Proximal methods in Banach spaces without monotonicity

Rolando Gárciga Otero* Alfredo N. Iusem†

February 14, 2006

Abstract

We introduce the concept of hypomonotone point-to-set operators in Banach spaces, with respect to a regularizing function. This notion coincides with the one given by Rockafellar and Wets in Hilbertian spaces, when the regularizing function is the square of the norm. We study the associated proximal mapping, which leads to a a hybrid proximal-extragradient and proximal-projection methods for non-monotone operators in reflexive Banach spaces. These methods allow for inexact solution of the proximal subproblems with relative error criteria. We then consider the notion of local hypomonotonicity and propose localized versions of the algorithms, which are locally convergent.

Key words: hypomonotone operator, regularizing function, proximal point algorithm, hybrid proximal-extragradient algorithm.

1 Introduction

We deal in this paper with methods for finding zeroes of point-to-set operators in Banach spaces, i.e., for solving the problem:

$$\text{Find} \quad x \in B \ \text{ such that } \ 0 \in T(x),$$

*Instituto de Economia da Universidade Federal de Rio de Janeiro, Avenida Pasteur 250, Rio de Janeiro, RJ, Brazil (**rgarciga@ie.ufrj.br**). The work of this author was partially supported by FAPERJ and CNPq.

†Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Rio de Janeiro, RJ, 22460-320, Brazil (**iusp@impa.br**). The work of this author was partially supported by CNPq grant no. 301280/86.

where $T : B \to \mathcal{P}(B^*)$ denotes an operator from a reflexive real Banach space B to parts of its topological dual B^*.

The proximal point algorithm, whose origins can be traced back to [13], attained its celebrated formulation in the work of Rockafellar [16], and is a relevant tool for solving this problem. In a Hilbert space H, the algorithm generates a sequence $\{x^k\} \subset H$, starting from some $x^0 \in H$, through the iteration

$$x^{k+1} = (I + \gamma_k T)^{-1}(x^k), \tag{1}$$

where $\{\gamma_k\}$ is a sequence of positive real numbers bounded away from zero. It has been proved in [16] that for a maximal monotone T, the sequence $\{x^k\}$ is weakly convergent to a zero of T when T has zeroes, and is unbounded otherwise. Such weak convergence is global, i.e. the result just announced holds in fact for any $x^0 \in H$.

The situation becomes considerably more complicated when T fails to be monotone. A survey of results on convergence of the proximal algorithm without monotonicity up to 1997 can be found in [12]. A new approach to the subject was taken in [15], which deals with a class of nonmonotone operators that, when restricted to a neighborhood of the solution set, are not far from being monotone. More precisely, it is assumed that, for some $\rho > 0$, the mapping $T^{-1} + \rho I$ (which is the inverse of the Yosida regularization of T), is monotone when restricted to a neighborhood of $\hat{S} \times \{0\}$, where $\hat{S}$ is a nonempty connected component of the solution set $S = T^{-1}(0)$. When this happens, T is said to be ρ-hypomonotone in such neighborhood, and the main convergence result of [15] states that a "localized" version of (1) generates a sequence that converges to a point in $\hat{S}$, provided x^0 is close enough to $\hat{S}$ and $\inf \gamma_k > 2\rho$. The approach in [15] was further developed in [11], where inexact versions of the algorithm are presented, allowing for constant relative errors, in the line of the hybrid proximal algorithms of [18, 19].

In the more general context of Banach spaces, an appropriate extension of (1) can be achieved (see e.g. [6, 4]), by replacing the identity operator by the Gâteaux derivative of a strictly convex G-differentiable function $f : B \to \mathbb{R}$, i.e.,

$$x^{k+1} = \left[(\gamma_k T + f')^{-1} \circ f' \right] (x^k). \tag{2}$$

It has been proved in [4] that if T is maximal monotone and has zeroes then $\{x^k\}$ is bounded and its weak accumulation points are zeros of T, provided that f satisfies some technical assumptions (see such assumptions in Section 2). Inexact versions of the algorithm, also in the spirit of hybrid proximal methods, have been studied in [10], establishing global convergence properties similar to those proved for the exact method. These results hold in principle for maximal monotone operators.

When T fails to be monotone, and the space is non-hilbertian, the situation is more complicated. This case, as far as the proximal method is concerned, has not been treated in the literature.

We describe next the two works we know about which deal with hypomonotone operators in Banach spaces. We mention first that in Hilbert spaces monotonicity of $T^{-1} + \rho I$ is equivalent to

$$\langle x - y, u - v \rangle \geq -\rho \|x - y\|^2, \tag{3}$$

for all $x,\ y \in B$ and all $u \in T(x), v \in T(y)$. In [3], the authors define ρ-hypomonotonicity of an operator $T : B \to \mathcal{P}(B^*)$ in a Banach space B by means of (3) (which is in fact the definition given in [17] for operators in finite dimensional spaces), and proceed to study those functions such that an $\epsilon-$localization of their subdifferential is hypomonotone, with the definition above.

In [1], a more restrictive definition of hypomonotone operator is given. T is said to be ρ-hyopomonotone if

$$\langle x - y, u - v \rangle \geq -\rho \left(\|x\| - \|y\|\right)^2, \tag{4}$$

for all $x,\ y \in B$ and all $u \in T(x), v \in T(y)$. This notion is then used for developing some non-proximal methods for solving variational inequalities with so defined hypomonotone operators.

The first objective of this work is the introduction of a notion of hypomonotonicity of an operator in a Banach space more general than those given by (3) or (4). Namely, we will define hypomonotonicity with respect to a regularizing function $f : B \to \mathbb{R}$. This notion will coincide with the one given by (3) for $f(x) = (1/2)\|x\|^2$. The additional flexibility will be quite welcome when dealing with proximal methods. The square of the norm, which is a natural regularization function in a Hilbert space, enjoys no special property in a Banach space. As discussed e.g. in [10], when $B = \ell_p$ or $B = \mathcal{L}^p(\Omega)$, the computations required by the proximal method become much easier when we take $f(x) = \|x\|_p^p$ rather than $f(x) = \|x\|_p^2$. In fact we will first introduce an appropriate Yosida regularization of T with parameter ρ related to the regularization function f (see Section 3), and then T will be said to be ρ-hypomonotone with respect to f when this Yosida regularization turns out to be monotone.

Application of the proximal algorithms, as given in [4], [10], to this Yosida regularization, leads to proximal-like algorithm for hypomonotone operators, with both exact and inexact versions (see Section 4). Our error criteria allow for constant relative errors, and its convergence properties are similar to those which hold for monotone case.

The analysis above refers to operators which satisfy the hypomonotonicity property in the whole space B. Such a behavior is not generic at all, but on the other hand most operators are ρ-hypomonotone (for some adequate ρ) in a generic sense, i.e. excepting in some "small" subset of their domains.

This fact leads us to study, in Section 5, the notion of local hypomonotonicy of T, meaning hypomonotonicity of the restriction of the graph of T to some subset. By so doing, we extend the approach taken in [11] for Hilbert spaces. Finally, we propose localized versions of the algorithm, which preserve the convergence properties in a neighborhood of $S_c \times \{0\}$, where S_c is a nonempty connected component of the solution set $S = T^{-1}(0)$, provided that the initial iterate x^0 is close enough to S_c, and that T is locally hypomonotone in such a neighborhood.

