Numerical Functional Analysis & Optimization
Vol. 18, No. 7&8, 1997, pp. 723-744.

ON A PROXIMAL POINT METHOD FOR CONVEX OPTIMIZATION IN BANACH SPACES

Dan Butnariu
Department of Mathematics and Computer Science
University of Haifa
31905 Haifa, Israel

Alfredo N. Iusem
Instituto de Matemática Pura e Aplicada
Estrada Dona Castorina 110, Jardim Botânico
Rio de Janeiro, R.J., CEP 22460-320, Brazil

ABSTRACT

We analyze the behavior of a parallel proximal point method for solving convex optimization problems in reflexive Banach spaces. Similar algorithms were known to converge under the implicit assumption that the norm of the space is Hilbertian. We extend the area of applicability of the proximal point method to solving convex optimization problems in Banach spaces on which totally convex functions can be found. This includes the class of all smooth uniformly convex Banach spaces. Also, our convergence results leave more flexibility for the choice of the penalty function involved in the algorithm and, in this way, allow simplification of the computational procedure.

Keywords: Convex optimization problem, Proximal point method, Totally convex function, Bregman distance, Bochner integral, Uniformly convex Banach space.

1. INTRODUCTION

The classical proximal point method for optimization, analyzed in details in [28], is devised to minimize a proper, lower semicontinuous, convex function $g : H \to (-\infty, +\infty]$ defined on a Hilbert space H. The iteration is of the form

$$x^{k+1} = \operatorname{argmin}\{g(x) + (\omega_k/2)\|x - x^k\|^2\}, \tag{1.1}$$

where $\{\omega_k\}_{k\in\mathbb{N}}$ is a bounded sequence of positive real numbers. If the function g is differentiable, then x^{k+1} is the unique solution of

$$g'(x) + \omega_k x = \omega_k x^k, \tag{1.2}$$

where $g'(x)$ denotes the derivative of g at x. If we have a constrained optimization problem, that is, if we have to minimize the function g over a closed, convex, nonempty subset C of H, then we have to replace in (1.1) the function g by the function $h := g + I_C$, where I_C is the indicator function of the set C, i.e. $I_C(x) = 0$ if $x \in C$ and $I_C(x) = +\infty$, otherwise. In this case the equation (1.2) becomes

$$g'(x) + N_C(x) + \omega_k x \ni \omega_k x^k, \tag{1.3}$$

where $N_C(x)$ is the normal cone to C at x.

One may attempt to extend the classical proximal point method (1.1) to a strictly convex and smooth Banach space B (that is, to a space with strictly convex and Gâteaux differentiable norm). Then one confronts the following difficulty: the Gâteaux derivative of the functional $x \to \|x\|^2$ is $2J(x)$, where J is the duality mapping from B to its dual B^*. Therefore, the first order differentiability condition of (1.1) is now

$$g'(x) + \omega_k J(x - x^k) = 0, \tag{1.4}$$

where $g'(x)$ is the Gâteaux differential of g at x. The nonlinearity of J makes equation (1.4) considerably harder to solve than (1.2). Note also that (1.4) is an equation in B^* rather than in B. A solution of this difficulty is the replacement in (1.1) of the square of the distance between x and x^k by the so called Bregman distance. The possibility of building proximal point type algorithms with Bregman distances instead of the norm induced distance was first studied in [15] in the finite dimensional context of $\mathbb{R}^n$ and for purposes different from ours. Given a function $f : B \to (-\infty, +\infty]$ which is Gâteaux differentiable and strictly convex on the interior $\mathcal{D}^\circ$ of its domain $\mathcal{D}$, the *Bregman distance with respect to* f (cf. [6] and [13]) is the function $D_f : \mathcal{D} \times \mathcal{D}^\circ \to \mathbb{R}$ defined by

$$D_f(y, x) = f(y) - f(x) - \langle f'(x), y - x\rangle, \tag{1.5}$$

where $\langle \cdot, \cdot \rangle$ stands for the duality pairing. The function D_f is not a distance in the usual sense of the term (in general, it is not symmetric and does not satisfy the triangle inequality). However, if we replace (1.1) by

