A proximal point method in nonreflexive Banach spaces

Alfredo N. Iusem* and Elena Resmerita†

September 22, 2009

Abstract

We propose an inexact version of the proximal point method and study its properties in nonreflexive Banach spaces which are duals of separable Banach spaces, both for the problem of minimizing convex functions and of finding zeroes of maximal monotone operators. By using surjectivity results for enlargements of maximal monotone operators, we prove existence of the iterates in both cases. Then we recover most of the convergence properties known to hold in reflexive and smooth Banach spaces for the convex optimization problem. When dealing with zeroes of monotone operators, our convergence result requests that the regularization parameters go to zero, as is the case for standard (non-proximal) regularization schemes.

Key words: Proximal point method, ε-subdifferential, ε-duality mapping, inexact Bregman distance.

AMS Subject classification: 90C48; 49J53.

1 Introduction

In this work we propose a proximal point method for certain problems defined in nonreflexive Banach spaces that are topological duals of separable Banach spaces. We consider separately the optimization case, namely the minimization of convex functions, and the operator case, namely the problem of finding zeroes of point-to-set maximal monotone operators.

An impressive amount of work has been devoted to proximal point methods for optimization and, more generally, for approximating zeros of maximal monotone operators, starting with the seminal works of Martinet [13] and Moreau [15], which were enlightened in the comprehensive study by Rockafellar [18], all of these and numerous others being set in Hilbert spaces.

Extensions of these methods to reflexive Banach spaces were carried out by Kassay [11], Butnariu and Iusem [5], Burachik and Scheimberg [4] and Iusem and Gárciga Otero [10], to name just a few of them. However, the complexity of the problems arising nowadays from practical situations require formulating them in atypical settings, such as in nonreflexive Banach spaces.

To our knowledge, the few proximal point methods investigated so far in nonreflexive Banach spaces address minimization of particular objective functions (e.g., quadratic) - see [16], [9], with the aim of denoising or deblurring images. Here, we combine the idea of [16] with a surjectivity result, shown in [7] and [12], in order to obtain a proximal point method for minimizing more general convex functions, with interesting convergence properties.

Thus, in the optimization case where the objective function is not necessarily quadratic, the regularizing term is a positive multiple of an inexact Bregman distance associated with the square of the norm; a solution is approached by a sequence of approximate minimizers of an auxiliary problem. In the operator case, the auxiliary problem consists of finding a zero of a regularized operator, namely the sum of the original one and a positive multiple of the ε-subdifferential of the square of the norm.

*Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, R.J., CEP 22460-320, Brazil (iusp@impa.br)

†Industrial Mathematics Institute, Johannes Kepler University, Altenbergerstraße 69, A-4040 Linz, Austria (elena.resmerita@jku.at)

Regarding the condition on the nonreflexive Banach space of being the dual of a separable Banach space, we mention that this is satisfied by large classes of spaces, including the cases of ℓ_∞ and $\mathcal{L}^\infty(\Omega)$, ℓ_1 and $BV(\Omega)$ (the space of functions of bounded variation) which appear quite frequently in a large range of applications - see the book of Meyer [14] in this respect.

In Section 2, we overview notions and results on which this work is based. Section 3 describes the proximal point method for convex optimization, while Section 4 deals with the similar method designed for finding zeroes of maximal monotone operators, both in the nonreflexive setting.

2 Preliminaries

Let X be a (possibly nonreflexive) Banach space, X^* its topological dual and $g : X \to \mathbb{R} \cup \{+\infty\}$ a proper, convex function. Denote the domain of g by $dom\, g = \{x \in X : g(x) < \infty\}$. For $\varepsilon \geq 0$, the ε-subdifferential of g at a point $x \in X$ is

$$\partial_\varepsilon g(x) = \{x^* \in X^* : g(y) - g(x) - \langle x^*, y - x\rangle \geq -\varepsilon, \quad \forall y \in dom\, g\}. \tag{1}$$

Given a point-to-set operator $U : X \to \mathcal{P}(X^*)$, $G(U)$ denotes the graph of U, i.e. $G(U) = \{(x, \xi) \in X \times X^* : \xi \in U(x)\}$. Consider $h : X \to \mathbb{R}$ defined as

$$h(x) = \frac{1}{2}\|x\|^2. \tag{2}$$

Given $\varepsilon > 0$, the so called normalized ε-duality operator $J_\varepsilon : X \to \mathcal{P}(X^*)$ is defined as the ε-subdifferential of h, i.e., $J_\varepsilon(x) = \partial_\varepsilon h(x)$. An equivalent definition for J_ε is

$$J_\varepsilon(x) = \left\{x^* \in X^* : \langle x^*, x\rangle + \varepsilon \geq \frac{1}{2}\|x^*\|^2 + \frac{1}{2}\|x\|^2\right\}. \tag{3}$$

We recall several properties of the ε-subdifferentials.