2 Preliminaries

From now on, $T : B \to \mathcal{P}(B^*)$ denotes an operator from a reflexive real Banach space B to parts of its topological dual B^*. The duality pairing in $B \times B^*$ is represented by $\langle \cdot, \cdot \rangle$, meaning that for any pair $(x, x^*) \in B \times B^*$, $\langle x, x^* \rangle = x^*(x)$. Moreover, by reflexivity, we identify B with its bidual B^{**} through the canonical inclusion $\mathbb{J} : B \to B^{**}$, defined at each $x \in B$ by $\langle \mathbb{J}(x), v \rangle = v(x)$, $\forall v \in B^*$. The norm of B is denoted by $\| \cdot \|$ and the norm of B^*, by $\| \cdot \|_*$. Convergence in the strong (respectively weak, weak*) topology of a sequence will be indicated by $\xrightarrow{s}$ (respectively $\xrightarrow{w}$, $\xrightarrow{w^*}$). We remind that T is monotone if

$$\langle x - y, u - v \rangle \geq 0, \tag{5}$$

for all $x, y \in B$, all $u \in T(x)$ and all $v \in T(y)$. The domain of T is the set $D(T) = \{x \in B \mid T(x) \neq \emptyset\}$ and the graph of T is the set

$$G(T) = \{(x, x^*) \in B \times B^* \mid x^* \in T(x)\}.$$

A monotone operator $T : B \to \mathcal{P}(B^*)$ is maximal monotone if its graph is not properly included in the graph of any other monotone operator. In this case $G(T)$ is *demiclosed* (see e.g. [14, page 105]), i.e.,

$$(x^k, v^k) \in G(T); \quad x^k \xrightarrow{w} x, \quad v^k \xrightarrow{s} v \quad (\text{or} \quad x^k \xrightarrow{s} x, \quad v^k \xrightarrow{w^*} v) \implies (x, v) \in G(T). \tag{6}$$

For regularization purposes, we will use a strictly convex and Gâteaux differentiable function $f : B \to \mathbb{R}$, with Gâteaux derivative denoted by f', and such that f'_* is continuous at 0 and $f'_*(0) = 0$. We will denote the family of functions satisfying these properties as $\mathcal{F}$ (or $\mathcal{F}_B$, when it is necessary to identify the Banach space). We remind that the *Fenchel (or convex) conjugate* of f is the function defined by $f_*(v) = \sup_x\{\langle x, v \rangle - f(x)\}$, at any $v \in B^*$.

The *Bregman distance* associated to f, $D_f : B \times B \to \mathbb{R}$, is given by

$$D_f(x, y) = f(x) - f(y) - \langle x - y, f'(y) \rangle, \tag{7}$$

and the *modulus of total convexity* $\nu_f : B \times \mathbb{R}_+ \to \mathbb{R}$, is defined as

$$\nu_f(x,t) = \inf\{D_f(y,x) \mid y \in B;\ ||y-x|| = t\}. \tag{8}$$

The function f is said to be *totally convex* if

$$\nu_f(x,t) > 0 \tag{9}$$

for all $x \in B$ and all $t > 0$. Total convexity first appeared (albeit under a different name), on p. 25 of [5] and it ensures the existence of the *Bregman projection* over a nonempty closed and convex set C ([7]):

$$\Pi_C^f(x) = \arg\min_{y \in C} D_f(y,x).$$

Uniqueness of the projection follows from strict convexity of f, which in turn follows from total convexity, since the domain f is the whole space B. Our convergence results requires some of the following assumptions on f:

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

H2: $\inf_{x \in C} \nu_f(x,t) > 0$, for all bounded set $C \subset B$ and all $t \in \mathbb{R}_{++}$.

H3: f' is uniformly continuous on bounded subsets of B.

H4: f' is onto.

Observe that if $f \in \mathcal{F}$ is a function satisfying H4, then the inverse operator of its subdifferential, $(\partial f)^{-1}$, has full domain, hence it is locally bounded (in fact, it is single valued). In particular f, which is strictly convex, is Legendre and f_* is Legendre too (see [2]), implying that f_* is strictly convex. Moreover, f_* belongs to $\mathcal{F}_{B^*}$ (provided that f' is continuous at 0), $(f')^{-1} = f'_*$, f_* satisfies H4, and

$$D_f(x,y) = D_{f_*}(f'(y), f'(x)) \tag{10}$$

for any $x,\ y \in B$.

Since our analysis requires regularizing functions enjoying some or all properties H1-H4 above, it is important to establish that such functions are available in a large class of Banach spaces. In fact, it has been proved in Proposition 2 of [10] that $f(x) = 1/r\|x\|^r$ belongs to $\mathcal{F}$ and satisfies H1–H4, in any uniformly convex and uniformly smooth Banach space for all $r > 1$. In such a case, $f_* = 1/s\|\cdot\|_*^s$, defined over B^*, where $1/s + 1/r = 1$, belongs to $\mathcal{F}_{B^*}$ and satisfies H1-H4 too.

We will need also the following identities which hold for any $x,\ y,\ w,\ z \in B$:

$$D_f(w,x) - D_f(w,y) = D_f(y,x) + \langle y - w, f'(x) - f'(y)\rangle, \tag{11}$$

and

$$D_f(w,x) - D_f(w,y) = D_f\,(z,x) - D_f\,(z,y) + \langle w - z, f'(y) - f'(x)\rangle, \tag{12}$$

known as the *three-point* (see [8]) and *four-point* (see [19]) properties respectively.

3 $f-$hypomonotonicity

Consider an operator $T : B \to \mathcal{P}(B^*)$ and a regularizing function $f : B \to \mathbb{R}$, belonging to $\mathcal{F}$, with conjugate $f_* : B^* \to \mathbb{R}$ and satisfying H4.

Definition 1. *T is ρ-hypomonotone with respect to f, or (f,ρ)-hypomonotone, if*

$$\langle x - y, u - v\rangle \geq -\rho[D_{f_*}(u,v) + D_{f_*}(v,u)], \quad \forall (x,u),\ (y,v) \in G(T).$$

An (f,ρ)-hypomonotone operator is maximal if its graph is not properly contained in the graph any (f,ρ)-hypomonotone operator.

Definition 2 (Yosida regularization of T). *The Yosida regularization of T, with respect to a function $f \in \mathcal{F}$ and a parameter $\rho > 0$, is the operator $T_\rho : B \to \mathcal{P}(B^*)$ defined by*

$$T_\rho = \left[T^{-1} + \rho f_*'\right]^{-1}. \tag{13}$$

Lemma 1. *If f belong to $\mathcal{F}$, then it holds that*

$$0 \in T_\rho(x) \Longleftrightarrow x \in T^{-1}(0)$$

for any $\rho > 0$. Moreover,

i) T is (f,ρ)-hypomonotone if and only if T_ρ is monotone,

ii) T is maximal (f,ρ)-hypomonotone if and only if T_ρ is maximal monotone.

Proof. Since $f_*'(0) = 0$, we have that

$$0 \in T_\rho(x) \Longleftrightarrow x \in T_\rho^{-1}(0) = T^{-1}(0) + \rho f_*'(0) = T^{-1}(0).$$

Concerning monotonicity, observe that T_ρ is monotone if and only if

$$\langle x - y, u - v\rangle \geq 0, \quad \forall x \in T_\rho^{-1}(u) \quad \text{and} \quad y \in T_\rho^{-1}(v).$$

Since $T_\rho^{-1}(u) = T^{-1}(u) + \rho f_*'(u)$ and $T_\rho^{-1}(v) = T^{-1}(v) + \rho f_*'(v)$, monotonicity of T_ρ is equivalent to

$$\langle (x - \rho f_*'(u)) - (y - \rho f_*'(v)) + \rho\left(f_*'(u) - f_*'(v)\right), u - v\rangle \geq 0$$

for all $x \in T_\rho^{-1}(u), y \in T_\rho^{-1}(v)$, which is in turn equivalent to

$$\langle (x - \rho f_*'(u)) - (y - \rho f_*'(v)), u - v\rangle \geq -\rho\langle f_*'(u) - f_*'(v), u - v\rangle$$

for all $x \in T_\rho^{-1}(u),\ y \in T_\rho^{-1}(v)$. Note that

$$x \in T_\rho^{-1}(u) \Longleftrightarrow \hat{x} = x - \rho f_*'(u) \in T^{-1}(u),$$

and

$$y \in T_\rho^{-1}(v) \Longleftrightarrow \hat{y} = y - \rho f_*'(v) \in T^{-1}(v).$$

Thus, monotonicity of T_ρ can also be written as

$$\langle \hat{x} - \hat{y}, u - v\rangle \geq -\rho\langle f_*'(u) - f_*'(v), u - v\rangle$$

for all $\hat{x} \in T^{-1}(u)$ and $\hat{y} \in T^{-1}(v)$, which, in turn is equivalent to

$$\langle \hat{x} - \hat{y}, u - v\rangle \geq -\rho\left[D_{f_*}(u, v) + D_{f_*}(v, u)\right]$$

for all $u \in T(\hat{x})$ and $v \in T(\hat{y})$, which is just (f, ρ)-hypomonotonicity of T. We have proved (i).