$$x^{k+1} = \operatorname{argmin}\{g(x) + \omega_k D_f(x, x^k)\}, \tag{1.6}$$

then, under the assumption that $\mathcal{D} = B$, the first order optimality condition becomes

$$g'(x) + \omega_k f'(x) = \omega_k f'(x^k). \tag{1.7}$$

Whenever convergence of the sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by (1.6) towards a minimum of g can be ensured, the equation (1.7) is more convenient than (1.4) because it leaves the

freedom of fitting the function f to the nature of the function g and of the space B in ways which makes resolution of (1.7) much simpler that resolution of (1.4). Also, note that the smoothness of B which is needed in order to make (1.4) an equation and not a more difficult to solve inclusion is no longer needed, that is, one may apply (1.6) to nonsmooth Banach spaces as long as a convenable function f on B can be found.

It should be observed that, if B is a Hilbert space H, then by taking in (1.6) the function $f(x) = \frac{1}{2}\|x\|^2$ one obtains exactly the classical proximal point method (1.1). However, if B is not a Hilbert space, but still uniformly convex and smooth, then, letting $f(x) = \frac{1}{2}\|x\|^2$ in (1.6), one can determine the iterates by the first order optimality condition

$$g'(x) + \omega_k J(x) = \omega_k J(x^k) \tag{1.8}$$

which, due to the nonlinearity of the duality mapping J, is not necessarily equivalent to the equation (1.4). This particular instance of the algorithm (1.6) was studied in [2] by using several results of Banach space geometry. Nevertheless, it must be remarked that the choice $f(x) = \frac{1}{2}\|x\|^2$ in some Banach spaces may make computation of the iterates (1.6) quite difficult. These computations can be simplified by an appropriate choice of f (see Section 4). For instance, if $B = \mathcal{L}^p$ or $B = l^p$ with $p \in (1, +\infty)$, then, by choosing $f(x) = \|x\|^p$, the equation (1.7) is much simpler than its version (1.8) occurring when one takes $f(x) = \frac{1}{2}\|x\|^2$.

The algorithm (1.6) has been considered in [17] (see also [18]) under the assumption that the function f is strongly convex, i.e.,

$$\langle f'(x) - f'(y), x - y\rangle > \gamma\|x - y\|^2, \tag{1.9}$$

for all $x, y \in B$ and for some $\gamma > 0$. Under this condition, the convergence analysis of (1.6) is almost identical to that in the Hilbert space case. The problem is that (1.9) is a too demanding condition. Strongly convex functions which are twice differentiable at least at some $z \in B$ (as implicitly required in the convergence analysis done in [17]) exist only if B is isomorphic to a Hilbert space (in the sense that the norm of B is equivalent to a Hilbertian norm). This fact was proved in [4] using the following argument: Let $f''(z) : B \times B \to \mathbb{R}$ be the continuous bilinear function which is the second derivative of f at the point z; Define $\|x\|_* = [f''(z)(x, x)]^{1/2}$; Then, by (1.9) and the bilinearity and continuity of $f''(z)$ it results that $\|x\|_*$ is a Hilbertian norm on B and

$$\sqrt{\gamma}\|x\| \leq \|x\|_* \leq \|f''(z)\|\,\|x\|.$$

This shows that the results in [17] and [18] hold only in Hilbert spaces. In particular, these results do not apply in spaces like $\mathcal{L}^p$, l^p and $W^{p,m}$ (Sobolev spaces) for $p \neq 2$.