Proposition 2.1. *Let g and h be two proper, convex functions on X and consider $x \in dom\, h \cap dom\, g$, $z^* \in X^*$, $\varepsilon > 0$, $\lambda > 0$. Then,*

i)

$$\partial_\varepsilon h(x) \neq \emptyset, \quad \forall \varepsilon > 0 \tag{4}$$

if and only if h is lower semicontinuous at x.

ii)

$$\cup_{\eta\in[0,\varepsilon]} \left[\partial_\eta g(x) + \partial_{\varepsilon-\eta} h(x)\right] \subset \partial_\varepsilon(g + h)(x). \tag{5}$$

iii)

$$\partial_\varepsilon(h + z^*)(x) = \partial_\varepsilon h(x) + z^*. \tag{6}$$

iv)

$$\partial_\varepsilon(\lambda h)(x) = \lambda \partial_{\frac{\varepsilon}{\lambda}} h(x). \tag{7}$$

Proof: See Theorem 2.4.2 and Theorem 2.4.4 in [21]. □

We need to introduce inexact Bregman distances with respect to the convex function h defined by (2) and to an ε-subgradient η of h. Given $\varepsilon \geq 0$, define $D^\varepsilon : X \times G(J_\varepsilon) \to [0, +\infty)$ as

$$D^\varepsilon(x, (y, \eta)) = h(x) - h(y) - \langle \eta, x - y\rangle + \varepsilon. \tag{8}$$

Clearly, we have $D^\varepsilon(x, (y, \eta)) \geq 0$ for all $(x, (y, \eta)) \in X \times G(J_\varepsilon)$, in view of (1). Also, when h is Fréchet differentiable, in which case $\eta = h'(y)$, we have $D^0(x(y, \eta)) = D(x, y)$, where D denotes the standard Bregman distance related to h (see e.g., [2], [6]). Note that, while $D(x, x) = 0$ for all $x \in X$, this is not the case for D^ε. In fact, if $\varepsilon > 0$, we have $D^\varepsilon(x, x) = \varepsilon > 0$ for all $x \in X$.

Given $\varepsilon \geq 0$ and a function $g : X \to \mathbb{R} \cup \{+\infty\}$, we say that $\bar{x} \in dom\, g$ is an ε-minimizer of g when

$$g(\bar{x}) \leq g(x) + \varepsilon \tag{9}$$

for all $x \in dom\, g$.

Our convergence theorem for the operator case requires the following well known property, called demi-closedness of the graph of maximal monotone operators.

Proposition 2.2. *Assume that X is the dual of a Banach space and let $T : X \to \mathcal{P}(X^*)$ be a maximal monotone operator. Consider a sequence $\{(y_k, \eta_k)\} \subset G(T)$. If $\{\eta_k\}$ converges to some point $\bar{\eta} \in X^*$ in the norm topology, and $\{y_k\}$ converges to some point $\bar{y} \in X$ in the weak* topology, then $(\bar{y}, \bar{\eta})$ belongs to $G(T)$.*

Proof: See, e.g., Proposition 4.2.1(i) in [3]. □

We shall also use in our analysis a surjectivity result for sums of maximal monotone operators and generalized duality operators. In order to introduce it, we need some background material. Let $T : X \to \mathcal{P}(X^*)$ be a maximal monotone operator, and consider X as a subset of the bidual space X^{**} through the canonical inclusion. T can be seen as an operator defined in X^{**}, and its monotonicity is obviously preserved, but not its maximality, in general. A standard argument using Zorn's Lemma establishes that there exist maximal monotone extensions of T as an operator defined on X^{**}. Operators for which such an extension is unique play a significant role in the theory. This happens when T is maximal monotone and X is reflexive, because in this case $X = X^{**}$. A class of operators for which the extension is unique was introduced in [7], and denoted as class (D). Its definition demands the introduction of the *monotone closure* $\bar{T}$ of T defined as

$$\bar{T} = \{(x^{**}, x^*) \in X^{**} \times X^* : \langle x^{**} - y, x^* - y^* \rangle \geq 0 \quad \forall (y, y^*) \in T\}.$$

Again, when X is reflexive, maximality of T implies that $T = \bar{T}$. Class (D) consists of those operators $T : X \to \mathcal{P}(X^*)$ such that T is "dense" in $\bar{T}$, with the following precise meaning: every point (x^{**}, x^*) in $\bar{T}$ is the limit of a bounded net $\{(x_i, x_i^*)\}_{i \in I}$ contained in T, where boundedness refers to the norm topology in $X \times X^*$ and the limit is taken with respect to the weak$^*\times$ strong topology in $X^{**} \times X^*$.

