Two Strong Convergence Theorems for a Proximal Method in Reflexive Banach Spaces

Simeon Reich and Shoham Sabach

Abstract. Two strong convergence theorems for a proximal method for finding common zeroes of maximal monotone operators in reflexive Banach spaces are established. Both theorems take into account possible computational errors.

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

2000 *Mathematics Subject Classification.* 47H05, 47J25.
Key words and phrases. Banach space, Bregman projection, Legendre function, monotone operator, proximal point algorithm, resolvent, totally convex function.

by the rule

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

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

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

This algorithm was first introduced by Martinet [**28**] and further developed by Rockafellar [**38**], 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 [**38**] asks if the sequence generated by (1.1) converges strongly. This question was answered in the negative by Güler [**24**], who presented an example of a subdifferential for which the sequence generated by (1.1) converges weakly but not strongly; see [**7**] for a more recent and simpler example. Quite a few results regarding the proximal point algorithm and its extensions can be found in the literature. See, for example, [**5, 6, 7, 10, 11, 15, 16, 19, 21, 22, 25, 26, 29, 30, 32, 33, 35, 39, 41, 43**]. We mention, in particular, the seminal papers [**41, 21, 5, 6**]. These papers introduce a new paradigm which has since led to many modifications. One such modification has been proposed by Bauschke and Combettes [**5**] (see also Solodov and Svaiter [**41**]), who have modified the proximal point algorithm in order to generate a strongly convergent sequence. They introduce, for example, the following algorithm (see [**5**, Corollary 6.1 (ii), p. 258] for a single operator and $\lambda_n = 1/2$):

$$\begin{cases} x_0 \in H, \\ y_n = R_{\lambda_n A}(x_n), \\ C_n = \{z \in H : \|y_n - z\| \le \|x_n - z\|\}, \\ Q_n = \{z \in H : \langle x_0 - x_n, z - x_n\rangle \le 0\}, \\ x_{n+1} = \mathrm{P}_{C_n \cap Q_n}(x_0), \qquad n = 0, 1, 2, \ldots. \end{cases} \tag{1.2}$$

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. The mapping $R_{\lambda A} = (I + \lambda A)^{-1}$ is the classical *resolvent* of the maximal monotone operator A. 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 $\mathrm{P}_{A^{-1}(0)}$. Wei and Zhou [**42**] generalize this result to those Banach spaces X which are both uniformly convex and uniformly smooth. They introduce the following algorithm:

$$
\begin{cases}
x_0 \in X, \\
y_n = J_{\lambda_n}(x_n), \\
C_n = \{z \in X : \phi(z, y_n) \le \phi(z, x_n)\}, \\
Q_n = \{z \in X : \langle Jx_0 - Jx_n, z - x_n \rangle \le 0\}, \\
x_{n+1} = \mathrm{Q}_{C_n \cap Q_n}(x_0), \qquad n = 0, 1, 2, \ldots,
\end{cases} \tag{1.3}
$$

where J is the normalized duality mapping of the space X, $J_\lambda(x) = (J + \lambda A)^{-1} J$ and $\phi(y, x) = \|y\|^2 - 2\langle Jx, y\rangle + \|x\|^2$. 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 $\mathrm{Q}_{A^{-1}(0^*)}$. In the present paper we extend Algorithms (1.2) and (1.3) to general reflexive Banach spaces using a well chosen convex function f. More precisely, we introduce the following algorithm:

$$
\begin{cases}
x_0 \in X, \\
y_n = \mathrm{Res}^f_{\lambda_n A}(x_n), \\
C_n = \{z \in X : D_f(z, y_n) \le D_f(z, x_n)\}, \\
Q_n = \{z \in X : \langle \nabla f(x_0) - \nabla f(x_n), z - x_n \rangle \le 0\}, \\
x_{n+1} = \mathrm{proj}^f_{C_n \cap Q_n}(x_0), \qquad n = 0, 1, 2, \ldots,
\end{cases} \tag{1.4}
$$

where $\{\lambda_n\}_{n\in\mathbb{N}}$ is a given sequence of positive real numbers, Res^f_A is the resolvent of A relative to f, introduced and studied in [**4**], ∇f is the gradient of f and proj^f_C is the Bregman projection of X onto C induced by f (see Section 2.4). 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 $f(x) = (1/2)\|x\|^2$. As a matter of fact, we propose two extensions of Algorithm (1.4) (see Algorithms (4.1) and (4.4)) which approximate a common zero of several maximal monotone operators and which allow computational errors. These algorithms are similar to but different from the one we have recently studied in [**34**]. They also differ from the algorithm in [**6**] in the definition of the sets C_n and in our taking into account possible computational errors. Our main results (Theorems 1 and 2) are formulated and proved in Section 4. The next section is devoted to several preliminary definitions and results. In section 3 we prove two auxiliary results which are used in the proofs of our main results in Section 4. The behavior of Algorithm (1.4) when the operator A is zero free is analyzed in Section 5 (see Theorem 3). The sixth section contains three corollaries of Theorems 1, 2 and 3. In the seventh and last section we present an application of Theorems 1, 2 and 3.

