AN INEXACT GENERALIZED PROXIMAL POINT ALGORITHM IN BANACH SPACES

G. J. P. da Silva[*]
Engenharia de Sistemas e Computação, COPPE-UFRJ,CP 68511,
Rio de Janeiro, RJ, CEP 21945-970, Brazil.
geci@cos.ufrj.br

B. F. Svaiter[†]
Instituto de Matemática Pura e Aplicada,
Estrada Dona Castorina 110,
Jardim Botânico,
Rio de Janeiro, RJ, CEP 22460-320, Brazil.
benar@impa.br

12/03/2001

Abstract

In this paper we prove weak convergence of an inexact proximal-like method for finding zeroes of a maximal monotone operator in reflexive Banach spaces. The method we consider is an inexact version of the proximal-like method proposed by Burachick and Scheimberg [9]. This inexact method can also be viwed as an extension of an inexact proximal-like method proposed by Solodov and Svaiter [29] in Hilbert Spaces with quadratic regularization.

Key words. Proximal point method, Banach spaces, maximal monotone operators, enalargement of a maximal monotone operator.
AMS subject classifications. 90C25, 90C30.

[*]Partially supported by PICDT/UFG-CAPES
[†]Partially supported by CNPq Grant 301200/93-9(RN) and by PRONEX–Optimization.

1 Introduction

Let B be a reflexive Banach space, with dual B^*, and $T : B \rightrightarrows B^*$ be a maximal monotone operator. We consider the problem

$$\text{find} \quad x \in B, \quad \text{such that} \quad 0 \in T(x). \tag{1.1}$$

Problem (1.1) is a general framework for many important real-world problems in economics and engineering. These include optimal control problems, shape optimization problems, and the problem of minimal area surface with obstacles, among many others. In many shape optimization problems[23], the function space is only a Banach and not a Hilbert space. An important motivation for analyzing the convergence properties of algorithms for (1.1) is related to the so-called *Mesh Independence Principle* [2, 1, 17, 21, 23]. The Mesh Independence Principle relies on infinite-dimensional convergence results for predicting the convergence properties of the discretized finite-dimensional method. Furthermore, it provides a theoretical foundation for the justification of refinement strategies and helps to design this refinement process. Since the focus is the infinite-dimensional solution, a fine discretization scheme has to be chosen, so that the discrete solution approximates the infinite-dimensional solution appropriately.

Recently, the following algorithm was studied in [9]. It employs a regularizing Bregman function f. Starting from $p^0 \in B$, a sequence $\{p^k\}$ is iteratively generated by the rule:

$$\begin{aligned} &p^{k+1} \text{ solves the problem} \\ &0 \in c_k T(x) + \nabla f(x) - \nabla f(p^k), \end{aligned} \tag{1.2}$$

where $\{c_k\}$ is a strictly positive sequence bounded away from zero. Additional properties of f, to insure well-definedness and convergence to a solution have been established in [9] and will be described later on. If B is a Hilbert space and $f(x) = 1/2\|x\|^2$, algorithm (1.2) reduces to the *classical* proximal point method in a Hilbert space [27]. So, scheme (1.2) can be seen as an extension to a Banach space of the proximal point method. From now on, we call scheme (1.2) a "generalized proximal point method", GPPM, for short.

Generalized proximal point methods were studied by many authors in the last decade. Some of these works deal with the optimization problem, that is, when the operator T is a subdifferential: [13], [14], [22]. Generalized proximal point methods for finding a zero of an arbitrary maximal monotone operator, have been studied in [3], [15], [18], for finite dimensional spaces, [5] for Hilbert spaces, and , [12], [4] for Banach spaces.

The convergence analysis of (1.2) presented in [9] considered the *exact* algorithm. For designing implementable algorithms for (1.1), it is essential to consider inexact versions of (1.2). This also holds for the implemetable discretized finite dimensional method for solving (1.1). In the case of the proximal point method, the classical analysis relies in a summable error tolerance[27]. The "summable error" approach has been extended to GPPM in finite dimensional spaces in [16]. Modifications of the proximal point method and generalize proximal point methods, admitting a *relative* error tolerance has been recently proposed [28, 29, 30]. A common feature of these modifications is that the proximal iteration is broken in two steps: first, the proximal subproblem is solved approximately; then a projection or extragradient step is performed. As these algorithms combine a proximal with an extragradient or projection step, we will call them Hybrid methods.

An outer approximation of the maximal monotone operator T, called T^ε was introduced in [6]. Properties of T^ε in Hilbert and Banach spaces were studied, respectively, in [7] and [8]. This outer approximation of T allows one to define approximated solutions of the generalized proximal subproblem, with an error in the inclusion. This inexact scheme for GPPM in a finite dimensional space was considered in [6], still in the summable error framework. In [29], a hybrid version of the proximal point method in a Hilbert space, using T^ε, and working with a relative error tolerance was proposed.

The hybrid scheme for GPPM, whith a relative error tolerance, has been already studied for finite dimensional spaces and Banach spaces respectively in [30, 20]. Even though, these works doesn't consider errors on the operator T by means of T^ε. The aim of this work is to extend the hybrid scheme/relative error framework to GPPM in Banach spaces, using the enlargement T^ε. Morover, we will also allow a summable (plus relative) error. We will also show that each step of the proposed method is equivalent to an approximated minimization of a convex merit function. Moreover, exact minimization of this funcions yields the exact solution of (1.2).