The motivation behind the definition of class (D) lies in a result in [17], stating that ∂g belongs to class (D) for all convex and lower semicontinuous $g : X \to \mathbb{R} \cup \{+\infty\}$.

It has been proved in [7] that when T belongs to class (D) the operator $\bar{T}$ is monotone, and hence it is the unique maximal monotone extension of T as an operator defined on X^{**}. The fact that for an operator T of class (D) the operator $T + \lambda J_\varepsilon$ is onto for all $\lambda > 0$ and all $\varepsilon > 0$ has also been established in [7].

Another class of operators, denoted as (NI), was introduced in [19]. An operator $T : X \to \mathcal{P}(X^*)$ belongs to class (NI) if

$$\inf_{(y,y^*) \in T} \langle y - x^{**}, y^* - x^* \rangle \leq 0 \quad \forall (x^{**}, x^*) \in X^{**} \times X^*.$$

An easy argument shows that class (D) is contained in class (NI). The uniqueness of the maximal monotone extension to X^{**}, as well as the above mentioned surjectivity result, have been extended to the class (NI) in [12]. We will need this fact, which we state next, in our analysis.

Proposition 2.3. *If T is a maximal monotone operator of class (NI), then the operator $T + \lambda J_\varepsilon$ is onto, for all $\lambda > 0$ and all $\varepsilon > 0$.*

Proof: See [12]. □

3 The proximal point method for convex optimization

Our aim is to approximate minimizers of a convex function $f : X \to \mathbb{R} \cup \{\infty\}$. We propose a proximal point method which employs inexact Bregman distances as regularizing terms.

From now on, we consider $h : X \to \mathbb{R}$ defined as $h(x) = \frac{1}{2}\|x\|^2$. Consider exogenous sequences $\{\varepsilon_k\}$, $\{\lambda_k\}$ of positive numbers satisfying the following two assumptions:

H1) The sequence $\{\varepsilon_k\}$ is summable, i.e., $\sum_{k=0}^{\infty} \varepsilon_k < \infty$,

H2) The sequence $\{\lambda_k\}$ is bounded above.

The number ε_k is some sort of error bound for the inexact minimization performed at the k-th iteration of the algorithm, while λ_k is the regularization parameter used in the same iteration.

Our algorithm, called Algorithm A in the sequel, is defined as follows:

Initialization

Take $(x_0, \xi_0) \in G(J_{\varepsilon_0})$.

Iterative step

Let $k \in \mathbb{N}$. Define $D^{\varepsilon_k}(x, (x_k, \xi_k)) = h(x) - h(x_k) - \langle \xi_k, x - x_k\rangle + \varepsilon_k$ and $\bar{\varepsilon}_k = \lambda_k \varepsilon_{k+1}$.

Determine $x_{k+1} \in dom\, f$ as an $\bar{\varepsilon}_k$-minimizer of the function $f_k(x)$ defined as

$$f_k(x) = f(x) + \lambda_k D^{\varepsilon_k}(x, (x_k, \xi_k)), \tag{10}$$

that is to say, in view of (9),

$$f(x_{k+1}) + \lambda_k D_h^{\varepsilon_k}(x_{k+1}, (x_k, \xi_k)) \le f(x) + \lambda_k D_h^{\varepsilon_k}(x, (x_k, \xi_k)) + \bar{\varepsilon}_k \tag{11}$$

for all $x \in dom\, f$.

Let $\xi_{k+1} \in J_{\varepsilon_{k+1}}(x_{k+1})$ and $\eta_{k+1} \in \partial f(x_{k+1})$ such that

$$\eta_{k+1} - \lambda_k(\xi_{k+1} - \xi_k) = 0. \tag{12}$$

We prove below that the sequence $\{x^k\}$ generated by Algorithm A is well defined.

