A PROXIMAL POINT METHOD FOR NONSMOOTH CONVEX OPTIMIZATION PROBLEMS IN BANACH SPACES

Y. I. ALBER[*], R. S. BURACHIK[**] AND A. N. IUSEM[**]

ABSTRACT. In this paper we show the weak convergence and stability of the proximal point method when applied to the constrained convex optimization problem in uniformly convex and uniformly smooth Banach spaces. In addition, we establish a nonasymptotic estimate of convergence rate of the sequence of functional values for the unconstrained case. This estimate depends on a geometric characteristic of the dual Banach space, namely its modulus of convexity. We apply a new technique which includes Banach space geometry, estimates of duality mappings, nonstandard Lyapunov functionals and generalized projection operators in Banach spaces.

1. INTRODUCTION

The proximal method, or more exactly, "the proximal point algorithm", is one of the most important successive approximation methods for finding fixed points of nonexpansive mappings in Hilbert spaces. This method, which is therefore not new, (see [12], [17], [18]), was earlier used for regularizing linear equations ([12], [13]), and seems to have been applied the first time to convex minimization by Martinet (see [15], [16]). The first important results (like approximate versions, linear and finite convergence) in the more general framework of maximal monotone operators in a Hilbert space are due

to Rockafellar [20]. Nowadays it is still the object of intensive investigation (see [14] for a survey on the method).

The proximal method can be seen as a regularization scheme in which the regularization parameter needs not approach zero, thus avoiding a possible ill behavior of the regularized problems.

We will state here some main properties of this algorithm and some applications to convex programming and maximal monotone inclusions.

Let H be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and Ω a closed and convex subset of H. Consider the problem

$$(1.1) \qquad\qquad\qquad \min_{x \in \Omega} f(x),$$

where $f : H \to \mathbb{R}$ is a convex functional. It is a familiar fact that in direct methods for solving (1.1), the tools for proving existence of solutions are the convexity properties of the functional. Based on the fact that

$$x^0 \in \operatorname{argmin}_{x \in H} f(x) \text{ iff } 0 \in \partial f(x^0),$$

we can pose the problem

$$(1.2) \qquad\qquad \text{Find } x^0 \in H \text{ such that } 0 \in T(x^0),$$

where $T : H \to \mathcal{P}(H)$ is a maximal monotone operator.

In the particular case in which $T = \partial f$, problem (1.2) is equivalent to (1.1). In this way, the theory of subdifferentials can transform our original problem (1.1) in the study of the range of the monotone operator $\partial f : H \to \mathcal{P}(H)$. Namely, we want to determine if $0 \in R(\partial f)$, where $R(T)$ stands for the range of an operator T.

Replacing ∂f by an arbitrary monotone operator T, we transform an optimization problem into a more general one, involving monotone operators.

The proximal method for (1.1) in a Hilbert space H generates a sequence $\{x^k\} \subset H$ in the following way:

$$(1.3) \qquad x^0 \in H, \ x^{k+1} = \operatorname{argmin}_{x \in H}(f(x) + \lambda_k \|x - x^k\|^2),$$

where

$$(1.4) \qquad\qquad\qquad \lambda_k \in (0, \lambda], \ \lambda > 0.$$

Martinet showed in [15] that if $\{x^k\}$ is bounded, then it converges weakly to a minimizer of f. In [20] Rockafellar studied the convergence properties of this algorithm, when applied to problem (1.2). In this case, the sequence $\{x^k\}$ is defined by:

$$x^0 \in H, \ x^{k+1} \in H \text{ such that } 0 \in Tx^{k+1} + \lambda_k(x^{k+1} - x^k).$$

From now on, unless explicit mention, the parameters λ_k will be taken as in (1.4). As we can easily see, in the case of $T = \partial f$, this procedure reduces to (1.3). It is shown in [20] that $\{x^k\}$ converges weakly to a zero of T, provided that the set of zeroes is nonempty. Rockafellar also proves

linear convergence rate if either one of the following conditions is satisfied:
(a) T is strongly monotone.
(b) T^{-1} is Lipschitz continuous at 0 and $\{x^k\}$ bounded.
Finally, he furnishes a criterion for convergence in a finite number of iterations, which requires the nonemptiness of the interior of the set $T^{-1}(0)$. As a particular case of the latter result, we have finite convergence if H is finite dimensional, $T = \partial f$, and f is polyhedral convex (i.e., the epigraph of f is a polyhedral convex set). In the optimization case, we observe directly from (1.3) that the objective function for each subproblem is coercive (remember that $g : H \to \mathbb{R}$ is coercive if and only if $\lim_{\|x\| \to +\infty} g(x)/\|x\| = +\infty$). In particular, this property implies boundedness of the level sets of g, which ensures existence of solutions of each subproblem. The uniqueness, on the other hand, is ensured by the strict convexity of the objective function given in (1.3). So this algorithm is a true regularization and the sequence $\{x^k\}$ is well defined.

The structure of the iteration in (1.3) suggests the possibility of adding other kind distances to the function f. For instance, a *Bregman distance* D_g, where $g : H \to \mathbb{R}$ is a strictly convex function, satisfying some additional properties (see, e.g. [6]). D_g is defined as:

$$(1.5) \qquad D_g(x, y) := g(x) - g(y) - \langle \nabla g(y), x - y \rangle.$$

The proximal point method with $Dg(x, x^k)$ substituting for $\|x - x^k\|^2$ in (1.3) has been analyzed in [7] for convex optimization in finite dimensional spaces and in [6] for the variational inequality problem in a Hilbert space, i.e., find $x^* \in \Omega \subset H$ such that there exists $y^* \in T(x^*)$ with

$$(1.6) \qquad \langle y^*, x - x^* \rangle \geq 0,$$

for all $x \in \Omega$.

All previous results apply to a Hilbert space. In a Banach space, the iteration (1.3) has as optimality conditions

$$(1.7) \qquad \partial f(x) + \lambda_k J(x - x^k) = 0,$$

where J is the duality mapping, defined in section 3. If the Banach space is not hilbertian, J is not a linear operator, so that (1.7) is not equivalent to

$$(1.8) \qquad \partial f(x) + \lambda_k J(x) = \lambda_k J(x^k).$$

(1.8) is more natural that (1.7) (e.g., with constant λ_k, the left hand side of (1.8) is the same in all iterations).

In this paper we study the proximal point method in Banach spaces, where x^{k+1} is the solution of (1.8). It is easy to check that this is equivalent to

$$x^{k+1} = \mathrm{argmin}_{x \in H}\{f(x) + \frac{\lambda_k}{2}D_g(x, x^k)\},$$

where D_g is as in (1.5) with $g(x) = \|x\|^2$. We emphasize that this is not the iteration given by (1.3), because, if the space is not a Hilbert one, then

$D_g(x, y) \neq \|x - y\|^2$. When B is finite dimensional, but $\| \cdot \|$ is not an euclidean norm, (e.g. $\|x\|_p = (\sum_{j=1}^n |x_j|^p)^{\frac{1}{p}}$, $p \in (1, \infty)$, $p \neq 2$), the method considered here is a particular case of the proximal point method with Bregman distances studied in [7]. We give a full convergence analysis in an arbitrary uniformly convex and uniformly smooth Banach space. We present also stability and convergence rate results.

2. Previous results in a Hilbert space

Let H be a Hilbert space and $f : H \to \mathbb{R} \cup \{\infty\}$ be a proper, closed and convex functional. We recall below known results due to Güler (see [10]) about the convergence properties of the proximal method, applied to the minimization of f. From now on we will write $f^* := f(x^*)$, where x^* is any minimizer of f.

Theorem 2.1. [10] *Consider the sequence defined by (1.3), where λ_k is taken as in (1.4). Define $\sigma_k := \sum_{i=1}^k \lambda_i^{-1}$, by convention, $\sigma_0 = 0$. Suppose that the set of minimizers of f, which we call X^*, is not empty and take $x^* \in X^*$. Then $\{x^k\}$ converges weakly to a minimizer of f. In this conditions, the following global convergence estimate holds:*

$$f(x^k) - f^* \leq \inf_{z \in X^*} \frac{\|x^0 - z\|}{2\sigma_k}.$$

Consequently, if $\sigma_k \to \infty$, then the sequence $f(x^k) \downarrow f^* =: \inf_{z \in H} f(z)$.

The next lemma shows that the proximal point method can be defined in terms of the metric projection operator P_Ω in a Hilbert space, and therefore method (2.2) below shares all well known properties of the proximal point method.