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, int dom f, is nonempty, f is Gâteaux differentiable (see below) on int dom f, and

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

(L2) The interior of the domain of f^*, int dom f^*, is nonempty, f^* is Gâteaux differentiable on int 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 [**8**, 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 dom $f^* = X^*$.

2.2. A property of gradients. For any convex $f : X \to (-\infty, +\infty]$ we denote by dom f the set $\{x \in X : f(x) < +\infty\}$. For any $x \in$ 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 $\lim_{t\to 0}\left(f(x+ty)-f(x)\right)/t$ 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 1 (*cf.* [**34**, Proposition 2]). *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^*.*

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

$$D_f(y,x) := f(y) - f(x) - f^{\circ}(x, y-x), \tag{2.1}$$

is called the *Bregman distance with respect to f* (*cf.* [**18**]). If f is a Gâteaux differentiable function, then the Bregman distance has the following important property, called the *three point identity*: for any $x,y,z \in \operatorname{int}\operatorname{dom} f$,

$$D_f(x,y) + D_f(y,z) - D_f(x,z) = \left\langle \nabla f(z) - \nabla f(y), x-y \right\rangle. \tag{2.2}$$

Recall that, according to [**13**, Section 1.2, p. 17] (see also [**12**]), 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

$$\upsilon_f(x,t) := \inf\left\{ D_f(y,x) \ : \ y \in \operatorname{dom} f, \ \ \|y-x\| = t \right\}, \tag{2.3}$$

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

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

Examples of totally convex functions can be found, for example, in [**13, 17**]. The following proposition summarizes some properties of the modulus of total convexity.

Proposition 2 (*cf.* [**13**, 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 [**15**, Proposition 2.3, p. 39] and [**44**, Theorem 3.5.10, p. 164].

Proposition 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 results.

Proposition 4 (*cf.* [**36**, 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 [**17**]) 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 5 (*cf.* [**13**, 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 [**4**], 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 $\operatorname{prox}^f_{\iota_C}$ is called the *Bregman projection* onto C with respect to f (*cf.* [**9**]) and we denote it by proj^f_C. 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 $\operatorname{proj}^f_K(x) \in K$ satisfying

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

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

Proposition 6 (*cf.* [**17**, 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

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

For the next technical result we need to define, for any $\lambda > 0$, the *Yosida approximation* of A by

$$A_\lambda = \left(\nabla f - \nabla f \circ \operatorname{Res}^f_{\lambda A}\right)/\lambda.$$

We have the following properties of the Yosida approximation A_λ.

Proposition 7: *For any $\lambda > 0$ and for any $x \in X$, we have*

(i) $\left(\operatorname{Res}^f_{\lambda A}(x), A_\lambda(x)\right) \in \operatorname{graph} A$;

(ii) $0^* \in Ax$ *if and only if* $0^* \in A_\lambda x$.

Proof. *(i)* Indeed,

$$\begin{aligned}\operatorname{Res}^f_{\lambda A}(x) = (\nabla f + \lambda A)^{-1} \circ \nabla f(x) &\Leftrightarrow \nabla f(x) \in (\nabla f + \lambda A) \circ \operatorname{Res}^f_{\lambda A}(x) \\ &\Leftrightarrow \left(\nabla f - \nabla f \circ \operatorname{Res}^f_{\lambda A}\right)(x)/\lambda \in A\left(\operatorname{Res}^f_{\lambda A}(x)\right) \\ &\Leftrightarrow A_\lambda(x) \in A\left(\operatorname{Res}^f_{\lambda A}(x)\right).\end{aligned}$$

(ii) Indeed,

$$\begin{aligned}0^* \in Ax &\Leftrightarrow 0^* \in \lambda Ax \Leftrightarrow \nabla f(x) \in (\nabla f + \lambda A)(x) \\ &\Leftrightarrow x \in (\nabla f + \lambda A)^{-1} \circ \nabla f(x) \Leftrightarrow \nabla f(x) \in \nabla f\left(\operatorname{Res}^f_{\lambda A}(x)\right) \\ &\Leftrightarrow 0^* \in \left(\nabla f - \nabla f \circ \operatorname{Res}^f_{\lambda A}\right)(x) \Leftrightarrow 0^* \in \lambda A_\lambda x \Leftrightarrow 0^* \in A_\lambda x.\end{aligned}$$

□

Now we can prove the following important property of the resolvent.

Proposition 8: *Let $A : X \to 2^{X^*}$ be a maximal monotone operator such that $A^{-1}(0^*) \neq \varnothing$. Then*

$$D_f\left(u, \operatorname{Res}^f_{\lambda A}(x)\right) + D_f\left(\operatorname{Res}^f_{\lambda A}(x), x\right) \leq D_f(u, x)$$

for all $\lambda > 0$, $u \in A^{-1}(0^*)$ and $x \in X$.