Proposition 3.1. *Let X be a Banach space and $f : X \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex and lower semicontinuous function. Then for each $k \in \mathbb{N}$, there exists an $\bar{\varepsilon}_k$-minimizer x_{k+1} of the function f_k defined by* (10) *and there exist subgradients $\xi_{k+1} \in J_{\varepsilon_{k+1}}(x_{k+1})$ and $\eta_{k+1} \in \partial f(x_{k+1})$ such that* (10) *holds.*

Proof: We proceed by induction on k. The definition of the initial iterate x_0, as given in the initialization step, offers no difficulty. For the inductive step, we assume that the pair $(x_k, \xi_k) \in G(J_{\varepsilon_k})$ is given and show that an appropriate x_{k+1}, i.e. an $\bar{\varepsilon}_k$-minimizer of f_k, indeed exists. Observe first that, in view of (8), convexity of h implies convexity of $D^{\varepsilon_0}(\cdot, (x_0, \xi_0))$, and then (10) and convexity of f imply convexity of f_0. Note that the first order optimality condition for the problem of approximate minimization of f_0, sufficient by convexity, is

$$0 \in \partial_{\bar{\varepsilon}_k} f_k(x). \tag{13}$$

Thus, it suffices to show that the inclusion given by (13) has solutions. Recall that the subdifferential of a lower semicontinuous convex function is a maximal monotone operator, and, as we have already commented upon, it belongs to class (D) (both facts were proved in [17]), and "a fortiori", to class (NI). Therefore, Proposition 2.3 can be applied and yields that the operator $\partial f + \lambda_k J_{\varepsilon_{k+1}}$ is surjective. Hence there exists $x_{k+1} \in X$ such that

$$\lambda_k \xi_k \in [\partial f + \lambda_k J_{\varepsilon_{k+1}}](x_{k+1}), \tag{14}$$

i.e., there exist $\xi_{k+1} \in J_{\varepsilon_{k+1}}(x_{k+1})$ and $\eta_k + 1 \in \partial f(x_k + 1)$ which satisfy $\eta_{k+1} - \lambda_k(\xi_{k+1} - \xi_k) = 0$. We claim that such an x_{k+1} is an $\bar{\varepsilon}_k$-minimizer of f_k because, due to (5), (7) and (6), we get from (14)

$$0 \in \partial f(x_{k+1}) + \lambda_k \left[J_{\varepsilon_{k+1}}(x_{k+1}) - \xi_k\right] = \partial f(x_{k+1}) + \lambda_k [\partial_{\varepsilon_{k+1}} h(x_{k+1}) - \xi_k] =$$

$$\partial f(x_{k+1}) + \lambda_k \left[\partial_{\varepsilon_{k+1}} D^{\varepsilon_k}(\cdot,(x_k,\xi_k))(x_{k+1})\right] = \partial f(x_{k+1}) + \partial_{\bar{\varepsilon}_k}\left[\lambda_k D^{\varepsilon_k}(\cdot,(x_k,\xi_k))\right](x_{k+1}) \subset$$
$$\partial_{\bar{\varepsilon}_k}(f + \lambda_k D^{\varepsilon_k}(\cdot,(x_k,\xi_k))(x_{k+1}) = \partial_{\bar{\varepsilon}_k} f_k(x_{k+1}). \quad (15)$$

Hence, x_{k+1} satisfies (13), the claim is established and the inductive step is complete. □

Proximal point methods based on some distances (metric and Bregman, for instance) in reflexive Banach spaces enjoy the following features:

- the objective function decreases along the iterations,
- the distance from the iterates to any solution decreases along the iterations.

Due to the lack of reflexivity of the space, and the inexact nature of the Bregman distances, which seems to be unavoidable in this context, our method has weaker monotonicity properties. We establish next some of them.

Proposition 3.2. *Let X be a Banach space and $f : X \to \mathbb{R} \cup \{+\infty\}$ a proper, convex and lower semicontinuous function. Assume that z is a minimizer of f. Define $\beta_k = D^{\varepsilon_k}(z,(x_k,\xi_k))$, $\gamma_k = D^{\varepsilon_k}(x_{k+1},(x_k,\xi_k))$, with D^{ε_k} as in (8). If H1 and H2 hold, then the sequence $\{x_k\}$ generated by Algorithm A has the following properties:*

i)

$$f(x_{k+1}) \le f(x_k) + \lambda_k(\varepsilon_k + \varepsilon_{k+1}), \quad (16)$$

ii)

$$\beta_{k+1} - \beta_k + \gamma_k + \frac{f(x_{k+1}) - f(z)}{\lambda_k} \le \varepsilon_{k+1}, \quad (17)$$

iii) The sequence $\{\beta_k\}$ is bounded,

iv) The sequence $\{\gamma_k\}$ is summable,

v) The sequence $\{f(x_k) - f(z)\}$ is summable.