Lemma 2.2. *Let $f : H \to \mathbb{R}$ be a closed and convex function, $\Omega \subset H$ a closed and convex set and P_Ω the orthogonal projection onto Ω. For fixed $a \in H$ and λ positive, consider the following two problems:*

Problem (1): *Find $x \in \Omega$ such that*

$$(2.1) \qquad x := argmin_{z \in \Omega}(f(z) + (\lambda/2)\|z - a\|^2).$$

Problem (2): *Find $x \in \Omega$ such that there exists $y \in H$ satisfying*

$$(2.2) \qquad \begin{cases} 0 \in \partial f(x) + \lambda(y - a), \\ x = P_\Omega(y). \end{cases}$$

The sets of solutions of Problem (1) and Problem (2) coincide.

Proof. Let x be a solution of Problem (2). We claim that there exists $u \in \partial f(x)$ such that

$$(2.3) \qquad \langle u + \lambda(x - a), z - x \rangle \geq 0, \text{ for all } z \in \Omega.$$

Observe that (2.3) are precisely the Karush-Khun-Tucker conditions for Problem (1) (see [4]). Since x is a solution of Problem (2), there exist

$y \in H$ and $u \in \partial f(x)$ such that $x = P_\Omega(y)$ and $0 = u + \lambda(y - a)$. Therefore,

$$\langle u + \lambda(x - a), z - x \rangle = \langle \lambda(x - y), z - x \rangle = \lambda \langle P_\Omega(y) - y, z - P_\Omega(y) \rangle \geq 0 \,,$$

and (2.3) is established, therefore x is a solution of (2.1). Reciprocally, if x solves Problem (1), then there exists $w \in N_\Omega(x)$, the normality operator associated to the set Ω, such that

$$0 = u + w + \lambda(x - a) \quad \text{and} \quad u \in \partial f(x),$$

which implies

(2.4) $$0 = u + \lambda(x + \lambda^{-1}w - a).$$

Take $y = x + \lambda^{-1}w$. Then

(2.5) $$P_\Omega(y) = x \quad \text{and} \quad \partial f(x) = \partial f(P_\Omega(y)).$$

So $u \in \partial f(P_\Omega(y))$. Hence, by (2.4), $0 \in \partial f(x) + \lambda(y - a)$, which, together with (2.5), implies that x solves Problem (2). ∎

The following theorem is a direct consequence of the lemma above and Theorem 2.1.

Theorem 2.3. *Consider the sequence $\{x^k\}$ given by*

1) *Take $x^0 \in \Omega$.*
2) *Given $x^k \in \Omega$, find $y \in H$ and $x^{k+1} \in \Omega$ such that:*

$$\begin{cases} 0 \in \partial f(P_\Omega(y)) + \lambda_k(y - x^k), \\ x^{k+1} = P_\Omega(y). \end{cases}$$

Then if problem (1.1) has solutions, it holds that
(i) the sequence $\{x^k\}$ is well defined and bounded,
(ii) $\limsup_k \lambda_k^{-1}(f(x^{k+1}) - f^*) = 0$,
(iii) if $\lambda_k < \lambda$, then all weak accumulation points of $\{x^k\}$ are solutions,
(iv) $f(x^k) \downarrow f^*$ as $k \to \infty$,
(v) if $\lambda_k > \bar{\lambda} > 0$, then $\lim_{k \to \infty} \|x^{k+1} - x^k\| = 0$,
(vi) if $\lambda_k > \bar{\lambda} > 0$, then there exist a sequence $\{u^k \in \partial f(x^k)\}$ for all k such that $\lim_k \langle u^k, x^k - x^* \rangle = 0$,
(vii) the whole sequence converges weakly to a solution, i.e., there exists a unique weak accumulation point.

All these statements are easy consequences of Lemma 2.2 and the results in [10] and [20]. We will show in Section 4, Theorem 4.1, that all these results are also valid in a Banach space, and hence Theorem 2.3 is a particular case of Theorem 4.1. We present an estimate of convergence rate in Section 4.3.

3. THE BANACH SPACE CONCEPT

Let B be a uniformly convex and uniformly smooth Banach space, (see [1] and [5]). The operator $J : B \to B^*$ is the *normalized duality mapping* associated to B, determined by the equalities

$$\langle Jx, x \rangle = \|Jx\|_{B^*} \|x\|_B = \|x\|^2,$$

where $\langle \cdot, \cdot \rangle$ stands for the usual dual product in B, $\| \cdot \|$ is the norm in the Banach space B and $\| \cdot \|_{B^*}$ is the norm in the dual space B^*. For later use, we state the following lemma, whose proof can be found in [1].

Lemma 3.1. *Let B be as above. Let $\delta_B(\varepsilon) : [0, 2] \to [0, 1]$ denote the modulus of convexity and $\rho_B(\tau) : [0, \infty) \to [0, \infty)$ the modulus of smoothness of the Banach space B. If $x,\, y \in B$ are such that $\|x\| \leq C$ and $\|y\| \leq C$ then*

$$(3.1) \qquad \langle Jx - Jy, x - y \rangle \geq (2L)^{-1} C^2 \delta_B \left(\frac{\|x - y\|}{2C} \right),$$

$$\langle Jx - Jy, x - y \rangle \geq (2L)^{-1} C^2 \delta_{B^*} \left(\frac{\|Jx - Jy\|_{B^*}}{2C} \right),$$

and

$$(3.2) \qquad \|Jx - Jy\|_{B^*} \leq 8C h_B(16L \|x - y\| / C), \quad h_B(\tau) = \rho_B(\tau) / \tau,$$

where L is the constant in Figiel's inequalities (see [8], p. 128).

We recall next the analytical expressions of the duality mapping $J(\cdot)$ in the uniformly smooth and uniformly convex Banach spaces l^p, L^p and Sobolev spaces W_m^p, $p \in (1, \infty)$, (see [1]).

For l^p, $Jx = \|x\|_{l^p}^{2-p} y \in l^q$, where $x = \{x_1, x_2, \dots\}$, $p^{-1} + q^{-1} = 1$ and $y = \{x_1 |x_1|^{p-2}, x_2 |x_2|^{p-2}, \dots\}$.

For L^p, $Jx = \|x\|_{L^p}^{2-p} |x|^{p-2} x \in L^q$, where $p^{-1} + q^{-1} = 1$.

For W_m^p, $Jx = \|x\|_{W_m^p}^{2-p} \sum (-1)^{|t|} D^t (|D^t x|^{p-2} D^t x) \in W_{-m}^q$, where $p^{-1} + q^{-1} = 1$, and D^t is the weak derivative operator of order t.

For a convex and closed set $\Omega \subset B$, we define the *normality operator* $N_\Omega : B \to \mathcal{P}(B^*)$, in the following way:

$$N_\Omega(x) = \begin{cases} \{\phi \in B^* \,|\, \langle \phi, z - x \rangle \leq 0 \text{ for any } z \in \Omega\} & \text{if } x \in \Omega, \\ \emptyset & \text{if } x \notin \Omega. \end{cases}$$

It is easy to check that $N_\Omega(\cdot)$ is the subdifferential of the indicator function $\chi_\Omega(x)$ associated to the set Ω, i.e.

$$\chi_\Omega(x) = \begin{cases} 0 & \text{if } x \in \Omega, \\ +\infty & \text{if } x \notin \Omega. \end{cases}$$

The function χ_Ω is convex and closed, hence $N_\Omega(\cdot)$ is a maximal monotone operator.

We follow in the sequel the theory developed in [1], where a generalized projection operator $\Pi_\Omega(\cdot)$ in a Banach space B is introduced.

Take the Lyapunov functional $W : B \times B \to \mathbb{R}_+$ given by

$$(3.3) \qquad W(x, z) = \|x\|^2 - 2\langle Jx, z \rangle + \|z\|^2.$$

It follows from the definition of J that $\nabla_z W(x, z) = 2(Jz - Jx)$. It also holds that $W(x, z) \geq 0$. In the sequel we will need a property of $W(x, z)$ established in [1], namely

$$(3.4) \qquad (\|x\| - \|z\|)^2 \leq W(x, z) \leq (\|x\| + \|z\|)^2.$$

The generalized projection operator

$$\begin{aligned} \Pi_\Omega : B &\to \Omega \subset B \\ x &\mapsto \Pi_\Omega(x) := argmin_{z \in \Omega} W(x, z), \end{aligned}$$

has also been introduced in [1], where the following lemma is proved.

Lemma 3.2. (i) *The operator $\Pi_\Omega(\cdot)$ is well defined in any uniformly convex and uniformly smooth Banach space.*
(ii) *In the conditions of the definitions above, the following inequality holds for any fixed $x \in B$ and any $z \in \Omega$*

$$\langle Jx - J(\Pi_\Omega(x)), \Pi_\Omega(x) - z \rangle \geq 0.$$

Remark 3.3. *We emphasize that in a Hilbert space $W(x, z) = \|x - z\|^2$, $\Pi_\Omega(\cdot)$ coincides with the classical metric projection $P_\Omega(\cdot)$, and the inequality (ii) in Lemma 3.2 reduces to the Kolmogorov criterion which characterizes the metric projection $P_\Omega(x)$:*

$$\langle x - P_\Omega(x), P_\Omega(x) - z \rangle \geq 0,$$

for all $z \in \Omega$.

