Journal of Nonlinear and Convex Analysis
Volume 15, Number 2, 2014, 379–397

Yokohama Publishers
ISSN 1880-5221 ONLINE JOURNAL

SPLITTING METHODS FOR FINDING ZEROES OF SUMS OF MAXIMAL MONOTONE OPERATORS IN BANACH SPACES

ALFREDO N. IUSEM* AND BENAR F. SVAITER†

Honoring Simeon Reich on his 65th birthday

Abstract. We introduce a general scheme for finding zeroes of the sum of two maximal monotone operators in a reflexive Banach space X. It generates a sequence in the product space $X \times X^*$, where X^* is the dual of X. It is essentially a projection method, in the sense that in each iteration a hyperplane is constructed, separating the current iterate from a generalized solution set, whose projection onto X in indeed the solution set of the problem, and then the next iterate is taken as the projection of the current one onto this separating hyperplane. In order to construct such hyperplane, two proximal-like steps are taken from the current iterate, each one using only one of the two maximal monotone operators. Thus, the resulting procedure is a *splitting method*, which solves subproblems involving only one of the two operators. Similarly to other methods designed for Banach spaces, auxiliary functions, giving rise to *Breman distances* and *Bregman projections*, are used in both the proximal-like step and in the projection step of the scheme. A full convergence analysis is presented.

1. Introduction

Let X be a Banach space and X^* the associated dual space. Consider set-valued maximal monotone operators $A, B : X \to \mathcal{P}(X^*)$. The problem of interest, to be denoted as problem P, consists of finding $z^* \in X$ such that

$$0 \in A(z^*) + B(z^*). \tag{1.1}$$

It is frequently the case that the problems of finding zeroes of A and B separately are relatively easy, or at least considerably easier than solving P.

A prototypical instance of this situation is the *Variational Inequality Problem* VIP(T,C), associated to a maximal monotone $T : X \to \mathcal{P}(X^*)$ and a closed and convex subset C of X. It consists of finding $z^* \in C$ such that there exists $u^* \in T(x^*)$ satisfying $\langle z^* - z, u^* \rangle \geq 0$ for all $z \in C$. The solutions of VIP(T,C) are precisely the zeroes of $T + N_C$, where N_C is the normal operator associated to C, known to be maximal monotone.

For problems of the above mentioned type, it is natural to consider iterative algorithms which at each step solve subproblems involving either A or B, but not both. Such algorithms are called *splitting methods*, or *decomposition methods*. Methods

---

2010 *Mathematics Subject Classification.* 47H05, 90C25, 49M27.

Key words and phrases. Splitting methods, projection methods, reflexive Banach spaces, Douglas-Rachford method.

*The work of this author author was partially supported by CNPq grant no. 301280/86.

†The work of this author author was partially supported by CNPq grants no. 474944/2010-7, 303583/2008-8 and FAPERJ grant E-26/110.821/2008.

of this kind were originally developed in a linear algebra context, i.e., for the case in which X is finite dimensional and A, B are single-valued affine operators (not necessarily monotone), so that (1.1) reduces to solving a system of linear equations. In fact, the most classical iterative methods for solving systems of linear equations, namely Jacobi's and Gauss-Seidel's, as well as their more advanced versions SOR and JOR, can be cast in the framework of splitting methods (see [11]).

Moving now to the realm of nonlinear operators, special attention has been given to the case in which both A and B are *maximal monotone*, which allows for much stronger results, both in terms of existence of the iterates and convergence of the generated sequence. In the sequel, we will deal exclusively with a pair (A, B) of maximal monotone operators.

Next we comment on splitting methods for solving problem P in the particular case in which X is a Hilbert space. Three basic families of splitting methods for this problem were identified in [14]:

i) The *Douglas/Peaceman-Rachford* family, whose iteration is given by:

$$y^k = [2(I + \xi B)^{-1} - I]x^k,$$
$$z^k = [2(I + \xi A)^{-1} - I]y^k,$$
$$x^{k+1} = (1 - \rho_k)x^k + \rho_k z^k, \tag{1.2}$$

where $\xi > 0$ is a fixed scalar, and $\{\rho_k\} \subset (0, 1]$ is a sequence of relaxation parameters.

ii) The *double backward* splitting method, with iteration given by:

$$y^k = (I + \lambda_k B)^{-1}x^k,$$
$$x^{k+1} = (I + \lambda_k A)^{-1}y^k,$$

where $\{\lambda_k\} \subset \mathbb{R}_{++}$ is a sequence of regularization parameters.

iii) The *forward-backward* splitting method, with iteration given by:

$$y^k \in (I - \lambda_k A)x^k,$$
$$x^{k+1} = (I + \lambda_k B)^{-1}y^k,$$

with λ_k as in (ii).

Note first that all these are splitting methods, in the sense that each sub-step requires solving an inclusion involving only A or B. The maximal monotonicity of A and B ensures that y^k and z^k in case (i) are uniquely determined, and so the same happens with x^{k+1}; this is also the case for y^k, x^{k+1} in case (ii), but for the forward-backward method (iii) y^k fails to be uniquely determined. In connection with (i), we mention that the well known *Peaceman-Rachford* method corresponds to taking $\rho_k = 1$ for all k in (1.2); taking $\rho_k = 1/2$ for all k in (1.2) produces the classical *Douglas-Rachford* method. Convergence results for these two special cases were established in [20], under some additional hypotheses on A, B for the case of Peaceman-Rachford. Convergence results for the general scheme (i), in the case in which $\{\rho_k\}$ is contained in a compact subset of $(0, 1)$, can be found in [12]. See also [13] and [18] for additional insights on the scheme presented in (i).

The convergence analysis of the double backward scheme given by (ii), which can be found in [19] and [21], establishes much weaker convergence properties (without additional assumptions on A, B, besides maximal monotonicity): the sequence $\{\lambda_k\}$ must converge to 0 in a particular way, and the sequence which is proved to converge to a zero of $A + B$ is not $\{x^k\}$, but rather an "ergodic" average of $\{x^k\}$.

The forward-backward scheme (iii) is computationally less demanding, since it requires the solution of only one inclusion per iteration (it can be seen indeed as a generalization of the projected gradient method for convex optimization). On the other hand, the standard convergence analysis for this method (see [23]), requires that A be single-valued and furthermore co-coercive, and the parameters λ_k must have an upper bound related to A.

A substantial progress in this area was achieved in [14], which presents a new scheme, generating a sequence in the product space $X \times X$, for which quite solid convergence results were established. Several of the previously known splitting method turned out to be special cases of the scheme developed in [14], while others, e.g. Douglas-Rachford, were identified as "excluding limiting" cases of this scheme, corresponding to values of the parameters lying in the boundary of the region for which convergence was established. Later on, a convergence analysis of Douglas-Rachford method along the lines of [14] was presented in [22]. In general, the convergence results in [14] and [22] proved to be stronger than those in the previous literature.

The scheme in [14], which is the departure point for the method in this paper, is essentially a projection method in the space $X \times X$. Consider the set $S_e(A, B) \subset X \times X$ defined as $S_e(A, B) = \{(z, w) : -w \in A(z), w \in B(z)\}$. Clearly, given $(z, w) \in S_e(A, B)$, one has that $0 = -w + w \in A(z) + B(z)$, i.e., the first component of a pair in $S_e(A, B)$ solves P. The basic ingredient of the scheme in [14] consists of, given a pair $(z, w) \notin S_e(A, B)$, constructing a hyperplane in $X \times X$ which separates (z, w) from $S_e(A, B)$. The parameters of such hyperplane are obtained by solving two inclusions, one involving only A and the other one only B, thus ensuring the splitting nature of the algorithm. Then, the iterative scheme works by taking the orthogonal projection of the current iterate (z^k, w^k) onto the hyperplane which separates it from $S_e(A, B)$ as the next iterate (z^{k+1}, w^{k+1}).

Formally, the algorithm, to be refered as Algorithm ES in the sequel, proceeds as follows:

i) Start the method with $p^0 = (z^0, w^0) \in X \times X$, and choose an exogenous sequence of relaxation parameters $\{\rho_k\}$ contained in a compact subset of $(0, 2)$, and two exogenous sequences of regularization parameters $\{\lambda_k\}, \{\mu_k\}$ contained in a compact subset of $(0, \infty)$.

ii) Given $p^k = (z^k, w^k) \in X \times X$, find (x^k, b^k), $(y^k, a^k) \in X \times X$ satisfying:

$$x^k + \lambda_k b^k = z^k + \lambda_k w^k, \qquad b^k \in B(x^k),$$

$$y^k + \mu_k a^k = z^k - \mu_k w^k, \qquad a^k \in A(y^k).$$

iii) Define $\varphi_k : X \times X \to \mathbb{R}$ as $\varphi_k(z, w) = \langle z - x^k, b^k - w\rangle + \langle z - y^k, a^k + w\rangle$.

iv) Define the halfspace $H_k \subset X \times X$ as $H_k = \{(z, w) : \varphi_k(z, w) \leq 0\}$.

v) Compute $\bar{p}^k$, the orthogonal projection of p^k onto H_k.

vi) Define the next iterate as $p^{k+1} = p^k + \rho_k(\bar{p}^k - p^k)$.