In this paper we proceed to extend the above discussed results. We consider the optimization problem

$$\text{minimize } g(x) \text{ such that } x \in C, \tag{1.10}$$

where C is a nonempty, convex and closed subset of the reflexive, separable Banach space B and $g : C \to \mathbb{R}$ is a proper lower semicontinuous convex function which is

bounded from below. We associate to these data a function $f : B \to (-\infty, +\infty]$ with closed domain $\mathcal{D} := dom(f)$ such that $C \subset \mathcal{D}^\circ := int(dom(f))$, and f is totally convex and Fréchet differentiable on $\mathcal{D}^\circ$. For each real number $\omega > 0$ define the operator $T_\omega : C \to C$ by

$$T_\omega(x) = \operatorname{argmin}\{g(y) + \omega D_f(y, x);\ y \in C\}, \tag{1.11}$$

where D_f is the Bregman distance with respect to f defined by (1.5). The operators T_ω, $\omega \in (0, +\infty)$, are well-defined as we show below (see Lemma 1). The iterations we consider are of the form

$$x^0 \in C \text{ and } x^{k+1} = \int_0^b T_\omega(x^k) d\lambda_k(\omega), \tag{1.12}$$

where b is a positive real number, λ_k denotes a complete probability measure over a σ-algebra $\mathcal{A}_k$ on $(0, b]$ which contains all Borel subsets of this interval and the integral is in the sense of Bochner (see [25]). We call this method of generating sequences $\{x^k\}_{k\in\mathbb{N}}$ *the parallel proximal point method.*

We aim at showing that, under quite undemanding conditions concerning the Banach space B and the optimization problem (1.10), sequences $\{x^k\}_{k\in\mathbb{N}}$ generated in reflexive Banach spaces via the proximal point method exist and the corresponding sequences $\{g(x^k)\}_{k\in\mathbb{N}}$ converge nonincreasingly to the optimal value of the problem (1.10). Moreover, when some additional requirements are satisfied, the sequences $\{x^k\}_{k\in\mathbb{N}}$ converge weakly, and sometimes even strongly, to optimal solutions of (1.10). Observe that the iterative procedure (1.6) (and, then, (1.1)) is the particular case of (1.12) where $C = B$ and the measures λ_k are concentrated in the points ω_k, i.e., $\lambda_k(A) = 1$ if $\omega_k \in A \subseteq (0, b]$, and $\lambda_k(A) = 0$, otherwise. The particular version of (1.12) in which each of the probability measures λ_k is concentrated in a point $\omega_k \in (0, b]$ is called *the sequential proximal point method.*

The extension of the classical proximal point method incorporated in (1.12) is two-fold. The first significant extension lies in the requirement imposed upon the function f in order to guarantee convergence of the method. We require a condition much weaker than strong convexity, namely total convexity, which is described in Section 2. This condition is satisfied by the function $f(x) = \frac{1}{2}\|x\|^2$ when B is a Hilbert space, but also by the function $f(x) = \|x\|^s$ when $B = \mathcal{L}^p$ or $B = l^p$ with $s, p \in (1, +\infty)$, as shown in [22]. Since none of the later functions is strongly convex when $p \neq 2$ (as noted before, existence of twice differentiable strongly convex functions on a Banach space means that the space is equivalent to a Hilbert space), our approach covers algorithms which were not covered by the analysis done in [17] and [18] as shown above. The algorithm studied in [2] is another particular version of our method which occurs when B is uniformly convex and smooth and $f(x) = \|x\|^2$. We use a result in [1] in order to show (see Proposition 1) that, if B is uniformly convex and smooth, then the function $f(x) = \|x\|^s$ with $s \geq 2$ is totally convex and, therefore, our convergence results concerning (1.12) also extend the convergence results given in [2]. This extension is significant not only because it allows application of the proximal point method in new

environments, but also because it leaves much freedom for the choice of the function f involved in the algorithm and, thus, it may help reduce the computational effort required when the vectors $T_\omega(x^k)$ have to be computed (see Section 4).

The second extension incorporated in (1.12) consists of allowing averages of the vectors $T_\omega(x^k)$. Theorem 1 below shows that convergence towards the optimal value of (1.10) of the sequence $\{g(x^k)\}_{k\in\mathbb{N}}$ determined by the proximal point method can be ensured even with this modification. The introduction of the averages allows simultaneous use in the procedure of families of vectors $T_\omega(x^k)$ obtained by parallel computations. Corollary 1 in Section 2 shows that convergence of sequential proximal point algorithms can be ensured under less demanding condition. Proposition 2 in Section 4 emphasize the fact that there are problems of practical interest for the resolution of which the sequential proximal point method with appropriately chosen numbers ω_k ensures faster convergence of $\{g(x^k)\}_{k\in\mathbb{N}}$. However, in specific circumstances, the parallel proximal point method can be useful for avoiding computational difficulties caused by the use of approximate values of the iterates (see Section 4).