Concerning maximality, i.e. item (ii), assume that T is maximal (f, ρ)-hypomonotone and let $M : B \to \mathcal{P}(B^*)$ be a monotone operator satisfying $T_\rho(x) \subset M(x)$ for all $x \in D(T_\rho)$ (i.e., $T_\rho \subset M$). Then,

$$[T^{-1} + \rho f_*'](y) \subset M^{-1}(y)$$

for all $y \in D\left(T_\rho^{-1}\right) = D\left(T^{-1}\right)$. Thus, $T^{-1}(y) \subset [M^{-1} - \rho f_*'](y)$ for all $y \in D\left(T^{-1}\right)$ and $T(x) \subset \left[M^{-1} - \rho f_*'\right]^{-1}(x)$ for all $x \in D(T)$, i.e., $T \subset \hat{T}$, where $\hat{T} = \left[M^{-1} - \rho f_*'\right]^{-1}$ is (f, ρ)-hypomonotone. In fact, $\hat{T}_\rho = M$ is monotone by hypothesis and item (i) ensures (f, ρ)-hypomonotonicity of $\hat{T}$. Since T is maximal (f, ρ) hypomonotone, $T = \hat{T}$. That is to say, $T^{-1} = \hat{T}^{-1}$, or equivalently $T^{-1} + \rho f_*' = M^{-1}$, which in turn implies that $T_\rho = M$, as required. The converse statement is proved with a similar argument. $\square$

Observe now that for a given $x \in B$ and a parameter $\gamma > 0$, the proximal subproblem applied to T_ρ can be described as

$$0 \in \gamma T_\rho(y) + f'(y) - f'(x) \Longleftrightarrow y \in \left[\gamma T_\rho + f'\right]^{-1}\left(f'(x)\right).$$

Thus, it seems appropriate to define the resolvent as follows:

Definition 3. *Given an operator $T : B \to \mathcal{P}(B^*)$ and a parameter $\gamma > 0$, the resolvent of T with respect to a regularization function f is the operator $R_{T,\gamma} : B \to \mathcal{P}(B)$ defined by*

$$R_{T,\gamma} = [\gamma T + f']^{-1} \circ f'. \tag{14}$$

The following step consists of finding the resolvent of the Yosida regularization of the operator T, because T_ρ enjoys all the properties required for convergence of the proximal method.

Lemma 2. *For any $x \in B$ and $\gamma > 0$, the following statements are equivalent:*

i) $y \in R_{T_\rho,\gamma}(x)$,

ii) $y \in T_\rho^{-1}\left(\gamma^{-1}[f'(x) - f'(y)]\right)$,

iii) $u \in T_\rho(y)$ *and* $\gamma u + f'(y) - f'(x) = 0$,

iv) $u \in T(z),\ \gamma u + f'\left[z + \rho f'_*(u)\right] - f'(x) = 0$ *and* $y = z + \rho f'_*(u)$.

Proof. By definition of the resolvent (14),

$$\begin{aligned} y \in R_{T_\rho,\,\gamma}(x) &= \left[(\gamma T_\rho + f')^{-1} \circ f'\right](x) \\ &\Leftrightarrow\ f'(x) \in [\gamma T_\rho + f'](y) = \gamma T_\rho(y) + f'(y) \\ &\Leftrightarrow\ \gamma^{-1}[f'(x) - f'(y)] \in T_\rho(y) \\ &\Leftrightarrow\ y \in T_\rho^{-1}\left(\gamma^{-1}[f'(x) - f'(y)]\right). \end{aligned}$$

Let $u = \gamma^{-1}[f'(x) - f'(y)]$. Note that the inclusion above is the same as

$$y \in T_\rho^{-1}(u), u = \gamma^{-1}[f'(x) - f'(y)] \Leftrightarrow u \in T_\rho(y) \text{ and } \gamma u + f'(y) - f'(x) = 0.$$

Apply now the definition (13) of T_ρ, and get

$$\begin{aligned} &y \in T^{-1}(u) + \rho f'_*(u),\ \gamma u + f'(y) - f'(x) = 0 \Leftrightarrow \\ &y - \rho f'_*(u) \in T^{-1}(u),\ \gamma u + f'(y) - f'(x) = 0 \Leftrightarrow \\ &u \in T\left[y - \rho f'_*(u)\right],\ \gamma u + f'(y) - f'(x) = 0. \end{aligned}$$

Taking $z = y - \rho f'_*(u)$, the last inclusion and equation above can be written as

$$u \in T(z),\ \gamma u + f'\left[z + \rho f'_*(u)\right] - f'(x) = 0 \text{ and } y = z + \rho f'_*(u).$$

□

The result of Lemma 2 allows us to define a proximal-like method for f-hypomonotone operators. Let $\{\gamma_k\}$ be a sequence of real numbers satisfying $\gamma_k \geq \gamma > 0$.

Exact Proximal-Extragradient Method:

1. Given x^k, find $(z^k, v^k) \in B \times B^*$ satisfying

$$v^k \in T(z^k) \qquad \text{and} \qquad \gamma_k v^k + f'\left(z^k + \rho f_*'(v^k)\right) - f'(x^k) = 0. \tag{15}$$

2. Define x^{k+1} by

$$x^{k+1} = z^k + \rho f_*'(v^k). \tag{16}$$

Proposition 1. *Let T be a maximal (f, ρ)-hypomonotone operator, where $\rho > 0$ and f belongs to $\mathcal{F}$ and satisfies H4. Consider a sequence of regularization parameters $\{\gamma_k\}$ such that $\gamma_k \geq \gamma > 0$. Then*

i) The exact proximal-extragradient method given by (15) *and* (16) *is well defined,*

ii) if T has zeros and f also satisfies H1-H3, then the generated sequence $\{x^k\}$ is bounded and all its weak accumulation points are zeroes of T,

iii) if either T has a unique zero or f' is sequentially weak-to-weak* continuous, then the whole sequence converges weakly to a zero of T.

Proof. Observe that, in view of Lemma 2, x^{k+1}, belongs to $R_{T_\rho,\gamma_k}(x^k)$. Thus, $\{x^k\}$ is the sequence generated by the proximal point method given by (2) applied to the operator T_ρ. By Lemma 1, T_ρ is maximal monotone, by (f, ρ)-hypomonotonicity of T, and has zeroes if and only if T does. The results then follow from the convergence properties of the proximal point method for finding zeroes of maximal monotone operators in Banach spaces (see e.g. Theorem 2 in [10]). $\square$

4 Inexact versions of the method

In this section we provide versions of the method that allow for approximate solutions of the equation (15). We start with a hybrid proximal-extragradient algorithm. The error criteria admit constant relative errors, like in [18], [19]. Let $\{\gamma_k\}$ and $\{\rho_k\}$ be sequences of real numbers satisfying $\gamma_k \geq \gamma$ and $\rho_k \geq \rho$, for all $k \geq 0$ and some $\gamma > 0, \rho > 0$. The methods also need an exogenous constant $\sigma \in [0, 1)$ (the relative error constant).