The basic algorithm in [14] uses also a second exogenous sequence $\{\alpha_k\}$ of relaxation parameters, which we omit here, because it is absent from the algorithm to be developed in this paper; the scheme just presented corresponds to the case of $\alpha_k = 1$ for all k.

The main convergence result for the method above, established in Proposition 3 of [14], is the following: if A, B and $A+B$ are maximal monotone, and $A+B$ has zeroes, then the sequences $\{z^k\}, \{x^k\}$ and $\{y^k\}$ converge weakly to some zero z^* of $A+B$, the sequences $\{w^k\}$ and $\{b^k\}$ converge weakly to a point w^* such that $w^* \in B(z^*), -w^* \in A(z^*)$, and finally the sequence $\{a^k\}$ converges weakly to $-w^*$.

Later on, it was shown that the hypothesis of maximal monotonicity of $A+B$ (which in general does not follow from maximal monotonicity of A and B), can be removed; see [1].

The main purpose of this paper is to develop an algorithm based upon ES with good convergence properties for solving problem P in Banach spaces. To our knowledge, this is the first splitting algorithm for solving monotone inclusions in Banach spaces.

The main obstacle in pursuing this goal is the following: a basic property of the orthogonal projection onto a closed and convex set C in a Hilbert space is that, when moving from a point z to its orthogonal projection onto C, the norm-induced distance to *any* point in C decreases. This property is lost in Banach spaces, if we replace the orthogonal projection onto C by the *metric projection* $\Pi_C : X \to C$, defined as $\Pi_C(x) = \arg\min_{y\in C} \|x-y\|^2$. This failure is due to the fact that the derivative of the square of the norm in a nonhilbertian Banach space is not linear, while in a Hilbert space it is just twice the identity operator I.

In a Banach space, in order to recover the decreasing distance property of the orthogonal projection, one should minimize not the norm-induced distance, but rather the so called *Bregman distance*, introduced in [2], which can be defined as follows: Let $f(x) = \frac{1}{2}\|x\|^2$. Assume that X is such that f is Gâteaux differentiable, and define $D_f : X \times X \to \mathbb{R}$ as

$$D_f(x,y) = f(x) - f(y) - \langle x-y, f'(y)\rangle, \tag{1.3}$$

where $\langle\cdot,\cdot\rangle : X \times X^* \to \mathbb{R}$ denotes the duality coupling (i.e., $\langle z, w\rangle = w(z)$), and $f' : X \to X^*$ is the Gâteaux derivative of f. If we define now the *Bregman projection* P_f^C onto a closed and convex set $C \subset X$ as $P_f^C(x) = \arg\min_{y\in C} D_f(y,x)$, it happens to be the case that P_f^C enjoys several of the desirable properties of the orthogonal projections in Hilbert spaces, as we will explain in the following section (Bregman distances and projections have been defined also for the case in which f is not differentiable, see e.g. [8], but we will not be concerned with this issue in the sequel).

Another feature of nonhilbertian Banach spaces is that the square of the norm loses its privileged standing: when working, for instance, in $\mathcal{L}^p$ or ℓ_p spaces, calculations become simpler if we take $f(x) = \frac{1}{p}\|x\|^p$ in (1.3), instead of $f(x) = \frac{1}{2}\|x\|^2$. Thus, it has become customary to consider a rather general auxiliary function $f : X \to \mathbb{R}$ in (1.3), in order to define the Bregman distance and projection (see e.g. [4], [7], [15], [16]). The specific properties of f needed for convergence of the method, as well as examples of functions satisfying these properties, will be exhibited in Section 2.

Once the Bregman distance related to f is introduced in the projection step of Algorithm ES (item (v) above), one needs to match this step to the "proximal step", i.e., item (ii). It can be seen that the computation of x^k, y^k, a^k, b^k is akin to the performance of an iteration of the proximal point method starting from z^k using either the operator A or B. In a Hilbert space, the proximal resolvent $(I+\tau A)^{-1}$ of a maximal monotone operator A, with a positive regularization parameter τ, can also be seen as a sort of projection, in the sense that $(I+\tau A)^{-1}(z)$ is closer than z to any zero of A. Once again, this approximation property is lost in Banach spaces. In order to recover it, one must use instead the generalized resolvent $(f'+\tau A)^{-1}$, where $f : X \to \mathbb{R}$ enjoys the same properties that give a good behavior to the Bregman distance D_f and the Bregman projection P_f^C.

Note that in a "Banach version" of the projection step (item(v)) of Algorithm ES we need an auxiliary function defined on $X \times X^*$, while for the proximal step (item(ii)), we need a function defined just on X.

We will present in Section 3 a method based on Algorithm ES, appropriate for a rather general class of Banach spaces. Its convergence behavior will be established in Section 4. We will prove a convergence theorem rather close to the above described Proposition 3 in [14], thus recovering most of the strength of the convergence properties which hold for Algorithm ES in Hilbert spaces. We remark that the proofline of our convergence analysis is quite diffferent from (and in fact much simpler than) that in [14], despite the additional complications resulting from working in Banach spaces, as compared to Hilbert ones.

2. Preliminaries

We begin with some material related to the Bregman distances and projections to be used in our algorithm. Most of the results presented in this section were established in [6], [8], [17] and [22].

In this section, $f : X \to \mathbb{R}$ is strictly convex, lower semicontinuous and Gâteaux differentiable, and $f' : X \to X^*$ is its Gâteaux derivative. We will denote the family of such functions as $\mathcal{F}(X)$.

The *Bregman distance* $D_f : X \times X^* \to \mathbb{R}$ is defined as

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

We start with two elementary properties of Bregman distances.

Proposition 2.1.

i) $D_f(x,y) \geq 0$ *for all* $x, y \in X$, *and* $D_f(x,y) = 0$ *if and only if* $x = y$.

ii) $D_f(x,y) + D_f(y,x) = \langle x-y, f'(x) - f'(y)\rangle$ *for all* $x, y \in X$.

Proof. Item (i) follows from the strict convexity of f, and item(ii) is an immediate consequence of (2.1). □

The next result is known as the *Four-point Lemma* for Bregman distances.

Lemma 2.2. *Take* $f \in \mathcal{F}(X)$. *Then*

$$D_f(w,z) - D_f(w,x) - D_f(y,z) + D_f(y,x) = \langle w-y, f'(x) - f'(z)\rangle \tag{2.2}$$

for all $w, x, y, z \in X$.

Proof. Follows easily from (2.1). □

We will use in the sequel the *modulus of total convexity* $\nu_f : X \times \mathbb{R}_+ \to \mathbb{R}$ defined as

$$\nu_f(x,t) = \inf_{y \in \{y \in X : \|y-x\|=t\}} D_f(y,x),$$

with D_f as in (2.1). If $f \in \mathcal{F}(X)$ is such that $\nu_f(x,t) > 0$ for all $x \in X$ and all $t > 0$, then f is said to be *totally convex*. In finite dimensional spaces total convexity is equivalent to strict convexity, but in infinite dimensional spaces total convexity is more demanding that strict convexity, though less demanding than uniform convexity (see [8]).

The methods we analyze in this paper use, as an auxiliary device, functions $f \in \mathcal{F}(X)$, $\mathcal{F}(X^*)$ or $\mathcal{F}(X \times X^*)$ which satisfy some or all of the following assumptions:

H0: f is coercive, i.e. $\lim_{\|x\| \to \infty} \frac{f(x)}{\|x\|} = \infty$.
H1: The level sets of $D_f(x, \cdot)$ are bounded for all $x \in X$.
H2: $\inf_{x \in C} \nu_f(x,t) > 0$, for all bounded set $C \subset X$ and all $t \in \mathbb{R}_{++}$.
H3: f' is uniformly continuous on bounded subsets of X.
H4: f' is onto.
H5: f' is weak-to-weak* continuous.

It is important to exhibit functions which satisfy these properties in as large a class of Banach spaces as possible, and we focuse our attention on Banach spaces which are reflexive, uniformly convex and uniformly smooth, and on functions of the form $f_r(x) = \frac{1}{r}\|x\|^r$ with $r > 1$.