The technique on which our convergence analysis of the algorithm (1.12) is based (see Section 3) is implicitly exploiting the easily observable fact that the only common fixed points of the operators T_ω, $\omega \in (0, b]$, are the optimal solutions of the problem (1.10). This combines with the fact that the operators T_ω are *firmly nonexpansive with respect to* f (see Lemma 2 and its proof), that is,

$$\langle f'(T_\omega(x)),\, z - T_\omega(x)\rangle \geq \langle f'(x), z - T_\omega(x)\rangle, \tag{1.13}$$

whenever $x \in C$ and z is an optimal solution of (1.10). Condition (1.13) is equivalent to firm nonexpansivity in the special case when B is a Hilbert space and $f(x) = \|x\|^2$ and this justifies its name -- cf. [14]. Therefore, our analysis of the proximal point method is Sections 2 and 3 uses features commonly found in the convergence analysis of particular proximal point algorithms studied in [9], [15], [16], [19] and [21] in combination with arguments similar to those involved in the study of algorithms for computing common fixed points of families of operators (see [11]).

2. A CONVERGENCE ANALYSIS OF THE PROXIMAL POINT METHOD

In this section we present a series of conditions which ensure that the sequences generated by the proximal point method (1.12) exist and have interesting convergence properties. Our analysis of the proximal point method is done under the following assumptions concerning the basic data of the optimization problem (1.10) to which the method is applied:

(A) The Banach space B is reflexive;
(B) The feasible set C is nonempty, convex and closed and the function $g : C \to \mathbb{R}$ is convex, lower semicontinuous and bounded from below;
(C) There exists a convex function $f : B \to (-\infty, +\infty]$ with domain $\mathcal{D}$ such that $C \subset \mathcal{D}^\circ := int(\mathcal{D})$, f is Fréchet differentiable and totally convex on $\mathcal{D}^\circ$ and, for each $x \in C$, the level sets of the function $D_f(x, \cdot)$,

$$R_\alpha^f(z) := \{y \in C;\ D_f(z,y) \le \alpha\},$$

are bounded for all $\alpha \ge 0$.

Recall that the function $f : B \to (-\infty, +\infty]$ is called *totally convex on* $\mathcal{D}$ if, for each $x \in \mathcal{D}^\circ$, *the local modulus of convexity of* f *at* x defined by (see [10] and [11])

$$\nu_f(x,t) = \inf\Big\{D_f(y,x);\ y \in \mathcal{D},\ \|y-x\| = t\Big\}, \tag{2.1}$$

is positive for every $t \in (0, +\infty)$. Condition (C) above is essential to our convergence analysis of the proximal point method in Banach spaces. Thus, the question whether totally convex functions on a Banach space do exist is intrinsically related to the applicability of our results. This aspect is discussed in Section 4 where we show that on uniformly convex smooth Banach spaces there are meaningful pools of such functions which can be eventually used when the proximal point method is applied. The uniformly convex functions are among the totally convex functions and our analysis shows that using the proximal point method (1.12) with a uniformly convex function f leads to sequences $\{x^k\}_{k\in\mathbb{N}}$ with some special convergence properties. Note that the function f is called uniformly convex when its (global) modulus of convexity (see [29]) given by

$$\mu_f(t) = \inf\Big\{\frac{\alpha f(x) + (1-\alpha)f(y) - f[\alpha x + (1-\alpha)y]}{\alpha(1-\alpha)};\alpha \in (0,1), \|x-y\| = t\Big\},$$

is positive for all $t > 0$. The function f can be totally convex without being uniformly convex (cf. [10]) and, in some circumstances like in the case when B is one of the spaces $\mathcal{L}^p$ or l^p with $p \in (1,2)$, totally convex functions satisfying the requirements of the proximal point method although they are not uniformly convex, can be much easier found and dealt with in computations.