Proof. Let $\lambda > 0$, $u \in A^{-1}(0^*)$ *and* $x \in X$ be given. By the monotonicity of A, the three point identity (2.2) and Proposition 7(i), we have

$$\begin{aligned}D_f(u, x) &= D_f\left(u, \operatorname{Res}^f_{\lambda A}(x)\right) + D_f\left(\operatorname{Res}^f_{\lambda A}(x), x\right) \\ &\quad + \left\langle \nabla f \circ \operatorname{Res}^f_{\lambda A}(x) - \nabla f(x), u - \operatorname{Res}^f_{\lambda A}(x)\right\rangle \\ &= D_f\left(u, \operatorname{Res}^f_{\lambda A}(x)\right) + D_f\left(\operatorname{Res}^f_{\lambda A}(x), x\right) + \lambda\left\langle -A_\lambda x, u - \operatorname{Res}^f_{\lambda A}(x)\right\rangle \\ &\geq D_f\left(u, \operatorname{Res}^f_{\lambda A}(x)\right) + D_f\left(\operatorname{Res}^f_{\lambda A}(x), x\right).\end{aligned}$$

□

3. Auxiliary Results

In this section we prove two lemmata which are used in the proofs of our main results in Section 4.

Lemma 1: *Let $f : X \to \mathbb{R}$ be a totally convex function. If $\{D_f(x_n, x_0)\}_{n\in\mathbb{N}}$ is bounded, then the sequence $\{x_n\}_{n\in\mathbb{N}}$ is bounded too.*

Proof. Since the sequence $\{D_f(x_n, x_0)\}_{n\in\mathbb{N}}$ is bounded, there exists $M > 0$ such that $D_f(x_n, x_0) < M$ for any $n \in \mathbb{N}$. Therefore the sequence $\{\nu_f(x_0, \|x_n - x_0\|)\}_{n\in\mathbb{N}}$ is bounded by M too, because from the definition of the modulus of total convexity (see (2.3)) we get that

$$\nu_f(x_0, \|x_n - x_0\|) \leq D_f(x_n, x_0) \leq M. \tag{3.1}$$

Since the function f is totally convex, the function $\nu_f(x, \cdot)$ is strictly increasing and positive on $(0, \infty)$ (*cf.* Proposition 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(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 (3.1). Hence the sequence $\{x_n\}_{n\in\mathbb{N}}$ is indeed bounded, as claimed. □

Lemma 2: *Let $f : X \to \mathbb{R}$ be a totally convex function and let C be a nonempty, closed and convex subset of X. Suppose that the sequence $\{x_n\}_{n\in\mathbb{N}}$ is bounded and any weak subsequential limit of $\{x_n\}_{n\in\mathbb{N}}$ belongs to C. If $D_f(x_n, x_0) \leq D_f\left(\mathrm{proj}_C^f(x_0), x_0\right)$ for any $n \in \mathbb{N}$, then $\{x_n\}_{n\in\mathbb{N}}$ converges strongly to $\mathrm{proj}_C^f(x_0)$.*

Proof. Denote $\mathrm{proj}_C^f(x_0) = \tilde{u}$. The three point identity (see (2.2)) and the assumption $D_f(x_n, x_0) \leq D_f(\tilde{u}, x_0)$ yields

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

Since $\{x_n\}_{n\in\mathbb{N}}$ is bounded there is a weakly convergent subsequence $\{x_{n_i}\}_{i\in\mathbb{N}}$ and denote its weak limit by v. We know that $v \in C$. It follows from Proposition 6(ii) that

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

Hence

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

Proposition 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 $\tilde{u} = \mathrm{proj}_C^f(x_0)$, as claimed. $\square$

4. Two Strong Convergence Theorems

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

$$(4.1)\qquad \begin{cases} x_0 \in X, \\ \eta_n^i = \xi_n^i + \frac{1}{\lambda_n^i}\left(\nabla f(y_n^i) - \nabla f(x_n)\right), \qquad \xi_n^i \in A_i y_n^i, \\ w_n^i = \nabla f^*\left(\lambda_n^i \eta_n^i + \nabla f(x_n)\right), \\ C_n^i = \left\{z \in X : D_f\left(z, y_n^i\right) \leq D_f\left(z, w_n^i\right)\right\}, \\ C_n := \cap_{i=1}^N C_n^i, \\ Q_n = \left\{z \in X : \langle \nabla f(x_0) - \nabla f(x_n), z - x_n\rangle \leq 0\right\}, \\ x_{n+1} = \mathrm{proj}_{C_n \cap Q_n}^f(x_0), \qquad n = 0, 1, 2, \ldots, \end{cases}$$

Theorem 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. Assume further that f^* is bounded and uniformly Fréchet differentiable on bounded subsets of X^*. Then, for each $x_0 \in X$, there are sequences $\{x_n\}_{n\in\mathbb{N}}$ which satisfy (4.1). 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 $\lim_{n\to+\infty} \eta_n^i = 0^*$, then each such sequence $\{x_n\}_{n\in\mathbb{N}}$ converges strongly to $\mathrm{proj}_Z^f(x_0)$ as $n \to +\infty$.*

Proof. Note that $\mathrm{dom}\,\nabla f = X$ because $\mathrm{dom}\, f = X$ and f is Legendre. Hence it follows from [**4**, Corollary 3.14(ii), p. 606] that $\mathrm{dom}\,\mathrm{Res}_{\lambda A}^f = X$. We begin with the following claim.

Claim 1: *There are sequences $\{x_n\}_{n\in\mathbb{N}}$ which satisfy (4.1).*

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 (4.1) with $\eta_n^i = 0^*$ for all $i = 1, 2, \ldots, N$ and $n \in \mathbb{N}$.