2 Theoretical preliminaries

Let B be a real Banach space with dual B^*. An operator $T : B \rightrightarrows B^*$ is *monotone* if

$$\langle v - u, x - y \rangle \geq 0.$$

for all $x, y \in B$, $v \in T(x)$, $u \in T(y)$. The operator T is *maximal monotone* if it is monotone and for any monotone $T' : B \rightrightarrows B^*$,

$$T(x) \subseteq T'(x)$$

for all $x \in B$ implies $T' = T$.

Burachik, Iusem and Svaiter introduced in [6] an enlargement of a maximal monotone operator $T : B \rightrightarrows B^*$ as follows: for $x \in B$, $\varepsilon \geq 0$,

$$T^\varepsilon(x) = \{u \in B^* \mid \langle v - u, y - x \rangle \geq -\varepsilon, \forall y \in B,\ v \in T(y)\}. \tag{2.1}$$

Properties of this enlargement in Hilbert spaces and Banach spaces where studied respectivelly in [7, 8]. Trivially, from (2.1)

$$0 \leq \varepsilon_1 \leq \varepsilon_2 \quad \Longrightarrow T^{\varepsilon_1}(x) \subseteq T^{\varepsilon_2}(x) \ \ \forall x \in B.$$

Also, from the maximal monotonicity of T

$$T^0 = T.$$

So, for any $\varepsilon \geq 0$ and $x \in B$,

$$T(x) \subseteq T^\varepsilon(x).$$

An alternative notation for T^ε is

$$\begin{aligned} &\mathbf{B}_T : \mathbb{R}_+ \times B \rightrightarrows B^*, \\ &\mathbf{B}_T(\varepsilon, x) := \{u \in B^* \mid \langle v - u, y - x \rangle \geq -\varepsilon, \forall y \in B,\ v \in T(y)\}. \end{aligned} \tag{2.2}$$

Then $T^\varepsilon(x) = \mathbf{B}_T(\varepsilon, x)$.

Let $f : B \to \mathbb{R}$ be a Gateaux differentiable convex function. The *Bregman distance* with respect to f is the function $D_f : B \times B \to \mathbb{R}$ defined by:

$$D_f(x, y) := f(x) - f(y) - \langle \nabla f(y), x - y \rangle \tag{2.3}$$

where $\nabla f(\cdot)$ is the differential of f. As f is convex,

$$D_f(x,y) \geq 0,$$

for all $x, y \in B$. If f is strictlly convex, $D_f(x,y) = 0$ only for $x = y$. Even assuming strict convexity of f, $D_f(\cdot,\cdot)$ is not a distance in the usual sense of the term (in general, it is not symmetric and does not satisfy the triangular inequality). A useful poperty of the Bregman distance, called "the three point lemma"[14] is stated bellow.

Lemma 2.1 (three point lemma) *Given x, z and $y \in B$,*

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

In the sequel, we shall use the following implication of Lemma 2.1, which seems to deserve the name of "four point corollary".

Corollary 2.2 (four point corollary) *For any x, x', y, $y' \in B$,*

$$D_f(x,y') + D_f(x',y) = D_f(x,y) + D_f(x',y') + \langle \nabla f(y) - \nabla f(y'), x - x' \rangle.$$

According [11], the *modulus of convexity* of f is $\nu_f : B \times \mathbb{R}_+ \to \mathbb{R}_+$

$$\nu_f(z,t) := \inf\{D_f(x,z) |\ ||x - z|| = t\},$$

The function f is *totally convex* in B if $\nu_f(z,t) > 0$ for any $z \in B$ and $t > 0$. A stronger condition will be needed for the regularization function. The function f is *uniformly totally convex* [9] or sequential consistency (see [11]) if for any bounded set $K \subset B$, and $t > 0$,

$$\inf_{x \in K} \nu_f(x,t) > 0.$$

Observe that uniform total convexity implies total convexity, and this latter implies strict convexity. The result below, which will be useful in the sequel, has been proved in [11] (Lemma 2.1.2).

Lemma 2.3 *Let $f : B \to \mathbb{R}$ be a Gateaux differentiable convex function. The following conditions are equivalents:*

(i) *f is uniformly totally convex.*

(ii) *If $\{x^k\}$, $\{y^k\}$ are sequences in B such that $\lim_{k\to\infty} D_f(y^k, x^k) = 0$ and $\{x^k\}$ is bounded, then $\lim_{k\to\infty} ||x^k - y^k|| = 0$.*

Consider the following set of assumptions on f:

A1 The right level sets of $D_f(y, \cdot)$:

$$R(y, \alpha) := \{z \in B : D_f(y, z) \leq \alpha\}$$

are bounded for all $\alpha \geq 0$ and for all $y \in B$.

A2 $\nabla f : B \to B^*$ is onto (zone coerciveness).

A3 ∇f is uniformly continuous in bounded sets.

A4 f is uniformly totally convex.

Regarding the generality/restrictivity of the assumptions above, when $B = L^p$ or $B = \ell_p$ with $p > 1$, the family of functions $f(x) = ||x||_p^s$, $s > 1$, satisfies the conditions above (see [19]). If B is a uniformly smooth and uniformly convex Banach space, then, for all $s > 1$, $f(x) = \|x\|^s$ satisfies the conditions above (see [20]).