Our results on the validity of H0–H5 for $f_r(x) = \frac{1}{r}\|x\|^r$ are summarized in the following proposition.

Proposition 2.3.

i) *If X is a reflexive, uniformly smooth and uniformly convex Banach space, then $f_r(x) = \frac{1}{r}\|x\|^r$ satisfies H0, H1, H2, H3 and H4 for all $r > 1$.*
ii) *If X is a Hilbert space, then $f_2(x) = \frac{1}{2}\|x\|^2$ satisfies H5. If $X = \ell_p$ ($1 < p < \infty$) then $f_p(x) = \frac{1}{p}\|x\|_p^p$ satisfies H5.*

Proof. i) For H1-H4, see Proposition 2 in [17], in whose proof several results from [9] are invoked. The result is immediate for H0, since $r > 1$.
ii) In the case of a Hilbert space, f_2' is the identity, which is certainly weak-to-weak continuous. The result for f_p in ℓ_p has been proved in Proposition 8.2 of [3].

□

We refer to [10] for the definitions of uniformly smooth and uniformly convex Banach spaces. We mention that the spaces ℓ_p, $\mathcal{L}^p[\alpha,\beta]$ and the Sobolev spaces $W^{p,m}$ (in all cases with $1 < p < \infty$), are uniformly smooth and uniformly convex.

Unfortunately, it has been proved in [8] that for $X = \ell_p$ or $X = \mathcal{L}^p[\alpha,\beta]$ with $1 < p < \infty$, the function $f_r(x) = \frac{1}{r}\|x\|_p^r$ does not satisfy H5, excepting in the two cases considered in Proposition 2.3(ii). We remark that, as it will be seen, properties H0–H4 are required for establishing existence and uniqueness of the iterates of the algorithm under consideration, boundedness of the generated sequences and optimality of their weak cluster points, while H5 is required only for uniqueness of the weak cluster points of such sequences. We mention also that the factor $\frac{1}{r}$ in the definition of f_r is inessential for Proposition 2.3, whose results trivially hold for all positive multiples of $\|\cdot\|^r$.

We discuss next some properties of functions satisfying some of the assumptions above. Given Banach spaces X, Y with norms $\|\cdot\|_X$, $\|\cdot\|_Y$, we consider the Banach space $X \times Y$ with the product norm $\|(x, y)\| = \|x\|_X + \|y\|_Y$.

Proposition 2.4. *Let X, Y be real Banach spaces. Take $f \in \mathcal{F}(X)$ and $g \in \mathcal{F}(Y)$. Define $h : X \times Y \to \mathbb{R}$ as $h(x, y) = f(x) + g(y)$. Then for $i = 1, ..., 5$, if both f and g satisfy Hi then h also satisfies Hi.*

Proof. See Proposition 3 in [17]. □

Proposition 2.5. *Take $f : X \to \mathbb{R}$ Gâteaux differentiable. If f satisfies H3, then both f and f' are bounded on bounded subsets of X.*

Proof. See Proposition 4 in [17]. □

Proposition 2.6. *If $f \in \mathcal{F}(X)$ satisfies H2 then, for all $\{x^k\}$, $\{y^k\} \subset X$ such that $\{x^k\}$ (or $\{y^k\}$) is bounded and $\lim_{k\to\infty} D_f(y^k, x^k) = 0$, it holds that $x^k - y^k \xrightarrow[k\to\infty]{s} 0$.*

Proof. See Proposition 5 in [17]. □

Proposition 2.7. *Let $T : X \to \mathcal{P}(X)$ be maximal monotone. Take $f \in \mathcal{F}(X)$ satisfying H4. Then, for all $z \in X^*$ there exists a unique $x \in X$ such that $z \in f'(x) + T(x)$.*

Proof. See [6], Corollary 3.1. □

Proposition 2.7 can be rephrased as saying that under its assumptions the operator $(f' + T)^{-1}$ is single-valued and its domain is the whole space X^*. Since $(f'+T)^{-1}$ is clearly maximal monotone, it follows from its single-valuedness that it is continuous on X^* (se, e.g., Theorem 4.6.4 in [5]). This operator is the proximal resolvent associated to T and f. In a Hilbert space, if we take $f(x) = \frac{1}{2}\|x\|^2$, so that f' is the identity operator, this proximal resolvent is called the *Moreau-Yoshida transform.*

The next result deals with the existence of Bregman projections.

Proposition 2.8. *If $f : X \to \mathbb{R}$ is totally convex and $C \subset X$ is closed and convex, then for all $u \in X$ there exists a unique $\bar{v} \in C$ which solves the problem min $D_f(v, u)$ subject to $v \in C$.*

Proof. See 2.1.5. in [8]. □

Given f and C as in Proposition 2.8, we define the *Bregman projection onto C*, $P_f^C : X \to C$, in the following way: $P_f^C(u)$ is the only solution $\bar{v}$ of the problem $\min D_f(v, u)$ subject to $v \in C$. Our next result deals with the basic property of the Bregman projections onto hyperplanes.

Lemma 2.9. *Take a totally convex $f \in \mathcal{F}(X)$. Then for all $v \in X^* \setminus \{0\}$, $\tilde{y} \in X$, $x \in H^+$, $\bar{x} \in H^-$, it holds that $D_f(\bar{x}, x) \geq D_f(\bar{x}, z) + D_f(z, x)$, where $z = P_f^H(x)$ and H, H^+ and H^- are defined as $H = \{y \in X : \langle y - \tilde{y}, v\rangle = 0\}$, $H^+ = \{y \in X : \langle y - \tilde{y}, v\rangle \geq 0\}$ and $H^- = \{y \in X : \langle y - \tilde{y}, v\rangle \leq 0\}$.*

Proof. See Lemma 1 in [17]. □

We end this section with a result on the graph of the sum of two maximal monotone operators, taken from [22].

Lemma 2.10. *If $S, T : X \to \mathcal{P}(X^*)$ are maximal monotone operators, $\{x^k\}_{k\in K}$, $\{y^k\}_{k\in K}$ are bounded nets in X, and $\{u^k\}_{k\in K}, \{v^k\}_{k\in K}$ are bounded nets in X^* such that:*

i) *$u^k \in S(x^k), v^k \in T(y^k)$ for all $k \in K$,*
ii) *The net $\{x^k - y^k\}_{k\in K}$ is strongly convergent to 0,*
iii) *The net $\{u^k + v^k\}_{k\in K}$ is strongly convergent to a point $\bar{s} \in X^*$,*
iv) *The nets $\{x^k\}_{k\in K}$ and $\{y^k\}_{k\in K}$ both converge weakly to some $\bar{x} \in X$,*
v) *The nets $\{u^k\}_{k\in K}$ and $\{v^k\}_{k\in K}$ converge weakly to points $\bar{u}, \bar{v} \in X^*$ respectively,*

then $\bar{u} \in S(\bar{x})$, $\bar{v} \in T(\bar{x})$.

Proof. See Lemma 5 in [22]. □

We mention that Lemma 5 in [22] deals with m, rather than 2, maximal monotone operators. The statement of our Lemma 2.10 corresponds to the one in [22] for the case of $m = 2$.

3. A splitting algorithm in Banach spaces

We assume from now on that X is a reflexive Banach space. We consider set-valued maximal monotone operators $A, B : X \to \mathcal{P}(X^*)$ and problem P, as defined in (1.1).

We present now Algorithm BS (Banach Splitting) for finding zeroes of $A + B$.

i) **Initialization:** Start with any initial iterate $(z^0, w^0) \in X \times X^*$. Choose:
   a) constants $\bar{\rho} \in (0, 1]$ and $\underline{\theta}, \bar{\theta} \in \mathbb{R}$ such that $0 < \underline{\theta} \leq \bar{\theta}$,
   b) sequences of regularization parameters $\{\lambda_k\}, \{\mu_k\} \subset [\underline{\theta}, \bar{\theta}]$,
   c) auxiliary functions $f \in \mathcal{F}(X), g \in \mathcal{F}(X^*)$.
ii) **Proximal step:** Given $(z^k, w^k) \subset X \times X^*$, find $x^k, y^k \in X$, $a^k, b^k \in X^*$ such that:

$$a^k \in A(y^k), \qquad b^k \in B(x^k), \tag{3.1}$$

$$f'(x^k) + \lambda_k b^k = f'(z^k) + \lambda_k w^k, \tag{3.2}$$

$$f'(y^k) + \mu_k a^k = f'(z^k) - \mu_k w^k, \tag{3.3}$$

iii) **Projection step:** Define

$$\gamma_k = \langle x^k, b^k\rangle + \langle y^k, a^k\rangle, \tag{3.4}$$

$$\delta_k = \langle z^k, a^k + b^k\rangle + \langle x^k - y^k, w^k\rangle. \tag{3.5}$$

If $\gamma_k = \delta_k$ then stop. Otherwise, choose as the next iterate (z^{k+1}, w^{k+1}) any pair $(z, w) \in X \times X^*$ satisfying:

$$f'(z) = f'(z^k) + \eta(a^k + b^k), \tag{3.6}$$

$$g'(w) = g'(w^k) + \eta(x^k - y^k), \tag{3.7}$$

$$\gamma_k \leq \langle z, a^k + b^k \rangle + \langle x^k - y^k, w \rangle \leq (1 - \bar{\rho})\gamma_k + \bar{\rho}\delta_k \tag{3.8}$$

for some $\eta \in \mathbb{R}$.