Consider now the following two algorithms:
1) Take $x^0 \in \Omega$.

2) Given x^k, define x^{k+1} by the inclusion

$$(3.5) \qquad 0 \in (\partial f + \lambda N_\Omega + \lambda J)x^{k+1} - \lambda Jx^k,$$

with λ positive, and

1) Take $z^0 \in \Omega$.

2) Given z^k, define z^{k+1} by the system

$$(3.6) \qquad \begin{cases} 0 \in \partial f(\Pi_\Omega(y)) + \lambda(Jy - Jz^k), \\ z^{k+1} = \Pi_\Omega(y), \end{cases}$$

with λ positive.

In a similar way as we established the equivalence of the algorithms in Lemma 2.2, we will show next that the sequences defined by (3.5) and (3.6) coincide when they start from the same point. Observe that the solution

of (3.5) exists and is unique by coerciveness and strict monotonicity of J and monotonicity of N_Ω (see [21], Corollary 32.35). We point out also that existence of the iterates in (3.6) is not obvious at all. Nevertheless, as a by-product of the following lemma, we will show the existence of solution of all the subproblems (3.6).

Lemma 3.4. *In the algorithms (3.5) and (3.6), if $z^0 = x^0$, then z^k is well defined and $z^k = x^k$ for any $k \geq 0$.*

Proof. By Lemma 3.2 and the definition of normality operator, for any $y \in B$,

$$(3.7) \qquad Jy - J(\Pi_\Omega(y)) \in N_\Omega(\Pi_\Omega(y)).$$

We proceed by induction. The result holds for $k = 0$ by hypothesis. Suppose that z^k is well defined and $z^k = x^k$; we must show that z^{k+1} exists and $z^{k+1} = x^{k+1}$.

By definition of x^{k+1},

$$(3.8) \qquad 0 \in (\partial f + \lambda N_\Omega + \lambda J)x^{k+1} - \lambda Jx^k.$$

We remark that x^{k+1} is uniquely determined by the previous inclusion because of strict monotonicity of J. By (3.8) there exist $u^{k+1} \in \partial f(x^{k+1})$ and $w^{k+1} \in N_\Omega(x^{k+1})$ such that

$$(3.9) \qquad 0 = u^{k+1} + \lambda(w^{k+1} + Jx^{k+1} - Jx^k).$$

Since J is onto, there exists $y \in B$ such that

$$(3.10) \qquad Jy = w^{k+1} + Jx^{k+1}.$$

We claim that
a) $\Pi_\Omega(y) = x^{k+1}$,

b) $0 \in \partial f(\Pi_\Omega(y)) + \lambda(Jy - Jz^k)$.

Since $x^{k+1} \in \Omega$, for proving (a) it will be enough to show that

$$\langle Jy - Jx^{k+1}, z - x^{k+1} \rangle \leq 0,$$

for any $z \in \Omega$. In fact, by (3.10) and the properties of w^{k+1},

$$\langle Jy - Jx^{k+1}, z - x^{k+1} \rangle = \langle w^{k+1}, z - x^{k+1} \rangle \leq 0,$$

which implies (a).

Now we proceed to prove (b): by (3.9), (3.10) and the induction hypothesis,

$$(3.11) \qquad 0 = u^{k+1} + \lambda(Jy - Jx^k) = u^{k+1} + \lambda(Jy - Jz^k),$$

which implies (b). Now, using (a) and (3.11) we conclude that the system (3.6) has a solution z^{k+1}, and this solution coincides with x^{k+1}. The lemma is complete. ∎

4. Convex functionals in a Banach space

The following results deal with convergence of the sequence given by (3.6) in a uniformly convex and uniformly smooth Banach space B. Under such conditions, we get boundedness of the sequence and optimality of any weak accumulation point. Weak convergence of the whole sequence is established for a special kind of Banach spaces, namely those in which there exists a weak-to-weak continuous duality operator J.

4.1. **Constrained minimization problem: convergence analysis.** Let Ω be a closed and convex subset of B. Consider the problem

$$(4.1) \qquad \min_{x \in \Omega} f(x),$$

where $f : B \to \mathbb{R}$ is a convex functional. Define the sequence $\{x^k\}$ as:
1) Take $x^0 \in \Omega$.

2) Given x^k, define x^{k+1} by the system

$$(4.2) \qquad \begin{cases} 0 \in \partial f(\Pi_\Omega(y)) + \lambda_k(Jy - Jx^k), \\ x^{k+1} = \Pi_\Omega(y), \end{cases}$$

with λ_k positive.

Theorem 4.1. *Consider the sequence $\{x^k\}$ given by (4.2). Suppose that the set of minimizers of f, which we call X^*, is not empty and fix $x^* \in X^*$. Then it holds that*
(i) The sequence $\{x^k\}$ is well defined and bounded,
(ii) $\limsup_k \lambda_k^{-1}(f(x^{k+1}) - f^*) = 0$,
(iii) if $\lambda_k < \lambda$, then all weak accumulation points are solutions,
(iv) if $\lambda_k < \lambda$, then $f(x^k) \downarrow f^*$ as $k \to \infty$,
(v) if $\lambda_k > \bar{\lambda} > 0$, then $\lim_{k \to \infty} \|x^{k+1} - x^k\| = 0$,
(vi) if $\lambda_k > \bar{\lambda} > 0$, then there exist a sequence $\{u^k\}$, with each $u^k \in \partial f(x^k)$, such that

$$\lim_{k \to \infty} \langle u^k, x^k - x^* \rangle = 0,$$

(vii) if B is such that there exists a weak-to-weak continuous duality mapping, then the whole sequence converges weakly to a solution, i.e., there exists a unique weak accumulation point.

Proof. (i) As we mentioned before, in order to prove welldefinedness of the iterates of algorithm (4.2), it is enough to observe that by Lemma 3.4 they coincide with the iterates of algorithm (3.5), which is well defined, as discussed before. In order to prove boundedness, we will show that the sequence $W(x^k, x^*)$ is decreasing, in which case the result will follow from boundedness of the level sets of the function $W(\cdot, x^*)$. Recall that W is given by (3.3) and x^* is any solution of problem (4.1). By the definition of $W(x, \xi)$, (4.2) and the projection properties, we get

$$W(x^{k+1}, x^*) \leq W(x^k, x^*) - W(x^k, x^{k+1})$$

$$-\frac{2}{\lambda_k} \langle u^{k+1}, x^{k+1} - x^* \rangle,$$

where $u^{k+1} \in \partial f(x^{k+1})$. As $W(x^k, x^{k+1}) \geq 0$, the previous equation yields

$$(4.3) \qquad W(x^{k+1}, x^*) \leq W(x^k, x^*) - \frac{2}{\lambda_k} \langle u^{k+1}, x^{k+1} - x^* \rangle.$$

At this point we use the fact that $x^* \in X^*$ and the gradient inequality to obtain

$$(4.4) \qquad W(x^{k+1}, x^*) \leq W(x^k, x^*) - \frac{2}{\lambda_k}(f(x^{k+1}) - f^*) \leq W(x^k, x^*).$$

Hence we proved that $W(x^k, x^*)$ is a decreasing sequence, which is also bounded below, and therefore it converges. Using (3.4), we get

$$(\|x^k\| - \|x^*\|)^2 \leq W(x^k, x^*) \leq W(x^0, x^*) \leq (\|x^0\| + \|x^*\|)^2,$$

which implies

$$\|x^k\| \leq C_0,$$

where C_0 is any real number larger than $\|x^0\| + 2\|x^*\|$. This establishes the first statement.

(ii) This result is a consequence of (4.4) and the fact that

$$\lim_{k \to \infty} (W(x^{k+1}, x^*) - W(x^k, x^*)) = 0.$$

(iii) Take a sequence $\{x^{k_j}\}$ which converges weakly to $\hat{x}$. We will show that $\hat{x}$ is a minimizer of f. By (ii) and the assumption on λ_k we obtain that

$$\lim_{k \to \infty} f(x^{k+1}) = f^*.$$

By weak lower semicontinuity of f,

$$f(\hat{x}) \leq \liminf_{k_j} f(x^{k_j}) = f^*.$$

Therefore $\hat{x}$ is also a minimizer of f.