Proof: i) The definition of x_{k+1}, namely (11), yields

$$f(x_{k+1}) + \lambda_k D^{\varepsilon_k}(x_{k+1},(x_k,\xi_k)) \le f(x_k) + \lambda_k D^{\varepsilon_k}(x_k,(x_k,\xi_k)) + \bar{\varepsilon}_k. \quad (18)$$

Since $D^{\varepsilon_k}(x_k,(x_k,\xi_k)) = \varepsilon_k$ and $D^{\varepsilon_k}(x_{k+1},(x_k,\xi_k)) \ge 0$, (16) follows from (18), in view of the definition of $\bar{\varepsilon}_k$.

ii) From the definition of the inexact Bregman distance, i.e. (8), we obtain

$$\beta_k - \beta_{k+1} - \gamma_k = D^{\varepsilon_k}(z,(x_k,\xi_k)) - D^{\varepsilon_{k+1}}(z,(x_{k+1},\xi_{k+1})) - D^{\varepsilon_k}(x_{k+1},(x_k,\xi_k))$$
$$= \langle \xi_k - \xi_{k+1}, x_{k+1} - z\rangle - \varepsilon_{k+1}. \quad (19)$$

Rearranging (19), we get

$$\beta_{k+1} - \beta_k + \gamma_k - \langle \xi_k - \xi_{k+1}, z - x_{k+1}\rangle = \varepsilon_{k+1} \quad (20)$$

Note that equality (12) implies

$$\lambda_k(\xi_k - \xi_{k+1}) \in \partial f(x_{k+1}). \quad (21)$$

In view of the definition of subgradient,

$$\langle \lambda_k(\xi_k - \xi_{k+1}), z - x_{k+1}\rangle \le f(z) - f(x_{k+1}). \quad (22)$$

Dividing both sides of (22) by λ_k and replacing the result in (20) yields (17), establishing (ii).

[iii)-v)] Summing (17) with k between 0 and ℓ we get

$$\beta_{\ell+1} + \sum_{k=0}^{\ell} \gamma_k + \sum_{k=0}^{\ell} \frac{f(x_{k+1}) - f(z)}{\lambda_k} \leq \beta_0 + \sum_{k=0}^{\ell} \varepsilon_{k+1} \leq \beta_0 + \sum_{k=0}^{\infty} \varepsilon_k, \tag{23}$$

Let $\hat{\lambda}$ be an upper bound of the sequence $\{\lambda_k\}$, which exists by virtue of assumption H2. Define $\theta = \sum_{k=0}^{\infty} \varepsilon_k$. Note that θ is well defined by assumption H1. Since $f(x_{k+1}) - f(z) \geq 0$ because z is a minimizer of f, we obtain from (23)

$$\beta_{\ell+1} + \sum_{k=0}^{\ell} \gamma_k + \frac{1}{\hat{\lambda}} \sum_{k=0}^{\ell} [f(x_{k+1}) - f(z)] \leq \beta_0 + \theta, \tag{24}$$

for all $\ell \in \mathbb{N}$, and then items (iii), (iv) and (v) follow immediately by taking limits in (24) with $\ell \to \infty$. ▯

We prove next that $\{x_k\}$ is a minimizing sequence for f, and that all its weak* cluster points are minimizers.

Theorem 3.3. *Let X be a Banach space which is the dual of a separable Banach space and $f : X \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex and weakly* lower semicontinuous function. Assume that f has minimizers. If H1 and H2 hold, then the sequence $\{x_k\}$ generated by Algorithm A is bounded, $\lim_{k\to\infty} f(x_k) = \min_{x\in X} f(x)$, and all cluster points of $\{x_k\}$ in the weak* topology of X are minimizers of f.*

Proof. Note that weak* lower semicontinuity of f yields also its lower semicontinuity, because the norm topology of X is finer than the weak* topology of X (see, e.g., [1, Proposition 2.3, p. 34]). Let z be a minimizer of the function f. Thus, Proposition 3.2 holds. It follows from Proposition 3.2(v) that

$$\lim_{k\to\infty} f(x_{k+1}) = f(z), \tag{25}$$

establishing the second statement of the theorem.

By Proposition 3.2(iii), $\{\beta_k\}$ is bounded, and hence there exists $\sigma > 0$ such that

$$\sigma \geq \beta_k = D^{\varepsilon_k}(z, (x_k, \xi_k)) = \frac{\|z\|^2}{2} - \frac{\|x_k\|^2}{2} - \langle \xi_k, z - x_k \rangle + \varepsilon_k \geq \frac{\|z\|^2}{2} - \langle \xi_k, z \rangle + \frac{\|\xi_k\|^2}{2}, \tag{26}$$

using (3) in the last inequality. Thus,

$$\frac{\|\xi_k\|^2}{2} \leq \langle \xi_k, z \rangle + \sigma \leq \|\xi_k\| \, \|z\| + \sigma, \tag{27}$$