We have presented Algorithm BS without any assumption on f, g besides the fact that they belong to $\mathcal{F}(X), \mathcal{F}(X^*)$ respectively. Along the course of Section 4, we will add the additional assumptions on f, g required for each convergence result. Now, we will comment on several features of the algorithm, and compare it to Algorithm ES.

Remark 3.1. Note that (3.1)–(3.3) reduce to finding $x^k \in (f' + \lambda_k B)^{-1}(f'(z^k) + \lambda_k w^k)$ and $y^k \in (f' + \mu_k A)^{-1}(f'(z^k) - \mu_k w_k)$. When f satisfies $H4$, existence and uniqueness of x^k, u^k are easy consequences of Proposition 2.7 (see Proposition 4.1 in Section 4). We mention that for a maximal monotone operator $T : X \to \mathcal{P}(X^*)$, the iteration $v^{k+1} = (f' + \lambda_k T)^{-1}(v^k)$, with λ_k as in our case, defines the *proximal point method* for finding zeroes of T. Though our method is slightly different, because of the presence of the second terms in the right hand sides of (3.2) and (3.3), it seems reasonable to call this step "Proximal".

Remark 3.2. In connection with the reformulation of the Proximal step as that of finding $x^k \in (f' + \lambda_k B)^{-1}(f'(z^k) + \lambda_k w^k)$ and $y^k \in (f' + \mu_k A)^{-1}(f z^k) - \mu_k w_k)$, note that both inclusions are independent of each other, and that the first one involves only the operator A, while the second one uses only B. Since the Projection step requires neither A nor B, Algorithm BS is indeed a "bona fide" splitting method.

Remark 3.3. We show now that in a large class of Banach spaces, under a sensible choice of f, g, the Projection step (3.6)–(3.8) reduces to finding a real number satisfying two nonlinear inequalities, and hence this step is computationally much less demanding than the Proximal step, which requires solution of two nonlinear inclusions in X. If f, g satisfy H4, then f' and g' are invertible, and their inverses are related to their Fenchel conjugates f^*, g^* through the well known identities $(f')^{-1} = (f^*)'$, $(g')^{-1} = (g^*)'$, which follow easily from the definitions of f^*, g^*, namely, $f^*(u) = \sup_{z \in X}\{\langle z, u \rangle - f(z)\}$, $g^*(v) = \sup_{w \in X^*}\{\langle v, w \rangle - g(w)\}$. So, we can rewrite (3.6) and (3.7) as:

$$z = (f^*)'[f'(z^k) + \eta(a^k + b^k)], \tag{3.9}$$

$$w = (g^*)'[g'(w^k) + \eta(x^k - y^k)], \tag{3.10}$$

and thus z, w are given by closed formulae on the already available data z^k, w^k, x^k, y^k, a^k and b^k and the real unknowkn η. Replacing now z and w in (3.8) by the right hand sides of (3.9) and (3.10), the Projection step reduces to finding $\eta \in \mathbb{R}$ satisfying the double inequality in the new version of (3.8), and then replacing the obtained value of η in (3.9), (3.10), in order to get the next iterates z^{k+1}, w^{k+1} as the right hand sides of (3.9) and (3.10) respectively.

A further simplification is possible if we take the Fenchel conjugate f^* as the regularizing function g for the dual space X^*. Recalling that in our reflexive setting $(f^*)^* = f$, under this choice of g (3.10) becomes

$$w = f'[(f^*)'(w^k) + \eta(x^k - y^k)]. \tag{3.11}$$

In order to make this choice of g, one needs to ascertain that f^* inherits the "good" properties of f. This is the case in our main setting. If X is uniformly smooth and uniformly convex, then the same holds for X^* (in fact, uniform smoothness of X implies uniform convexity of X^*, and uniform convexity of X implies uniform smoothness of X^*, see [10]). If we take now $f(z) = \frac{1}{r}\|z\|^r$, with $r > 1$, then a simple computation shows that $f^*(w) = \frac{1}{s}\|w\|_*^s$, where $s = r/(r-1) > 1$ and $\|\cdot\|_*$ denotes the dual norm in X^*. We already mentioned that in this family of Banach spaces such an f satisfies H0-H4, and also H5 in the case of $X = \ell_p$ and $r = p$. Since $(\ell_p)^* = \ell_q$ with $q = p/(p-1)$, it follows that the choice $g = f^*$ does ensure the good properties of g in all these cases.

Remark 3.4. Now we compare the Projection step of our method with steps (iii)-(vi) of Algorithm ES in Section 1. The details of the following argument will be presented in Lemma 4.3.

Define $h : X \times X^* \to \mathbb{R}$ as $h(z, w) = f(z) + g(w)$. In view of Proposition 2.4, if f, g enjoy some of the good properties H0-H5, so does h. Take a hyperplane $H \subset X \times X^*$ of the form $H = \{(z, w) : \langle z, c\rangle + \langle d, w\rangle = \sigma\}$, with $c \in X^*$, $d \in X$ and $\sigma \in \mathbb{R}$. In view of the convexity of D_h in its first argument, the Bregman projection of (z^k, w^k) onto H with respect to h is determined by the first order optimality conditions for the problem

$$\min D_h((z, w), (z^k, w^k)) \quad \text{s.t. } \langle z, c\rangle + \langle d, w\rangle = \sigma,$$

which are:

$$f'(z) = f'(z^k) + \eta c, \tag{3.12}$$

$$g'(w) = g'(w^k) + \eta d, \tag{3.13}$$

$$\langle z, c\rangle + \langle d, w\rangle = \sigma, \tag{3.14}$$

where $\eta \in \mathbb{R}$ is the Lagrange multiplier of the affine constraint. If we look now at (3.6), (3.7), taking $c = a^k + b^k$, $d = x^k - y^k$, and compare with (3.12)–(3.14), we realize that the pair (z^{k+1}, w^{k+1}) is the Bregman projection of (z^k, w^k) onto the hyperplane $\widehat{H}_k \subset X \times X^*$ defined as

$$\widehat{H}_k = \{(z, w) : \langle (z, w), (a^k + b^k, x^k - y^k)\rangle = \sigma_k\},$$

with

$$\sigma_k = \langle z^{k+1}, a^k + b^k\rangle + \langle x^k - y^k, w^{k+1}\rangle. \tag{3.15}$$

Now, consider again $\varphi_k : X \times X \to \mathbb{R}$ defined as $\varphi_k(z, w) = \langle z - x^k, b^k - w\rangle + \langle z - y^k, a^k + w\rangle$, note that φ is affine, because terms involving $\langle z, w\rangle$ cancel, and define the hyperplane $\bar{H}_k \subset X \times X$ as $\bar{H}_k = \{(z, w) : \varphi_k(z, w) = 0\}$, so that $\bar{H}_k$ is the limiting hyperplane of the halfspace H_k defined in step (iv) of Algorithm ES.

If we call now $\rho_k = (\sigma_k - \gamma_k)/(\delta_k - \gamma_k)$, it can be checked that $\rho_k \in [\bar{\rho}, 1]$, and that $\widehat{H}_k$ is a relaxed hyperplane parallel to $\bar{H}_k$ lying between (z^k, w^k) and $\bar{H}_k$. For $\rho_k = 1$ we get $\widehat{H}_k = \bar{H}_k$, and for $\rho_k = 0$ we would have $(z^k, w^k) \in \widehat{H}_k$, but this case cannot occur because $\bar{\rho} > 0$, by virtue of which $\widehat{H}_k$ strictly separates (z^k, w^k) from the extended solution set $S_e(A, B)$, which is essential for convergence of the method.