(*iv*) By (*iii*), it is enough to prove that $f(x^k)$ is decreasing. In order to prove this we use the gradient inequality and (4.2)

(4.5)
$$f(x^k) \geq f(x^{k+1}) + \lambda_k \langle Jx^k - Jy, x^k - x^{k+1} \rangle$$
$$= f(x^{k+1}) + \lambda_k (\langle Jx^{k+1} - Jy, x^k - x^{k+1} \rangle$$
$$+ \langle Jx^k - Jx^{k+1}, x^k - x^{k+1} \rangle).$$

Observe that the last two terms in the rightmost side of (4.5) are nonnegative; the first one because of the properties of the generalized projection and the fact that $x^k \in \Omega$, and the second one because of the monotonicity of J. Hence we get $f(x^k) \geq f(x^{k+1})$.

(*v*) By *(iv)* we obtain that

$$\lim_{k\to\infty} \lambda_k \langle Jx^k - Jy, x^k - x^{k+1} \rangle = 0.$$

Therefore the assumption on λ_k yields

$$\lim_{k\to\infty} \langle Jx^k - Jy, x^k - x^{k+1} \rangle = 0.$$

On the other hand,

$$\langle Jx^k - Jy, x^k - x^{k+1} \rangle = \langle Jx^k - Jx^{k+1}, x^k - x^{k+1} \rangle$$
$$+ \langle Jx^{k+1} - Jy, x^k - x^{k+1} \rangle.$$

The second term on the right hand side of the previous equation is nonnegative by the projection properties. For the first one, we use (3.1) with $x = x^k$ and $y = x^{k+1}$:

$$\langle Jx^k - Jx^{k+1}, x^k - x^{k+1} \rangle \geq (2L)^{-1} C_0^2 \delta_B \left(\frac{\|x^k - x^{k+1}\|}{2C_0} \right),$$

where we are using the fact that $\{x^k\}$ is bounded by C_0. Now the properties of $\delta_B(\cdot)$, namely the fact that it is an increasing and continuous function such that $\delta_B(0) = 0$, imply that

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

(*vi*) From (3.2)

$$\|Jx^{k+1} - Jx^k\|_{B^*} \leq 8C_0 h_B(16LC_0^{-1}\|x^{k+1} - x^k\|),$$

By virtue of the fact that $h_B(\tau)$ tends to 0 as τ tends to 0, which holds for any uniformly smooth Banach space, we have

(4.6)
$$\lim_{k\to\infty} \|Jx^k - Jx^{k+1}\|_{B^*} = 0.$$

Let now $u^k := Jx^k - Jy$, where x^k and y are taken as in (4.2). We know, by the definition of the algorithm, that $u^k \in \partial f(x^{k+1})$. In order to prove (*vi*) it is enough to show that

$$\lim_{k\to\infty} \langle u^k, x^* - x^{k+1} \rangle = \lim_{k\to\infty} \langle Jx^k - Jy, x^* - x^{k+1} \rangle = 0.$$

As $x^* \in X^*$,

$$0 \leq \langle u^k, x^{k+1} - x^* \rangle = \langle Jy - Jx^k, x^* - x^{k+1} \rangle = \langle Jy - Jx^{k+1}, x^* - x^{k+1} \rangle$$
$$+ \langle Jx^{k+1} - Jx^k, x^* - x^{k+1} \rangle \leq \langle Jx^{k+1} - Jx^k, x^* - x^{k+1} \rangle$$
$$\leq \|x^* - x^{k+1}\| \ \|Jx^{k+1} - Jx^k\|_{B^*},$$

where we have used the gradient inequality, the definition of the algorithm, the projection properties and Cauchy-Schwarz.

Now, using in the previous chain of inequalities (4.6) and the boundedness of $\{x^k\}$, we obtain the desired result.

(vii) Consider the duality mapping J in (4.2), which is weak-to-weak continuous. We will show that there exists only one weak accumulation point. Suppose there are two points z_1, z_2, which are weak limits of subsequences of $\{x^k\}$. By part (i) and (iii), we know that there exist nonnegative numbers l_1 and l_2 such that

$$\lim_{k \to \infty} W(x^k, z_1) = l_1 \quad \text{and} \quad \lim_{k \to \infty} W(x^k, z_2) = l_2.$$

Then

(4.7)
$$\lim_{k \to \infty} W(x^k, z_1) - W(x^k, z_2) = l_1 - l_2$$
$$= \|z_1\|^2 - \|z_2\|^2 + 2 \lim_{k \to \infty} \langle Jx^k, z_2 - z_1 \rangle.$$

Let $l := \lim_{k \to \infty} \langle Jx^k, z_2 - z_1 \rangle$. Let $\{x^{k_j}\}$ and $\{x^{l_j}\}$ be subsequences converging weakly to z_1 and z_2 respectively. Then, taking $k = k_j$ in (4.7) and using the weak-to-weak continuity of J, we get that $l = \langle Jz_1, z_2 - z_1 \rangle$. Repeating the same argument with $k = l_j$ in (4.7), we get $l = \langle Jz_2, z_2 - z_1 \rangle$. Hence, $\langle Jz_2 - Jz_1, z_2 - z_1 \rangle = 0$. According to Lemma 3.1 and the properties of the duality mapping, we conclude that $z_1 = z_2$, which establishes the uniqueness of the weak accumulation point. The proof of item (vii) and of the theorem is now complete. ∎

We have proved that existence of solutions of Problem 4.1 is sufficient to guarantee convergence of the sequence generated by Method (4.2). The next lemma shows that it is also a necessary condition. More precisely, we will prove that the sequence $\{x^k\}$ is unbounded when X^* is empty.

Theorem 4.2. *Under the hypotheses of Theorem 4.1, X^* is nonempty if and only if the sequence $\{x^k\}$ is bounded.*

Proof. Theorem 4.1 provides the proof of the "only if" part. We proceed to prove the "if" part. Suppose that $\{x^k\}$ is bounded. Then its weak closure $\overline{\{x^k\}}$ is bounded and there exists a closed, convex and bounded set D such that

(4.8)
$$\overline{\{x^k\}} \subset D^o,$$

where D^o is the interior of D. It follows that any weak accumulation point of $\{x^k\}$ belongs to D^o. Now we apply method (4.2) to the function $\tilde{f} = f + \chi_D$, where, as before, χ_D denotes the indicator function of the set D. If $\{\tilde{x}^k\}$

is the sequence generated by the algorithm for $\tilde{f}$, then $\tilde{x}^{k+1}$ is uniquely determined by the following system in unknowns $\tilde{x}^{k+1}$, $\tilde{y}$ (see Lemma 3.4).

$$(4.9) \qquad \begin{cases} 0 \in (\partial f + N_D)(\Pi_\Omega(\tilde{y})) + \lambda_k(J\tilde{y} - J\tilde{x}^k), \\ \tilde{x}^{k+1} = \Pi_\Omega(\tilde{y}). \end{cases}$$

We will first prove by induction that the sequence $\{\tilde{x}^k\}$ coincides with the sequence $\{x^k\}$, resulting from applying method (4.2) to f rather than $\tilde{f}$, when $x^0 = \tilde{x}^0$. Suppose that $\tilde{x}^k = x^k$. x^{k+1} is uniquely determined by the following system in unknowns x^{k+1}, y:

$$(4.10) \qquad \begin{cases} 0 \in \partial f(\Pi_\Omega(y)) + \lambda_k(Jy - Jx^k), \\ x^{k+1} = \Pi_\Omega(y). \end{cases}$$

Since x^{k+1} belongs to D^o by (4.8), we have that $N_D(\Pi_\Omega(y)) = N_D(x^{k+1}) = 0$, which, together with (4.10) and the induction hypothesis, gives

$$0 \in (\partial f + N_D)(\Pi_\Omega(y)) + \lambda_k(Jy - J\tilde{x}^k).$$

Therefore, $\Pi_\Omega(y)$ is the unique solution of (4.9), and, as a consequence, $\tilde{x}^{k+1} = \Pi_\Omega(y) = x^{k+1}$. The induction step is complete.