It is obvious that C_n^i are closed and convex sets for any $i = 1, 2, \ldots, N$. Hence C_n is also closed and convex. It is also obvious that Q_n is a closed and convex set. Let $u \in Z$. For any $n \in \mathbb{N}$ we have from Proposition 8 that

$$D_f\left(u, y_n^i\right) = D_f\left(u, \mathrm{Res}_{\lambda_n^i A_i}^f w_n^i\right) \leq D_f\left(u, w_n^i\right),$$

which implies that $u \in C_n^i$. Since this holds for any $i = 1, 2, \ldots, N$, it follows that $u \in C_n$. Thus $Z \subset C_n$ for any $n \in \mathbb{N}$. On the other hand it is obvious that $Z \subset Q_0 = X$. Thus $Z \subset C_0 \cap Q_0$, and therefore $x_1 = \mathrm{proj}_{C_0\cap Q_0}^f(x_0)$ is well defined. Now suppose that $Z \subset C_{n-1} \cap Q_{n-1}$ for some $n \geq 1$. Then it follows that there exists $x_n \in C_{n-1} \cap Q_{n-1}$ such that $x_n = \mathrm{proj}_{C_{n-1}\cap Q_{n-1}}^f(x_0)$ since $C_{n-1} \cap Q_{n-1}$ is a nonempty, closed and convex subset of X. So from Proposition 6(ii) we have

$$\langle \nabla f(x_0) - \nabla f(x_n), y - x_n \rangle \leq 0,$$

for any $y \in C_n \cap Q_n$. Hence we obtain that $Z \subset Q_n$. Therefore $Z \subset C_n \cap Q_n$ and hence $x_{n+1} = \mathrm{proj}^f_{C_n \cap Q_n}(x_0)$ is well defined. Consequently, we see that $Z \subset C_n \cap Q_n$ for any $n \in \mathbb{N}$. Thus the sequence we constructed is indeed well defined and satisfies (4.1), as claimed.

From now on we fix an arbitrary sequence $\{x_n\}_{n \in \mathbb{N}}$ satisfying (4.1). It is clear from the proof of Claim 1 that $Z \subset C_n \cap Q_n$ for each $n \in \mathbb{N}$.

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

It follows from the definition of Q_n and Proposition 6(ii) that $\mathrm{proj}^f_{Q_n}(x_0) = x_n$. Furthermore, by Proposition 6(iii), for each $u \in Z$, we have

$$\begin{aligned} D_f(x_n, x_0) &= D_f\left(\mathrm{proj}^f_{Q_n}(x_0), x_0\right) \\ &\leq D_f(u, x_0) - D_f\left(u, \mathrm{proj}^f_{Q_n}(x_0)\right) \\ &\leq D_f(u, x_0). \end{aligned} \tag{4.2}$$

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 by Lemma 1 the sequence $\{x_n\}_{n \in \mathbb{N}}$ is bounded too, as claimed.

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

It follows from the definition of Q_n and Proposition 6(ii) that $\mathrm{proj}^f_{Q_n}(x_0) = x_n$. Since $x_{n+1} \in Q_n$, it follows from Proposition 6(iii) that

$$D_f\left(x_{n+1}, \mathrm{proj}^f_{Q_n}(x_0)\right) + D_f\left(\mathrm{proj}^f_{Q_n}(x_0), x_0\right) \leq D_f(x_{n+1}, x_0)$$

and hence

$$D_f(x_{n+1}, x_n) + D_f(x_n, x_0) \leq D_f(x_{n+1}, x_0). \tag{4.3}$$

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 (4.3) it follows that

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

Proposition 5 now implies that $\lim_{n \to +\infty}(x_{n+1} - x_n) = 0$. Since

$$w^i_n = \nabla f^*\left(\lambda^i_n \eta^i_n + \nabla f(x_n)\right)$$

and ∇f^* is uniformly continuous on bounded subsets of X^* by Proposition 1, it follows that

$$\lim_{n \to +\infty}\left(w^i_n - x_n\right) = 0$$

for any $i = 1, 2, \ldots, N$, and hence

$$\lim_{n\to+\infty} D_f\left(x_n, w_n^i\right) = 0.$$

For any $i = 1, 2, \ldots, N$, the three point identity (see (2.2)) implies that

$$D_f\left(x_{n+1}, w_n^i\right) = D_f\left(x_{n+1}, x_n\right) - D_f\left(x_n, w_n^i\right) + \left\langle \nabla f(x_n) - \nabla f(w_n^i), x_{n+1} - x_n \right\rangle.$$

Therefore

$$\lim_{n\to+\infty} D_f\left(x_{n+1}, w_n^i\right) = 0.$$

Next, for any $i = 1, 2, \ldots, N$, it follows from the inclusion $x_{n+1} \in C_n^i$ that

$$D_f\left(x_{n+1}, y_n^i\right) \le D_f\left(x_{n+1}, w_n^i\right).$$

Hence $\lim_{n\to+\infty} D_f\left(x_{n+1}, y_n^i\right) = 0$. Proposition 5 now implies that $\lim_{n\to+\infty}\left(y_n^i - x_{n+1}\right) = 0$. Therefore, for any $i = 1, 2, \ldots, N$, we have

$$\left\|y_n^i - x_n\right\| \le \left\|y_n^i - x_{n+1}\right\| + \left\|x_{n+1} - x_n\right\| \to 0.$$

This means that the sequence $\left\{y_n^i\right\}_{n\in\mathbb{N}}$ is bounded for any $i = 1, 2, \ldots, N$. Now let $\left\{x_{n_j}\right\}_{j\in\mathbb{N}}$ be a weakly convergent subsequence of $\left\{x_n\right\}_{n\in\mathbb{N}}$ 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 1 that

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

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

$$\left\langle \eta, z - v \right\rangle \ge 0$$

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* $\left\{x_n\right\}_{n\in\mathbb{N}}$ *converges strongly to* $\operatorname{proj}_Z^f(x_0)$.