4.1 Proximal-Extragradient Method

Algorithm 1

1. Given x^k, find $(z^k, v^k) \in B \times B^*$ satisfying

$$v^k \in T(z^k), \qquad \gamma_k v^k + f'\left(z^k + \rho_k f'_*(v^k)\right) - f'(x^k) = e^k \tag{17}$$

   and

$$D_f\left(z^k + \rho_k f'_*(v^k), f'_*\left[f'(x^k) - \gamma_k v^k\right]\right) \leq \sigma D_f\left(z^k + \rho_k f'_*(v^k), x^k\right). \tag{18}$$

2. Define x^{k+1} by

$$x^{k+1} = f'_*\left[f'(x^k) - \gamma_k v^k\right]. \tag{19}$$

It is easy to check that for $\sigma = 0$, in which case both e^k and the left hand side of (18) vanish, we recover the exact method given by (15) and (16). The method is inexact because $z^k + \rho_k f'_*(v^k)$ needs not to be equal to $f'_*\left[f'(x^k) - \gamma_k v^k\right]$ (indeed, e^k is the error in the solution of the proximal equation); it is enough to request that the Bregman distance between these two points does not exceed a σ-fraction of the Bregman distance between $z^k + \rho_k f'_*(v^k)$ and the previous iterate x^k. In (19), the direction v^k, belonging to $T(z^k)$, is used to move away from $f'(x^k)$, in an extragradient fashion (adapted to the geometry of Banach spaces through the use of the auxiliary function f), and x^{k+1} is obtained by solving the equation $f'(x) = f'(x^k) - \gamma_k v^k$.

We proceed now to the convergence analysis of this algorithm.

Proposition 2. *Assume that T is a maximal (f, ρ)-hypomonotone operator and that $f \in \mathcal{F}$ satisfies H4. Then*

i) the algorithm described by (17)-(19) is well defined, i.e., the requested z^k, v^k and x^{k+1} always exist,

ii) for any $\bar{x} \in T^{-1}(0)$ it holds that

$$\begin{aligned} D_f(\bar{x}, x^{k+1}) - D_f(\bar{x}, x^k) &= (\sigma - 1) D_f(z^k + \rho_k f'_*(v^k), x^k) + \\ &+ (\rho - \rho_k)\gamma_k \left[D_{f_*}(0, v^k) + D_{f_*}(v^k, 0)\right] \end{aligned} \tag{20}$$

for all $k \geq 0$.

Proof. Note that the exact solution of the hypomonotone subproblem (i.e. the solution of (17) corresponding to $e^k = 0$), exists, because it is the solution of (15), which has solutions by Proposition 1, in view of the maximal (f, ρ)-hypomonotonicity of T. Note also that when $e^k = 0$ the left-hand side of (18) vanishes, so that the exact solution satisfies the error criterion, establishing (i). For (ii), let $\bar{x} \in T^{-1}(0)$, $y^k = z^k + \rho_k f_*'(v^k)$, and apply the four-point equality (12) to get

$$\begin{aligned}
D_f(\bar{x}, x^{k+1}) \quad - \quad & D_f(\bar{x}, x^k) = D_f\left(y^k, x^{k+1}\right) - D_f\left(y^k, x^k\right) + \langle \bar{x} - y^k, f'(x^k) - f'(x^{k+1}) \rangle \\
= \quad & D_f\left(y^k, x^{k+1}\right) - D_f\left(y^k, x^k\right) + \langle \bar{x} - z^k - \rho_k f_*'(v^k), \gamma_k v^k \rangle \\
\leq \quad & (\sigma - 1) D_f\left(y^k, x^k\right) + \gamma_k \langle \bar{x} - z^k, v^k \rangle - \gamma_k \rho_k \langle f_*'(v^k), v^k \rangle \\
\leq \quad & (\sigma - 1) D_f\left(y^k, x^k\right) + \gamma_k \rho \left[D_{f_*}(0, v^k) + D_{f_*}(v^k, 0)\right] - \gamma_k \rho_k \langle f_*'(v^k), v^k \rangle \\
= \quad & (\sigma - 1) D_f\left(y^k, x^k\right) + (\rho - \rho_k) \gamma_k \left[D_{f_*}(0, v^k) + D_{f_*}(v^k, 0)\right],
\end{aligned}$$

where the first inequality follows from the definition of x^{k+1}, in (19) and (18), and the second inequality is the (f, ρ)-hypomonotonicity property of T for the pairs $(\bar{x}, 0)$ and (z^k, v^k), both of which belong to the graph of T. $\square$

Our main convergence result for the inexact proximal-extragradient method follows.

Theorem 1. *Assume that T is a maximal (f, ρ)-hypomonotone operator where $f \in \mathcal{F}$ satisfies H1-H4. Take a constant $\sigma \in [0, 1)$, and exogenous sequences $\{\gamma_k\}$, $\{\rho_k\}$, satisfying $\gamma_k \geq \gamma$ for some $\gamma > 0$, and $\bar{\rho} \geq \rho_k \geq \rho$ for some $\bar{\rho}$. Let $\{x^k\}$ be the sequence generated defined by (17)-(19). If T has zeros then $\{x^k\}$ is bounded and all its weak accumulation points are zeroes of T.*

Proof. Let $\bar{x}$ be any zero of T. Define $y^k = z^k + \rho_k f_*'(v^k)$. Since $\sigma \in [0, 1)$, $\gamma_k > 0$ and $\rho_k \geq \rho$, Proposition 2 gives

$$D_f(\bar{x}, x^{k+1}) - D_f(\bar{x}, x^k) \leq 0.$$

Then, $\{D_f(\bar{x}, x^k)\}$ is a nonnegative and nonincreasing sequence, hence convergent, and $\{x^k\}$ is bounded, because $\{x^k\} \subset \{x \mid D_f(\bar{x}, x) \leq D_f(\bar{x}, x^0)\}$, which is bounded by assumption H1 on f. Moreover,

$$\sum_{k=0}^{n} \left((\rho_k - \rho)\gamma_k \left[D_{f_*}(0, v^k) + D_{f_*}(v^k, 0)\right] + (1 - \sigma) D_f(y^k, x^k)\right) = D_f(\bar{x}, x^0) - D_f(\bar{x}, x^{n+1}).$$

It follows that $\sum_{k=0}^{\infty} (\rho_k - \rho) \left[D_{f_*}(0, v^k) + D_{f_*}(v^k, 0)\right] < +\infty$ and $\sum_{k=0}^{\infty} (1 - \sigma) D_f(y^k, x^k) < +\infty$. In particular, $\lim_{k \to \infty} D_f(y^k, x^k) = 0$. Thus, $y^k - x^k \xrightarrow{s} 0$, because f is uniformly convex on bounded sets by H2. In view of (18)-(19), we also have $D_f\left(y^k, x^{k+1}\right) \leq \sigma D_f\left(y^k, x^k\right)$. Thus, $\lim_{k \to \infty} D_f(y^k, x^{k+1}) = 0$ and $y^k - x^{k+1} \xrightarrow{s} 0$, so that $x^k - x^{k+1} \xrightarrow{s} 0$. Equation

