P_C is defined by $\|x - \mathrm{P}_C x\| = \inf\{\|x - z\| : z \in C\}$. This mapping is called the *metric projection* of H onto C. They prove that if $A^{-1}(0)$ is nonempty and $\liminf_{n\to+\infty}\lambda_n > 0$, then the sequence generated by (1.2) converges strongly to $\mathrm{P}_{A^{-1}(0)}$. Kamimura and Takahashi [25] generalize this result to those Banach spaces X which are both uniformly convex and uniformly smooth. They introduce the following algorithm [21]:

$$\begin{cases} x_0 \in X, \\ 0^* = v_n + \frac{1}{\lambda_n}\left(Jy_n - Jx_n\right), \qquad v_n \in Ay_n, \\ H_n = \left\{z \in X : \langle v_n, z - y_n\rangle \leq 0\right\}, \\ W_n = \left\{z \in X : \langle Jx_0 - Jx_n, z - x_n\rangle \leq 0\right\}, \\ x_{n+1} = \mathrm{Q}_{H_n\cap W_n}(x_0), \qquad n = 0, 1, 2, \ldots, \end{cases} \tag{1.3}$$

where J is the normalized duality mapping of the space X. Here, for each nonempty, closed and convex subset C of X, Q_C is a certain generalization of the metric projection P_C in H. They prove that if $A^{-1}(0^*)$ is nonempty and $\liminf_{n\to+\infty}\lambda_n > 0$, then the sequence generated by (1.3) converges strongly to $\mathrm{Q}_{A^{-1}(0^*)}$. Other developments regarding the proximal point algorithm can be found, for example, in [4, 6, 9, 10, 14, 16, 18, 20, 23, 24, 27, 29, 30, 31, 35, 38].

In the present paper we study an extension of algorithms (1.2) and (1.3) to all reflexive Banach spaces using a well-chosen convex function f. More precisely, we consider the following algorithm introduced by Gárciga Otero and Svaiter [21]:

$$
(1.4)\qquad \begin{cases} x_0 \in X, \\ 0^* = \xi_n + \frac{1}{\lambda_n}\left(\nabla f(y_n) - \nabla f(x_n)\right), \qquad \xi_n \in Ay_n, \\ H_n = \left\{z \in X : \langle \xi_n, z - y_n\rangle \le 0\right\}, \\ W_n = \left\{z \in X : \langle \nabla f(x_0) - \nabla f(x_n), z - x_n\rangle \le 0\right\}, \\ x_{n+1} = \mathrm{proj}^f_{H_n \cap W_n}(x_0), \qquad n = 0, 1, 2, \ldots, \end{cases}
$$

where $\{\lambda_n\}_{n\in\mathbb{N}}$ is a given sequence of positive real numbers, ∇f is the gradient of f and proj^f_C is the Bregman projection (see Section 2.4) of X onto C induced by f. Algorithm (1.4) is more flexible than (1.3) because it leaves us the freedom of fitting the function f to the nature of the operator A (especially when A is the subdifferential of some function) and of the space X in ways which make the application of (1.4) simpler than that of (1.3). It should be observed that if X is a Hilbert space H, then using in (1.4) the function $f(x) = (1/2)\|x\|^2$, one obtains exactly algorithm (1.2). If X is not a Hilbert space, but still a uniformly convex and uniformly smooth Banach space X, then setting $f(x) = (1/2)\|x\|^2$ in (1.4), one obtains exactly (1.3). We also note that the choice $f(x) = (1/2)\|x\|^2$ in some Banach spaces may make the computations in algorithm (1.3) quite difficult. These computations can be simplified by an appropriate choice of f. For instance, if $X = \ell^p$ or $X = L^p$ with $p \in (1, +\infty)$, and $f(x) = (1/p)\|x\|^p$ in (1.4), then the computations become simpler than those required in (1.3), which corresponds to $f(x) = (1/2)\|x\|^2$. We propose an extension of algorithm (1.4) (see algorithm (3.1)) which approximates a common zero of several maximal monotone operators and which allows computational errors. Our main result (Theorem 3.1) is formulated and proved in Section 3. The next section is devoted to several preliminary definitions and results. The behavior of the algorithm when the operator A is zero free is analyzed in Section 4 (see Theorem 4.2). The fifth section contains three corollaries of Theorems 3.1 and 4.2. In the sixth and last section we present an application of Theorems 3.1 and 4.2.

2. Preliminaries

2.1. **Some facts about Legendre functions.** Legendre functions mapping a general Banach space X into $(-\infty, +\infty]$ are defined in [3]. According to [3, Theorems 5.4 and 5.6], since X reflexive, the function f is Legendre if and only if it satisfies the following two conditions:

(L1) The interior of the domain of f, $\operatorname{int}\operatorname{dom} f$, is nonempty, f is Gâteaux differentiable (see below) on $\operatorname{int}\operatorname{dom} f$ and

$$\operatorname{dom}\nabla f = \operatorname{int}\operatorname{dom} f;$$

(L2) The interior of the domain of f^*, $\operatorname{int}\operatorname{dom} f^*$, is nonempty, f^* is Gâteaux differentiable on $\operatorname{int}\operatorname{dom} f^*$ and

$$\operatorname{dom}\nabla f^* = \operatorname{int}\operatorname{dom} f^*.$$

Since X is reflexive, we always have $(\partial f)^{-1} = \partial f^*$ (see [7, p. 83]). This fact, when combined with conditions (L1) and (L2), implies the following equalities:

$$\nabla f = (\nabla f^*)^{-1},$$

$$\operatorname{ran} \nabla f = \operatorname{dom} \nabla f^* = \operatorname{int} \operatorname{dom} f^*$$

and

$$\operatorname{ran} \nabla f^* = \operatorname{dom} \nabla f = \operatorname{int} \operatorname{dom} f.$$

Also, conditions (L1) and (L2), in conjunction with [3, Theorem 5.4], imply that the functions f and f^* are strictly convex on the interior of their respective domains.

Several interesting examples of Legendre functions are presented in [2] and [3]. Among them are the functions $\frac{1}{s}\|\cdot\|^s$ with $s \in (1, \infty)$, where the Banach space X is smooth and strictly convex and, in particular, a Hilbert space.

The function f is called *cofinite* if $\operatorname{dom} f^* = X^*$.