which shows that the sequence $\{\xi_k\}$ is bounded. Using now (27) and (3),

$$\frac{\|x_k\|^2}{2} \leq \langle \xi_k, x_k \rangle + \varepsilon_k \leq \|\xi_k\| \, \|x_k\| + \theta, \tag{28}$$

with $\theta = \sum_{k=0}^{\infty} \varepsilon_k < \infty$, in view of H1. It follows from (28) that the sequence $\{x_k\}$ is also bounded. Consider now any subsequence $\{x_{j_k}\}$ of $\{x_k\}$ which converges in the weak* topology (such a sequence does exist, because bounded sequences in X are weakly* sequentially relatively compact, by Banach-Alaoglu's Theorem), and let $\bar{x}$ be the weak* limit of $\{x_{j_k}\}$. In view of (25) and the lower semicontinuity of f with respect to the weak* topology, we get

$$f(\bar{x}) \leq \liminf_{k\to\infty} f(x_{j_k}) = f(z) < \infty,$$

showing that $\bar{x}$ is a minimizer of f and completing the proof. ▯

4 The proximal point method for finding zeroes of monotone operators

As it was made clear in Rockafellar's papers written in the seventies (e.g. [18]), the proximal point method can be naturally extended from the convex optimization context to the problem of finding zeroes of maximal monotone operators; it suffices to substitute a maximal monotone operator for the subdifferential of the objective function. Of course, the same extension is possible for our method in nonreflexive Banach spaces, but the absence of a real valued function which almost decreases along the iterations (cf. Proposition 3.2(i)), entails some complications for the convergence analysis, and the results become weaker. We are forced to demand that the sequence of regularization parameters $\{\lambda_k\}$ converge to 0. This is somewhat undesirable, because when λ_k is close to 0 the effect of the regularization term vanishes, so that if the original problem is ill-conditioned the same will occur with the k-th proximal subproblem for large k. Then one has to stop earlier the algorithm in order to obtain approximations which are as accurate and stable as possible. Such a proximal point method for ill-posed problems is dealt with in [8].

The fact that the generated sequence converges to a solution even with large values of the regularization parameter is one of the main advantages of the proximal method when compared with conventional regularization procedures (see e.g. [20]), but then nonreflexive Banach spaces are rather hard to deal with, and a convergence proof for the case of vanishing regularization parameters ensures that at least the proximal point for finding zeroes of maximal monotone operators behaves no worse than standard regularization schemes, in Tikhonov's sense.

In this section, X is, as before, a possibly nonreflexive Banach space, X^* its dual and $T : X \to \mathcal{P}(X^*)$ a maximal monotone operator of class (NI). The problem of interest consists of finding a zero of T, i.e. a point $\bar{x} \in X$ such that $0 \in T(\bar{x})$.

The proximal point method for this problem requires exogenous sequences of positive real numbers $\{\varepsilon_k\}$, $\{\lambda_k\}$ satisfying H1 of Section 3 and additionally:

H3) $\lim_{k\to\infty} \lambda_k = 0$.

The algorithm, to be denoted as Algorithm B in the sequel, is defined as follows.

Initialization

Take $(x_0, \xi_0) \in G(J_{\varepsilon_0})$.

Iterative step

For $k \in \mathbb{N}$, given (x_k, ξ_k), the next iterate is a pair $(x_{k+1}, \xi_{k+1}) \in G(J_{\varepsilon_{k+1}})$ such that

$$\lambda_k \xi_k \in T(x^{k+1}) + \lambda_k \xi_{k+1}. \tag{29}$$

The fact that the method is well defined is established with the same argument as in the optimization case.

Proposition 4.1. *If T is of class (NI) then the sequence $\{(x_k, \xi_k)\}$ generated by Algorithm B is well defined.*

Proof: Note that $T + \lambda_k J\varepsilon_{k+1}$ is surjective by Proposition 2.3. □

Now we establish the monotonicity properties of the algorithm.

Proposition 4.2. *Let $\{(x_k, \xi_k)\}$ be the sequence generated by algorithm B, consider a zero z of T, and define β_k, γ_k as in Section 3, i.e., $\beta_k = D^{\varepsilon_k}(z, (x_k, \xi_k))$, $\gamma_k = D^{\varepsilon_k}(x_{k+1}, (x_k, \xi_k))$, with D^{ε_k} as in* (8)*. Then*

i) The sequence $\{\beta_k\}$ is bounded,

ii) The sequence $\{\gamma_k\}$ is summable.