Consider now the problem

$$(4.11) \qquad \min(f + \chi_D)(x)$$

$$\text{s.t. } x \in \Omega.$$

Since D is closed, convex and bounded, and so the operator $\partial f + N_D$ has bounded domain, this problem is known to have solutions (see, e.g. [21], Corollary 32.35), and we can apply Theorem 4.1 to conclude that the sequence $\{\tilde{x}^k\}$ is weakly convergent to a solution of (4.11). Since $\{x^k\}$ and $\{\tilde{x}^k\}$ coincide, as proved above, we conclude that $\{x^k\}$ is weakly convergent to a solution $\hat{x}$ of (4.11). We prove next that $\hat{x}$ belongs to X^*, i.e. it is a solution of $\min f(x)$ subject to $x \in \Omega$. $\hat{x}$ belongs to the weak closure of $\{x^k\}$, and therefore to D^o, so that $N_D(\hat{x}) = 0$. In view of this fact and (4.11), we have

$$(4.12) \qquad 0 \in (\partial f + N_D + N_\Omega)(\hat{x}) = \partial f(\hat{x}) + N_\Omega(\hat{x}).$$

Inclusion (4.12) implies that $\hat{x}$ belongs to X^*, which is therefore nonempty. $\blacksquare$

4.2. Constrained minimization problem: stability analysis. We study in this section the convergence of the following inexact version of algorithm (4.2):

1) Take $\tilde{x}^0 \in \Omega$.
2) Given $\tilde{x}^k$, define $\tilde{x}^{k+1}$ by the system

$$(4.13) \qquad \begin{cases} 0 \in \partial_{\varepsilon_k} f(\Pi_\Omega(y)) + \lambda_k(Jy - J\tilde{x}^k), \\ \tilde{x}^{k+1} = \Pi_\Omega(y), \end{cases}$$

where

$$\partial_\varepsilon f(x) := \{u \in B^* \mid f(z) \geq f(x) + \langle u, z - x \rangle - \epsilon \;, \forall z \in B\}$$

is the ε-*subdifferential* of f at the point x (see, for instance, [11]). The parameters $\{\varepsilon_k\}$ and $\{\lambda_k\}$ are chosen in the following way:

Λ_1) $0 < \lambda_k < \lambda$,

Λ_2) $\varepsilon_k \geq 0$ and $\sum_{k=0}^\infty \lambda_k^{-1} \varepsilon_k < \infty$.

We will show next that for this choice of the parameters the method (4.2) is stable. We point out that the system given by (4.13) has always a solution as a straightforward consequence of the solvability of (4.2).

Theorem 4.3. *Take* $\{\tilde{x}^k\}$ *as in (4.13). In the conditions of Theorem 4.1, the algorithm given by (4.2) is stable, i.e.,*
(i)

$$\lim_{k \to \infty} f(\tilde{x}^k) = f^*,$$

and
(ii)

$$w - \lim_{k \to \infty} \tilde{x}^k = x^*,$$

where $w - \lim$ *stands for weak limit.*

Proof. Take $x^* \in X^*$. We consider as always the sequence $\{W(\tilde{x}^k, x^*)\}$. Let $W_k := W(\tilde{x}^k, x^*)$. Then, by definition of $W(\cdot, \cdot)$, we get

$$W_{k+1} \leq W_k + 2\langle J\tilde{x}^{k+1} - J\tilde{x}^k, \tilde{x}^{k+1} - x^* \rangle$$
$$\leq W_k + 2\langle J\tilde{x}^{k+1} - Jy, \tilde{x}^{k+1} - x^* \rangle + 2\langle Jy - J\tilde{x}^k, \tilde{x}^{k+1} - x^* \rangle$$
$$\leq W_k - 2\lambda_k^{-1} \langle u^{k+1}, \tilde{x}^{k+1} - x^* \rangle,$$

where we use the projection properties, and that $u^{k+1} \in \partial_{\varepsilon_k} f(\tilde{x}^{k+1})$ satisfies the inclusion (4.13). Applying the definition of ε-subdifferential in the previous inequality, we obtain

$$(4.14) \qquad W_{k+1} \leq W_k - 2\lambda_k^{-1}(f(\tilde{x}^{k+1}) - f^*) + 2\lambda_k^{-1}\varepsilon_k.$$

(4.14) and our assumption on the parameters $\{\varepsilon_k\}$ and $\{\lambda_k\}$ imply that the sequence $\{W_k\}$ is convergent and that $\lim_{k\to\infty} f(\tilde{x}^k) = f^*$, and then, with the same argument as in Theorem 4.1, we obtain weak convergence of $\{\tilde{x}^k\}$ to a minimizer. The theorem is proved. ∎

Remark 4.4. *In the particular case in which $\Omega = B$, we get stability results for the unconstrained problem.*

We recall now the definition of Hausdorff distance $\mathcal{H}$ between sets A_1 and A_2, defined as

$$\mathcal{H}(A_1, A_2) = \max\{\sup_{z_2 \in A_2} \inf_{z_1 \in A_1} \|z_1 - z_2\|, \sup_{z_2 \in A_1} \inf_{z_1 \in A_2} \|z_1 - z_2\|\}.$$

Note that if A_1 and A_2 are singletons, then $\mathcal{H}$ reduces to the usual distance.

Theorem 4.5. *Consider the sequence $\{\tilde{x}^k\}$ given by the following algorithm:*
1) *Take $\tilde{x}^0 \in \Omega$.*

2) *Given $\tilde{x}^k$, define $\tilde{x}^{k+1}$ by the system*

(4.15)
$$\begin{cases} 0 \in T^{\varepsilon_k}(\Pi_\Omega(y)) + \lambda_k(Jy - J\tilde{x}^k), \\ \tilde{x}^{k+1} = \Pi_\Omega(y), \end{cases}$$

where ε_k and λ_k are taken as in (4.13). Suppose that the operator $T^\varepsilon : B \to \mathcal{P}(B^*)$ is such that

(4.16)
$$\mathcal{H}(T^\varepsilon(x), \partial f(x)) \le \theta(\|x\|)\varepsilon,$$

for all $x \in \Omega$ and for some finite nondecreasing positive function $\theta(\cdot)$.

If problem (4.15) is solvable and if the sequence $\{\tilde{x}^k\}$ is bounded by C_0, then the whole sequence $\{\tilde{x}^k\}$ converges weakly to a minimizer, and the functional values converge to f^*.

Proof. Indeed, in analogy with Theorem 4.3, it follows from (4.15) that

$$W_{k+1} \le W_k + 2\lambda_k^{-1}\langle v^{k+1}, x^* - \tilde{x}^{k+1}\rangle,$$

where $v^{k+1} \in T^{\varepsilon_k}(\tilde{x}^{k+1})$ satisfies inequality (4.16). The previous inequality can be rewritten in the form

(4.17) $W_{k+1} \le W_k + 2\lambda_k^{-1}\langle u^{k+1}, x^* - \tilde{x}^{k+1}\rangle + 2\lambda_k^{-1}\langle v^{k+1} - u^{k+1}, x^* - \tilde{x}^{k+1}\rangle,$

where $u^{k+1} \in \partial f(\tilde{x}^{k+1})$. By the gradient inequality, we have

$$f^* - f(\tilde{x}^{k+1}) \ge \langle u^{k+1}, x^* - \tilde{x}^{k+1}\rangle,$$

which, together with (4.17), leads to

$$W_{k+1} \le W_k - 2\lambda_k^{-1}(f(\tilde{x}^{k+1}) - f^*) + 2\lambda_k^{-1}\langle u^{k+1} - v^{k+1}, x^* - \tilde{x}^{k+1}\rangle.$$

Using now the Cauchy-Schwartz inequality, (4.16) and the boundedness of the sequences $\{\tilde{x}^k\}$ and $\{\lambda_k\}$, the last expression becomes

$$W_{k+1} \le W_k - 2\lambda^{-1}(f(\tilde{x}^{k+1}) - f^*) + \theta(C_0)\bar{C}\lambda_k^{-1}\varepsilon_k,$$

where $\bar{C}$ is a bound for $2\|x^* - \tilde{x}^k\|$. Now the assertion is obtained by the same argument as in Theorem 4.3. ∎

We remark that when the set Ω is bounded, then the sequence $\{\tilde{x}^k\}$ is ensured to be also bounded.

4.3. Unconstrained minimization problem: convergence rate estimate. We study now the problem

$$\min_{x \in B} f(x),$$

where $f : B \to \mathbb{R}$ is a convex functional. We propose the following method for this problem:

Consider the sequence $\{x^k\}$ given by

1) Take $x^0 \in B$.

2) Given x^k, find $x^{k+1} \in B$, such that:

(4.18)
$$0 \in \partial f(x^{k+1}) + \lambda_k(Jx^{k+1} - Jx^k),$$

where the parameters λ_k are chosen such that $0 < \lambda_k \leq \lambda$.

The method (4.18) is the classical proximal method in a Banach space and also a particular case of (4.2) for $\Omega = B$. We will provide in this section a convergence rate estimate for the proximal point method in a Banach space. The lemmas we prove next are essential in the proof of Theorems 4.9 and 5.1 below.