2.2. **A property of gradients.** For any convex function $f : X \to (-\infty, +\infty]$ we denote by $\operatorname{dom} f$ the set $\{x \in X : f(x) < +\infty\}$. For any $x \in \operatorname{dom} f$ and $y \in X$, we denote by $f^\circ(x, y)$ the *right-hand derivative of f at x in the direction y*, that is,

$$f^\circ(x, y) := \lim_{t \searrow 0} \frac{f(x + ty) - f(x)}{t}.$$

The function f is said to be *Gâteaux differentiable at x* if $\lim_{t\to 0} (f(x + ty) - f(x))/t$ exists for any y. The function f is said to be *Fréchet differentiable at x* if this limit is attained uniformly in $\|y\| = 1$. Finally, f is said to be *uniformly Fréchet differentiable on a subset E of X* if the limit is attained uniformly for $x \in E$ and $\|y\| = 1$. We will need the following result.

Proposition 2.1. *If $f : X \to \mathbb{R}$ is uniformly Fréchet differentiable and bounded on bounded subsets of X, then ∇f is uniformly continuous on bounded subsets of X from the strong topology of X to the strong topology of X^*.*

Proof. If this result were not true, there would be bounded sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{y_n\}_{n\in\mathbb{N}}$, and a positive number ε such that $\|x_n - y_n\| \to 0$ and $\langle \nabla f(x_n) - \nabla f(y_n), w_n\rangle \geq 2\varepsilon$, where $\{w_n\}_{n\in\mathbb{N}}$ is a sequence in X with $\|w_n\| = 1$ for each $n \in \mathbb{N}$. Since f is uniformly Fréchet differentiable, there is a positive number δ such that

$$f(y_n + tw_n) - f(y_n) - t\langle \nabla f(y_n), w_n\rangle \leq \varepsilon t$$

for all $0 < t < \delta$ and $n \in \mathbb{N}$. We also have

$$\langle \nabla f(x_n), (y_n + tw_n) - x_n\rangle \leq f(y_n + tw_n) - f(x_n), \quad n \in \mathbb{N}.$$

In other words,

$$t\langle \nabla f(x_n), w_n\rangle \leq f(y_n + tw_n) - f(y_n) + \langle \nabla f(x_n), x_n - y_n\rangle + f(y_n) - f(x_n).$$

Hence

$$\begin{aligned} 2\varepsilon t \leq t\langle \nabla f(x_n) - \nabla f(y_n), w_n\rangle &\leq [f(y_n + tw_n) - f(y_n) - t\langle \nabla f(y_n), w_n\rangle] \\ &\quad + \langle \nabla f(x_n), x_n - y_n\rangle + f(y_n) - f(x_n) \\ &\leq \varepsilon t + \langle \nabla f(x_n), x_n - y_n\rangle + f(y_n) - f(x_n). \end{aligned}$$

Since ∇f is bounded on bounded subsets of X (see [12, Proposition 1.1.11, p. 17]), it follows that $\langle \nabla f(x_n), x_n - y_n \rangle$ converges to zero, while $f(y_n) - f(x_n) \to 0$ since f is uniformly continuous on bounded subsets (see [1, Theorem 1.8, p. 13]). But this would yield that $2\varepsilon t \leq \varepsilon t$, a contradiction. □

2.3. **Some facts about totally convex functions.** Let $f : X \to (-\infty, +\infty]$ be convex. The function $D_f : \operatorname{dom} f \times \operatorname{int} \operatorname{dom} f \to [0, +\infty]$ defined by

$$D_f(y, x) := f(y) - f(x) - f^\circ(x, y - x),$$

is called the *Bregman distance with respect to* f (*cf.* [17]). If f is a Gâteaux differentiable function, then the Bregman distance has the following important property, called the *three point identity*: for any $x, y, z \in \operatorname{int} \operatorname{dom} f$,

$$D_f(x, y) + D_f(y, z) - D_f(x, z) = \langle \nabla f(z) - \nabla f(y), x - y \rangle . \tag{2.1}$$

Recall that, according to [12, Section 1.2, p. 17] (see also [11]), the function f is called *totally convex at a point* $x \in \operatorname{int} \operatorname{dom} f$ if its *modulus of total convexity at* x, that is, the function $\upsilon_f : \operatorname{int} \operatorname{dom} f \times [0, +\infty) \to [0, +\infty]$, defined by

$$\upsilon_f(x, t) := \inf \{ D_f(y, x) \ : \ y \in \operatorname{dom} f, \ \ \|y - x\| = t \} , \tag{2.2}$$

is positive whenever $t > 0$. The function f is called *totally convex* when it is totally convex at every point $x \in \operatorname{int} \operatorname{dom} f$. In addition, the function f is called *totally convex on bounded sets* if $\upsilon_f(E, t)$ is positive for any nonempty bounded subset E of X and for any $t > 0$, where the *modulus of total convexity of the function* f on the set E is the function $\upsilon_f : \operatorname{int} \operatorname{dom} f \times [0, +\infty) \to [0, +\infty]$ defined by

$$\upsilon_f(E, t) := \inf \{ \upsilon_f(x, t) : x \in E \cap \operatorname{dom} f \} .$$

Examples of totally convex functions can be found, for example, in [12, 15]. The following proposition summarizes some properties of the modulus of total convexity.

Proposition 2.2 (*cf.* [12, Propostion 1.2.2, p. 18])**.** *Let f be a proper, convex and lower semicontinuous function. If $x \in \operatorname{int} \operatorname{dom} f$, then*

(i) *The domain of $\upsilon_f(x, \cdot)$ is an interval of the form $[0, \tau_f(x))$ or $[0, \tau_f(x)]$ with $\tau_f(x) \in (0, +\infty]$.*
(ii) *If $c \in [1, +\infty)$ and $t \geq 0$, then $\upsilon_f(x, ct) \geq c\upsilon_f(x, t)$.*
(iii) *The function $\upsilon_f(x, \cdot)$ is superadditive, that is, for any $s, t \in [0, +\infty)$, we have $\upsilon_f(x, s + t) \geq \upsilon_f(x, s) + \upsilon_f(x, t)$.*
(iv) *The function $\upsilon_f(x, \cdot)$ is increasing; it is strictly increasing if and only if f is totally convex at x.*

The following proposition follows from [14, Proposition 2.3, p. 39] and [39, Theorem 3.5.10, p. 164].

Proposition 2.3. *If f is Fréchet differentiable and totally convex, then f is cofinite.*

The next proposition turns out to be very useful in the proof of our main result.