Proof: As in the proof of Proposition 3.2, we have

$$\beta_k - \beta_{k+1} - \gamma_k = D^{\varepsilon_k}(z,(x_k,\xi_k)) - D^{\varepsilon_{k+1}}(z,(x_{k+1},\xi_{k+1})) - D^{\varepsilon_k}(x_{k+1},(x_k,\xi_k))$$
$$= \langle \xi_k - \xi_{k+1}, x_{k+1} - z\rangle - \varepsilon_{k+1}. \tag{30}$$

Since $\lambda_k(\xi_k - \xi_{k+1}) \in T(x_{k+1})$ by (29) and $0 \in T(z)$, we get from the monotonicity of T that $\langle \xi_k - \xi_{k+1}, x_{k+1} - z\rangle \geq 0$, and hence we obtain from (30)

$$\beta_k - \beta_{k+1} - \gamma_k \geq -\varepsilon_{k+1}. \tag{31}$$

It follows from (31) that

$$\gamma_k + \beta_{k+1} - \beta_k \leq \varepsilon_{k+1} \tag{32}$$

for all $k \in \mathbb{N}$. Summing (32) with k between 0 and ℓ, we get

$$\beta_{\ell+1} + \sum_{k=0}^{\ell} \gamma_k \leq \beta_0 + \sum_{k=0}^{\ell} \varepsilon_{k+1} \leq \beta_0 + \sum_{k=0}^{\infty} \varepsilon_k. \tag{33}$$

In view of H1, the rightmost expression in (33) is finite, implying boundedness of $\{\beta_k\}$ and summability of $\{\gamma_k\}$. □

We present next our convergence result for Algorithm B.

Theorem 4.3. *Let X be a Banach space which is the dual of a separable Banach space and $T : X \to \mathcal{P}(X^*)$ a maximal monotone operator of class (NI). Assume that T has zeroes. If H1 and H3 hold, then the sequence $\{(x_k,\xi_k)\}$ generated by Algorithm B is bounded, and all cluster points of $\{x_k\}$ in the weak* topology are zeroes of T.*

Proof: In view of Proposition 4.2(i), there exists $\sigma > 0$ such that $\sigma \geq \beta_k = D^{\varepsilon_k}(z,(x_k,\xi_k))$ for all k. As in the proof of Theorem 3.3, we get

$$\sigma \geq \frac{\|z\|^2}{2} - \langle \xi_k, z\rangle + \frac{\|\xi_k\|^2}{2},$$

using (3), and obtaining

$$\frac{\|\xi_k\|^2}{2} \leq \langle \xi_k, z\rangle + \sigma \leq \|\xi_k\|\,\|z\| + \sigma, \tag{34}$$

which implies that the sequence $\{\xi_k\}$ is bounded. Using now (34) and (3), we get

$$\frac{\|x_k\|^2}{2} \leq \langle \xi_k, x_k\rangle + \varepsilon_k \leq \|\xi_k\|\,\|x_k\| + \theta \tag{35}$$

where $\theta = \sum_{k=0}^{\infty} \varepsilon_k$, which is finite by H1. It follows from (35) that the sequence $\{x_k\}$ is also bounded, establishing the first statement of the theorem.

Since $\{\xi_k\}$ is bounded and $\lim_{k\to\infty} \lambda_k = 0$ by H3, we have

$$\lim_{k\to\infty} \lambda_k(\xi_k - \xi_{k+1}) = 0, \tag{36}$$

in the norm topology. Rewrite now (29) as

$$\lambda_k(\xi_k - \xi_{k+1}) \in T(x_{k+1}). \tag{37}$$

Since X is the dual of a separable Banach space, boundedness of $\{x_k\}$ implies weak* compactness of $\{x_k\}$, and hence it has weak* cluster points. Let $\bar{x}$ be one of them. Taking limits in (37) along a subsequence converging to $\bar{x}$, we invoke Proposition 2.2 and conclude from (36) that $0 \in T(\bar{x})$, i.e. $\bar{x}$ is a zero of T, completing the proof. □