Lemma 4.6. *Assume that $\{\alpha_k\}$ is a sequence of nonnegative real numbers satisfying the implicit recursive inequality*

$$(4.19) \qquad \alpha_{k+1} \leq \alpha_k - \mu\psi(\alpha_{k+1}), \ \ k = 0, 1, \ldots,$$

where $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ is a continuous and increasing function such that $\psi(0) = 0$ and μ is a positive constant. Then $\alpha_k \to 0$ as $k \to \infty$. Also, if $1 \in R(\psi)$, then there exists $\bar{k} \in [0, C_2]$ such that

$$(4.20) \qquad \alpha_k \leq \psi^{-1}(\frac{C_1}{k})$$

for all $k > \bar{k}$, where

$$C_1 = \exp\left(\mu^{-1}\psi^{-1}(1) + 1\right), \ C_2 = \exp\left(\mu^{-1}\alpha_0 + 1\right)$$

and $\psi^{-1}(\cdot)$ is the inverse function of ψ.

Proof. It follows directly from (4.19) that $\alpha_k \to 0$ as $k \to \infty$.
Consider now two alternatives for any fixed $k \in N$:

$(H_1):$ $\ \psi(\alpha_{k+1}) \leq (k+1)^{-1}.$

$(H_2):$ $\ \psi(\alpha_{k+1}) > (k+1)^{-1}.$

Denote by $N_1 := \{k \in N \,|\, \psi(\alpha_{k+1}) \leq (k+1)^{-1}\,\}$.
The proof of (4.20) will be performed in three steps.
1) We claim that the set N_1 is unbounded. Indeed, if this is not true, there exists N such that for all $k > N$, hypothesis H_2 is satisfied. By (4.19)

$$\alpha_k \leq \alpha_{k-1} - \mu\psi(\alpha_k) < \alpha_{k-1} - \mu k^{-1} < \alpha_N - \mu \sum_{i=N+1}^{k} i^{-1}.$$

If N_1 were bounded, then taking limits as $k \to \infty$ in the previous chain of inequalities, we obtain a contradiction. Indeed, the rightmost term goes to $-\infty$, while the leftmost one is nonnegative. Thus N_1 must be unbounded.

2) Let $\{k_j\}$ denote all the ordered elements of N_1. In the second step we will prove that

$$(4.21) \qquad \frac{k_{j+1}}{k_j + 1} \leq C_1,$$

where C_1 is as in the lemma. If $k_{j+1} = k_j + 1$ then (4.21) is trivially satisfied.

Otherwise, take $k \in [k_j + 1, k_{j+1} + 1]$. We emphasize that for k of the form $k_j + 1$ we have that

$$\psi(\alpha_k) \leq \frac{1}{k},$$

and, for k not of this form, it holds that

$$\psi(\alpha_k) > \frac{1}{k}.$$

Then, for all k such that $k_j + 2 \leq k \leq k_{j+1}$ we have:

$$\alpha_{k_j+2} \leq \alpha_{k_j+1} - \mu\psi(\alpha_{k_j+2}) < \alpha_{k_j+1} - \mu(k_j + 2)^{-1},$$

$$\alpha_{k_j+3} < \alpha_{k_j+1} - \mu((k_j + 2)^{-1} + (k_j + 3)^{-1}),$$

etc. Applying iteratively (4.19), we obtain

$$\alpha_{k_{j+1}} < \alpha_{k_j+1} - \mu \sum_{i=k_j+2}^{k_{j+1}} i^{-1}.$$

Therefore, denoting by $S_p := \sum_{i=1}^{p} i^{-1}$, we get

$$(4.22) \quad \mu(S_{k_{j+1}} - S_{k_j+1}) < \alpha_{k_j+1} - \alpha_{k_{j+1}} \leq \alpha_{k_j+1} \leq \psi^{-1}(\frac{1}{k_j+1}) \leq \psi^{-1}(1),$$

where the second inequality holds because $\{\alpha_k\}$ is a nonnegative sequence, the third one by definition of N_1, and the last one because ψ is an increasing function, and, consequently, ψ^{-1} is also increasing. In order to estimate the leftmost expression in (4.22), we use a result from [9]:

$$S_p = \ln p + \eta(p) + E,$$

where $\eta(p) \in (0, \frac{1}{2})$ and $E \approx 0,577$ is the Euler constant. Applying the previous equality in (4.22) we obtain,

$$\ln \frac{k_{j+1}}{k_j+1} \leq \mu^{-1}\psi^{-1}(1) + |\eta(k_j+1) - \eta(k_{j+1})| \leq \mu^{-1}\psi^{-1}(1) + 1.$$

It may thus be concluded that

$$\frac{k_{j+1}}{k_j+1} \leq \exp\left(\mu^{-1}\psi^{-1}(1) + 1\right),$$

and (4.21) is valid.

3) We will show now that

$$\alpha_k \leq \psi^{-1}(\frac{C_1}{k}),$$

for all $k_j + 1 \leq k \leq k_{j+1} + 1$. Indeed, for $k = k_j + 1$ we have that

$$\psi(\alpha_{k_j+1}) \leq (k_j + 1)^{-1},$$

and consequently

$$\alpha_{k_j+1} \leq \psi^{-1}(\frac{1}{k_j+1}) \leq \psi^{-1}(\frac{C_1}{k_j+1}),$$

because $C_1 \geq 1$. The same estimate is valid for $k = k_{j+1} + 1$.

If k is such that $k_j + 2 \leq k \leq k_{j+1}$, then $k_{j+1}/k \geq 1$ and

$$\alpha_k \leq \alpha_{k_j+1} \leq \psi^{-1}\left(\frac{1}{k_j + 1}\right)$$

$$\leq \psi^{-1}\left(\frac{k_{j+1}}{k(k_j + 1)}\right) \leq \psi^{-1}\left(\frac{C_1}{k}\right),$$

where we use that the sequence $\{\alpha_k\}$ is decreasing in the first inequality, definition of N_1 in the second one, the assumption on k and the properties of ψ in the third one, and step (2) in the last one.

We show next that the estimate (4.20) holds at least beginning from some $\bar{k}$, where $0 \leq \bar{k} \leq \exp(\mu^{-1}\alpha_0 + 1)$. If $k_1 = 0$ then (4.20) is satisfied by any $k > 1$. Otherwise, take $0 \leq k \leq k_1$. We know that hypothesis (H_2) holds for $0 < k \leq k_1$. This means that

$$0 \leq \alpha_k < \alpha_1 - \mu \sum_{i=2}^{k} i^{-1} \leq \alpha_0 - \mu \sum_{i=2}^{k} i^{-1}.$$

Therefore, using again [9], we obtain

$$\ln k + \eta(k) + E \leq \mu^{-1}\alpha_0 + 1,$$

which implies that

$$\ln k \leq \mu^{-1}\alpha_0 + 1.$$

This yields

(4.23) $$k \leq \exp(\mu^{-1}\alpha_0 + 1).$$

Then all the results above are valid for $k_1 \leq C_2$. The proof of the lemma is now complete. ∎

Remark 4.7. *If $1 \notin R(\psi)$, then there exists $0 < \beta < 1$ such that $\beta \in R(\psi)$. In this case the lemma remains valid, but the constants C_1 and C_2 should be changed.*

Next we consider the particular case of inequality (4.19) with $\psi(t) = t^2$, which allows us to improve upon our previous estimate.

Lemma 4.8. *If $\{\alpha^k\}$, $(k = 1, 2, \dots)$ is a sequence of nonnegative real numbers satisfying the implicit recursive inequality $\alpha_{k+1} \leq \alpha_k - \mu\alpha_{k+1}^2$ then it holds that $\alpha_k \leq C_1/k$ for all k, where $C_1 = \max\{2/\mu, \alpha_1\}$.*

Proof. We proceed by induction. The result holds for $k = 1$ by definition of C_1. Assume that it holds for k. Then, the recursive inequality and the induction hypothesis imply that

(4.24) $$\alpha_{k+1} \leq \frac{\sqrt{1 + 4\mu\alpha_k} - 1}{2\mu} \leq \frac{\sqrt{1 + 4\mu C_1/k} - 1}{2\mu},$$

and it suffices to prove that the rightmost expression in (4.24) is less than or equal to $C_1/(k + 1)$, which is equivalent, after some algebra, to

(4.25) $$\frac{1}{\mu}\left(1 + \frac{1}{k}\right) \leq C_1.$$

(4.25) holds because $k \geq 1$ and $C_1 \geq 2/\mu$. ∎

As we mentioned before, the convergence rate estimate that we present next is an application of Lemma 4.6.

Theorem 4.9. *Let $\{x^k\}$ be the sequence given by (4.18) with $0 < \bar{\lambda} \leq \lambda_k \leq \lambda$. Suppose that the set of minimizers of f, which we call X^*, is nonempty and take $x^* \in X^*$. Then results (i)-(vii) of Theorem 4.1 hold. Moreover, defining $u^k := f(x^k) - f^*$, there exists $\bar{k} \in [0, \exp(R_1^{-1}u^0 + 1)]$ such that*

$$(4.26) \qquad u^k \leq R_2^{-1}\delta_{B^*}^{-1}(\frac{C_1}{k})$$

for all $k > \bar{k}$. If C_0 is a bound for $\{x^k\}$ as in Theorem 4.1, then the constants above are given by the relations

$$C_1 = \exp\left(\frac{1}{R_1 R_2}\delta_{B^*}^{-1}(1) + 1\right),$$

$$R_1 = (2L)^{-1}\bar{\lambda}C_0^2, \ \ R_2 = (2\lambda C_0(C_0 + \|x^*\|))^{-1}.$$

Proof. Statements $(i) - (vi)$ hold because of our choice of λ_k and the fact that this algorithm is a particular case of (4.2) for $\Omega = B$.