Proposition 2.4 (*cf.* [32, Proposition 2.2, p. 3])**.** *If $x \in \operatorname{dom} f$, then the following statements are equivalent:*

(i) *The function f is totally convex at x;*

(ii) *For any sequence* $\{y_n\}_{n\in\mathbb{N}} \subset \operatorname{dom} f$*,*

$$\lim_{n\to+\infty} D_f(y_n, x) = 0 \Rightarrow \lim_{n\to+\infty} \|y_n - x\| = 0.$$

Recall that the function f is called *sequentially consistent* (see [15]) if for any two sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{y_n\}_{n\in\mathbb{N}}$ in X such that the first one is bounded,

$$\lim_{n\to+\infty} D_f(y_n, x_n) = 0 \Rightarrow \lim_{n\to+\infty} \|y_n - x_n\| = 0.$$

Proposition 2.5 (*cf.* [12, Lemma 2.1.2, p. 67])**.** *If* $\operatorname{dom} f$ *contains at least two points, then the function* f *is totally convex on bounded sets if and only if the function* f *is sequentially consistent.*

2.4. **The resolvent of A relative to f.** Let $A : X \to 2^{X^*}$ be an operator and assume that f Gâteaux differentiable. The operator

$$\operatorname{Prt}_A^f := (\nabla f + A)^{-1} : X^* \to 2^X$$

is called the *protoresolvent* of A, or, more precisely, the *protoresolvent of A relative to f*. This allows us to define the *resolvent* of A, or, more precisely, the *resolvent of A relative to f*, introduced and studied in [5], as the operator $\operatorname{Res}_A^f : X \to 2^X$ given by $\operatorname{Res}_A^f := \operatorname{Prt}_A^f \circ \nabla f$. This operator is single-valued when A is monotone and f is strictly convex on $\operatorname{int}\operatorname{dom} f$. If $A = \partial\varphi$, where φ is a proper, lower semicontinuous and convex function, then we denote

$$\operatorname{Prox}_\varphi^f := \operatorname{Prt}_{\partial\varphi}^f \quad \text{and} \quad \operatorname{prox}_\varphi^f := \operatorname{Res}_{\partial\varphi}^f.$$

If C is a nonempty, closed and convex subset of X, then the indicator function ι_C of C, that is, the function

$$\iota_C(x) := \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{if } x \notin C \end{cases}$$

is proper, convex and lower semicontinuous, and therefore $\partial\iota_C$ exists and is a maximal monotone operator with domain C. The operator $\operatorname{prox}_{\iota_C}^f$ is called the *Bregman projection* onto C with respect to f (*cf.* [8]) and we denote it by proj_C^f. Note that if X is a Hilbert space and $f(x) = \frac{1}{2}\|x\|^2$, then the Bregman projection of x onto C, *i.e.*, $\operatorname{argmin}\{\|y - x\| : y \in C\}$, is the metric projection P_C.

Recall that the Bregman projection of x onto the nonempty, closed and convex set $K \subset \operatorname{dom} f$, is the necessarily unique vector $\operatorname{proj}_K^f(x) \in K$ satisfying

$$D_f\left(\operatorname{proj}_K^f(x), x\right) = \inf\{D_f(y, x) : y \in K\}.$$

Similarly to the metric projection in Hilbert spaces, Bregman projections with respect to totally convex and differentiable functions have variational characterizations.

Proposition 2.6 (*cf.* [15, Corollary 4.4, p. 23])**.** *Suppose that* f *is totally convex on* $\operatorname{int}\operatorname{dom} f$*. Let* $x \in \operatorname{int}\operatorname{dom} f$ *and let* $K \subset \operatorname{int}\operatorname{dom} f$ *be a nonempty, closed and convex set. If* $\hat{x} \in K$*, then the following conditions are equivalent:*

(i) *The vector* $\hat{x}$ *is the Bregman projection of* x *onto* K *with respect to* f*;*

(ii) *The vector* $\hat{x}$ *is the unique solution of the variational inequality*

$$\langle \nabla f(x) - \nabla f(z), z - y\rangle \geq 0, \qquad \forall y \in K;$$

(iii) *The vector* $\hat{x}$ *is the unique solution of the inequality*

$$D_f(y,z) + D_f(z,x) \leq D_f(y,x), \qquad \forall y \in K.$$

3. A strong convergence theorem for a proximal-type algorithm

In this section we study the following algorithm when $Z := \bigcap_{i=1}^{N} A_i^{-1}(0^*) \neq \varnothing$:

$$\text{(3.1)} \qquad \begin{cases} x_0 \in X, \\ \eta_n^i = \xi_n^i + \frac{1}{\lambda_n^i}\left(\nabla f(y_n^i) - \nabla f(x_n)\right), \qquad \xi_n^i \in A_i y_n^i, \\ H_n^i = \left\{z \in X : \langle \xi_n^i, z - y_n^i\rangle \leq 0\right\}, \\ H_n := \cap_{i=1}^{N} H_n^i, \\ W_n = \{z \in X : \langle \nabla f(x_0) - \nabla f(x_n), z - x_n\rangle \leq 0\}, \\ x_{n+1} = \operatorname{proj}^f_{H_n \cap W_n}(x_0), \qquad n = 0, 1, 2, \ldots, \end{cases}$$

where, for each $i = 1, 2, \ldots, N$, $\left\{\lambda_n^i\right\}_{n\in\mathbb{N}}$ is a given sequence of positive real numbers and $\left\{\eta_n^i\right\}_{n\in\mathbb{N}}$ is the sequence of errors corresponding to the approximate solutions of the resolvent equation. Note that if $\eta_n^i = 0^*$, then

$$y_n^i = \operatorname{Res}^f_{\lambda_n^i A_i}(x_n).$$

Theorem 3.1. *Let* $A_i : X \to 2^{X^*}$, $i = 1, 2, \ldots, N$, *be* N *maximal monotone operators such that* $Z := \bigcap_{i=1}^{N} A_i^{-1}(0^*) \neq \varnothing$. *Let* $f : X \to \mathbb{R}$ *be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of* X. *Then, for each* $x_0 \in X$, *there are sequences* $\{x_n\}_{n\in\mathbb{N}}$ *which satisfy(3.1). If, for each* $i = 1, 2, \ldots, N$, $\liminf_{n\to+\infty} \lambda_n^i > 0$, *and the sequences of errors* $\left\{\eta_n^i\right\}_{n\in\mathbb{N}} \subset X^*$ *satisfy* $\lim_{n\to+\infty} \lambda_n^i \eta_n^i = 0^*$ *and* $\limsup_{n\to+\infty} \left\langle \eta_n^i, y_n^i\right\rangle \leq 0$, *then each such sequence* $\{x_n\}_{n\in\mathbb{N}}$ *converges strongly to* $\operatorname{proj}^f_Z(x_0)$ *as* $n \to +\infty$.

Proof. Note that $\operatorname{dom} \nabla f = X$ because $\operatorname{dom} f = X$ and f is Legendre. Hence it follows from [5, Corollary 3.14(ii), p. 606] that $\operatorname{dom} \operatorname{Res}^f_{\lambda A} = X$. We begin with the following claim.