We remark that in order to prove our convergence results assuming just H2, i.e. boundedness of the sequence $\{\lambda_k\}$, instead of H3, thus obtaining a method which converges with regularization coefficients which stay away from 0, it would be enough to prove that

$$\lim_{k\to\infty} (\xi_k - \xi_{k+1}) = 0. \tag{38}$$

One could expect to obtain this fact from the summability of the sequence γ_k, which in fact was not used in the analysis. This is what happens in the reflexive and smooth case, e.g. when X is uniformly convex and uniformly smooth, in which case h is Fréchet differentiable, $J := J_0$ is point-to-point, and $\xi_k = h'(x_k)$. In such a situation, $\lim_{k\to\infty} \gamma_k = 0$ implies that the Bregman distance between consecutive iterates goes to 0, which in turn implies that $\lim_{k\to\infty}[h'(x_{k+1}) - h'(x_k)] = 0$ (see a proof in [10] for the case of $\varepsilon_k = 0$ for all k). Unfortunately, nonreflexive Banach spaces fail to enjoy the smoothness properties required for this argument to work: in such a setting J is always point-to-set, and it seems hard to establish that $\{\xi_{k+1} - \xi_k\}$ goes to 0 in the norm topology just from the fact that x_{k+1} and x_k get arbitrarily close to each other, which follows, in some sense, from the summability of the γ_k's. On the other hand, we have no counterexample for (38), whose validity thus remains as an open problem.

As mentioned in the Introduction, there is a wide range of applications which fit into the theoretical setting of a nonreflexive Banach space that is the dual of a separable Banach space. Indeed, we intend to consider in the future the solution of several variational problems arising in image analysis based on the proximal method analyzed in this paper.

5 Acknowledgments

The authors thank Benar F. Svaiter (IMPA, Brasil) for pointing out to the authors the surjectivity results in nonreflexive Banach spaces and for useful discussions. Also, they are grateful to the referee for the suggestions which helped improving an earlier version of the manuscript. E.R. acknowledges support from the Austrian Science Foundation, Elise Richter scholarship (V82-N18 FWF).

References

[1] Barbu, V., Precupanu, T.: Convexity and Optimization in Banach Spaces. Reidel, Dordrecht (1986)

[2] Bregman, L.M.: The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. USSR Comp. Math. Math. Phys. 7, 200-217 (1967)

[3] Burachik, R.S., Iusem, A.N.: Set-valued Mappings and Enlargements of Monotone Operators. Springer, New York (2007)

[4] Burachik, R.S., Scheimberg, S.: A proximal point algorithm for the variational inequality problem in Banach spaces. SIAM J. Control Optim. 39, 1615-1632 (2001)

[5] Butnariu D., Iusem, A.N.: On a proximal point method for convex optimization in Banach spaces. Numer. Funct. Anal. Optim. 18, 723-744 (1997)

[6] Censor, Y., Iusem, A.N., Zenios, S.A.: An interior point method with Bregman functions for the variational inequality problem with paramonotone operators. Math. Programming 81, 373-400 (1998)

[7] Gossez, J.P.: Opérateurs monotones nonlineaires dans les espaces de Banach nonreflexifs. J. Math. Anal. Appl. 34, 371-395 (1971)

[8] Hanke, M., Groetsch, C.W.: Nonstationary iterated Tikhonov regularization. J. Optim. Theory Appl. 98, 37-53 (1998)

[9] He, L., Burger, M., Osher, S.: Iterative total variation regularization with non-quadratic fidelity. J. Math. Imaging Vision 26, 167-184 (2006)

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

[11] Kassay, G.: The proximal points algorithm for reflexive Banach spaces. Studia. Univ. Babes-Bolyai, Mathematica 30, 9-17 (1985)

[12] Marques Alves, M., Svaiter, B.: On the surjectivity properties of perturbations of maximal monotone operators in non-reflexive Banach spaces. Submitted (2008)

[13] Martinet, B.: Régularisation d'inéquations variationnelles par approximations successives. Rev. Francaise Inform. Rech. Oper. 3, 154-179 (1970)

[14] Meyer, Y.: Oscillatory Patterns in Image Processing and Nonlinear Evolution Equations. University Lecture Series, vol. 22, American Mathematical Society, Providence (2001)

[15] Moreau, J.: Proximité et dualité dans un espace hilbertien. Bull. Soc. Math. France 93, 273-299 (1965)

[16] Osher, S., Burger, M., Goldfarb, D., Xu, J., Yin, W.: An iterative regularization method for total variation based image restoration. Multiscale Modelling and Simulation 4, 460-489 (2005)

[17] Rockafellar, R.T.: On the maximal monotonicity of subdifferential mappings. Pacific J. Math. 33, 209-216 (1970)

[18] Rockafellar, R.T.: Monotone operators and the proximal point algorithm in convex programming. SIAM J. Control Optim. 14, 877-898 (1976)

[19] Simons, S.: The range of a monotone operator. J. Math. Anal. Appl. 199, 176-201 (1996)

[20] Tikhonov, A., Arsenin, V.: Solution of Ill-posed Problems. Wiley, New York (1977)

[21] Zalinescu, C.: Convex Analysis in General Vector Spaces. World Scientific, New Jersey (2002)