3 The inexact proximal point method in Banach spaces

From now on B is a reflexive Banach space, $T : B \rightrightarrows B^*$ is a maximal monotone operator and $f : B \to \mathbb{R}$ is a Gateaux differentiable convex function, satisfying assumptions **A1** to **A4**. Under this assuptions, each one of the proximal problems (1.2) have a unique solution. Convergence properties of the sequence generated by these exact solutions have been analyzed in [9].

Our main concern is a robust modification of the algorithm defined by (1.2), which shall allow for some error. So, error analysis of the proximal problems is the subject of this section.

Take $p \in B$, $c > 0$, and consider the generic proximal subproblem

$$\text{find} \quad x \in B, \quad \text{such that} \quad 0 \in cT(x) + \nabla f(x) - \nabla f(p) \tag{3.1}$$

This problem is equivalent to the "proximal system":

$$\text{find} \quad (x,v) \in B \times B^*, \quad \text{such that} \\ \begin{cases} v \in T(x) \quad \text{and} \\ cv + \nabla f(x) - \nabla f(p) = 0. \end{cases} \tag{3.2}$$

The inclusion on (3.2) will be relaxed with the use of T^ε. The equality will also be relaxed and deviation for equality will be measured by D_f. So, if

$$\begin{array}{l} v \in T^\varepsilon(x), \\ cv + \nabla f(x) - \nabla f(p) = r, \end{array} \tag{3.3}$$

we will use ε and $D_f(x, \nabla f^{-1}(\nabla f(x) - r))$ in the error mesure. More specifically, for x, v, ε as in (3.3), the error mesure we will use is

$$\begin{aligned} R_{p,c}(\varepsilon, x, v) &:= D_f(x, \nabla f^{-1}(\nabla f(x) - r)) + c\varepsilon, \\ &= D_f(x, \nabla f^{-1}(\nabla f(p) - cv)) + c\varepsilon. \end{aligned} \tag{3.4}$$

From the time being, $R_{p,c}$ make sense only in the graph of T^ε, which is defined as

$$G(T^\varepsilon) := \{(\varepsilon, x, v) \mid v \in T^\varepsilon(x)\}.$$

The critical reader will observe a small problem with our notation. In the above definition, ε has different meanings in each side of the equality. So, notation (2.2) is more consistent:

$$\begin{aligned} G(\mathbf{B}_T) &= \{((\varepsilon, x), v) \mid v \in \mathbf{B}_T(\varepsilon, x)\} \\ &\equiv \{(\varepsilon, x, v) \mid v \in \mathbf{B}_T(\varepsilon, x)\}. \end{aligned}$$

In the graph of $\mathbf{B}_T$, $R_{p,c}$ is a error measure for the proximal system.

Proposition 3.1 *If $v \in T^\varepsilon(x)$ then*

$$R_{p,c}(\varepsilon, x, v) \geq 0,$$

whit equality only at $\varepsilon = 0$ and (x,v) the exact solution of the proximal system (3.2).

Proof. Assume firs that $\varepsilon = 0$ and that (x,v) is the exact solution of (3.2) (note that $v \in T(x) = T^{\varepsilon=0}(x)$). Then $\nabla f(x) = \nabla f(p) - cv$ and we conlude that $R_{p,c}(\varepsilon, x, v) = 0$.

Conversely, if $R_{p,c}(\varepsilon,x,v) = 0$ then, since D_f and ε are nonegative we conclude that

$$\varepsilon = 0, \quad D_f(x, \nabla f^{-1}(\nabla f(p) - cv) = 0.$$

Hence $v \in T^{\varepsilon=0}(x) = T(x)$ and, since f is strictly convex, $x = \nabla f^{-1}(\nabla f(p) - cv)$ and $\nabla f(x) = \nabla f(p) - cv$. So, $\varepsilon = 0$ and (x,v) is the exact solution of (3.2). ∎

Lemma 3.2 *Let $\hat{x}$ be the (unique) solution of the proximal problem* (3.1)*. For any $v \in T^{\varepsilon}(x)$,*

$$D_f(x,\hat{x}) + D_f(\hat{x},z) \le R_{p,c}(\varepsilon,x,v),$$

where $z = \nabla f^{-1}(\nabla f(p) - cv)$.

Proof. Let $w \in T(\hat{x})$ such that $cw + \nabla f(\hat{x}) - \nabla f(p) = 0$. Using Lemma 2.1 with $z_1 = x$, $y_1 = z$ e $x_1 = \hat{x}$, we get

$$\begin{aligned} D_f(x,\hat{x}) + D_f(\hat{x},z) &= D_f(x,z) - \langle \nabla f(\hat{x}) - \nabla f(z), x - \hat{x}\rangle \\ &= D_f(x,z) - \langle \nabla f(\hat{x}) - \nabla f(p) + cv, x - \hat{x}\rangle \\ &= D_f(x,z) - c\langle v - w, x - \hat{x}\rangle \\ &\le D_f(x,z) + c\varepsilon = R_{p,c}(\varepsilon,x,v) \end{aligned}$$

where the inequality follows from fact that $w \in T(\hat{x})$ and $v \in T^{\varepsilon}(x)$. ∎

Algorithmic application of the error measure $R_{p,c}$ deppends on the following property of this function