Let $\tilde{u} = \operatorname{proj}_Z^f(x_0)$. Since $x_{n+1} = \operatorname{proj}_{C_n\cap Q_n}^f(x_0)$ and Z is contained in $C_n \cap Q_n$, we have $D_f\left(x_{n+1}, x_0\right) \le D_f\left(\tilde{u}, x_0\right)$. Therefore Lemma 2 implies that $\left\{x_n\right\}_{n\in\mathbb{N}}$ converges strongly to $\tilde{u} = \operatorname{proj}_Z^f(x_0)$, as claimed. This completes the proof of Theorem 1. □

We now present another result which is similar to Theorem 1, but with a different type of errors. More precisely, we study the following algorithm when

$Z := \bigcap_{i=1}^{N} A_i^{-1}(0^*) \neq \varnothing$:

$$\begin{cases} x_0 \in X, \\ y_n^i = \mathrm{Res}^f_{\lambda_n^i A_i}(x_n + e_n^i), \\ C_n^i = \left\{z \in X : D_f\left(z, y_n^i\right) \leq D_f\left(z, x_n + e_n^i\right)\right\}, \\ C_n := \cap_{i=1}^N C_n^i, \\ Q_n = \{z \in X : \langle \nabla f(x_0) - \nabla f(x_n), z - x_n\rangle \leq 0\}, \\ x_{n+1} = \mathrm{proj}^f_{H_n \cap W_n}(x_0), \qquad n = 0, 1, 2, \ldots, \end{cases} \tag{4.4}$$

Theorem 2: *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 (4.4). If, for each $i = 1, 2, \ldots, N$, $\liminf_{n\to+\infty} \lambda_n^i > 0$, and the sequences of errors $\left\{e_n^i\right\}_{n\in\mathbb{N}} \subset X$ satisfy $\lim_{n\to+\infty} e_n^i = 0$, then each such sequence $\{x_n\}_{n\in\mathbb{N}}$ converges strongly to $\mathrm{proj}^f_Z(x_0)$ as $n \to +\infty$.*

Proof. Note that $\mathrm{dom}\, \nabla f = X$ because $\mathrm{dom}\, f = X$ and f is Legendre. Hence it follows from [**4**, Corollary 3.14(ii), p. 606] that $\mathrm{dom}\, \mathrm{Res}^f_{\lambda A} = X$. We begin with the following claim.

Claim 1: *There are sequences $\{x_n\}_{n\in\mathbb{N}}$ which satisfy (4.4).*

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 (4.4) with $e_n^i = 0$ for all $i = 1, 2, \ldots, N$ and $n \in \mathbb{N}$.

It is obvious that C_n^i are closed and convex sets for any $i = 1, 2, \ldots, N$. Hence C_n is also closed and convex. It is also obvious that Q_n is a closed and convex set. Let $u \in Z$. For any $n \in \mathbb{N}$, we obtain from Proposition 8 that

$$D_f\left(u, y_n^i\right) = D_f\left(u, \mathrm{Res}^f_{\lambda_n^i A_i}(x_n + e_n^i)\right) \leq D_f\left(u, x_n + e_n^i\right),$$

which implies that $u \in C_n^i$. Since this holds for any $i = 1, 2, \ldots, N$, it follows that $u \in C_n$. Thus $Z \subset C_n$ for any $n \in \mathbb{N}$. On the other hand it is obvious that $Z \subset Q_0 = X$. Thus $Z \subset C_0 \cap Q_0$, and therefore $x_1 = \mathrm{proj}^f_{C_0\cap Q_0}(x_0)$ is well defined. Now suppose that $Z \subset C_{n-1} \cap Q_{n-1}$ for some $n \geq 1$. The it follows that there exists $x_n \in C_{n-1} \cap Q_{n-1}$ such that $x_n = \mathrm{proj}^f_{C_{n-1}\cap Q_{n-1}}(x_0)$ since $C_{n-1} \cap Q_{n-1}$ is a nonempty, closed and convex subset of X. So from Proposition 6(ii) we have

$$\langle \nabla f(x_0) - \nabla f(x_n), y - x_n\rangle \leq 0,$$

for any $y \in C_n \cap Q_n$. Hence we obtain that $Z \subset Q_n$. Therefore $Z \subset C_n \cap Q_n$ and hence $x_{n+1} = \mathrm{proj}^f_{C_n\cap Q_n}(x_0)$ is well defined. Consequently, we see that $Z \subset C_n \cap Q_n$ for any $n \in \mathbb{N}$. Thus the sequence we constructed is indeed well defined and satisfies (4.4), as claimed.