Claim 1: *There are sequences* $\{x_n\}_{n\in\mathbb{N}}$ *which satisfy (3.1).*

As a matter of fact, we will prove that, for each $x_0 \in X$, there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ which is generated by (3.1) with $\eta_n^i = 0^*$ for all $i = 1, 2, \ldots, N$ and $n \in \mathbb{N}$.

It is obvious that H_n^i are closed and convex sets for any $i = 1, 2, \ldots, N$. Hence H_n is also closed and convex. It is also obvious that W_n is a closed and convex set. Let $u \in Z$. Since $\operatorname{dom} \operatorname{Res}^f_{\lambda_0^i A_i} = X$, there exists $\left(y_0^i, \xi_0^i\right) \in X \times X^*$ such that $0^* = \xi_0^i + \frac{1}{\lambda_0^i}\left(\nabla f(y_0^i) - \nabla f(x_0)\right)$ $\left(y_0^i = \operatorname{Res}^f_{\lambda_0^i A_i}(x_0)\right)$ and $\xi_0^i \in A_i y_0^i$. Since A_i is monotone, it follows that

$$\left\langle \xi_0^i, y_0^i - u\right\rangle \geq 0,$$

which implies that $u \in H_0^i$. Since this holds for any $i = 1, 2, \ldots, N$, it follows that $u \in H_0$. It is also obvious that $u \in W_0 = X$. Thus $u \in H_0 \cap W_0$, and therefore $x_1 = \mathrm{proj}^f_{H_0 \cap W_0}(x_0)$ is well defined. Now suppose that $u \in H_{n-1} \cap W_{n-1}$ for some $n \geq 1$. Let $x_n = \mathrm{proj}^f_{H_{n-1} \cap W_{n-1}}(x_0)$. Again, there exists $\left(y_n^i, \xi_n^i\right) \in X \times X^*$ such that $0^* = \xi_n^i + \frac{1}{\lambda_n^i}\left(\nabla f(y_n^i) - \nabla f(x_n)\right)$ and $\xi_n^i \in A y_n^i$. The monotonicity of A_i implies that $u \in H_n^i$. Since this holds for any $i = 1, 2, \ldots, N$, it follows that $u \in H_n$. Now it follows from Proposition 2.6(ii) that

$$\begin{aligned} &\langle \nabla f(x_0) - \nabla f(x_n), u - x_n \rangle \\ &= \left\langle \nabla f(x_0) - \nabla f\left(\mathrm{proj}^f_{H_{n-1} \cap W_{n-1}}(x_0)\right), u - \mathrm{proj}^f_{H_{n-1} \cap W_{n-1}}(x_0) \right\rangle \\ &\leq 0, \end{aligned}$$

which implies that $u \in W_n$. Therefore $u \in H_n \cap W_n$, and hence $x_{n+1} = \mathrm{proj}^f_{H_n \cap W_n}(x_0)$ is well defined. Thus the sequence we constructed is indeed well defined and satisfies (3.1), as claimed.

From now on we fix an arbitrary sequence $\{x_n\}_{n \in \mathbb{N}}$ satisfying (3.1). It is clear from the proof of Claim 1 that $Z \subset H_n \cap W_n$ for each $n \in \mathbb{N}$.

Claim 2: *The sequence $\{x_n\}_{n \in \mathbb{N}}$ is bounded.*

It follows from the definition of W_n and Proposition 2.6(ii) that $\mathrm{proj}^f_{W_n}(x_0) = x_n$. Furthermore, by Proposition 2.6(iii), for each $u \in Z$, we have

$$\begin{aligned} D_f\left(x_n, x_0\right) &= D_f\left(\mathrm{proj}^f_{W_n}(x_0), x_0\right) \\ &\leq D_f\left(u, x_0\right) - D_f\left(u, \mathrm{proj}^f_{W_n}(x_0)\right) \\ &\leq D_f\left(u, x_0\right). \end{aligned} \tag{3.2}$$

Hence the sequence $\{D_f(x_n, x_0)\}_{n \in \mathbb{N}}$ is bounded by $D_f(u, x_0)$ for any $u \in Z$. Therefore the sequence $\{\nu_f(x_0, \|x_n - x_0\|)\}_{n \in \mathbb{N}}$ is bounded by $D_f(u, x_0)$, because from the definition of the modulus of total convexity (see (2.2)) and from (3.2) we get that

$$\nu_f(x_0, \|x_n - x_0\|) \leq D_f(x_n, x_0) \leq D_f(u, x_0). \tag{3.3}$$

Since the function f is totally convex, the function $\nu_f(x, \cdot)$ is strictly increasing and positive on $(0, \infty)$ (*cf.* Proposition 2.2(iv)). This implies, in particular, that $\nu_f(x, 1) > 0$ for all $x \in X$. Now suppose by way of contradiction that the sequence $\{x_n\}_{n \in \mathbb{N}}$ is not bounded. Then there exists a sequence $\{n_k\}_{k \in \mathbb{N}}$ of positive real numbers such that

$$\lim_{k \to +\infty} \|x_{n_k}\| = +\infty.$$

Consequently, $\lim_{k \to +\infty} \|x_{n_k} - x_0\| = +\infty$. This shows that the sequence $\{\nu_f(x_0, \|x_n - x_0\|)\}_{n \in \mathbb{N}}$ is not bounded. Indeed, there exists some $k_0 > 0$ such that $\|x_{n_k} - x_0\| > 1$ for any $k > k_0$ and then, by Proposition 2.2(ii), we see that

$$\nu_f(x_0, \|x_{n_k} - x_0\|) \geq \|x_{n_k} - x_0\| \cdot \nu_f(x_0, 1) \to +\infty,$$

because, as noted above, $\nu_f(x_0, 1) > 0$. This contradicts (3.3). Hence the sequence $\{x_n\}_{n \in \mathbb{N}}$ is indeed bounded, as claimed.

Claim 3: *Every weak subsequential limit of* $\{x_n\}_{n\in\mathbb{N}}$ *belongs to* Z.