Lemma 3.3 *Take $v \in T^{\varepsilon}(x)$ and define*

$$z = \nabla f^{-1}(\nabla f(p) - cv).$$

Then, for any $p^* \in T^{-1}(0)$,

$$D_f(p^*,p) \ge D_f(p^*,z) + D_f(x,p) - R_{p,c}(\varepsilon,x,v).$$

Proof. Using Corollary 2.2 with $x_1 = p^*$, $x_1' = x$, $y_1 = z$ and $y_1' = p$, we get

$$\begin{aligned} D_f(p*,p) &= D_f(p*,z) + D_f(x,p) - D_f(x,z) + c\langle 0 - v, p^* - x\rangle \\ &\ge D_f(p*,z) + D_f(x,p) - D_f(x,z) - c\varepsilon \\ &= D_f(p*,z) + D_f(x,p) - R_{p,c}(\varepsilon,x,v) \end{aligned}$$

where the inequality follows from fact that $p^* \in T^{-1}(0)$ and $v \in T^{\varepsilon}(x)$. ■

Now we can state formally our algorithm.

Algorithm 1 *Choose any $p^0 \in B$.*
For $k = 0, 1, \ldots$

a *Choose some $c_k > 0$ and find (x^k, v^k) an approximated solution of the proximal system*

$$\begin{cases} v \in T(x) \quad \textit{and} \\ c_k v + \nabla f(x) - \nabla f(p^k) = 0. \end{cases} \tag{3.5}$$

b *Set*

$$p^{k+1} = \nabla f^{-1}(\nabla f(p^k) - c_k v^k) \tag{3.6}$$

Our erro tolerance citerion is

$$\begin{array}{l} v^k \in T^{\varepsilon_k}(x^k), \\ R_{p^k, c_k}(\varepsilon_k, x^k, v^k) \le \sigma D_f(x^k, p^k) + \alpha_k, \end{array} \tag{3.7}$$

with σ, α_k nonegative numbers staisfying

$$0 \le \sigma < 1, \quad \sum_{k=0}^{\infty} \alpha_k < +\infty. \tag{3.8}$$

4 Convergence analysis

In this section, $\{c_k\}$, $\{p^k\}$, $\{(x^k, v^k)\}$, $\{\varepsilon_k\}$ are as described in Algorithm 1, (3.7), (3.8).

The main result we shall prove are as follows. If problem (1.1) has solutions, then the sequence $\{p^k\}$ is bounded and all its weak cluster points are solutions to (1.1). Under additional assumptions on f, $\{p^k\}$ converges weakly. If problem (1.1) has not solution, the sequence $\{x^k\}$ is unbounded.

Proposition 4.1 *If $p^* \in T^{-1}(0)$, then for any k*

$$D_f(p^*, p^k) \ge D_f(p^*, p^{k+1}) + (1 - \sigma) D_f(x^k, p^k) - \alpha_k.$$

Moreover, $\{D_f(p^*, p^k)\}$ converges.

Proof. Apply Lemma 3.3, (3.6) and (3.7) to obtain the first inequality. ∎

Proposition 4.2 *If $T^{-1}(0) \neq \emptyset$ then*

(i) *The sequence $\{p^k\}$ is bounded;*

(ii) $\sum_{k=0}^{\infty} D_f(x^k, p^k) < \infty$;

(iii) $\sum_{k=0}^{\infty} c_k \varepsilon_k < \infty$;

(iv) $\sum_{k=0}^{\infty} D_f(x^k, p^{k+1}) < \infty$;

(v) *the sequence $\{x^k\}$ is bounded;*

(vi) $\lim_{k\to\infty} \|x^k - p^k\| = \lim_{k\to\infty} \|x^k - p^{k+1}\| = 0$.

Proof. Let $p^* \in T^{-1}(0)$. We already know that (Lemma 3.3) $D_f(p^*, p^k)$ is decreasing. Using also assumption **A1** we conlude that (i) holds. To proceed with the proof take an integer n and add the inequality stated on Lemma 3.3 for $k = 0$ to $k = n$. After cancealing the terms that apear on bothe sides of the resulting inquality we get

$$D_f(p^*, p^0) \geq D_f(p^*, p^{p+1}) + (1 - \sigma) \sum_{k=0}^{n} D_f(x^k, p^k).$$

As n is arbitrary, D_f is nonegative and $1 - \sigma$ is strictly positive, we conclude that (ii) also holds.

Items (iii) and (iv) follows from (ii) and (3.4). As D_f is nonegative, from items (ii), (iv) we have

$$\begin{aligned} \lim_{k\to\infty} D_f(x^k, p^k) &= 0, \\ \lim_{k\to\infty} D_f(x^k, p^{k+1}) &= 0. \end{aligned}$$

Using also item (i), assumption **A4** and Lemma 2.3 we conclude that item (vi) also holds. ∎

Lemma 4.3 *Suppose that $T^{-1} \neq \emptyset$. If*

$$\limsup_{j\to\infty} c_{k_j} > 0, \quad p^{k_j} \xrightarrow{w} \overline{p},$$

then $\overline{p} \in T^{-1}(0)$.

Proof. Using item (vi) of Proposition 4.2 we have

$$\lim_{j\to\infty} \|p^{k_j} - p^{k_j+1}\| = 0.$$

As $\{p^{k_j}\}$ is bounded, ∇f is uniformly continuous on bounded sets (assumption **A3**) and $\limsup_{j\to\infty} c_{k_j} > 0$, using also (3.3) we get

$$\lim_{j\to\infty} \|v^{k_j}\| = 0.$$

Note that by item (iii) of Proposition 4.2, $\{\varepsilon_{k_j}\}$ converges to 0.Using again item (vi) of Proposition 4.2 we have $x^{k_j} \xrightarrow{w} \overline{p}$.