From now on we fix an arbitrary sequence $\{x_n\}_{n\in\mathbb{N}}$ satisfying (4.4). It is clear from the proof of Claim 1 that $Z \subset C_n \cap Q_n$ for each $n \in \mathbb{N}$.

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

It follows from the definition of Q_n and Proposition 6(ii) that $\operatorname{proj}^f_{Q_n}(x_0) = x_n$. Furthermore, by Proposition 6(iii), for each $u \in Z$, we have

$$\begin{aligned} D_f(x_n, x_0) &= D_f\left(\operatorname{proj}^f_{Q_n}(x_0), x_0\right) \\ &\le D_f(u, x_0) - D_f\left(u, \operatorname{proj}^f_{Q_n}(x_0)\right) \\ &\le D_f(u, x_0). \end{aligned} \tag{4.5}$$

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 by Lemma 1 the sequence $\{x_n\}_{n\in\mathbb{N}}$ is bounded too, as claimed.

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

It follows from the definition of Q_n and Proposition 6(ii) that $\operatorname{proj}^f_{Q_n}(x_0) = x_n$. Since $x_{n+1} \in Q_n$, it follows from Proposition 6(iii) that

$$D_f\left(x_{n+1}, \operatorname{proj}^f_{Q_n}(x_0)\right) + D_f\left(\operatorname{proj}^f_{Q_n}(x_0), x_0\right) \le D_f(x_{n+1}, x_0)$$

and hence

$$D_f(x_{n+1}, x_n) + D_f(x_n, x_0) \le D_f(x_{n+1}, x_0). \tag{4.6}$$

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 (4.6) it follows that

$$\lim_{n\to+\infty} D_f(x_{n+1}, x_n) = 0. \tag{4.7}$$

Proposition 5 now implies that $\lim_{n\to+\infty}(x_{n+1} - x_n) = 0$. For any $i = 1, 2, \ldots, N$, it follows from the definition of the Bregman distance (see (2.1)) that

$$\begin{aligned} D_f\left(x_n, x_n + e^i_n\right) = f(x_n) - f\left(x_n + e^i_n\right) - \left\langle \nabla f(x_n + e^i_n), x_n - \left(x_n + e^i_n\right)\right\rangle = \\ f(x_n) - f\left(x_n + e^i_n\right) + \left\langle \nabla f(x_n + e^i_n), e^i_n\right\rangle. \end{aligned}$$

The function f is bounded on bounded subsets of X and therefore ∇f is bounded on bounded subsets of X (see [**13**, Proposition 1.1.11, p. 17]). In addition, f is uniformly Fréchet differentiable and therefore f is uniformly continuous on bounded subsets (see [**1**, Theorem 1.8, p. 13]). Hence, since $\lim_{n\to+\infty} e^i_n = 0$, it follows that

$$\lim_{n\to+\infty} D_f\left(x_n, x_n + e^i_n\right) = 0. \tag{4.8}$$

For any $i = 1, 2, \ldots, N$, it follows from the three point identity (see (2.2)) that

$$D_f\left(x_{n+1}, x_n + e_n^i\right) = D_f\left(x_{n+1}, x_n\right) + D_f\left(x_n, x_n + e_n^i\right) + \left\langle \nabla f(x_n) - \nabla f(x_n + e_n^i), x_{n+1} - x_n \right\rangle.$$

Since $\lim_{n\to+\infty}\left(x_{n+1} - x_n\right) = 0$ and ∇f is bounded on bounded subsets of X, (4.7) and (4.8) imply that

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

For any $i = 1, 2, \ldots, N$, it follows from the inclusion $x_{n+1} \in C_n^i$ that

$$D_f\left(x_{n+1}, y_n^i\right) \le D_f\left(x_{n+1}, x_n + e_n^i\right).$$

Hence $\lim_{n\to+\infty} D_f\left(x_{n+1}, y_n^i\right) = 0$. Proposition 5 now implies that $\lim_{n\to+\infty}\left(y_n^i - x_{n+1}\right) = 0$. Therefore, for any $i = 1, 2, \ldots, N$, we have

$$\left\|y_n^i - x_n\right\| \le \left\|y_n^i - x_{n+1}\right\| + \left\|x_{n+1} - x_n\right\| \to 0.$$

This means that the sequence $\left\{y_n^i\right\}_{n\in\mathbb{N}}$ is bounded for any $i = 1, 2, \ldots, N$. Now let $\left\{x_{n_j}\right\}_{j\in\mathbb{N}}$ be a weakly convergent subsequence of $\left\{x_n\right\}_{n\in\mathbb{N}}$ 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 $i = 1, 2, \ldots, N$. Let $\xi_n^i \in Ay_n^i$, since $\liminf_{n\to+\infty} \lambda_n^i > 0$ and $\lim_{n\to+\infty} e_n^i = 0$, it follows from Proposition 1 that

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

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

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

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

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

for all $(z, \eta) \in \text{graph}\,(A_i)$. 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* $\left\{x_n\right\}_{n\in\mathbb{N}}$ *converges strongly to* $\text{proj}_Z^f(x_0)$.