(19) also gives $\gamma_k v^k = f'(x^k) - f'(x^{k+1})$. Since f' is uniformly continuous on bounded sets by H3, we obtain $\gamma_k v^k \xrightarrow{s} 0$, which in turn implies $v^k \xrightarrow{s} 0$, because $\gamma_k \geq \gamma > 0$ for all k. Taking into account that $y^k = z^k + \rho_k f'_*(v^k)$, f'_* is continuous at zero and $\rho_k \leq \bar{\rho}$, it follows that $y^k - z^k \xrightarrow{s} 0$. Then, the bounded sequences $\{x^k\}$, $\{z^k\}$ and $\{y^k\}$ have the same weak accumulation points. Observe now that $v^k \in T(z^k)$ is equivalent to

$$z^k \in T^{-1}(v^k) \Leftrightarrow z^k + \rho f'_*(v^k) \in \left[T^{-1} + \rho f'_*\right](v^k) \Leftrightarrow v^k \in \left[T^{-1} + \rho f'_*\right]^{-1}(z^k + \rho f'_*(v^k)).$$

In view of (13), this is also equivalent to

$$v^k \in T_\rho(z^k + \rho f'_*(v^k)) = T_\rho(y^k + (\rho - \rho_k) f'_*(v^k)) = T_\rho(\tilde{y}^k),$$

where $\tilde{y}^k = y^k + (\rho - \rho_k) f'_*(v^k)$. It follows that $\tilde{y}^k - y^k \xrightarrow{s} 0$, so that the bounded sequences $\{x^k\}$ and $\{\tilde{y}^k\}$ also share the same weak accumulation points. Let x^∞ be a weak accumulation point of $\{x^k\}$. By demiclosedness of the graph of the maximal monotone operator T_ρ (see(6)), we get $0 \in T_\rho(x^\infty)$, which implies $0 \in T(x^\infty)$, in view of Lemma 1. $\square$

4.2 Proximal-Projection Method

In the following algorithm, the extragradient step (19) is replaced by a Bregman projection onto a hyperplane separating the current iterate from the set of zeroes of T. As in the case of proximal-extragradient method, e^k is the error in the solution of the proximal equation, and σ is the maximal relative error admitted. The approximate solution z^k, together with the vector $v^k \in T(z^k)$, allow us to construct the hyperplane H_k with the announced separating property, and the next iterate is the Bregman projection of the current one onto H_k. The algorithm requires a constant $\sigma \in [0, 1)$ (relative error constant), and exogenous sequences $\{\gamma_k\}$ and $\{\rho_k\}$ satisfying $\gamma_k \geq \gamma$ and $\bar{\rho} \geq \rho_k \geq \rho$, for all $k \geq 0$ and some $\bar{\rho}$, $\gamma > 0$. It is formally defined as:

Algorithm 2

1. Given x^k, find $(z^k, v^k) \in B \times B^*$ satisfying

$$v^k \in T(z^k), \qquad \gamma_k v^k + f'\left(z^k + \rho_k f'_*(v^k)\right) - f'(x^k) = e^k \tag{21}$$

and

$$\langle z^k + \rho_k f'_*(v^k) - x^k, e^k \rangle \leq \sigma D_f\left(z^k + \rho_k f'_*(v^k), x^k\right). \tag{22}$$

2. Define x^{k+1} as

$$x^{k+1} = \Pi^f_{H_k}(x^k) = \arg\min_{x \in H_k} D_f(x, x^k), \tag{23}$$

where

$$H_k = \{x \in B \mid \langle x - \left[z^k + \rho_k f'_*(v^k)\right], v^k \rangle \leq 0\}. \tag{24}$$

We proceed now to the convergence analysis of this algorithm.

Proposition 3. *Assume that T is a maximal (f,ρ)-hypomonotone operator, where $f \in \mathcal{F}$ satisfies H4 and is totally convex (see* (8), (9)*). Consider $\sigma \in [0,1)$, and $\{\rho_k\},\{\gamma_k\}$ as in the statement of the method. Then*

i) the algorithm described in (21)-(23) *is well defined:*

- *a) for any k the proximal-like subproblem* (21) *has a solution satisfying* (22)*,*
- *b) $H_k \neq \emptyset$, and therefore the Bregman projection onto H_k* (23) *exists,*
- *c) $T^{-1}(0) \subset H_k$,*
- *d) if $x^k \neq z^k + \rho_k f'_*(v^k)$ then x^k belongs to H_k^+, where*

$$H_k^+ = \{x \in B \mid \langle x - \left[z^k + \rho_k f'_*(v^k)\right], v^k\rangle > 0\},$$

ii) for any $k \geq 0$,

$$\begin{aligned} D_f(x^{k+1},x^k) - D_f\left(x^{k+1}, z^k + \rho_k f'_*(v^k)\right) &\geq (1-\sigma) D_f\left(z^k + \rho_k f'_*(v^k), x^k\right) + \\ &+ \langle x^{k+1} - x^k, e^k\rangle, \end{aligned} \tag{25}$$

iii) if $\bar{x} \in T^{-1}(0)$ then

- *a) the sequence $\{D_f(\bar{x},x^k)\}$ is nonincreasing and convergent,*
- *b) $\sum_{k=0}^{+\infty} D_f(x^{k+1},x^k)$ is convergent.*

Proof. The proof of item (i-a) is similar to the proof of Proposition 2. For proving (i-b), observe that $z^k + \rho_k f'_*(v^k)$ belongs to H_k. For proving (i-c), take $\bar{x} \in T^{-1}(0)$. Then

$$\begin{aligned} \langle \bar{x} - \left[z^k + \rho_k f'_*(v^k)\right], v^k\rangle &= \langle \bar{x} - z^k, v^k\rangle - \rho_k \langle f'_*(v^k), v^k\rangle \\ &\leq \rho\left[D_{f_*}(0,v^k) + D_{f_*}(v^k,0)\right] - \rho_k\langle f'_*(v^k), v^k\rangle \\ &= (\rho - \rho_k)\left[D_{f_*}(0,v^k) + D_{f_*}(v^k,0)\right] \leq 0, \end{aligned}$$

where the first inequality follows from the (f,ρ)-hypomonotonicity of T and the second one from the fact that $\rho_k \geq \rho$.

Item (i-d) is a consequence of the error criterion given in (22): defining $y^k = z^k + \rho_k f'_*(v^k)$, we get, in view of (21),

$$\begin{aligned} \gamma_k \langle x^k - \left[z^k + \rho_k f'_*(v^k)\right], v^k\rangle &= \langle x^k - y^k, f'(x^k) - f'(y^k) + e^k\rangle \\ &= \left[D_f(x^k,y^k) + D_f(y^k,x^k)\right] - \langle y^k - x^k, e^k\rangle \\ &\geq \left[D_f(x^k,y^k) + D_f(y^k,x^k)\right] - \sigma D_f(y^k,x^k) \\ &= (1-\sigma) D_f(y^k,x^k) + D_f(x^k,y^k). \end{aligned}$$

Hence $\langle x^k - \left[z^k + \rho_k f'_*(v^k)\right], v^k\rangle > 0$, unless $y^k = x^k$.