As $\{v^{k_j}\}$ converges strongly to $0 \in B^*$, $\{\varepsilon_{k_j}\}$ converges to 0, $\{x^{k_j}\}$ converges weakly to $\overline{p}$ and $v^{k_j} \in T^{\varepsilon_{k_j}}(x^{k_j})$, using also [8] we conclude that $\overline{p} \in T^{-1}(0)$. ■

Theorem 4.4 *If $T^{-1}(0) \neq \emptyset$ and $\{c_k\}$ is bounded away form zero, then the sequences $\{p^k\}$, $\{x^k\}$ are bounded, have the same weak cluster points, and these points are solutions of* (1.1).

Proof. By Proposition 4.2, item (i), $\{p^k\}$ is bounded. Using item (vi) of the same proposition, we conclude that $\{x^k\}$ is also bounded and that $\{x^k\}$, $\{p^k\}$ have the same weak cluster points. So it is enought to show that all weak cluster points of $\{x^k\}$ are solutions of (1.1).

Using again item (vi) of Proposition 4.2 we have

$$\lim_{k\to\infty} \|p^k - p^{k+1}\| = 0.$$

As $\{p^k\}$ is bounded, ∇f is uniformly continuous on bounded sets (assumption **A3**) and $\{c_k\}$ is bounded away from zero, using also (3.3) we get

$$\lim_{k\to\infty} \|v^k\| = 0.$$

Note that by item (iii) of Proposition 4.2, $\{\varepsilon_k\}$ converges to 0. Let z be a weak cluster point of $\{x^k\}$ (and $\{p^k\}$). Then there exist a subsequence $\{x^{k_j}\}$ converging weakly to z. As $\{v^{k_j}\}$ converges strongly to $0 \in B^*$, $\{\varepsilon_{k_j}\}$ converges to 0 and $v^{k_j} \in T^{\varepsilon_{k_j}}(x^{k_j})$, using also [8] we conclude that $0 \in T(z)$. ■

To obtain weak convergence of the whole sequences $\{p^k\}$ and $\{x^k\}$ to a solution, addtional assumptions shall be made on f. Recall that $G : B \to B^*$ is *weak-to-weak* continuous* if

$$p^k \overset{w}{\to} p \Rightarrow \nabla G(p^k) \overset{w}{\to} \nabla G(p).$$

Theorem 4.5 *If $T^{-1}(0) \neq \emptyset$, $\{c_k\}$ is bounded away form zero, and ∇f is weak-to-weak* continuous, then $\{p^k\}$ and $\{x^k\}$ converges weakly to a solution of* (1.1).

Proof. By item (vi) of Proposition 4.2, it is enough to prove that $\{p^k\}$ converges weakly. By Theorem 4.4, $\{p^k\}$ is bounded and its weak cluster belongs to $T^{-1}(0)$. Let z, $\bar{z}$ be weak cluster points of $\{p^k\}$. Then there exist two subsequences $\{p^{k_j}\}$ and $\{p^{k_r}\}$ converging weakly to z and $\bar{z}$ respectively, that is:

$$p^{k_j} \overset{w}{\to} z,$$
$$p^{k_r} \overset{w}{\to} \bar{z}.$$

Direc calculation yelds

$$D_f(z, p^k) - D_f(\bar{z}, p^k) = f(z) - f(\bar{z}) + \langle \nabla f(p^k), \bar{z} - z \rangle.$$

Since $0 \in T(z)$, $0 \in T(\bar{z})$, by Lemma 3.3 we have that the sequences $\{D_f(z, p^k)\}$ and $\{D_f(\bar{z}, p^k)\}$ are convergent. Therefore, using the above equality we conclude that $\{\langle \nabla f(p^k), \bar{z} - z \rangle\}$ converges to some $\beta \in \mathbb{R}$. In particular

$$\lim_{j\to\infty} \langle \nabla f(p^{k_j}), \bar{z} - z \rangle = \lim_{r\to\infty} \langle \nabla f(p^{k_r}), \bar{z} - z \rangle = \beta.$$

using the weak-to-weak* continuity of ∇f, we get that

$$\langle \nabla f(z), \bar{z} - z \rangle = \langle \nabla f(\bar{z}), \bar{z} - z \rangle = \beta.$$

Therefore,

$$\langle \nabla f(z) - \nabla f(\bar{z}), \bar{z} - z \rangle = 0,$$

wich implies $z = \bar{z}$. As $\{p^k\}$ is a bounded sequence with only one weak cluster point, it converges weakly. ■

Remark 4.6 (i) *The assumption that* ∇f *is a weak-to-weak** *continuous mapping is fulfilled when* $B = \ell_p$ *and* $f(x) = \frac{1}{p}||x||_p^p$.

The result below establishes that boundeness of the sequence $\{x^k\}$ is a sufficient and necessary condition for existence of solutions of the problem (1.1).