It follows from the definition of W_n and Proposition 2.6(ii) that $\mathrm{proj}_{W_n}^f(x_0) = x_n$. Since $x_{n+1} \in W_n$, it follows from Proposition 2.6(iii) that

$$D_f\left(x_{n+1}, \mathrm{proj}_{W_n}^f(x_0)\right) + D_f\left(\mathrm{proj}_{W_n}^f(x_0), x_0\right) \le D_f\left(x_{n+1}, x_0\right)$$

and hence

$$D_f\left(x_{n+1}, x_n\right) + D_f\left(x_n, x_0\right) \le D_f\left(x_{n+1}, x_0\right). \tag{3.4}$$

Therefore the sequence $\{D_f(x_n, x_0)\}_{n\in\mathbb{N}}$ is increasing and since it is also bounded (see Claim 2), $\lim_{n\to+\infty} D_f(x_n, x_0)$ exists. Thus from (3.4) it follows that

$$\lim_{n\to+\infty} D_f\left(x_{n+1}, x_n\right) = 0. \tag{3.5}$$

Proposition 2.5 now implies that $\lim_{n\to+\infty}\left(x_{n+1} - x_n\right) = 0$. For any $i = 1, 2, \ldots, N$, it follows from the three point identity (see (2.1)) that

$$\begin{aligned}
&D_f\left(x_{n+1}, x_n\right) - D_f\left(y_n^i, x_n\right)\\
&= D_f\left(x_{n+1}, y_n^i\right) + \left\langle \nabla f(x_n) - \nabla f(y_n^i), y_n^i - x_{n+1}\right\rangle\\
&\ge \left\langle \nabla f(x_n) - \nabla f(y_n^i), y_n^i - x_{n+1}\right\rangle = \left\langle \lambda_n^i\left(\xi_n^i - \eta_n^i\right), y_n^i - x_{n+1}\right\rangle\\
&= \lambda_n^i\left\langle \xi_n^i, y_n^i - x_{n+1}\right\rangle - \lambda_n^i\left\langle \eta_n^i, y_n^i - x_{n+1}\right\rangle \ge -\lambda_n^i\left\langle \eta_n^i, y_n^i - x_{n+1}\right\rangle
\end{aligned}$$

because $x_{n+1} \in H_n^i$. We now have

$$\begin{aligned}
D_f\left(y_n^i, x_n\right) &\le D_f\left(x_{n+1}, x_n\right) + \left\langle \lambda_n^i\eta_n^i, y_n^i - x_{n+1}\right\rangle\\
&= D_f\left(x_{n+1}, x_n\right) + \lambda_n^i\left\langle \eta_n^i, y_n^i\right\rangle - \left\langle \lambda_n^i\eta_n^i, x_{n+1}\right\rangle\\
&\le D_f\left(x_{n+1}, x_n\right) + \lambda_n^i\left\langle \eta_n^i, y_n^i\right\rangle + \left\|\lambda_n^i\eta_n^i\right\|_* \left\|x_{n+1}\right\|.
\end{aligned}$$

Hence

$$\begin{aligned}
\limsup_{n\to+\infty} D_f\left(y_n^i, x_n\right) &\le \limsup_{n\to+\infty} D_f\left(x_{n+1}, x_n\right)\\
&\quad + \limsup_{n\to+\infty} \lambda_n^i\left\langle \eta_n^i, y_n^i\right\rangle + \limsup_{n\to+\infty}\left\|\lambda_n^i\eta_n^i\right\|_* \left\|x_{n+1}\right\|.
\end{aligned}$$

Since $\lim_{n\to+\infty}\lambda_n^i\eta_n^i = 0^*$, $\limsup_{n\to+\infty}\left\langle \eta_n^i, y_n^i\right\rangle \le 0$, and $\lim_{n\to+\infty} D_f\left(x_{n+1}, x_n\right) = 0$ (by (3.5)), we see that $\limsup_{n\to+\infty} D_f\left(y_n^i, x_n\right) \le 0$. Hence $\lim_{n\to+\infty} D_f\left(y_n^i, x_n\right) = 0$. Proposition 2.5 now implies that $\lim_{n\to+\infty}\left(y_n^i - x_n\right) = 0$. Now let $\left\{x_{n_j}\right\}_{j\in\mathbb{N}}$ be a weakly convergent subsequence of $\{x_n\}_{n\in\mathbb{N}}$ and denote its weak limit by v. Then $\left\{y_{n_j}^i\right\}_{j\in\mathbb{N}}$ also converges weakly to v for any $i = 1, 2, \ldots, N$. Since $\liminf_{n\to+\infty}\lambda_n^i > 0$ and $\lim_{n\to+\infty}\eta_n^i = 0^*$, it follows from Proposition 2.1 that

$$\xi_n^i = \frac{1}{\lambda_n^i}\left(\nabla f(x_n) - \nabla f(y_n^i)\right) + \eta_n^i \to 0^*, \tag{3.6}$$

for any $i = 1, 2, \ldots, N$. Since $\xi_n^i \in Ay_n^i$ and A_i is monotone, it follows that

$$\left\langle \eta - \xi_n^i, z - y_n^i\right\rangle \ge 0$$

for all $(z, \eta) \in \operatorname{graph}(A_i)$. This, in turn, implies that

$$\langle \eta, z - v \rangle \geq 0$$

for all $(z, \eta) \in \operatorname{graph}(A_i)$. Therefore, using the maximal monotonicity of A_i, we now obtain that $v \in A_i^{-1}(0^*)$ for each $i = 1, 2, \ldots, N$. Thus $v \in Z$ and this proves Claim 3.

Claim 4: *The sequence* $\{x_n\}_{n\in\mathbb{N}}$ *converges strongly to* $\operatorname{proj}_Z^f(x_0)$.

Let $\tilde{u} = \operatorname{proj}_Z^f(x_0)$. Since $x_{n+1} = \operatorname{proj}_{H_n \cap W_n}^f(x_0)$ and Z is contained in $H_n \cap W_n$, we have $D_f(x_{n+1}, x_0) \leq D_f(\tilde{u}, x_0)$. The three point identity (see (2.1)) yields

$$\begin{aligned} D_f(x_n, \tilde{u}) &= D_f(x_n, x_0) + D_f(x_0, \tilde{u}) - \langle \nabla f(\tilde{u}) - \nabla f(x_0), x_n - x_0 \rangle \\ &\leq D_f(\tilde{u}, x_0) + D_f(x_0, \tilde{u}) - \langle \nabla f(\tilde{u}) - \nabla f(x_0), x_n - x_0 \rangle \\ &= \langle \nabla f(\tilde{u}) - \nabla f(x_0), \tilde{u} - x_0 \rangle - \langle \nabla f(\tilde{u}) - \nabla f(x_0), x_n - x_0 \rangle \\ &= \langle \nabla f(\tilde{u}) - \nabla f(x_0), \tilde{u} - x_n \rangle . \end{aligned}$$