Summarizing this discussion, the pair (z^{k+1}, w^{k+1}) generated by Algorithm BS can be seen as the Bregman projection of the pair (z^k, w^k) onto a hyperplane lying between (z^k, w^k) and the limiting hyperplane of the halfspace H_k used in step (v) of Algorithm ES, corresponding to a relaxation parameter $\rho_k \in [\bar{\rho}, 1]$. In this sense, Algorithm BS is similar to Algorithm ES, but there are three differences worth mentioning:

a) In Algorithm ES, first the orthogonal projection $\bar{p}^k$ of (z^k, w^k) onto H_k is computed (step (v)), and the relaxation is performed afterward (step(vi)), while in Algorithm BS the hyperplane is (implicitly) relaxed and the Bregman projection is computed after the relaxation. Since orthogonal projections onto hyperplanes in Hilbert spaces are affine, the order of the operations relaxation-projection is irrelevant (in both cases the same point is finally obtained). The nonlinear nature of Bregman projections in nonhilbertian Banach spaces makes the order relevant indeed, and the one selected in Algorithm BS is esssential for the good behavior of the method.

b) In Algorithm ES the relaxation parameters ρ_k are contained in a compact subset of $(0, 2)$ while in Algorithm BS they are (implicitly) restricted to a compact subset of $(0, 1]$. Again this is a consequence of the nonlinear nature of Bregman projections in nonhilbertian spaces; over-relaxed projections (i.e. with $\rho_k > 1$), do not enjoy the decreasing distance property, and thus must be excluded.

c) In Algorithm ES the relaxation parameter ρ_k is exogenously given, while in Algorithm BS no relaxation parameter is explicitly employed, but instead we can take any pair (z^{k+1}, w^{k+1}) satisfying the double inequality in (3.8). This is a significant advance: if we specify a relaxation parameter ρ_k beforehand, (3.8) becomes an equality, i.e. a nonlinear equation in the real variable η to be exactly solved; our formulation, with the two inequalities in (3.8), is akin to admitting inexact solutions of the nonlinear equation. On the other hand, this advantage would not be significant in Algorithm ES, because in the hilbertian environment there is no equation to solve: the orthogonal projection is given by an affine operator with well determined parameters, and hence it is not worthwhile to admit inexactness in its computation.

Remark 3.5. We have seen in the previous remark that the next iterate in Algorithm BS can be seen as the Bregman Projection with respect to the auxiliary function h of the current iterate onto a certain hyperplane. The fact that we have taken the auxiliary function h in $X \times X^*$ as a separable one, of the form $h(z, w) = f(z) + g(w)$, is inessential for the analysis. The same convergence results can be established with any auxiliary function defined on $X \times X^*$ and enjoying the required properties among H0-H5, possibly unrelated to the auxilary function f used in the Proximal step. The advantage of the separable auxiliary function becomes clear when we look at the first order optimality conditions related to the computation of the Bregman projection: in the separable case, we get (3.6)–(3.8), which can be further simplified to (3.9)–(3.10), and even to (3.11) by choosing $g = f^*$; the use of a nonseparable auxiliary function would lead to a system considerably more involved than (3.6)–(3.8). For this reason, we prefered to present Algorithm BS only with a separable auxiliary function in the product space.

4. Convergence analysis

We proceed to the convergence analysis of Algorithm BS. We start by establishing that the generated sequence is well defined. From now on, we define $h : X \times X^* \to \mathbb{R}$ as $h(z, w) = f(z) + g(w)$, where f, g are the auxiliary functions chosen in item (c) of the Initialization of Algorithm BS.

Proposition 4.1. *If f satisfies H2 and H4, and g satisfies H2, then the sequence $\{(z^k, w^k)\}$ generated by Algorithm BS is well defined, in the sense that, given the k-th iterate (z^k, w^k), there exists always a pair $(z^{k+1}, w^{k+1}) \in X \times X^*$ satisfying the algorithm prescriptions. Also, $\gamma_k \le \delta_k$ for all k, with γ_k, δ_k as defined by* (3.4) *and* (3.5)*.*

Proof. We consider first the Proximal step. An elementary algebraic manipulation shows that (3.1)–(3.3) is equivalento to finding $x^k \in (f' + \lambda_k B)^{-1}(f'(z^k) + \lambda_k w^k)$ and $y^k \in (f' + \mu_k A)^{-1}(f z^k) - \mu_k w_k)$. Since A, B are maximal monotone and λ_k, μ_k are positive, we get that $\lambda_k B$ and $\mu_k A$ are also maximal monotone. Proposition 2.7 and the fact that f satisfies H4 imply that x^k, y^k, a^k and b^k are uniquely determined by (3.1)–(3.3).

We move on now to the Projection step. It suffices to show that

$$\gamma_k \le (1 - \bar{\rho})\gamma_k + \bar{\rho}\delta_k, \tag{4.1}$$

and that there exist z, w such that the leftmost inequality in (3.8) holds with equality. Note that, since $\bar{\rho} \in (0, 1]$, the inequality in (4.1) is equivalent to stating that $\gamma_k \le \delta_k$. Using (3.4) and (3.5), and some elementary algebra, this inequality turns out to be equivalent to

$$\langle z^k - x^k, b^k - w^k \rangle + \langle z^k - y^k, a^k + w^k \rangle \ge 0. \tag{4.2}$$

If we use (3.2) and (3.3) for writing $b^k - w^k$ and $a^k + w^k$ in terms of z^k, x^k and y^k, and replace the result in the left hand side of (4.2), we get

$$\begin{aligned} \langle z^k - x^k, b^k - w^k \rangle + \langle z^k - y^k, a^k + w^k \rangle &= \frac{1}{\lambda_k}\langle z^k - x^k, f'(z^k) - f'(x^k) \rangle \\ &\quad + \frac{1}{\mu_k}\langle z^k - y^k, f'(z^k) - f'(y^k) \rangle \\ &\ge 0, \end{aligned} \tag{4.3}$$

using the convexity of f and the positivity of λ_k, μ_k in the inequality of (4.3). We have proved that (4.2) holds, and therefore $\gamma_k \le \delta_k$, establishing the final statement of the proposition.

Now, we show that there exists a pair $(z, w) \in X \times X^*$ satisfying the system consisting of (3.6), (3.7) and (3.8) with equality in its leftmost inequality, i.e.

$$\langle z, a^k + b^k \rangle + \langle x^k - y^k, w \rangle = \langle x^k, b^k \rangle + \langle y^k, a^k \rangle \tag{4.4}$$

Using (2.1), it is easy to check that these three equations are just the first order optimality conditions for the problem of minimizing $D_h((z, w), (z^k, w^k))$ subject to (4.4). These first order conditions are not only necessary but also sufficient, in view of the convexity of both $D_h(\cdot, (z^k, w^k))$ and the hyperplane in $X \times X^*$ defined by (4.4), which we will call $\bar{H}_k$. Thus, the issue boils down to proving that there exists the Bregman projection of (z^k, w^k) onto $\bar{H}_k$ with respect to the auxiliary function h. By Proposition 2.4, h satisfies H2, which implies total convexity. Since $\bar{H}_k$ is clearly closed and convex, the existence of the required pair (z, w) follows from Proposition 2.8. □

It can be seen that the hyperplane $\bar{H}_k$ defined by (4.4) coincides with the hyperplane defined in Remark 3.4 of Section 3, though this fact is not needed in our proofs.

Next we look at the stopping criterion in the Projection step of Algorithm BS. We recall that $S_e(A, B) \subset X \times X^*$ is defined as $S_e(A, B) = \{(z, w) : -w \in A(z), w \in B(z)\}$.

Proposition 4.2. *If Algorithm BS stops at step k, then (z^k, w^k) belongs to $S_e(A, B)$, i.e., z^k solves problem P.*

Proof. If BS stops at iteration k, then $\gamma_k = \delta_k$, in which case, looking at the proof of Proposition 4.1, we have equality in (4.2) and (4.3), i.e.

$$\frac{1}{\lambda_k}\langle z^k - x^k, f'(z^k) - f'(x^k)\rangle + \frac{1}{\mu_k}\langle z^k - y^k, f'(z^k) - f'(y^k)\rangle = 0.$$

In view of the positivity of λ_k, μ_k and the strict convexity of f, we conclude that $z^k = x^k = y^k$. Replacing x^k and y^k by z^k in (3.2) and (3.3) we get that $w^k = b^k$, $-w^k = a^k$. Looking now at (3.1), we conclude that

$$0 = -w^k + w^k = a^k + b^k \in A(y^k) + B(x^k) = A(z^k) + b(z^k),$$

i.e., z^k is a zero of $A + B$, thus solving problem P. □

Next we prove the distance reducing property of Algorithm BS, i.e., that the Bregman distance related to h from the iterates to any point in $S_e(A, B)$ decreases with the iteration count. This is a consequence of the properties of the Bregman projections, and is the driving mechanism leading to the optimality of the weak cluster points of the sequence generated by the algorithm.