Theorem 4.7 *If* $T^{-1}(0) = \emptyset$, *and* $\{c_k\}$ *is bounded away from zero, then* $\{x^k\}$ *is unbounded.*

Proof. Suppose that $T^{-1}(0) = \emptyset$ and $\{x^k\}$ is bounded. Take some $R > 0$ such that $\{x^k\}$ is contained in the interior of a ball with center in 0 and radius R. Define

$$S := \{x \in B : \ \|x\| \leq 2R\},$$

and let N_S be the normal cone operator tor the set S:

$$N_S = \partial I_S,$$

where $I_S : B \to (-\infty, +\infty]$ is the indicator function for S: it is zero on S and $+\infty$ elsewhere. Then N_S is maximal monotone and $dom(T) \cap int(dom(N_S)) \neq \emptyset$. Therefore, the operator

$$T_1 := T + N_S$$

is maximal monotone [26]. Moreover, T_1 has a bounded domain and, hence, it has a zero [27]. So the problem

$$\text{find} \quad x \in B \quad \text{such that} \quad 0 \in T_1(x), \tag{4.1}$$

has a solution.

Now we claim that for every index k,

$$v^k \in T_1^{\varepsilon_k}(x^k) \tag{4.2}$$

Take any $x \in B$ and $w \in T_1(x)$. Then there exist u, $v \in B^*$ such that

$$w = u + v,$$

$$u \in T(x) \qquad \text{and} \qquad v \in N_S(x).$$

Note that $x^k \in S$ for all k. Therefore, by definition of the normal operator, it holds that

$$\langle v, x - x^k \rangle \geq 0.$$

Taking into account that $v^k \in T^{\varepsilon_k}(x^k)$, we further obtain

$$\langle w - v^k, x - x^k \rangle = \langle u - v^k, x - x^k \rangle + \langle v, x - x^k \rangle \geq -\varepsilon_k,$$

which proves (4.2). Therefore $\{p^k\}$, $\{(x^k, v^k)\}$ and $\{\varepsilon_k\}$ can be considered as sequences generated by Algorithm 1 applied to problem (4.1) which has solutions. As $\{x^k\}$ is bounded (and B is reflexive), there exist some $\bar{x}$ weak cluster point of $\{x^k\}$. Since $\|\cdot\|$ is weakly lower semicontinuous, $\|\bar{x}\| \leq R$. According to the Theorem 4.4, $\overline{x}$, is a solution of (4.1), that is

$$0 \in T_1(\overline{x}) = T(\overline{x}) + N_S(\overline{x}),$$

Since $||\overline{x}|| \leq R$, $\overline{x}$ is in the interior of the set S. Therefore $N_S(\overline{x}) = \{0\}$ and $0 \in T(\overline{x})$. Then $T^{-1}(0) \neq \emptyset$, in contradiction with our hypothesis. It follows that $\{x^k\}$ is unbounded. ■

5 A convex Variational Principle for the Proximal Subproblems

Recall that the generic poximal problem (3.1)

$$\text{find} \quad x \in B, \quad \text{such that} \quad 0 \in cT(x) + \nabla f(x) - \nabla f(p)$$

is equivalent to the generic proximal system (3.2)

$$\begin{array}{l} \text{find} \quad (x, v) \in B \times B^*, \quad \text{such that} \\ \begin{cases} v \in T(x) \quad \text{and} \\ cv + \nabla f(x) - \nabla f(p) = 0. \end{cases} \end{array}$$

In each iteration of Algorithm 1, (for $p = p^k$, $c = c^k$) one shall find an approximated solution of the problem (3.2): That is, some $\varepsilon \geq 0$, $x \in B$, $v \in B^*$ such that $v \in T^\varepsilon(x)$ and

$$R_{p,c}(\varepsilon, x, v) \leq \sigma D_f(x, p) + \alpha,$$

with $\alpha = \alpha_k$. So, the proximal iteration is implicitly or explicitly related to the following minimization problem

$$\left\{ \begin{array}{ll} \min & R_{p,c}(\varepsilon, x, v) \\ \text{s.t.} & \varepsilon \geq 0, x \in B, v \in B^*, \\ & v \in T^\varepsilon(x). \end{array} \right. \tag{5.1}$$

Note that the approximation criteria depends on ε, x, v, while the iteration rule (3.6) depends only on x, v. Define $\lambda_T : B \times B^* \to \mathbb{R} \cup \{+\infty\}$,

$$\lambda_T(x, v) := \inf\{\varepsilon \geq 0 \mid v \in T^\varepsilon(x)\}. \tag{5.2}$$

We are addopting the convention that the inf of an empty set is $+\infty$. Since T^ε is increasing in ε and is demiclosed, $\lambda_T(\cdot)$ is lower semicontinuous and

$$\forall \varepsilon \in \mathbb{R}, \quad v \in T^\varepsilon(x) \Leftrightarrow \lambda_T(x, v) \leq \varepsilon.$$

In particular, if $\lambda_T(x, v) < \infty$ then $v \in T^{\lambda_T(x,v)}(x)$. Deffine $S_{p,c} : B \times B^* \to \mathbb{R} \cup \{+\infty\}$,

$$S_{p,c}(x, v) := D_f(x, \nabla f^{-1}(\nabla f(p) - cv)) + c\lambda_T(x, v). \tag{5.3}$$

we have from the above equations,

$$S_{p,c}(x, v) = \min\{R_{p,c}(\varepsilon, x, v) \mid v \in T^\varepsilon(x)\}$$

Now, problem (5.1) may be written as

$$\left\{ \begin{array}{ll} \min & S_{p,c}(x, v), \\ \text{s.t.} & (x, v) \in B \times B^*. \end{array} \right. \tag{5.4}$$

and condition (3.7) becomes

$$S_{p^k,c^k}(x^k, v^k) \leq \sigma D_f(x^k, p^k) + \alpha_k.$$

We claim that $S_{p,c}$ is a proper closed convex merit function for the proximal system. Yhe merit function $S_{p,c}$, for the special case where B is a Hilbert space and $f(x) = 1/2\|x\|^2$ was studied in [31]. From Proposition 3.1 and Lemma 3.2 we have:

Proposition 5.1 *For any $(x, v) \in B \times B^*$, $S_{p,c}(x, v) \geq 0$, with equality only at the solution of the proximal system* (3.2).