Let $\tilde{u} = \text{proj}_Z^f(x_0)$. Since $x_{n+1} = \text{proj}_{C_n\cap Q_n}^f(x_0)$ and Z is contained in $C_n \cap Q_n$, we have $D_f\left(x_{n+1}, x_0\right) \le D_f\left(\tilde{u}, x_0\right)$. Therefore Lemma 2 implies that $\left\{x_n\right\}_{n\in\mathbb{N}}$ converges strongly to $\tilde{u} = \text{proj}_Z^f(x_0)$, as claimed. This completes the proof of Theorem 2. □

5. Zero Free Operators

This section concerns the case where our two algorithms are applied to a single zero free operator A. In this case both our algorithms take the form

$$(5.1)\qquad \begin{cases} x_0 \in X, \\ \eta_n = \xi_n + \frac{1}{\lambda_n}\left(\nabla f(y_n) - \nabla f(x_n)\right), \qquad \xi_n \in Ay_n, \\ w_n = \nabla f^*\left(\lambda_n \eta_n + \nabla f(x_n)\right), \\ C_n = \left\{z \in X : D_f\left(z, y_n\right) \leq D_f\left(z, x_n\right)\right\}, \\ Q_n = \left\{z \in X : \langle \nabla f(x_0) - \nabla f(x_n), z - x_n \rangle \leq 0\right\}, \\ x_{n+1} = \mathrm{proj}^f_{C_n \cap Q_n}(x_0), \qquad n = 0, 1, 2, \ldots, \end{cases}$$

and

$$(5.2)\qquad \begin{cases} x_0 \in X, \\ y_n = \mathrm{Res}^f_{\lambda_n A}(x_n + e_n), \\ C_n = \left\{z \in X : D_f\left(z, y_n\right) \leq D_f\left(z, x_n + e_n\right)\right\}, \\ Q_n = \left\{z \in X : \langle \nabla f(x_0) - \nabla f(x_n), z - x_n \rangle \leq 0\right\}, \\ x_{n+1} = \mathrm{proj}^f_{C_n \cap Q_n}(x_0), \qquad n = 0, 1, 2, \ldots, \end{cases} .$$

We first recall the following lemma (see [**34**, Lemma 1]):

Lemma 3: *If $A : X \to 2^{X^*}$ is a maximal monotone operator with bounded domain, then $A^{-1}\left(0^*\right) \neq \varnothing$.*

Now we can prove that the generation of an infinite sequence by Algorithm (5.1) or (5.2) does not depend on the zero set $A^{-1}\left(0^*\right)$ of A being not empty.

Theorem 3. *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. In case of Algorithm (5.1) assume, in addition, that f^* is bounded and uniformly Fréchet differentiable on bounded subsets of X^*. Then, for each $x_0 \in X$, there are sequences $\left\{x_n\right\}_{n\in\mathbb{N}}$ which satisfy either (5.1) or (5.2). If $\liminf_{n\to+\infty} \lambda_n > 0$, and either the sequence of errors $\left\{\eta_n\right\}_{n\in\mathbb{N}} \subset X^*$ satisfies $\lim_{n\to+\infty} \eta_n = 0^*$ or the sequence of errors $\left\{e_n\right\}_{n\in\mathbb{N}} \subset X$ satisfies $\lim_{n\to+\infty} e_n = 0$, then either $A^{-1}\left(0^*\right) \neq \varnothing$ and each such sequence $\left\{x_n\right\}_{n\in\mathbb{N}}$ converges strongly to $\mathrm{proj}^f_{A^{-1}(0^*)}(x_0)$ or $A^{-1}\left(0^*\right) = \varnothing$ and each such sequence $\left\{x_n\right\}_{n\in\mathbb{N}}$ satisfies $\lim_{n\to+\infty} \left\|x_n\right\| = +\infty$.*

Proof. In view of Theorem 1 and Theorem 2, we only need to consider the case where $A^{-1}\left(0^*\right) = \varnothing$. First of all we prove that in this case, for each $x_0 \in X$, there is a sequence $\left\{x_n\right\}_{n\in\mathbb{N}}$ which satisfies either (5.1) with $\eta_n = 0$ or (5.2) with $e_n = 0$ for all $n \in \mathbb{N}$.

We prove this by induction. We first check that the initial step ($n = 0$) is well defined. Indeed, 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 $x = \mathrm{Res}^f_{\lambda_0 A}(x_0)$ and this problem does have a solution since $\mathrm{dom}\,\mathrm{Res}^f_{\lambda A} = X$ (see Proposition 3 and [**4**, Theorem 3.13(iv), p. 606]). Now note that $Q_0 = X$. Since C_0 cannot be empty ($y_0 \in C_0$), the next iterate x_1 can be generated; it is the Bregman projection of x_0 onto $C_0 = Q_0 \cap C_0$.