The following result emphasizes the most important property of sequences $\{x^k\}_{k\in\mathbb{N}}$ generated by the proximal point method. Namely, it shows that the sequence $\{g(x^k)\}_{k\in\mathbb{N}}$ converges decreasingly to the optimal value of the optimization problem, even if the sequence $\{x^k\}_{k\in\mathbb{N}}$ itself does not necessarily converge. Also, it gives a lower estimate of number $g(x^k) - g(x^{k+1})$, the *descending jump of* g at step k, which shows how much closer we are to the optimal value of the problem after the k-th step of the algorithm.

Theorem 1. *Suppose that the problem* (1.10) *has optimal solutions. Then,*

(I) *For any initial point* $x^0 \in C$, *the sequence* $\{x^k\}_{k\in\mathbb{N}}$ *generated by the proximal point method* (1.12) *exists, is contained in* C, *the sequence* $\{g(x^k)\}_{k\in\mathbb{N}}$ *converges nonincreasingly and*

$$g(x^k) - g(x^{k+1}) \ge \int_0^b \omega D_f(T_\omega(x^k), x^k)\, d\lambda_k(\omega). \tag{2.2}$$

(II) *If, for any optimal solution $z \in C$ of the problem* (1.10), *the function $D_f(z, \cdot)$ is convex on C, then the sequence $\{g(x^k)\}_{k\in\mathbb{N}}$ converges to the minimal value of g over C, i.e.,*

$$\lim_{k\to\infty} g(x^k) = \inf\{g(x);\ x \in C\}. \tag{2.3}$$

In this case, the sequence $\{x^k\}_{k\in\mathbb{N}}$ is bounded, has weak accumulation points and each weak accumulation point of it is an optimal solution of the problem (1.10). *Moreover, the following limit exists and*

$$\lim_{k\to\infty} D_f(x^{k+1}, x^k) = 0. \tag{2.4}$$

If f is also uniformly convex, then

$$\lim_{k\to\infty} \|x^k - x^{k+1}\| = 0. \tag{2.5}$$

A restrictive condition involved in Theorem 1(II) is the requirement that $D_f(z, \cdot)$ should be convex on C when z is an optimal solution of (1.10). This is needed in the proof of that result (see the proof of Lemma 4 in Section 3 below) in order to ensure that the firm nonexpansivity (1.13) is transferable from the operators T_ω to the operator

$$\boldsymbol{T}_k(x) = \int_0^b T_\omega(x) d\lambda_k(\omega).$$

If λ_k is concentrated in one point ω_k, then $\boldsymbol{T}_k$=T_{ω_k} and, therefore, the firm nonexpansivity of $\boldsymbol{T}_k$ with respect to f is a direct result of the firm nonexpansivity of $T_{\omega_k}$with respect to f which is proved in Lemma 2 without involving the requirement that $D_f(z, \cdot)$ should be convex. Thus, we have the following result showing that for the sequential proximal point algorithm the basic property (2.3) can be guaranteed under less restrictive conditions than that of the parallel proximal point method.

Corollary 1. *Suppose that the problem* (1.10) *has optimal solutions. Then, for any sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by the sequential proximal point method, the sequence $\{g(x^k)\}_{k\in\mathbb{N}}$ converges nonincreasingly,*

$$g(x^k) - g(x^{k+1}) \geq \omega_k D_f(T_{\omega_k}(x^k), x^k), \tag{2.6}$$

and the equations (2.3) *and* (2.4) *hold. Moreover, $\{x^k\}_{k\in\mathbb{N}}$ has weak accumulation points and all its weak accumulation points are optimal solutions of the problem* (1.10).

Theorem 1 and its corollary give sufficient conditions for the proximal point method approximation of the optimal value of the problem (1.10). Does the sequence $\{x^k\}_{k\in\mathbb{N}}$ converge (weakly or strongly) to an optimal solutions of the given problem? The next result shows that this indeed happens whenever the function f involved in the algorithm satisfies some additional requirements.