Proposition 5.2 *Let $\hat{x}$ be the exact solution of the proximal problem* (3.1)*. Then for any $(x, v) \in B \times B^*$,*

$$D_f(x, \hat{x}) + D_f(\hat{x}, z) \leq S_{p,c}(x, v),$$

where $z = \nabla f^{-1}(\nabla f(p) - cv)$.

Now we will prove convexity and lower semicontinuity of $S_{p,c}$. Recall that the *conjugate* of f is the function $f^* : B^* \to \mathbb{R} \cup \{\pm\infty\}$,

$$f^*(v) := \sup_{x \in B} \langle v, x \rangle - f(x).$$

Trivially, f^* is convex and lowersemicontinuous on B^*. As f is a proper closed convex function, f^* is also proper, that is, $f^* > -\infty$ everywhere. Define now $\widetilde{D}_f : B \times B^* \to \mathbb{R} \cup \{+\infty\}$,

$$\widetilde{D}_f(x, v) = f(x) + f^*(v) - \langle x, v \rangle. \tag{5.5}$$

Note that $\widetilde{D}_f(x, v)$ is convex in x, convex in v, but in general is not jointlly convex. Trivially, $\widetilde{D}_f$ is lower semicontinuous in $B \times B^*$. Since f is also differentiable, using the gradient inequality we have

$$f^*(\nabla f(y)) = \langle y, \nabla f(y) \rangle - f(y),$$

for any $y \in B$. Therefore, using also (5.5) we conlude that

$$D_f(x, y) = \widetilde{D}_f(x, \nabla f(y))$$

for any $x, y \in B$. As ∇f is onto an one to one, the converse also holds, that is, for any $x \in B$, $v \in B^*$,

$$\widetilde{D}_f(x, v) = D_f(x, (\nabla f)^{-1}(v)).$$

Now we get another expression for $S_{p,c}$

$$S_{p,c}(x, v) := \widetilde{D}_f(x, \nabla f(p) - cv) + c\lambda_T(x, v). \tag{5.6}$$

Since the graph of T^ε is demiclosed, λ_T is lower semicontinuous. We already observed that $\widetilde{D}_f$ is lower semicontinuous. Using the above expression for $S_{p,c}$ we conclude that this function is lower semicontinuous. Now only the convexity is missing.

Theorem 5.3 *The function $S_{p,c}$ is convex.*

Proof. From (5.5) and (5.6) we get

$$S_{p,c}(x,v) := f(x) + f^*(\nabla f(p) - cv) - \langle x, \nabla f(p)\rangle + c\Big[\langle v, x\rangle + \lambda_T(x,v)\Big].$$

Define $g, h : B \times B^* \to \mathbb{R} \cup \{\infty\}$,

$$\begin{aligned} g(x,v) &:= f(x) + f^*(\nabla f(p) - cv) - \langle x, \nabla f(p)\rangle, \\ h(x,v) &:= \langle v, x\rangle + \lambda_T(x,v). \end{aligned}$$

The function g is trivially convex. According to [10], h is also. For the sake of completeness, we will prove convexity of h. This is equivalent to convexity of the epigraph of h, $P(h)$. Take $(x^1, v^1, t_1), (x^2, v^2, t_2) \in P(h)$, $p, q \geq 0$, $p + q = 1$. Define

$$(\bar{x}, \bar{v}, \bar{t}) := p(x^1 + q, v^1, t_1) + q(x^2, v^2, t_2).$$

Define also

$$\varepsilon_1 := t_1 - \langle x^1, v^1\rangle, \quad \varepsilon_2 := t_2 - \langle x^2, v^2\rangle.$$

Then, $\lambda_T(x^1, v^1) \leq \varepsilon_1$, $\lambda_T(x^2, v^2) \leq \varepsilon_2$. Hence $v^1 \in T^{\varepsilon_1}(x^1)$, $v^2 \in T^{\varepsilon_2}(x^2)$. Using the tranportation formula for T^ε [8] we get

$$pv^1 + qv^2 \in T^{\bar{\varepsilon}}(px^1 + qx^2),$$

whith

$$\bar{\varepsilon} = p\varepsilon_1 + q\varepsilon_2 + pq\langle v^1 - v^2, x^1 - x^2\rangle.$$

So, $\lambda_T(\bar{x}, \bar{v}) \leq \bar{\varepsilon}$, and $h(\bar{x}, \bar{v}) \leq \langle \bar{v}, \bar{x}\rangle + \bar{\varepsilon}$. Direct calculation yelds

$$\bar{t} = \langle \bar{v}, \bar{x}\rangle + \bar{\varepsilon},$$

and we conclude that $(\bar{x}, \bar{v}, \bar{t}) \in P(h)$. As g and h are convex, $S_{p,c}$ is convex. ∎

References

[1] Allgower, E. L. and Böhmer, K. Application of the Mesh Independence Principle to mesh refinement strategies. *SIAM Journal of Numerical Analysis* **24** (1987) 1335-1351.