For proving (ii), apply the three point equality (11) and get

$$\begin{aligned} D_f(x^{k+1},x^k) - D_f(x^{k+1},y^k) &= D_f(y^k,x^k) + \langle y^k - x^{k+1}, f'(x^k) - f'(y^k)\rangle \\ &= D_f(y^k,x^k) + \langle y^k - x^{k+1}, \gamma_k v^k\rangle - \langle y^k - x^{k+1}, e^k\rangle \\ &= D_f(y^k,x^k) - \langle y^k - x^k, e^k\rangle - \langle x^k - x^{k+1}, e^k\rangle \\ &\geq (1-\sigma)D_f(y^k,x^k) + \langle x^{k+1} - x^k, e^k\rangle. \end{aligned}$$

Finally, in order to prove (iii), take $\bar{x} \in T^{-1}(0)$. Note that $\bar{x} \in H_k$ by (i-c), and so, since x^{k+1} is a solution of the optimization problem given in (23), it satisfies the first order optimality conditions, namely

$$\langle x^{k+1} - \bar{x}, f'(x^k) - f'(x^{k+1})\rangle \geq 0. \tag{26}$$

Using now (26) and the three point equality, we get

$$D_f(\bar{x},x^k) - D_f(\bar{x},x^{k+1}) = D_f(x^{k+1},x^k) + \langle x^{k+1} - \bar{x}, f'(x^k) - f'(x^{k+1})\rangle \geq D_f(x^{k+1},x^k).$$

Thus, $\{D_f(\bar{x},x^k)\}$ is nonincreasing and bounded, hence convergent. Moreover,

$$\sum_{k=0}^{n} D_f(x^{k+1},x^k) \leq \sum_{k=0}^{n}\left[D_f(\bar{x},x^k) - D_f(\bar{x},x^{k+1})\right] = D_f(\bar{x},x^0) - D_f(\bar{x},x^{n+1}),$$

which ensures convergence of $\sum_{k=0}^{\infty} D_f(x^{k+1},x^k)$. □

The following theorem contains our main result concerning the proximal-projection algorithm.

Theorem 2. *Assume that T is a maximal (f,ρ)-hypomonotone operator, where f belongs to $\mathcal{F}$ and satisfies H1-H4. Consider $\{\gamma_k\}, \{\rho_k\}$ and σ as in the statement of the method. Let $\{x^k\}$ be the sequence generated by* (21)-(23). *Then*

i) f T has zeros then $\{x^k\}$ is bounded,

ii) if additionally $e^k \xrightarrow{s} 0$, then all weak accumulation points of $\{x^k\}$ are zeroes of T.

Proof. Let $\bar{x}$ be a zero of T. Define $y^k = z^k + \rho_k f'_*(v^k)$. From Proposition 3(iii), we get that $\{x^k\}$ is bounded, establishing (i). We also get that $x^{k+1} - x^k \xrightarrow{s} 0$. Moreover, Proposition 3(ii) ensures that

$$D_f(x^{k+1},x^k) - D_f\left(x^{k+1},y^k\right) \geq (1-\sigma)D_f\left(y^k,x^k\right) + \langle x^{k+1} - x^k, e^k\rangle. \tag{27}$$

Take now limits with $k \to \infty$ in (27). Since σ belongs to $[0,1)$ and $\{e^k\}$ is bounded, by the assumption of item (ii), we have $\lim_{k\to\infty} D_f\left(x^{k+1}, y^k\right) = 0$ and $\lim_{k\to\infty} D_f\left(y^k, x^k\right) = 0$. Hence, $y^k - x^k \xrightarrow{s} 0$, because f satisfies H2. Observe now that (22) implies

$$\gamma_k v^k + f'(y^k) - f'(x^k) = e^k. \tag{28}$$

Since f' is uniformly continuous on bounded sets by H3, and $\gamma_k \geq \gamma > 0$ for all k, we obtain from (28) $v^k \xrightarrow{s} 0$, as a consequence of the assumption that $e^k \xrightarrow{s} 0$. Thus, $\{x^k\}$ is a bounded sequence with the same weak accumulation points as $\{y^k\}$, and we also have $v^k \in T(z^k)$, $v^k \xrightarrow{s} 0$ and $y^k - z^k = \rho_k f'_*(v^k) \xrightarrow{s} 0$. The remainder of the proof uses the final argument in the proof of Theorem 1. □

We mention a remarkable weakness of this convergence result, in comparison with the similar result for the inexact proximal-extragradient method, i.e., Theorem 1. In this case, we need to include among the assumptions the convergence to 0 of the error term e^k. Such assumption was not needed in Theorem 1, where in fact such convergence to 0 was a consequence of the result itself. We have been unable to get a similar result in this case, and the question whether such assumption is indeed essential is left as an open problem for future research.

5 Local f-hypomonotonicity

Up to this point, our results apply to operators which are (f,ρ)-hypomonotone in the whole space. The class of such operators is of course much larger than the class of monotone operators, but in some respect it is not large enough.

We observe that the set of zeroes of an (f,ρ)-hypomonotone operator, with $f \in \mathcal{F}$, is closed and convex, and hence also weakly closed, because according to Lemma 1, it coincides with the set of zeroes of a maximal monotone operator, namely T_ρ, and it is well known that the set of zeroes of a maximal monotone operator enjoys such properties. In fact, this result indicates that (f,ρ)-hypomonotonicity is not generic: in the one dimensional case, for instance, only functions whose set of zeroes is either empty or an interval can be (f,ρ)-hypomonotone. On the other hand, most well behaved operators are locally (f,ρ)-hypomonotone near a zero of T (see, e.g. in [11] the discussion of this genericity for the case of Hilbert spaces). We will consider thus operators which are (f,ρ)-hypomonotone only in a certain subset W of $B \times B^*$, in the following sense.

Recall that a set-valued operator $T : B \to \mathcal{P}(B^*)$ can be identified with its graph. By so doing, the (f,ρ)-hypomonotonicity of T becomes a property of the graph of T, as a subset of $B \times B^*$. In fact, we can say that an arbitrary subset Z of $B \times B^*$ is (f,ρ)-hypomonotone when

$$\langle x-y, u-v\rangle \geq -\rho[D_{f_*}(u,v)+D_{f_*}(v,u)], \quad \forall (x,u),\ (y,v) \in Z.$$

Definition 4. *Given $\rho > 0$, a regularizing function $f \in \mathcal{F}$ and a subset W of $B \times B^*$, an operator $T : B \to \mathcal{P}(B^*)$ is said to be*

i) (f,ρ)-hypomonotone in W if and only if $T \cap W$ is (f,ρ)-hypomonotone.

ii) maximal (f,ρ) hypomonotone in W if and only if T is (f,ρ)-hypomonotone in W, and additionally $T \cap W = T' \cap W$, whenever $T' \subset B \times B^*$ is (f,ρ)-hypomonotone and $T \cap W \subset T' \cap W$.

For the sake of simplicity, we will take W of the form $W = U \times V$, with $U \subset B$, $V \subset B^*$. In order to obtain any convergence results, we need to assume that U contains some zero of T. Furthermore, for convenience, we will assume that U is closed and convex. In such a case, it happens that $T^{-1}(0) \cap U$ is also closed and convex, and that the Bregman Π^f projection onto this set is well defined, when f is totally convex.

In fact, we need some additional assumptions on the set W where the local (f,ρ)-hypomonotonicity holds. The reason is the following: in order to get any meaningful result, we need that whenever an iterate belongs to W the next one also does. This will be a consequence of the fact that the Bregman distance from the iterates to any zero of T decreases (see Propositions 2(ii) and 3(ii)), but then we will need that W contains a whole neighborhood of $(T^{-1} \cap U) \times \{0\} \subset B \times B^*$. In other words, U will have to contain an open set around a closed and convex subset of the set of zeroes of T. These assumptions on $W = U \times V$ will appear as hypotheses (i) and (ii) in Lemma 3 below.

Finally, we will need also two additional technical assumptions on f, or more precisely on its conjugate f_*: the first one requires that if $x, y \in B^*$ are close to each other with respect to D_{f_*}, then $x - y$ must be close to 0 in the same sense. Formally, we have:
H5: There exists a nondecreasing function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ such that, for all $\delta > 0$ and all $x, y \in B^*$ it holds that

$$D_{f_*}(\delta(x-y), 0) \leq \phi(\delta) D_{f_*}(x,y).$$

The second one, which relates f_* with its derivative f'_*, is the following:
H6: There exists a nondecreasing function $\psi : \mathbb{R}_{++} \to \mathbb{R}_{++}$ such that, if $f_*(x) \leq \psi(\lambda)$ then $\|f'_*(x)\| \leq \lambda$.