For establishing (4.26) we show first that

$$(4.27) \qquad u^{k+1} - u^k \leq -R_1\delta_{B^*}(R_2 u^{k+1}),$$

which will allow us to apply Lemma 4.6 to the sequence $\alpha_k = u^k$, with $\mu = R_1$ and $\psi(t) = \delta_{B^*}(R_2 t)$.

By the definition of the method and the gradient inequality, there exists $w^{k+1} \in \partial f(x^{k+1})$ such that

$$(4.28) \quad u^{k+1} - u^k \leq \langle w^{k+1}, x^{k+1} - x^k \rangle = -\lambda_k\langle Jx^{k+1} - Jx^k, x^{k+1} - x^k \rangle.$$

Using Lemma 3.1 and the fact that the sequence is bounded by C_0, the previous inequality becomes

$$(4.29)$$
$$u^{k+1} - u^k \leq -\frac{\lambda_k C_0^2}{2L}\delta_{B^*}\left(\frac{\|Jx^{k+1} - Jx^k\|_{B^*}}{2C_0}\right)$$
$$= -\frac{\lambda_k C_0^2}{2L}\delta_{B^*}(\frac{\|w^{k+1}\|_{B^*}}{2\lambda_k C_0}) \leq -\frac{\bar{\lambda}C_0^2}{2L}\delta_{B^*}(\frac{\|w^{k+1}\|_{B^*}}{2\lambda C_0})$$

On the other hand,

$$f(x^{k+1}) - f^* \leq \langle w^{k+1}, x^{k+1} - x^* \rangle \leq \|x^{k+1} - x^*\|\|w^{k+1}\|_{B^*}$$
$$\leq (C_0 + \|x^*\|)\|w^{k+1}\|_{B^*}.$$

The previous chain of inequalities implies that

$$(4.30) \qquad \|w^{k+1}\|_{B^*} \geq \frac{u^{k+1}}{C_0 + \|x^*\|},$$

which, together with (4.29) and the fact that $\delta_{B^*}(\cdot)$ is an increasing function, gives (4.27).

Now all the assertions of the theorem follow from Lemma 4.6 and the fact that $\psi^{-1}(z) = R_2^{-1}\delta_{B^*}^{-1}(z)$, and the proof is complete. ∎

We recall that spaces ℓ^p, L^p and the Sobolev spaces W_m^p are uniformly convex and uniformly smooth for all $p \in (1, \infty)$ and, denoting any of these spaces by B, we have

$$\delta_B(\varepsilon) \geq 8^{-1}(p-1)\varepsilon^p, \quad \rho_B(\tau) \leq p^{-1}\tau^2$$

for $p \in [2, \infty)$, and

$$\delta_B(\varepsilon) \geq p^{-1}2^{-p}\varepsilon^2, \quad \rho_B(\tau) \leq p^{-1}\tau^p$$

for $p \in (1, 2]$.

In the examples above, as well as in other spaces considered by Pisier in [19], it holds that $\delta_B(\varepsilon) \geq C\varepsilon^\gamma$, where $\gamma \geq 2$ and C is a constant. Under such assumption, we get from (4.29) and (4.30)

$$u^{k+1} - u^k \leq -\frac{\lambda_k C_0^2 C \|w^{k+1}\|_{B^*}^\gamma}{L 2^{\gamma+1} \lambda_k^\gamma C_0^\gamma}$$

$$= -\frac{C\|w^{k+1}\|_{B^*}^\gamma}{2^{\gamma+1} L \lambda_k^{\gamma-1} C_0^{\gamma-2}} \leq -\frac{C}{2^{\gamma+1} L \lambda^{\gamma-1} C_0^{\gamma-2}} \frac{(u^{k+1})^\gamma}{(C_0 + \|x^*\|)^\gamma}.$$

Therefore, it follows from Lemma 4.6 that there exists $\bar{k} \in [0, \exp(R_3 u^0 + 1)]$ such that

$$f(x^k) - f^* \leq (\frac{C_1}{k})^{\frac{1}{\gamma}},$$

for all $k > \bar{k}$, where

$$C_1 = \exp(R_3), R_3 = C^{-1}2^{\gamma+1}L\lambda^{\gamma-1}C_0^{\gamma-2}(C_0 + \|x^*\|)^\gamma.$$

For such spaces, we do not need a positive lower bound for λ_k.

Remark 4.10. *Recalling that $u^0 = f(x^0) - f^*$, (4.23) gives a precise relation between k_1, the first element of N_1, and the initial data. As a corollary, we conclude that if the functional value at the initial point is close to the minimum value f^* of f, then k_1 cannot be too large.*

In Theorem 4.9 we obtained estimate (4.26) for an arbitrary uniformly convex and uniformly smooth Banach space B. Using Lemma 4.8 we can improve the mentioned estimate for a very wide family of Banach spaces, namely the ones which satisfy that $\delta_{B^*}(\varepsilon) \geq C\varepsilon^2$, $C = $ const. This family includes the Hilbert spaces, because

$$\frac{\varepsilon^2}{8} \leq \delta_H(\varepsilon) \leq \frac{\varepsilon^2}{4}.$$

Theorem 4.11. *Suppose that all conditions of Theorem 4.9 hold with $0 < \lambda_k \leq \lambda$. Moreover, assume that $\delta_{B^*}(\varepsilon) \geq C\varepsilon^2$, $C = $ const. Then*

$$(4.31) \quad f(x^k) - f^* \leq \frac{C_1}{k}, \ C_1 = \max\{16C^{-1}L\lambda(C_0 + \|x^*\|)^2, f(x^0) - f^*\}$$

for all $k \geq 1$.

Proof. Since $\delta_{B^*}(\varepsilon) \geq C\varepsilon^2$, we have from (4.29)

$$u^{k+1} - u^k \leq -\frac{C}{8\lambda L(C_0 + \|x^*\|)^2}(u^{k+1})^2.$$

Now (4.31) follows from Lemma 4.8 and the fact that $u^1 \leq u^0$. ∎

Corollary 4.12. *Consider the spaces ℓ^p, L^p and W_m^p. Suppose that $2 \leq p < \infty$. If $0 < \lambda_k \leq \lambda$, then, under the hypotheses of Theorem 4.9, we have*

$$f(x^k) - f^* \leq \frac{C_1}{k}, \ C_1 = \max\{16q2^q\lambda L(C_0 + \|x^*\|)^2, \ f(x^0) - f^*\}$$

for all $k \geq 1$, where $q = p/(p-1)$.

The proof follows directly from the fact that when $2 \leq p < \infty$, the corresponding dual space $B^* = \ell^q$, L^q, W_m^q satisfies

$$\delta_{B^*}(\varepsilon) \geq q^{-1}2^{-q}\varepsilon^2, \ q = p/(p-1), \ 1 < q \leq 2.$$

An estimate similar to (4.31) was obtained in [10] for the particular case of a Hilbert space. Lemma 4.8 allows us to obtain a much simpler proof for this case. In Hilbert spaces, J is the identity operator, and the method (4.18) takes the form

$$0 \in \partial f(x^{k+1}) + \lambda_k(x^{k+1} - x^k).$$

Using (4.30), we have

$$u^{k+1} - u^k \leq \langle w^{k+1}, x^{k+1} - x^k \rangle =$$

$$\leq -\frac{\|w^{k+1}\|^2}{\lambda_k} \leq -\frac{(u^{k+1})^2}{\lambda(C_0 + \|x^*\|)^2}.$$

Now Lemma 4.8 immediately gives the following estimate:

$$f(x^k) - f^* \leq \frac{C_1}{k}, \ C_1 = \max\{2\lambda(C_0 + \|x^*\|)^2, \ f(x^0) - f^*\}$$

for all $k \geq 1$.

5. CONVERGENCE ANALYSIS FOR UNIFORMLY CONVEX FUNCTIONALS

We will establish now strong convergence in the case of uniformly convex functionals. In this section, a uniformly convex functional will be understood as a functional f such that there exists a continuous and increasing function $\Psi : [0, \infty] \to [0, \infty]$ satisfying $\Psi(0) = 0$, $\lim_{t\to\infty} \Psi(t) = \infty$ and

(5.1) $$\langle u, x - x^* \rangle \geq \Psi(\|x - x^*\|).$$

for all $x \in B$, and all $u \in \partial f(x)$, where x^* is the minimizer of f on Ω.