Lemma 4.3. *Assume that problem P has solutions and that f satisfies H2. Take any pair $(\bar{z}, \bar{w}) \in S_e(A, B)$. Let $\{(z^k, w^k)\}$ be the sequence generated by Algorithm BS. Define $\bar{p} := (\bar{z}, \bar{w}), p^k := (z^k, w^k)$. Then*

$$D_h(\bar{p}, p^k) \geq D_h(\bar{p}, p^{k+1}) + D_h(p^{k+1}, p^k). \tag{4.5}$$

Proof. Observe that the existence of solutions of problem P is equivalent to nonemptiness of $S_e(A,B)$. As announced in Remark 3.4, we will show now that the p^{k+1} is the Bregman projection of p^k with respect to h onto a hyperplane which separates p^k from $S_e(A,B)$. The result will then be a consequence of Lemma 2.9.

Let us define, as in (3.15), $\sigma_k = \langle z^{k+1}, a^k + b^k\rangle + \langle x^k - y^k, w^{k+1}\rangle$, and consider the hyperplane $\widehat{H}_k \subset X \times X^*$ given by

$$\widehat{H}_k = \{(z,w) : \langle z, a^k + b^k\rangle + \langle x^k - y^k, w\rangle = \sigma_k\}. \tag{4.6}$$

Note that σ_k has been defined precisely so that p^{k+1} belongs to $\widehat{H}_k$. Note also that (3.6), (3.7) and (4.6) are the first order optimality conditions for the problem $\min_{p\in\widehat{H}_k} D_h(p,p^k)$. Since these conditions are sufficient, by virtue of the convexity of $\widehat{H}_k$ and of $D_h(\cdot,p^k)$, we have proved that p^{k+1} is the Bregman projection of p^k onto $\widehat{H}_k$ with respect to the auxilary function h. Now we must check that $S_e(A,B)$ and p^k lie on opposite sides of the hyperplane $\widehat{H}_k$. These two inclusions are a consequence of the selection of x^k, y^k, a^k and b^k in the Proximal step of Algorithm BS.

Define the halfspaces H_k^+, H_k^- as

$$H_k^+ = \{(z,w) : \langle z, a^k + b^k\rangle + \langle x^k - y^k, w\rangle \geq \sigma_k\}, \tag{4.7}$$

$$H_k^- = \{(z,w) : \langle z, a^k + b^k\rangle + \langle x^k - y^k, w\rangle \leq \sigma_k\}. \tag{4.8}$$

Take any pair $(\bar{z},\bar{w}) \in S_e(A,B)$. In order to establish that $S_e(A,B) \subset H_k^-$, it suffices to verify that (4.8) holds with $(z,w) = (\bar{z},\bar{w})$. Look now at the leftmost inequality in (3.8) with $(z,w) = (z^{k+1},w^{k+1})$, which holds indeed because (3.8) defines the next iterate (z^{k+1},w^{k+1}). Taking into account (3.15), such inequality can be rewritten as

$$\gamma_k \leq \sigma_k. \tag{4.9}$$

We claim that

$$\langle \bar{z}, a^k + b^k\rangle + \langle x^k - y^k, \bar{w}\rangle\} \leq \gamma_k. \tag{4.10}$$

In view of (3.4), (4.10) is equivalent to

$$\langle \bar{z}, a^k + b^k\rangle + \langle x^k - y^k, \bar{w}\rangle\} \leq \langle x^k, b^k\rangle + \langle y^k, a^k\rangle. \tag{4.11}$$

Adding and substracting $\langle \bar{z},\bar{w}\rangle$ in the right hand side of (4.11), an elementary algebraic manipulation shows that (4.11) is equivalent to

$$\langle \bar{z} - y^k, -\bar{w} - a^k\rangle + \langle \bar{z} - x^k, \bar{w} - b^k\rangle \geq 0. \tag{4.12}$$

Since $-\bar{w} \in A(\bar{z}), \bar{w} \in B(\bar{z})$ by definition of $S_e(A,B)$, and $a^k \in A(y^k), b^k \in B(x^k)$ by (3.1), the monotonicity of A, B implies that (4.12) holds, establishing the claim, i.e. the validity of (4.10). Combining (4.10) with (4.9), we conclude that (4.8) holds with $(z,w) = (\bar{z},\bar{w})$, i.e. that $S_e(A,B) \subset H_k^-$.

Now we prove that p^k belongs to H_k^+. We must verify that (4.7) holds with $(z,w) = (z^k,w^k)$. By the Projection step of Algorithm BS, (3.8) holds with $(z,w) =$

(z^{k+1}, w^{k+1}). Taking into account (3.15), the rightmost inequality in (3.8) with $(z, w) = (z^{k+1}, w^{k+1})$ is equivalent to

$$\sigma_k \leq (1 - \bar{\rho})\gamma_k + \bar{\rho}\delta_k. \tag{4.13}$$

Since $\bar{\rho} \leq 1$ by item (a) in the Initialization of Algorithm BS, and $\gamma_k \leq \delta_k$ by Proposition 4.1, we get from (4.13) that

$$\sigma_k \leq \delta_k. \tag{4.14}$$

In view of (4.7) and (4.14), in order to check that (z^k, w^k) belongs to H_k^+ it suffices to check that

$$\langle z^k, a^k + b^k \rangle + \langle x^k - y^k, w^k \rangle \} \geq \delta_k,$$

which holds (indeed, with equality) by virtue of (3.5). We have established that $p^k \in H_k^+$, $S_e(A, B) \subset H_k^-$ and $p^{k+1} = P_h^{\widehat{H}_k}(p^k)$. Since h is totally convex because it satisfies H2, we are precisely within the hypotheses of Lemma 2.10, and hence (4.5) holds true. □

We remark that only (3.1) is needed for proving that $S_e(A, B) \subset H_k^-$, while the fact that p^k belongs to H_k^+ is a consequence of (3.2) and (3.3), used in Proposition 4.1 for proving that $\gamma_k \leq \delta_k$. We also mention that the fact that $\widehat{H}_k$ can be written as a relaxed hyperplane with relaxation parameter ρ_k, as explained in Remark 3.4 in Section 3, is not needed in the convergence analysis.

Now all the pieces are in order for our convergence theorem. Note that up to now only properties H2 and H4 of the auxiliary functions have been invoked. The remaining properties, namely H0, H1, H3 and H5, will be used in the proof of the theorem.

Theorem 4.4. i) *Assume that Problem P has solutions and that f and g satisfy H0-H4. Then the sequences $\{z^k\}$ and $\{w^k\}$ are bounded, the corresponding differences between consecutive iterates, $\{z^k - z^{k+1}\}$ and $\{w^k - w^{k+1}\}$, converge strongly to 0, and all weak cluster points of $\{(z^k, w^k)\}$ belong to $S_e(A, B)$, so that all weak cluster points of $\{z^k\}$ are zeroes of $A + B$.*

ii) *If additionally f and g satisfy H5, then the sequences $\{z^k\}, \{x^k\}$ and $\{y^k\}$ converge weakly to some zero z^* of $A + B$, the sequences $\{w^k\}$ and $\{b^k\}$ converge weakly to a point w^* such that $w^* \in B(z^*), -w^* \in A(z^*)$, and the sequence $\{a^k\}$ converges weakly to $-w^*$.*

Proof. Assume first that f, g satisfy H0-H4. Define $h : X \times X^* \to \mathbb{R}$ as $h(z, w) = f(z) + g(w)$. By Proposition 2.3, h satisfies H1-H4.

Take any pair $(\bar{z}, \bar{w}) \in S_e(A, B)$, which is nonempty because P has solutions. Again we take $\bar{p} = (\bar{z}, \bar{w}), p^k = (z^k, w^k)$. In view of Lemma 4.3, (4.5) holds. Since D_h is nonnegative by Proposition 2.1(i), it follows that $\{D_h(\bar{p}, p^k)\} \subset \mathbb{R}$ is nonincreasing and nonnegative, hence convergent. Define $\zeta = D_h(\bar{p}, p^0)$. It follows that $D_h((\bar{p}, p^k) \leq \zeta$ for all k. Since h satisfies H1, $\{p^k\}$ is bounded, and hence $\{z^k\}$ and $\{w^k\}$ are bounded.