We mention that H6 holds for $f(x) = (1/r)\|x\|^r$ $(r > 1)$ in any uniformly smooth and uniformly convex Banach space, with $\psi(t) = [(r-1)/r]t^r$, which is increasing.

In connection with H5, we consider now the spaces $B = \mathcal{L}^p(\Omega)$ or $B = \ell_p$, with $1 < p \leq 2$, where the standard regularizing function is given by $f(x) = (1/p)\|x\|_p^p$. In this case we have

$B^* = \mathcal{L}^q(\Omega)$ or $B^* = \ell_q$, and $f_*(y) = (1/q)\,\|y\|_q^q$, with $1/p + 1/q = 1$. In this setting, it has been proved in [9] that the modulus of total convexity of f_*, as in (8), satisfies $\nu_{f_*}(y,t) \geq (2^{1-q}/q)\,t^q$ for all $y \in B^*$. It follows also from the definition of Bregman distance that if $f_*'(0) = 0$ then $D_{f_*}(y,0) = f_*(y)$ for all $y \in B^*$. Thus

$$\begin{aligned} D_{f_*}(x,y) &\geq \nu_{f_*}(y, \|x-y\|) \geq \frac{2^{1-q}}{q}\|x-y\|_q^q = \frac{2^{1-q}}{q}\delta^{-q}\|\delta(x-y)\|_q^q \\ &= 2^{1-q}\delta^{-q} D_{f_*}(\delta(x-y),0), \end{aligned}$$

i.e., H5 holds with $\phi(t) = 2^{q-1}t^q$, which is increasing.

We present next a localization lemma, which says that if T is locally (f,ρ)-hypomonotone on a set $W = U \times V$ satisfying some regularity properties, and if $x \in U$ is close enough to a zero of T belonging to U, then, the vectors u, y, z associated to the resolvent $R_{T_\rho,\gamma}(x)$ as in Lemma 2(iv) will be such that u belongs to V and y, z belong to U. The assumption on V is that it contains a level set of $D_{f_*}(\cdot,0)$. Regarding U, the requirement that it contains an open set around its intersection with $T^{-1}(0)$, will be expressed in the following way: it contains all points of the form $a+b$, where the distance from a to $U \cap T^{-1}(0)$ does not exceed some value $\beta > 0$ and the norm of b does not exceed some $\eta > 0$.

It is convenient to introduce some notation for the sublevel sets of $D_{f_*}(\cdot,0)$. For $\alpha \geq 0$, define

$$L_{f_*}(\alpha) = \{y \in B^* \mid f_*(v) = D_{f_*}(v,0) \leq \alpha\}.$$

Lemma 3. *Let T be maximal (f,θ)-hypomonotone in a subset $U \times V$ of $B \times B^*$, where $f \in \mathcal{F}$ satisfies H4, H5 and H6. Assume that T has a nonempty set of zeroes S, and that U and V satisfy the following two conditions:*

i) $S \cap U$ is nonempty and U is closed and convex,

ii) there exist $\alpha, \beta, \eta \in \mathbb{R}_{++}$ such that

$$L_{f_*}(\alpha) \subset V, \tag{29}$$

$$\{a + b \in B \mid D_f(S \cap U, a) \leq \beta, \|b\| \leq \eta\} \subset U. \tag{30}$$

Fix some $\delta > 0$ and some $\mu \geq \theta$, and take ε satisfying

$$\varepsilon \leq \min\left\{\beta, \frac{\alpha}{\phi(1/\delta)}, \frac{\psi(\eta/\mu)}{\phi(1/\delta)}\right\}, \tag{31}$$

with ϕ as in H5 and ψ as in H6. If $x \in B$ is such that $D_f(S \cap U, x) \leq \varepsilon$, then there exist $y \in U$, $z \in U$ and $u \in V$, such that *a)* $u \in T(z)$, $\delta u + f'[z + \theta f'_*(u)] - f'(x) = 0$, $y = z + \theta f'_*(u)$,

b) $D_f(S \cap U, y) \le \varepsilon$.

Proof. One comment is in order before starting with the formal proof: there is not much difficulty in finding vectors u, y, z satisfying (a): they will be the vectors associated to the resolvent $R_{T_\theta,\delta}(x)$. The main difficulty is to establish that they remain in W (or, more precisely, that $u \in V$ and $y, z \in U$). It is at this point that we need H5, H6 and the definition of ε.

Take $x \in B$ such that $D_f(S \cap U, x) \le \varepsilon$. Since T is maximal (f, θ)-hypomonotone in $U \times V$, T_θ is also maximal monotone in $U \times V$. Identifying T_θ with its graph, we get that $T_\theta \cap (U \times V)$ is a monotone subset of $B \times B^*$. Thus, there exists some maximal monotone subset of $B \times B^*$, say $\hat{T}_\theta$, which contains $T_\theta \cap (U \times V)$. Since $\hat{T}_\theta$ is maximal monotone, its resolvent $R_{\hat{T}_\theta,\delta}$ is well defined. Take $y \in R_{\hat{T}_\theta,\delta}(x)$. In view of Lemma 2, there exists $u \in B^*$ such that $u \in \hat{T}_\theta(y)$ and

$$\delta u + f'(y) - f'(x) = 0. \tag{32}$$

Take any $\bar{x} \in S \cap U$ and apply the three point equality (11) to get

$$\begin{aligned} D_f(\bar{x}, x) - D_f(\bar{x}, y) &= D_f(y, x) + \langle y - \bar{x}, f'(x) - f'(y) \rangle \\ &= D_f(y, x) + \delta \langle u, y - \bar{x} \rangle \ge D_f(y, x), \end{aligned} \tag{33}$$

where the inequality follows from monotonicity of $\hat{T}_\theta$ and the fact that $\bar{x}$ is a zero of T, and hence of $\hat{T}_\theta$. In view of (33),

$$\inf_{\bar{x} \in S \cap U} D_f(\bar{x}, x) \ge \inf_{\bar{x} \in S \cap U} [D_f(\bar{x}, y) + D_f(y, x)] \ge D_f(y, x) + \inf_{\bar{x} \in S \cap U} D_f(\bar{x}, y). \tag{34}$$

Since $D_f(S \cap U, x) \le \varepsilon$, we conclude that $D_f(S \cap U, y) \le \varepsilon$, establishing (b). Using now assumption (ii) and the definition of ε, we obtain that $y \in U$.

Next, invoking (34) and (10), we get

$$\varepsilon \ge D_f(y, x) = D_{f_*}(f'(x), f'(y)).$$

In view of (32), $u = \delta^{-1}(f'(x) - f'(y))$. Since f satisfies H4 and H5, we have

$$f_*(u) = D_{f_*}(u, 0) = D_{f_*}\left(\delta^{-1}(f'(x) - f'(y)), 0\right) \le \phi(1/\delta)\varepsilon \le \alpha, \tag{35}$$

using (31) in the inequality. It follows from (29) that u belongs to V. Hence, (y, u) belongs to $\hat{T}_\theta \cap (U \times V)$. Since $\hat{T}_\theta$ coincides with T_θ in $U \times V$, it follows that $u \in T_\theta(y)$. Now we use Lemma 2(iv), and obtain that there exists $z \in B$ such that $u \in T(z)$, $\delta u + f'[z + \theta f'_*(u)] - f'(x) = 0$

and $y = z + \theta f'_*(u)$. Observe that $z = y - \theta f'_*(u)$, with $y \in \{x \in B \mid D_f(S \cap U, x) \le \varepsilon\}$. Since $\varepsilon \le \beta$ by (31), we get that $y \in \{a \in B \mid D_f(S \cap U, a) \le \beta\}$. Let now $b = -\theta f'_*(u)$. Then,

$$\|b\| = \theta \|f'_*(u)\| . \tag{36}$$

Note that

$$f_*(u) \le \phi\left(\frac{1}{\delta}\right)\varepsilon \le \psi\left(\frac{\eta}{\mu}\right) \le \psi\left(\frac{\eta}{\theta}\right), \tag{37}$$

using (35) in the first inequality, (31) in the second one, and the facts that ψ is nondecreasing and that $\theta \le \mu$ in the third one. It follows from (37) and H6 that $\|f'_*(u)\| \le \eta/\theta$, and therefore, by (36), $\|b\| \le \eta$. We have written z as $y + b$ with $y \in \{a \in B \mid D_f(S \cap U, a) \le \beta\}$ and $\|b\| \le \eta$. We conclude from (30) that z belongs to U, completing the proof. ☐

Our local convergence result for our Algorithms 1 and 2 applied to locally (f,ρ)-hypomonotone operators is contained in the following theorem.