Now let $\{x_{n_i}\}_{i\in\mathbb{N}}$ be a weakly convergent subsequence of $\{x_n\}_{n\in\mathbb{N}}$ and denote its weak limit by v. We already know (see Claim 3) that $v \in Z$. It follows from Proposition 2.6(ii) that

$$\limsup_{i\to+\infty} D_f(x_{n_i}, \tilde{u}) \leq \langle \nabla f(\tilde{u}) - \nabla f(x_0), \tilde{u} - v \rangle \leq 0.$$

Hence

$$\lim_{i\to+\infty} D_f(x_{n_i}, \tilde{u}) = 0.$$

Proposition 2.4 now implies that $x_{n_i} \to \tilde{u}$. It follows that the whole sequence $\{x_n\}_{n\in\mathbb{N}}$ converges strongly to $\tilde{u} = \operatorname{proj}_Z^f(x_0)$, as claimed. This completes the proof of Theorem 3.1. □

4. Zero free operators

Suppose now that the operators A_i, $i = 1, 2, \ldots, N$, have no common zero. If $\{x_n\}_{n\in\mathbb{N}}$ is a sequence satisfying (3.1), then $\lim_{n\to+\infty} \|x_n\| = +\infty$. This is because if $\{x_n\}_{n\in\mathbb{N}}$ were to have a bounded subsequence, then it would follow from Claim 3 in the proof of Theorem 3.1 that the operators A_i, $i = 1, 2, \ldots, N$, did share a common zero. In the case of a single zero free operator A, we can prove that such a sequence always exists.

To this end, we first recall the duality mapping of the space X, *i.e.,* the mapping $J : X \to 2^{X^*}$ which is defined by

$$Jx = \left\{ \xi \in X^* : \langle \xi, x \rangle = \|x\|^2 = \|\xi\|_*^2 \right\}.$$

We continue with the following lemma.

Lemma 4.1. *If* $A : X \to 2^{X^*}$ *is a maximal monotone operator with a bounded effective domain, then* $A^{-1}(0^*) \neq \varnothing$.

Proof. Let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence of positive numbers which converges to zero. The operator $A+\varepsilon_n J$ is surjective for any $n\in\mathbb{N}$ because A is a maximal monotone operator (see [19, Theorem 3.11, p. 166]). Therefore, for any $n\in\mathbb{N}$, there exists $x_n\in\operatorname{dom}A$ such that $0^*\in(A+\varepsilon_n J)\,x_n$. Consequently, for any $n\in\mathbb{N}$, there are $\xi_n\in Ax_n$ and $\eta_n\in Jx_n$ such that $\xi_n+\varepsilon_n\eta_n=0^*$. Therefore we have

$$\|\xi_n\|_*=\varepsilon_n\|\eta_n\|_*=\varepsilon_n\|x_n\|\to 0$$

because $\{x_n\}_{n\in\mathbb{N}}$ is a bounded sequence. Hence there exists a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ of $\{x_n\}_{n\in\mathbb{N}}$ which converges weakly to some $x_0\in X$. Since A is monotone we have

$$\langle\zeta-\xi_{n_k},v-x_{n_k}\rangle\geq 0,\quad k\in\mathbb{N}, \tag{4.1}$$

for any $(v,\zeta)\in\operatorname{graph}A$. Letting $k\to+\infty$ in (4.1), we obtain $\langle\zeta,v-x_0\rangle\geq 0$ for all $(v,\zeta)\in\operatorname{graph}A$ and from the maximality of A it follows that $x_0\in A^{-1}(0^*)$. Hence $A^{-1}(0^*)\neq\varnothing$, as claimed. □

Theorem 4.2. *Let $A:X\to 2^{X^*}$ be a maximal monotone operator. Let $f:X\to\mathbb{R}$ be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of X. Then, for each $x_0\in X$, there are sequences $\{x_n\}_{n\in\mathbb{N}}$ which satisfy (3.1) with $N=1$. If $\liminf_{n\to+\infty}\lambda_n>0$, and the sequence of errors $\{\eta_n\}_{n\in\mathbb{N}}\subset X^*$ satisfies $\lim_{n\to+\infty}\lambda_n\eta_n=0^*$ and $\limsup_{n\to+\infty}\langle\eta_n,y_n\rangle\leq 0$, then either $A^{-1}(0^*)\neq\varnothing$ and each such sequence $\{x_n\}_{n\in\mathbb{N}}$ converges strongly to $\operatorname{proj}^f_{A^{-1}(0^*)}(x_0)$ or $A^{-1}(0^*)=\varnothing$ and each such sequence $\{x_n\}_{n\in\mathbb{N}}$ satisfies $\lim_{n\to+\infty}\|x_n\|=+\infty$.*

Proof. In view of Theorem 3.1, we only need to consider the case where $A^{-1}(0^*)=\varnothing$. First of all we prove that in this case, for each $x_0\in X$, there is a sequence $\{x_n\}_{n\in\mathbb{N}}$ which satisfies (3.1) with $\eta_n=0^*$ for all $n\in\mathbb{N}$.

We prove this by induction. We first check that the initial step ($n=0$) is well defined. The problem

$$0^*\in Ax+\frac{1}{\lambda_0}\left(\nabla f(x)-\nabla f(x_0)\right)$$

always has a solution (y_0,ξ_0) because it is equivalent to the problem $x=\operatorname{Res}^f_{\lambda_0 A}(x_0)$ and this problem does have a solution since $\operatorname{dom}\operatorname{Res}^f_{\lambda A}=X$ (see Proposition 2.3 and [5, Theorem 3.13(iv), p. 606]). Now note that $W_0=X$. Since H_0 cannot be empty, the next iterate x_1 can be generated; it is the Bregman projection of x_0 onto $H_0=W_0\cap H_0$.

Note that whenever x_n is generated, y_n and ξ_n can further be obtained because the proximal subproblems always have solutions. Suppose now that x_n and (y_n,ξ_n) have already been defined for $n=0,\ldots,\hat{n}$. We have to prove that $x_{\hat{n}+1}$ is also well defined. To this end, take any $z_0\in\operatorname{dom}A$ and define

$$\rho=\max\{\|y_n-z_0\|:n=0,\ldots,\hat{n}\}$$

and

$$h(x)=\begin{cases}0, & \|x-z_0\|\leq\rho+1\\ +\infty, & \text{otherwise.}\end{cases}$$

Then $h : X \to (-\infty, +\infty]$ is a proper, convex and lower semicontinuous function, its subdifferential ∂h is maximal monotone (see [28, Theorem 2.13, p. 124]), and

$$A' = A + \partial h$$

is also maximal monotone (see [33]). Furthermore,

$$A'(z) = A(z) \quad \text{for all } \|z - z_0\| < \rho + 1.$$

Therefore $\xi_n \in A' y_n$ for $n = 0, \ldots, \hat{n}$. We conclude that x_n and (y_n, ξ_n) also satisfy the conditions of Theorem 3.1 applied to the problem $0^* \in A'(x)$. Since A' has a bounded effective domain, this problem has a solution by Lemma 4.1. Thus it follows from Claim 1 in the proof of Theorem 3.1 that $x_{\hat{n}+1}$ is well defined. Hence the whole sequence $\{x_n\}_{n \in \mathbb{N}}$ is well defined, as asserted.