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 $z_0 \in \mathrm{dom}\,A$ and define

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

and

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

Then $h : X \to (-\infty, +\infty]$ is a proper, convex and lower semicontinuous function, its subdifferential ∂h is maximal monotone (see [**31**, Theorem 2.13, p. 124]), and

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

is also maximal monotone (see [**37**]). 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 Theorems 1 and 2 applied to the problem $0^* \in A'(x)$. Since A' has a bounded effective domain, this problem has a solution by Lemma 3. Thus it follows from Claim 1 in the proofs of Theorems 1 and 2 that $x_{\hat{n}+1}$ is well defined in both Algorithms (5.1) and (5.2). 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 proofs of Theorems 1 and 2 that A had a zero. Therefore if $A^{-1}(0^*) = \varnothing$, then $\lim_{n\to+\infty}\|x_n\| = +\infty$, as asserted. □

6. Consequences of the Strong Convergence Theorems

Algorithm (1.4) is a special case of Algorithm (5.1) when $\eta_n = 0$ for all $n \in \mathbb{N}$, and a special case of Algorithm (5.2) when $e_n = 0$ for all $n \in \mathbb{N}$. Hence as a direct consequence of Theorems 1, 2 and 3 we obtain the following result:

Corollary 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 1, 2 and 3 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 [**14**, 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 1, 2 and 3 hold in this context and lead us to the following two results which, in some sense, complement Theorem 3.1 in [**42**] (see also Theorem 3.5 in [**29**]).

Corollary 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 $\mathrm{Q}_{A^{-1}(0^*)}(x_0)$ as $n \to +\infty$, or $A^{-1}(0^*) = \varnothing$ and $\lim_{n\to+\infty} \|x_n\| = +\infty$.*

Corollary 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 in the presence of computational errors as in Theorems 1, 2 and 3.

7. An Application of the Strong Convergence Theorems

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

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

Applying Theorems 1, 2 and 3 to the subdifferential of g, we obtain an algorithm for finding minimizers of g.

Proposition 9. *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*

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

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 because g is a proper, convex and lower semicontinuous function (see [**31**, Theorem 2.13, p. 124]). Since the zero set of ∂g coincides with the set of minimizers of g, Proposition 9 follows immediately from Theorems 1, 2 and 3. □

Note that in this case

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

is equivalent to

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

8. Acknowledgements

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. Both authors thank the referee for several helpful comments.

References

[1] Ambrosetti, A. and Prodi, G.: A primer of nonlinear analysis, *Cambridge University Press*, Cambridge, 1993.

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

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

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

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

[6] Bauschke, H. H. and Combettes, P. L.: Construction of best Bregman approximations in reflexive Banach spaces, *Proc. Amer. Math. Soc.* **131** (2003), 3757–3766.

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

[8] Bonnans, J. F. and Shapiro, A.: Perturbation analysis of optimization problems, *Springer Verlag*, New York, 2000.

[9] Bregman, L. M.: 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.

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

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

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

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

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

[15] Butnariu, D., Iusem, A. N. and Zălinescu, C.: 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.

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

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

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

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

[20] Cioranescu, I.: Geometry of Banach spaces, duality mappings and nonlinear problems, *Kluwer Academic Publishers*, Dordrecht, 1990.

[21] Combettes, P. L.: Strong convergence of block-iterative outer approximation methods for convex optimization, *SIAM J. Control Optim.* **38** (2000), 538–565.

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

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

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

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

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

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

[28] Martinet, B.: Régularisation d'inéquations variationelles par approximations successives, *Revue Française d'Informatique et de Recherche Opérationelle* **4** (1970), 154–159.

[29] Nakajo, K. and Takahashi, W.: Strong convergence theorems for nonexpansive mappings and nonexpansive semigroups, *J. Math. Anal. Appl.* **279** (2003), 372–379.

[30] Nevanlinna, O. and Reich, S.: Strong convergence of contraction semigroups and of iterative methods for accretive operators in Banach spaces, *Israel J. Math.* **32** (1979), 44–58.

[31] Pascali, D. and Sburlan, S.: Nonlinear mappings of monotone type, *Sijthoff & Nordhoff International Publishers,* Alphen aan den Rijn*,* 1978.

[32] Reich, S.: Weak convergence theorems for nonexpansive mappings in Banach spaces, *J. Math. Anal. Appl*. **67** (1979), 274–276.

[33] Reich, S.: 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, 313–318.

[34] Reich, S. and Sabach, S.: A strong convergence theorem for a proximal-type algorithm in reflexive Banach spaces, preprint, 2009.

[35] Reich, S. and Zaslavski, A. J.: Infinite products of resolvents of accretive operator, *Topol. Methods Nonlinear Anal.* **15** (2000), 153–168.

[36] Resmerita, E.: On total convexity, Bregman projections and stability in Banach spaces, *J. Convex Anal.* **11** (2004), 1–16.

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

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

[39] Rockafellar, R. T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming, *Math. Oper. Res.* **1** (1976), 97–116.

[40] Rockafellar, R. T. and Wets, R. J.-B.: Variational analysis, *Springer Verlag*, Berlin, 1998.

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

[42] Wei, L. and Zhou, H. Y.: Projection scheme for zero points of maximal monotone operators in Banach spaces, *J. Math. Res. Exposition* **28** (2008), 994–998.

[43] Yao, J. C. and Zeng, L. C.: An inexact proximal-type algorithm in Banach spaces, *J. Optim. Theory Appl.* **135** (2007), 145–161.

[44] Zălinescu, C.: Convex analysis in general vector spaces, *World Scientific Publishing*, Singapore, 2002.

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