Also, since (4.5) implies that

$$D_h(p^{k+1}, p^k) \leq D_h(\bar{p}, p^k) - D_h(\bar{p}, p^{k+1}), \tag{4.15}$$

we have that $\{D_h(p^{k+1}, p^k)\}$ converges to 0, because the right hand side of (4.15) is the difference between consecutive terms of a convergent sequence. By Proposition 2.6, $p^k - p^{k+1} \xrightarrow[k\to\infty]{\text{s}} 0$, and therefore,

$$z^k - z^{k+1} \xrightarrow[k\to\infty]{\text{s}} 0, \qquad w^k - w^{k+1} \xrightarrow[k\to\infty]{\text{s}} 0. \tag{4.16}$$

Next we will use H0 to get boundedness of $\{x^k\}, \{b^k\}, \{y^k\}$ and $\{a^k\}$. Let $q^k = f'(z^k) + \lambda_k w^k$, so that (3.2) becomes

$$f'(x^k) + \lambda_k b^k = q^k. \tag{4.17}$$

Since $\{z^k\}$ is bounded and f satisfies H3, $\{f'(z^k)\}$ is bounded by Proposition 2.5. Since $\lambda_k \leq \bar{\theta}$, boundeness of $\{w^k\}$ implies boundedness of $\{q^k\}$. Substracting $f'(x^0) + \lambda_k b^k$ from both sides of (4.17) and computing the duality product with $x^k - x^0$, we get

$$\langle x^k - x^0, q^k\rangle = \langle x^k - x^0, f'(x^k) - f'(x^0)\rangle + \lambda_k \langle x^k - x^0, b^k - b^0\rangle \geq$$

$$\langle x^k - x^0, f'(x^k) - f'(x^0)\rangle = D_f(x^k, x^0) + D_f(x^0, x^k) \geq D_f(x^k, x^0), \tag{4.18}$$

using nonnegativity of λ_k, monotonicity of B and (3.1) in the first inequality, Proposition 2.1(ii) in the second equality and Proposition 2.1(i) in the second inequality. From (4.18) and (2.1) we get

$$\begin{aligned} f(x^k) &\leq \langle x^k - x^0, q^k\rangle + f(x^0) + \langle f'(x^0), x^k - x^0\rangle \\ &\leq f(x^0) + \left\|x^k - x^0\right\| \left[\left\|q^k\right\| + \|f'(x^0)\|\right], \end{aligned} \tag{4.19}$$

using the Cauchy-Schwartz inequality in the second inequality of (4.19). Therefore

$$\frac{f(x^k)}{\|x^k\|} \frac{\|x^k\|}{\|x^k - x^0\|} = \frac{f(x^k)}{\|x^k - x^0\|} \leq \frac{f(x^0)}{\|x^k\|} + \left[\left\|q^k\right\| + \|f'(x^0)\|\right]. \tag{4.20}$$

We claim that (4.20) implies boundedness of $\{x^k\}$. Suppose, for the sake of contradiction, that $\{x^k\}$ has an unbounded subsequence. Then, the left hand side of (4.20) tends to ∞ along such subsequence, because f satisfies H0 and $\lim_{k\to\infty} \|x^k\| / \|x^k - x^0\| = 1$, while in the right hand side, the first term converges to 0, and the second one remains bounded, by boundedness of $\{q^k\}$. The resulting contradiction implies that the claim holds. From (4.17) we get

$$b^k = \frac{1}{\lambda_k}[q^k - f'(x^k)].$$

Since $\lambda_k \geq \underline{\theta} > 0$, we obtain, from Proposition 2.5 and the boundedness of $\{q^k\}$ and $\{x^k\}$, that $\{b^k\}$ is bounded. A similar argument, starting from (3.3), establishes boundedness of $\{y^k\}$ and $\{a^k\}$.

Now, we combine (3.8), (3.4) and (3.5) to get

$$\begin{aligned} \langle z^{k+1}, a^k + b^k\rangle + \langle x^k - y^k, w^{k+1}\rangle \;\leq\; & (1-\bar{\rho})[\langle x^k, b^k\rangle + \langle y^k, a^k\rangle] \\ & +\bar{\rho}[\langle z^k, a^k + b^k\rangle + \langle x^k - y^k, w^k\rangle] \end{aligned} \tag{4.21}$$

Multiplying (4.21) by -1 and adding $\langle z^k, a^k + b^k\rangle + \langle x^k - y^k, w^k\rangle$ to both sides, we obtain

$$\langle z^k - z^{k+1}, a^k + b^k\rangle + \langle x^k - y^k, w^k - w^{k+1}\rangle \tag{4.22}$$

$$
\begin{aligned}
&\geq (1-\bar{\rho})[(\langle z^k, a^k+b^k\rangle + \langle x^k-y^k, w^k\rangle) - (\langle x^k, b^k\rangle + \langle y^k, a^k\rangle)\\
&= (1-\bar{\rho})[\langle z^k-y^k, a^k+w^k\rangle + \langle z^k-x^k, b^k-w^k\rangle]\\
&= (1-\bar{\rho})\left[\frac{1}{\mu_k}\langle z^k-y^k, f'(z^k)-f'(y^k)\rangle + \frac{1}{\lambda_k}\langle z^k-x^k, f'(z^k)-f'(x^k)\rangle\right]\\
&= (1-\bar{\rho})\left[\frac{1}{\mu_k}\Big(D_f(z^k,y^k)+D_f(y^k,z^k)\Big) + \frac{1}{\lambda_k}\Big(D_f(z^k,x^k)+D_f(x^k,z^k)\Big)\right]\\
&\geq \frac{1-\bar{\rho}}{\bar{\theta}}\Big[(D_f(z^k,y^k)+D_f(y^k,z^k)+D_f(z^k,x^k)+D_f(x^k,z^k)\Big],
\end{aligned}
$$

using some elementary algebra in the first equality, (3.2) and (3.3) in the second equality, Proposition 2.1(ii) in the third equality, and the upper bound for $\{\lambda_k\}, \{\mu_k\}$ given in item (b) in the initialization of Algorithm BS in the second inequality.

In view of (4.16) and the boundedness of $\{x^k\}, \{y^k\}, \{a^k\}$ and $\{b^k\}$, the leftmost expression on (4.22) converges to 0 as $k \to \infty$. Using now Proposition 2.1(i), we get $\lim_{k\to\infty} D_f(x^k, z^k) = \lim_{k\to\infty} D_f(y^k, z^k) = 0$, and then, using Proposition 2.6, we conclude that

$$z^k - x^k \xrightarrow[k\to\infty]{s} 0, \qquad z^k - y^k \xrightarrow[k\to\infty]{s} 0, \tag{4.23}$$

which implies

$$x^k - y^k \xrightarrow[k\to\infty]{s} 0, \tag{4.24}$$

From (4.23), boundedness of $\{x^k\}, \{y^k\}$ and $\{z^k\}$, together with the fact that f satisfies H3, we get $f'(z^k) - f'(x^k) \xrightarrow[k\to\infty]{s} 0$, $f'(z^k) - f'(y^k) \xrightarrow[k\to\infty]{s} 0$, so that, taking into account (3.2) and (3.3), we obtain that $\lambda_k(b^k - w^k) \xrightarrow[k\to\infty]{s} 0$, $\mu_k(a^k + w^k) \xrightarrow[k\to\infty]{s} 0$, implying, since $\lambda_k, \mu_k \leq \bar{\theta}$, that

$$b^k - w^k \xrightarrow[k\to\infty]{s} 0, \qquad a^k + w^k \xrightarrow[k\to\infty]{s} 0, \tag{4.25}$$

and therefore

$$a^k + b^k \xrightarrow[k\to\infty]{s} 0. \tag{4.26}$$

By reflexivity of X, the sequences $\{z^k\}, \{w^k\}, \{x^k\}, \{y^k\}, \{a^k\}$ and $\{b^k\}$ have weak cluster points. Let z^* be a cluster point of $\{z^k\}$, and $\{z^{j_k}\}$ a subsequence of $\{x^k\}$ weakly convergent to z^*. In view of (4.23), we have

$$x^{j_k} \xrightharpoonup[k\to\infty]{w} z^*, \qquad y^{j_k} \xrightharpoonup[k\to\infty]{w} z^*. \tag{4.27}$$

Without loss of generality, i.e. refining the subsequence if needed, we can assume that there exist $a^*, b^*, w^* \in X^*$ such that

$$a^{j_k} \xrightharpoonup[k\to\infty]{w} a^*, \qquad b^{j_k} \xrightharpoonup[k\to\infty]{w} b^*, \qquad w^{j_k} \xrightharpoonup[k\to\infty]{w} w^*. \tag{4.28}$$

Now we check the assumptions of Lemma 2.10 with $S = B, T = A$, $x^k = x^{j_k}$, $y^k = y^{j_k}$, $u^k = b^{j_k}$ and $v^k = a^{j_k}$. Assumption (i) holds by (3.1), (ii) follows from (4.24), (iii) from (4.26), (iv) from (4.27) and (v) from (4.28). We conclude from Lemma 2.10 that $a^* \in A(z^*), b^* \in B(z^*)$. On the other hand, (4.26) implies that $a^* = -b^*$, so that $0 = a^* + b^* \in A(z^*) + B(z^*)$, i.e. z^* is a solution of problem P. Now, from (4.25) and (4.28) we get $w^* = b^* = -a^*$, so that $-w^* \in A(z^*), w \in B(z^*)$, i.e., (z^*, w^*) belongs to $S_e(A, B)$, completing the proof of item (i).