Theorem 3. *Let T be maximal (f,ρ)-hypomonotone in $U \times V \subset B \times B^*$, where $f \in \mathcal{F}$ satisfies H1-H6, and U, V satisfy conditions (i) and (ii) of Lemma 3. Assume that the set S of zeroes of T is nonempty. Take an exogenous constant $\sigma \in [0,1)$ and exogenous sequences $\{\gamma_k\}$, $\{\rho_k\}$ such that $\gamma_k \ge \gamma > 0$, $\bar\rho \ge \rho_k \ge \rho > 0$ for some $\gamma, \bar\rho$. Define ε as*

$$\varepsilon = \min\left\{\beta, \frac{\alpha}{\phi(1/\gamma)}, \frac{\psi(\eta/\bar\rho)}{\phi(1/\gamma)}\right\}. \tag{38}$$

If $D_f(S \cap U, x^0) \le \varepsilon$,

a) for all k there exist $z^k \in U$, $v^k \in V$, $e^k \in B^*$ and $x^{k+1} \in U$ satisfying (17)-(18) *and* (19) *in the case of Algorithm 1, and* (21)-(22) *and* (23), *in the case of Algorithm 2, such that $D_f(S \cap U, x^{k+1}) \le \varepsilon$.*

(b) Any sequence $\{x^k\}$ constructed as in item (a) is bounded and any weak accumulation point of $\{x^k\}$ belongs to $S \cap U$ (provided that $e^k \xrightarrow{s} 0$ in the case Algorithm 2).

Proof. a) We proceed by induction. By inductive hypothesis, $D_f(S \cap U, x^k) \le \varepsilon$. We intend to apply Lemma 3 with $x = x^k$, $\delta = \gamma_k$, $\theta = \rho_k$ and $\mu = \bar\rho$. We proceed to check that the hypotheses of this lemma are satisfied. First, note that if T is $(f,\hat\rho)$-hypomonotone and $\tilde\rho \ge \hat\rho$, then T is also $(f,\tilde\rho)$-hypomonotone. Since $\rho_k \ge \rho$ for all k, and T is (f,ρ)-hypomonotone by hypothesis, we obtain that T is (f,ρ_k)-hypomonotone for all k, and hence by choosing $\theta = \rho_k$ we remain within the assumptions of the lemma. Additionally, it suffices to check that ε, defined as in (38), satisfies (31) with $\delta = \gamma_k$ and $\mu = \bar\rho$, for which it is enough to verify that $\phi(1/\gamma_k) \leq \phi(1/\gamma)$, and that $\psi(\eta/\bar{\rho}) \leq \psi(\eta/\rho_k)$. These inequalities hold because $\gamma_k \geq \gamma$, $\rho_k \leq \bar{\rho}$ and ϕ, ψ are nondecreasing by H5, H6 respectively. Thus, we invoke the Lemma 3 with these values of x, δ, θ, μ and conclude that there exist $y, z \in U$ and $u \in T(z) \cap V$ satisfying the required equations (e.g. with $e^k = 0$). Take $v^k = u$, $z^k = z$, and consider x^{k+1} defined from v^k, z^k as in (19) for Algorithm 1, and (23) for Algorithm 2. It remains to establish that in both cases $D_f(S \cap U, x^{k+1}) \leq \varepsilon$. Take $\bar{x} \in S \cap U$. For the case of Algorithm 1, since both $(\bar{x}, 0)$ and (z^k, v^k) belong to $G(T) \cap (U \times V)$, and T is (f, ρ)-hypomonotone in $U \times V$, we get that

$$D_f(\bar{x}, x^{k+1}) - D_f(\bar{x}, x^k) \leq 0. \tag{39}$$

following exactly the steps in the proof of Proposition 2 which lead to (20). In the case of Algorithm 2, we also have that $(\bar{x}, 0)$ and (z^k, v^k) belong to $G(T) \cap (U \times V)$, where T is (f, ρ)-hypomonotone, and we also obtain (39). In both cases, we get from (39)

$$D_f(S \cap U, x^{k+1}) = \inf_{\bar{x} \in S \cap U} D_f(\bar{x}, x^{k+1}) \leq \inf_{\bar{x} \in S \cap U} D_f(\bar{x}, x^k) = D_f(S \cap U, x^k) \leq \varepsilon.$$

establishing the result.

b) By (a), the sequence $\{(x^k, v^k)\}$ remains in $W = U \times V$. Since T is (f, ρ)-hypomonotone in this set, thus results of Theorem 1 for Algorithm 1 and Theorem 2 for Algorithm 2 hold also locally, and so all weak accumulation points of $\{x^k\}$ belong to S. Note also that, in view of (a), the whole sequence $\{x^k\}$ is contained in U, which is closed and convex, and therefore weakly closed. Thus all weak accumulation points of $\{x^k\}$ belong to U, completing the proof. $\square$

References

[1] Alber, Y., Butnariu, D., and Ryazantseva, I. Regularization and resolution of monotone variational inequalities with operators given by hypomonotone approximations. *J. Nonlinear Convex Anal. 6*, 1 (2005), 23–53.

[2] 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*, 4 (2001), 615–647.

[3] Bernard, F., and Thibault, L. Prox-regularity of functions and sets in Banach spaces. *Set-Valued Anal. 12*, 1-2 (2004), 25–47.

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

[5] Butnariu, D., Censor, Y., and Reich, S. Iterative averaging of entropic projections for solving stochastic convex feasibility problems. *Comput. Optim. Appl. 8*, 1 (1997), 21–39.

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

[7] Butnariu, D., Iusem, A. N., and Resmerita, E. Total convexity for powers of the norm in uniformly convex Banach spaces. *J. Convex Anal. 7*, 2 (2000), 319–334.

[8] Chen, G., and Teboulle, M. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM J. Optim. 3*, 3 (1993), 538–543.

[9] Isnard, C. A., and Iusem, A. N. On mixed Hölder-Minkowski inequalities and total convexity of certain functions in $L^p(\Omega)$. *Math. Inequal. Appl. 3*, 4 (2000), 519–537.

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

[11] Iusem, A. N., Pennanen, T., and Svaiter, B. F. Inexact variants of the proximal point algorithm without monotonicity. *SIAM J. Optim. 13*, 4 (2003), 1080–1097 (electronic).

[12] Kaplan, A., and Tichatschke, R. Proximal point methods and nonconvex optimization. *J. Global Optim. 13*, 4 (1998), 389–406. Workshop on Global Optimization (Trier, 1997).

[13] Krasnosel′skiĭ, M. A. Two remarks on the method of successive approximations. *Uspehi Mat. Nauk (N.S.) 10*, 1(63) (1955), 123–127.

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

[15] Pennanen, T. Local convergence of the proximal point algorithm and multiplier methods without monotonicity. *Math. Oper. Res. 27*, 1 (2002), 170–191.

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

[17] Rockafellar, R. T., and Wets, R. J.-B. *Variational analysis*, vol. 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1998.

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

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