Theorem 5.1. *Suppose that f is a uniformly convex functional and take $\Psi(\cdot)$ as in (5.1). Then, the sequence generated by (4.2) with $0 < \lambda_k \leq \lambda$ converges strongly to x^* and there exists $\bar{k} \in [0, \exp(2^{-1}\lambda D_0 + 1)]$ such that*

(5.2) $$\|x^k - x^*\| \leq 4C_0\delta_B^{-1}(\frac{1}{8C_0^2}\tilde{\Psi}^{-1}(\frac{C_1}{k})),$$

for all $k > \bar{k}$, where C_0 is a bound for $\{x^k\}$, $\tilde{\Psi}(t) = \Psi(4C_0\rho_B^{-1}(t/4C_0^2))$, $D_0 = (\|x^0\| + \|x^*\|)^2$ and

$$C_1 = \exp\left(\frac{\lambda}{2}\tilde{\Psi}^{-1}(1) + 1\right).$$

Proof. By (4.3), the gradient inequality and definition of $\Psi(\cdot)$, we get

(5.3) $$W(x^{k+1}, x^*) \leq W(x^k, x^*) - (2/\lambda_k)\Psi(\|x^{k+1} - x^*\|).$$

Then the sequence $\{W(x^k, x^*)\}$ is nonnegative and decreasing, hence convergent. We show next that the properties of $\Psi(\cdot)$ and the previous inequality imply the strong convergence of $\{x^k\}$. Indeed, the following property of $W(\cdot, \cdot)$ can be found in [1], Theorem 7.5:

(5.4) $$4C\rho_B^{-1}[W(x,y)/(4C^2)] \leq \|x - y\| \leq 4C\delta_B^{-1}[W(x,y)/(8C^2)],$$

for any $x, y \in B$, such that $\|x\| \leq C$, $\|y\| \leq C$. Applying the leftmost inequality in (5.4) to (5.3), and using the assumption on C_0, we have

$$W(x^{k+1}, x^*) \leq W(x^k, x^*) - 2\lambda_k^{-1}\Psi(4C_0\rho_B^{-1}[W(x^{k+1}, x^*)/(4C_0^2)].$$

Let $W_k := W(x^k, x^*)$. Then the previous inequality can be rewritten as

$$W_{k+1} \leq W_k - 2\lambda_k^{-1}\tilde{\Psi}(W_{k+1}) \leq W_k - 2\lambda^{-1}\tilde{\Psi}(W_{k+1}).$$

Observe that $\tilde{\Psi}$ is continuous and $\tilde{\Psi}(0) = 0$, because $\rho_B^{-1}(0) = 0$. Then we are in conditions of Lemma 4.6 with $\alpha_k = W_k$, $\psi(t) = \tilde{\Psi}(t)$ and $\mu = 2/\lambda$. This lemma implies that

$$W_k \leq \tilde{\Psi}^{-1}\left(\frac{C_1}{k}\right),$$

for all $k \geq \bar{k}$, where $\bar{k} \leq \exp(2^{-1}\lambda W_0 + 1)$. Combining now (5.4) with the previous inequality, we get

$$8C_0^2\delta_B\left(\frac{\|x^k - x^*\|}{4C_0}\right) \leq \tilde{\Psi}^{-1}\left(\frac{C_1}{k}\right),$$

which gives

$$\|x^k - x^*\| \leq 4C_0\delta_B^{-1}\left(\frac{1}{8C_0^2}\tilde{\Psi}^{-1}\left(\frac{C_1}{k}\right)\right).$$

The constant D_0 is obtained applying the formula (3.4) in the bound for $\bar{k}$ and thus this establishes (5.2). ∎

Remark 5.2. *In the same way as in Section 4, the estimate (5.2) can be improved for the special case in which $\tilde{\Psi}(t) = t^2$. It is also obvious that if $\tilde{\Psi}(t) = t$, then W_k converges to 0 with a linear convergence rate.*

Remark 5.3. *In [2], explicit versions of gradient type methods were considered for $\{\lambda_k\}$ and $\{\varepsilon_k\}$ such that $\lim_{k\to\infty} \lambda_k^{-1} = 0$ and $0 \leq \varepsilon_k \leq \bar{C}\lambda_k^{-1}$, for some constant $\bar{C}$. The authors proved in this reference all the results of Theorem 4.1 in the case of a Banach space in which $\delta_B(\varepsilon) \geq \bar{C}\varepsilon^2$. The same assumption on the parameter $\{\lambda_k\}$ is still necessary for uniformly convex*

functionals in an arbitrary uniformly convex and uniformly smooth Banach space (see [3]*).*

Remark 5.4. *It is not difficult to show that the proximal point algorithm (4.2) can be written as*

$$(5.5) \qquad x^{k+1} = \pi_\Omega(Jx^k - \lambda_k u^{k+1}),$$

where $u^{k+1} \in \partial f(x^{k+1})$, *and* $\pi_\Omega : B^* \to \Omega \subset B$ *is generalized projection operator introduced in* [1].

Acknowledgment. The first author wishes to thank the Institute for Pure and Applied Mathematics (IMPA) in Rio de Janeiro, Brazil, where he was a visiting professor while this paper was written.

REFERENCES

[1] Ya. Alber, *Metric and generalized projection operators in Banach spaces: properties and applications*, in "Theory and Applications of Nonlinear Operators of Monotone and Accretive Type", (ed. A. Kartsatos), Marcel Dekker, New York, 1996, 15–50.

[2] Ya. Alber, A. N. Iusem and M. Solodov, *Minimization of nonsmooth convex functionals in Banach spaces*, J. Convex Anal., to appear.

[3] Ya. Alber and A. Notik *On the iterative method for variational inequalities with nonsmooth unbounded operators in Banach space*, J. Math. Anal. Appl. **188** (1994), 928–939.

[4] M. Avriel, *Nonlinear programming, analysis and methods*, Prentice Hall, New Jersey, 1976.

[5] F. E. Browder, *Nonlinear operators and nonlinear equations of evolution in Banach spaces*, Proc. Sympos. Pure Math., **#18**, AMS, 1976.

[6] R. S. Burachik and A. N. Iusem, *A generalized proximal point algorithm in a Hilbert space*, SIAM J. Optim., to appear.

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

[8] T. Figiel, *On the moduli of convexity and smoothness*, Studia Math. **56** (1976), 121–155.

[9] I. S. Gradshteyn and M. Ryzhik, *Table of integrals, series and products*, Academic Press, New York, 1965.

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

[11] J. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms*, Springer-Verlag, Berlin, 1993.

[12] M. A. Krasnoselskii, *Two observations about the method of successive approximations*, Uspehi Mat. Nauk, **10** (1955), 123–127.

[13] A. V. Kryanev, *The solution of incorrectly posed problems by methods of successive approximations*, Soviet Math. **14** (1973), 673–676.

[14] B. Lemaire, *The proximal algorithm*, in "International Series of Numerical Mathematics", (ed. J. P. Penot), Birkhäuser, Basel, **87** (1989), 73–87.

[15] B. Martinet, *Régularisation d'inéquations variationelles par approximations successives*, Rev. Française Informat. Recherche Opérationelle, **4** (1970), 154–159.

[16] B. Martinet, *Algorithmes pour la résolution de problèmes d'optimisation et minimax*, Thèse d'État, Université de Grenoble, 1972.

[17] J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, **93** (1965), 273–299.

[18] Z. Opial, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bull. Amer. Math. Soc. **73** (1967), 591–597.
[19] G. Pisier, *Martingales with values in uniformly convex spaces*, Israel J. Math. **20** (1975), 329–350.
[20] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim. **14** (1976), 877–898.
[21] E. Zeidler, *Functional analysis and its applications*, **II/B**, Springer-Verlag, Berlin, 1985.

Y. I. ALBER
INSTITUTO DE MATEMÁTICA PURA E APLICADA
ESTRADA DONA CASTORINA 110
RIO DE JANEIRO, RJ, CEP 22460-320, BRAZIL

E-mail address: alberya@techunix.technion.ac.il

R. S. BURACHIK
DEPARTAMENTO DE MATEMÁTICA
PONTÍFICIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO
RUA MARQUÉS DE SÃO VICENTE 225
RIO DE JANEIRO, RJ, CEP 22453-030, BRAZIL

E-mail address: regina@impa.br

A. N. IUSEM
INSTITUTO DE MATEMÁTICA PURA E APLICADA
ESTRADA DONA CASTORINA 110
RIO DE JANEIRO, RJ, CEP 22460-320, BRAZIL