If $\{x_n\}_{n \in \mathbb{N}}$ were to have a bounded subsequence, then it would follow from Claim 3 in the proof of Theorem 3.1 that A had a zero. Therefore if $A^{-1}(0^*) = \varnothing$, then $\lim_{n \to +\infty} \|x_n\| = +\infty$, as asserted. $\square$

Remark 4.3. In both Theorems 3.1 and 4.2 we can replace the assumptions that $\liminf_{n \to +\infty} \lambda_n > 0$ and f is uniformly Fréchet differentiable on bounded subsets of X with the assumption that $\lim_{n \to +\infty} \lambda_n = +\infty$. This is because in this case $\{\nabla f(x_n) - \nabla f(y_n)\}_{n \in \mathbb{N}}$ is bounded and therefore (3.6) continues to hold.

5. Consequences of the strong convergence theorem

Algorithm (1.4) is a special case of algorithm (3.1) when $N = 1$ and $\eta_n = 0^*$ for all $n \in \mathbb{N}$. Hence as a direct consequence of Theorem 3.1 we obtain the following result (*cf.* [21]) .

Corollary 5.1. *Let $A : X \to 2^{X^*}$ be a maximal monotone operator. Let $f : X \to \mathbb{R}$ be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of X, and suppose that $\liminf_{n \to +\infty} \lambda_n > 0$. Then for each $x_0 \in X$, the sequence $\{x_n\}_{n \in \mathbb{N}}$ generated by (1.4) is well defined, and either $A^{-1}(0^*) \neq \varnothing$ and $\{x_n\}_{n \in \mathbb{N}}$ converges strongly to $\mathrm{proj}^f_{A^{-1}(0^*)}(x_0)$ as $n \to +\infty$, or $A^{-1}(0^*) = \varnothing$ and $\lim_{n \to +\infty} \|x_n\| = +\infty$.*

Notable corollaries of Theorems 3.1 and 4.2 occur when the space X is both uniformly smooth and uniformly convex. In this case the function $f(x) = \frac{1}{2}\|x\|^2$ is Legendre (*cf.* [3, Lemma 6.2, p.24]) and uniformly Fréchet differentiable on bounded subsets of X. According to [13, Corollary 1(ii), p. 325], f is sequentially consistent since X is uniformly convex and hence f is totally convex on bounded subsets of X. Therefore Theorems 3.1 and 4.2 hold in this context and lead us to the following two results which, in some sense, complement Theorem 8 in [25] (see also Theorem 1 in [37]).

Corollary 5.2. *Let X be a uniformly smooth and uniformly convex Banach space and let $A : X \to 2^{X^*}$ be a maximal monotone operator. Then, for each $x_0 \in X$, the sequence $\{x_n\}_{n \in \mathbb{N}}$ generated by (1.3) is well defined. If $\liminf_{n \to +\infty} \lambda_n > 0$, then either $A^{-1}(0^*) \neq \varnothing$ and $\{x_n\}_{n \in \mathbb{N}}$ converges strongly to $Q_{A^{-1}(0^*)}(x_0)$ as $n \to +\infty$, or $A^{-1}(0^*) = \varnothing$ and $\lim_{n \to +\infty} \|x_n\| = +\infty$.*

Corollary 5.3. *Let X be a Hilbert space and let $A : X \to 2^X$ be a maximal monotone operator. Then, for each $x_0 \in X$, the sequence $\{x_n\}_{n\in\mathbb{N}}$ generated by (1.2) is well defined. If $\liminf_{n\to\infty} \lambda_n > 0$, then either $A^{-1}(0) \neq \varnothing$ and $\{x_n\}_{n\in\mathbb{N}}$ converges strongly to $\mathrm{P}_{A^{-1}(0)}(x_0)$ as $n \to +\infty$, or $A^{-1}(0) = \varnothing$ and $\lim_{n\to+\infty} \|x_n\| = +\infty$.*

These corollaries also hold, of course, in the presence of computational errors as in Theorems 3.1 and 4.2.

6. An application of the strong convergence theorem

Let $g : X \to (-\infty, +\infty]$ be a proper, convex and lower semicontinuous function. Recall that the subdifferential ∂g of g is defined for any $x \in X$ by

$$\partial g(x) := \{\xi \in X^* : \langle \xi, y - x\rangle \leq g(y) - g(x) \quad \forall y \in X\}.$$

Using Theorem 3.1 and the subdifferential of g, we obtain an algorithm for finding a minimizer of g.

Proposition 6.1. *Let $g : X \to (-\infty, +\infty]$ be a proper, convex and lower semicontinuous function which attains its minimum over X. If $f : X \to \mathbb{R}$ is a Legendre function which is bounded, uniformly Fréchet differentiable, and totally convex on bounded subsets of X, and $\{\lambda_n\}_{n\in\mathbb{N}}$ is a positive sequence with $\liminf_{n\to+\infty} \lambda_n > 0$, then, for each $x_0 \in X$, the sequence $\{x_n\}_{n\in\mathbb{N}}$ generated by*

$$\begin{cases} x_0 \in X, \\ 0^* = \xi_n + \frac{1}{\lambda_n}\left(\nabla f(y_n) - \nabla f(x_n)\right), \qquad \xi_n \in \partial g(y_n), \\ H_n = \{z \in X : \langle \xi_n, z - y_n\rangle \leq 0\}, \\ W_n = \{z \in X : \langle \nabla f(x_0) - \nabla f(x_n), z - x_n\rangle \leq 0\}, \\ x_{n+1} = \operatorname{proj}^f_{H_n\cap W_n}(x_0), \qquad n = 0, 1, 2, \ldots, \end{cases}$$

converges strongly to a minimizer of g as $n \to +\infty$. If g does not attain its minimum over X, then $\lim_{n\to+\infty} \|x_n\| = +\infty$.