[2] Allgower, E. L., Böhmer, K., Potra, F. A. and Rheimboldt, W. C. A Mesh Independence Principle for operator equations and their discretizations. *SIAM Journal of Numerical Analysis* **23** (1986) 160-169.

[3] Auslender, A. and Haddou, M. An interior proximal method for convex linearly constrained problems and its extension to varational inequalities. *Mathematical Programming* **71** (1995) 77-100.

[4] Burachik, R. S., Butnariu, D. and Iusem, A. N. Iterative methods for solving stochastic convex feasibility problems and applications. Submitted

[5] Burachik, R. S. and Iusem, A. N. Generalized proximal point methods for the variational inequality problem in a Hilbert space. *SIAM Journal on Control* **8** (1998) 197-216.

[6] Burachik, R. S., Iusem, A. N. and Svaiter, B. F. Enlargement of maximal monotone operators with application to variational inequalities. *Set-Valued Analysis* **5** (1997)159-180.

[7] Burachik, R. S., Sagastizábal, C. A. and Svaiter, B. F. ε-enlargement of maximal monotone operators with application to variational inequalities. in: M. Fukushima and L. Qi (eds),*Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, Kluwer Acad. Publ., Dordrecht, 1997.

[8] Burachik, R.S., Svaiter, B.F. Epsilon-Enlargements of maximal monotone operators in Banach Spaces *Set-Valued Analysis* **7** (1999) 117-132.

[9] Burachik, R.S., Scheimberg, S. A proximal point algorithm for the variational inequality problem in Banach spaces. Technical Report ES-470/98, COPPE-UFRJ, Rio de Janeiro (1998).

[10] R.S. Burachik, B. F. Svaiter. *Maximal monotone operators, convex functions and a special family of enlargements.* Submitted to Set-Valued Analysis.

[11] Butnariu, D., Iusem, A.N. *Totally convex functions for fixed points computation and infinite dimensional optimization.* Kluwer, Dordrecht, The Netherlands (2000).

[12] Butnariu, D. and Iusem, A. N. On a proximal point methods for convex optimization in Banach space. *Numerical Functional Analysis and Optimization* **18** (1997) 723-744.

[13] Censor, Y. and Zenios, S. A. The proximal minimization algorithm with D-functions. *Journal of Optimization Theory and Applications* **73** (1992) 451-464.

[14] Chen, G. and Teboulle, M. Convergence analysis of a proximal-like optimization algorithm using Bregman functions. *SIAM Journal of Optimization* **3** (1993) 538-543.

[15] Eckstein, J. Nonlinear proximal point algorithm using Bregman functions with applications to convex programming. *Mathematics of Operations Research* **18** (1993) 202-226.

[16] Eckstein, J. Approximate iterations in Bregman-functions-based proximal algorithms. *Mathematical Programming* **55** (1998) 113-123.

[17] Heinkenschloss M. Mesh Independence for Nonlinear least squares problems with norm constraints. *SIAM Journal of Optimization* **3** (1993) 81-117.

[18] Iusem, A. N. On some properties of generalized proximal point methods for quadratic and linear programming. *Journal of Optimization Theory and Applications* **85** (1995) 593-612.

[19] Iusem, A.N., Isnard, C.A. On mixed Hölder-Minkowski inequalities and total convexity of certain functions in $\mathcal{L}^p(\Omega)$ (to be published in *Mathematical Inequalities and Applications*).

[20] Iusem, A.N., Otero, R.G. Inexact versions of proximal point and augmented Lagrangian algorithms in Banach spaces (to be published).

[21] Kelley, C. T. and Sachs, E. W. A Mesh Independence of Newton-like methods for infinite dimensional problems. *Journal of Integral Equations and Applications* **3** (1991) 549-573.

[22] Kiwiel, K. The proximal minimization methods with generalized Bregman functions. *SIAM Journal on Contrl and Optimization* **35** (1997) 1142-1168.

[23] Laumen, M. Newton's Mesh Independence Principle for a class of optimization shape design problems. *SIAM Journal on Control and Optmization* **37** (1999) 1070-1088.

[24] Pascali, D., Sburlan, S. *Nonlinear Mappings of Monotone Type* . Editura Academiei, Bucarest (1978).

[25] Rockafellar, R.T. Local boundedness of nonlinear monotone operators. *Michigan Mathematical Journal* **16**(1969) 4.

[26] Rockafellar, R.T. On the maximality of sums of nonlinear monotone operators. *Transactions of The American Mathematical Society* **149**(1970) 75-88.

[27] Rockafellar, R.T. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization* **14** (1976) 877-898.

[28] Solodov, M.V., Svaiter, B.F., A hybrid projection-proximal point algorithm, *Journal of Convex Analysis*, **6** (1999), 59–70.

[29] Solodov, M.V., Svaiter, B.F. A hybrid aproximate extragradient- proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Analysis* **7**(1999) 323-345.

[30] Solodov, M.V., Svaiter, B.F. An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions. *Mathematics of Operations Research* **25** (2000) 214-230.

[31] Solodov, M.V., Svaiter, B.F. Error bounds for proximalpoint subproblems and inexact proximal algorithms. *Mathematical Programming* **88** (2000) 371-389.