Now we proceed to prove item (ii), assuming that f and g satisfy H5, and hence h satisfies H5 by Proposition 2.4. We have already shown that $\{p^k\} = \{(z^k, w^k)\}$ is bounded. We will establish next that $\{p^k\}$ has a unique weak cluster point. Assume that both $\widetilde{p}$ and $\widehat{p}$ are cluster points of $\{p^k\}$, and let $\{p^{i_k}\}, \{p^{j_k}\}$ be subsequences of $\{p^k\}$ which converge weakly to $\widetilde{p}$, $\widehat{p}$ respectively. We have proved in item (i) that both $\widehat{p}$ and $\widetilde{p}$ belong to $S_e(A, B)$. In view of Lemma 4.3, we get from (4.5) that both $\{D_h(\widehat{p}, p^k)\}$ and $\{D_h(\widetilde{p}, p^k)\}$ are nonnegative and nonincreasing, hence convergent, i.e. there exist $\widehat{\beta}, \widetilde{\beta} \in \mathbb{R}$ such that

$$\lim_{k\to\infty} D_h(\widehat{p}, p^k) = \widehat{\beta}, \quad \lim_{k\to\infty} D_h(\widetilde{p}, p^k) = \widetilde{\beta}. \tag{4.29}$$

Now, using (2.2) in Lemma 2.2, we get

$$\begin{aligned} \left|\langle h'(p^{i_k}) - h'(p^{j_k}), \widehat{p} - \widetilde{p}\rangle\right| &= \left|\left[D_h(\widehat{p}, p^{i_k}) - D_h(\widehat{p}, p^{j_k})\right] - \left[D_h(\widetilde{p}, p^{i_k}) - D_h(\widetilde{p}, p^{j_k})\right]\right| \\ &\le \left|D_h(\widehat{p}, p^{i_k}) - D_h(\widehat{p}, p^{j_k})\right| + \left|D_h(\widetilde{p}, p^{i_k}) - D_h(\widetilde{p}, p^{j_k})\right|. \end{aligned} \tag{4.30}$$

In view of (4.29), both terms in the rightmost expression of (4.30) converge to 0 as $k \to \infty$, so that

$$\begin{aligned} 0 &= \lim_{k\to\infty} \left|\langle h'(p^{i_k}) - h'(p^{j_k}), \widehat{p} - \widetilde{p}\rangle\right| = \left|\langle h'(\widetilde{p}) - h'(\widehat{p}), \widehat{p} - \widetilde{p}\rangle\right| \\ &= \langle h'(\widetilde{p}) - h'(\widehat{p}), \widetilde{p} - \widehat{p}\rangle = D_h(\widehat{p}, \widetilde{p}) + D_h(\widetilde{p}, \widehat{p}) \ge D_h(\widetilde{p}, \widehat{p}) \ge 0, \end{aligned} \tag{4.31}$$

using property H5 of h in the second equality, convexity of h in the second one, and Proposition 2.1(ii) in the third one. It follows from (4.31) that $D_h(\widetilde{p}, \widehat{p}) = 0$, so that $\widehat{p} = \widetilde{p}$ by Proposition 2.1(i). We have proved that $\{p^k\}$ has a unique cluster point, and so both $\{z^k\}$ and $\{w^k\}$ are weakly convergent, say to z^* and w^* respectively. By item (i), (z^*, w^*) belongs to $S_e(A, B)$, and hence z^* solves problem P. The weak convergence of $\{x^k\}$, $\{y^k\}$, $\{a^k\}$ and $\{b^k\}$, as well as the value of their weak limits, follow then from (4.23) and (4.25). □

References

[1] H. H. Bauschke, *A note on the paper by Eckstein and Svaiter "General projective splitting methods for sums of monotone operators"*, SIAM Journal on Control and Optimization **48** (2009), 2513–2515.

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

[3] F. E. Browder, *Nonlinear operators and nonlinear equations of evolution in Banach spaces*, Proceedings of Symposia in Pure Mathematics, American Mathematical Society **18** (1976).

[4] R. S. Burachik and A. N. Iusem, *A generalized proximal point algorithm for the variational inequality problem in a Hilbert space*, SIAM Journal on Optimization **8** (1998), 197–216.

[5] R. S. Burachik and A. N. Iusem, *Set-Valued Mappings and Enlargements of Monotone Operators*. Springer, Berlin, 2007.

[6] R. S. Burachik an S. Scheimberg, *A proximal point algorithm for the variational inequality problem in Banach spaces*, SIAM Journal on Control and Optimization **39** (2001), 1615–1632.

[7] D. Butnariu and A. N. Iusem, *On a proximal point method for convex optimization in Banach spaces*, Numerical Functional Analysis and Optimization **18** (1997), 723–744.

[8] D. Butnariu and A. N. Iusem, *Totally Convex Functions for Fixed Point Computation and Infinite Dimensional Optimization*. Kluwer, Dordrecht, 2000.

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

[10] I. Ciorănescu, *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems*, Kluwer, Dordrecht, 1990.
[11] J. W. Demmel, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, 1997.
[12] J. Eckstein and D. Bertsekas, *On the Douglas-Ratford splitting method and the proximal point algorithm for maximal monotone operators*, Mathematical Programming **55** (1992), 293–318.
[13] J. Eckstein and M. C. Ferris, *Operator-splitting methods for monotone affine variational inequalities, with a parallel application to optimal control*, INFORMS Journal on Computing **10** (1998), 208–235.
[14] J. Eckstein and B. F. Svaiter, *A family of projective splitting methods for the sum of two maximal monotone operators*, Mathematical Programming **111** (2008), 173–199.
[15] R. Gárciga Otero and A. N. Iusem,*Proximal methods with penalization effects in Banach spaces*, Numerical Functional Analysis and Optimization **25** (2004), 69–91.
[16] R. Gárciga Otero and A. N. Iusem, *Proximal methods in Banach spaces without monotonicity*, Journal of Mathematical Analysis and Applications **330** (2007), 433–450.
[17] A. N. Iusem and R. Gárciga Otero, *Inexact versions of proximal point and augmented Lagrangian algorithms in Banach spaces*, Numerical Functional Analysis and Optimization **22** (2001), 609–640.
[18] J. Lawrence and J. E. Spingarn, *On fixed points of nonexpansive piecewise isometric mappings*, Proceedings of the London Mathematical Society **55** (1987), 605–624.
[19] P.-L. Lions, *Une méthode itérative de résolution d'une inéquation variationelle*, Israel Journal of Mathematics **31** (1978), 204–208.
[20] P.-L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM Journal on Numerical Analysis **16** (1979), 164–179.
[21] G. B. Passty, *Ergodic convergence to a zero of the sum of monotone operators in Hilbert space*, Journal of Mathematical Analysis and Applications **72** (1979), 383–390.
[22] B. F. Svaiter, *On weak convergence of the Douglas-Rachford method*, SIAM Journal on Control and Optimization **49** (2011), 280–287.
[23] P. Tseng, *A modified forward-backward splitting method for maximal monotone mappings SIAM Journal on Control and Optimization* **38** (2000), 431–446.

Manuscript received April 1, 2013
revised May 28, 2013

Alfredo N. Iusem
Instituto de Matématica Pura e Aplicada (IMPA), Estrada Dona Castorina 110, Rio de Janeiro, RJ, CEP 22460-320, Brazil
E-mail address: `iusp@impa.br`

Benar F. Svaiter
Instituto de Matématica Pura e Aplicada, (IMPA), Estrada Dona Castorina 110, Rio de Janeiro, RJ, CEP 22460-320, Brazil
E-mail address: `benar@impa.br`