Proof. The subdifferential ∂g of g is a maximal monotone operator since g is a proper, convex and lower semicontinuous function (see [28, Theorem 2.13, p. 124]). Since the zero set of ∂g coincides with the set of minimizers of g, Proposition 6.1 follows immediately from Theorems 3.1 and 4.2. □

Note that in this case

$$y_n = \arg\min_{x\in X}\left\{g(x) + \frac{1}{\lambda_n} D_f(x, x_n)\right\}$$

is equivalent to

$$0^* \in \partial g(y_n) + \frac{1}{\lambda_n}\left(\nabla f(y_n) - \nabla f(x_n)\right).$$

References

[1] A. Ambrosetti and G. Prodi, *A Primer of Nonlinear Analysis*, Cambridge University Press, Cambridge, 1993.

[2] H. H. Bauschke and J. M. Borwein, *Legendre functions and the method of random Bregman projections*, J. Convex Anal. **4** (1997), 27–67.

[3] H. H. Bauschke, J. M. Borwein and P. L. Combettes, *Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces*, Commun. Contemp. Math. **3** (2001), 615–647.

[4] H. H. Bauschke and P. L. Combettes, *A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces*, Math. Oper. Res. **26** (2001), 248–264.

[5] H. H. Bauschke, J. M. Borwein and P. L. Combettes, *Bregman monotone optimization algorithms*, SIAM J. Control Optim. **42** (2003), 596–636.

[6] H. H. Bauschke, E. Matoušková and S. Reich, *Projection and proximal point methods: convergence results and counterexamples*, Nonlinear Anal. **56** (2004), 715–738.

[7] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer Verlag, New York, 2000.

[8] L. M. Bregman, *A relaxation method for finding the common point of convex sets and its application to the solution of problems in convex programming*, USSR Comput. Math. and Math. Phys. **7** (1967), 200–217.

[9] H. Brézis and P.-L. Lions *Produits infinis de résolvantes*, Israel J. Math. **29** (1978), 329–345.

[10] R. E. Bruck and S. Reich, *Nonexpansive projections and resolvents of accretive operators*, Houston J. Math. **3** (1977), 459–470.

[11] D. Butnariu, Y. Censor and S. Reich, *Iterative averaging of entropic projections for solving stochastic convex feasibility problems*, Computational Optimization and Applications **8** (1997), 21–39.

[12] D. Butnariu and A. N. Iusem, *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization*, Kluwer Academic Publishers, Dordrecht, 2000.

[13] D. Butnariu, A. N. Iusem and E. Resmerita, *Total convexity for powers of the norm in uniformly convex Banach spaces*, J. Convex Anal. **7** (2000), 319–334.

[14] D. Butnariu, A. N. Iusem and C. Zălinescu, *On uniform convexity, total convexity and convergence of the proximal point and outer Bregman projection algorithms in Banach spaces*, J. Convex Anal. **10** (2003), 35–61.

[15] D. Butnariu and E. Resmerita, *Bregman distances, totally convex functions and a method for solving operator equations in Banach spaces*, Abstr. Appl. Anal. 2006, Art. ID 84919, 1–39.

[16] D. Butnariu and G. Kassay, *A proximal-projection method for finding zeroes of set-valued operators*, SIAM J. Control Optim. **47** (2008), 2096–2136.

[17] Y. Censor and A. Lent, *An iterative row-action method for interval convex programming*, J. Optim. Theory Appl. **34** (1981), 321–353.

[18] Y. Censor and S. A. Zenios, *Proximal minimization algorithm with D-functions*, J. Optim. Theory Appl. **73** (1992), 455–468.

[19] I. Cioranescu, *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems*, Kluwer Academic Publishers, Dordrecht, 1990.

[20] J. Eckstein, *Nonlinear proximal point algorithms using Bregman functions, with application to convex programming*, Math. Oper. Res. **18** (1993), 202–226.

[21] R. Gárciga Otero and B. F. Svaiter, *A strongly convergent hybrid proximal method in Banach spaces*, J. Math. Anal. Appl. **289** (2004), 700–711.

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

[23] S. Kamimura and W. Takahashi, *Approximating solutions of maximal monotone operators in Hilbert spaces*, J. Approx. Theory **106** (2000), 226–240.

[24] S. Kamimura and W. Takahashi, *Weak and strong convergence of solutions to accretive operator inclusions and applications*, Set-Valued Anal. **8** (2000), 361–374.

[25] S. Kamimura and W. Takahashi, *Strong convergence of a proximal-type algorithm in a Banach space*, SIAM J. Optim. **13** (2002), 938–945.

[26] B. Martinet, *Régularisation d'inéquations variationelles par approximations successives*, Revue Française d'Informatique et de Recherche Opérationelle **4** (1970), 154–159.
[27] O. Nevanlinna and S. Reich, *Strong convergence of contraction semigroups and of iterative methods for accretive operators in Banach spaces*, Israel J. Math. **32** (1979), 44–58.
[28] D. Pascali and S. Sburlan, *Nonlinear Mappings of Monotone Type*, Sijthoff & Nordhoff International Publishers, Alphen aan den Rijn, 1978.
[29] S. Reich, *Weak convergence theorems for nonexpansive mappings in Banach spaces*, J. Math. Anal. Appl. **67** (1979), 274–276.
[30] S. Reich, *A weak convergence theorem for the alternating method with Bregman distances*, in Theory and Applications of Nonlinear Operators of Accretive and Monotone Type, Marcel Dekker, New York, 1996, pp. 313–318.
[31] S. Reich and A. J. Zaslavski, *Infinite products of resolvents of accretive operator*, Topol. Methods Nonlinear Anal. **15** (2000), 153–168.
[32] E. Resmerita, *On total convexity, Bregman projections and stability in Banach spaces*, J. Convex Anal. **11** (2004), 1–16.
[33] R. T. Rockafellar, *On the maximality of sums of nonlinear monotone operators*, Trans. Amer. Math. Soc. **149** (1970), 75–88.
[34] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim. **14** (1976), 877–898.
[35] R. T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res. **1** (1976), 97–116.
[36] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer Verlag, Berlin, 1998.
[37] M. V. Solodov and B. F. Svaiter, *Forcing strong convergence of proximal point iterations in a Hilbert space*, Math. Program. **87** (2000), 189–202.
[38] J. C. Yao and L. C. Zeng, *An inexact proximal-type algorithm in Banach spaces*, J. Optim. Theory Appl. **135** (2007), 145–161.
[39] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific Publishing, Singapore, 2002.

Manuscript received April 13, 2009
revised October 2, 2009

Simeon Reich
Department of Mathematics, The Technion – Israel Institute of Technology, 32000 Haifa, Israel
E-mail address: `sreich@tx.technion.ac.il`

Shoham Sabach
Department of Mathematics, The Technion – Israel Institute of Technology, 32000 Haifa, Israel
E-mail address: `ssabach@tx.technion.ac.il`