Theorem 2. *Suppose that the problem* (1.10) *has optimal solutions. If either* $D_f(z, \cdot)$ *is convex for each optimal solution* z *of* (1.10), *or each probability measure* λ_k *is concentrated in some point* $\omega_k \in (0, b]$, *then*

(I) *The sequence* $\{x^k\}_{k\in\mathbb{N}}$ *generated by the proximal point method converges weakly to an optimal solution of the problem* (1.10) *whenever one of the following conditions is satisfied:*

(a) *For any two sequences* $\{u^k\}_{k\in\mathbb{N}}$ *and* $\{v^k\}_{k\in\mathbb{N}}$ *contained in* C *and converging weakly to* u *and* v, *respectively,*

$$\liminf_{k\to\infty}\langle f'(u^k) - f'(v^k), u - v\rangle = 0 \tag{2.7}$$

implies $u = v$;

(b) *The function* $f' : \mathcal{D}^\circ \to B^*$ *is sequentially weakly-to-weak* continuous on* C;

(c) *For any sequence* $\{u^k\}_{k\in\mathbb{N}}$ *contained in* C *and converging weakly to some point* u, *we have*

$$\limsup_{k\to\infty}\langle f'(u^k), u^k - u\rangle \le 0. \tag{2.8}$$

(II) *The sequence* $\{x^k\}_{k\in\mathbb{N}}$ *converges strongly to an optimal solution of the problem* (1.1) *whenever one of the following requirements is satisfied:*

(d) *The space* B *has finite dimension;*

(e) *The set* C *is compact;*

(f) *The function* f *is uniformly convex and condition* (c) *above holds.*

As noted in Section 1, the basic feature of the proximal point method is reducing the resolution of the original problem (1.10) to solving a series of optimization problems for finding values $T_\omega(x^k)$ of the operators T_ω. Clearly, finding a value $T_\omega(x^k)$ means solving a convex optimization problem in the form (1.11). This problem has only one optimal solution (see Lemma 1 in Section 3 below) even if problem (1.10) may have infinitely many optimal solutions. Finding the optimal solution of the problem (1.11) amounts to finding the unique point $x \in C$ such that

$$0 \in \partial[g + \omega D_f(\cdot, x^k)](x). \tag{2.9}$$

Analyzing the proof of Lemma 2 in the next section one can easily see that finding a point $x \in C$ which satisfies (2.9) is equivalent to finding a point $x \in C$ such that

$$\omega \cdot [f'(x^k) - f'(x)] \in \partial g(x). \tag{2.10}$$

In general, solving the inclusion (2.10) may not be easy. An example in Section 4 illustrates the fact that the freedom of choosing the function f given by the results above allows considerable simplification of this problem.

3. PROOFS OF THEOREMS 1 AND 2

The proofs of Theorems 1 and 2 consist of a series of lemmas. The first of the lemmas shows that the operators T_ω given by (1.11) are well-defined.

Lemma 1. *For each $x \in C$ and for any $\omega \in (0,+\infty)$, there exists an unique minimizer of the function $\Phi(\omega,x,\cdot) : C \to \mathbb{R}$ given by*

$$\Phi(\omega, x, y) = g(y) + \omega D_f(y, x).$$

Proof. The function $\Phi(\omega,x,\cdot)$ is convex because g, as well as $D_f(\cdot, x)$, are convex. Since f is totally convex it is strictly convex on $\mathcal{D}^\circ$ (cf. [11]) and, consequently, so is $D_f(\cdot, x)$. Hence, $\Phi(\omega,x,\cdot)$ is strictly convex and, therefore, it has at most one minimization point over C. Note that $\Phi(\omega,x,\cdot)$ is bounded from below because g and $D_f(\cdot, x)$ are such. Let $\{y^k\}_{k\in\mathbb{N}}$ be a sequence in C such that

$$\lim_{k\to\infty} \Phi(\omega,x, y^k) = \inf\{\Phi(\omega,x, y);\ y \in C\}. \tag{3.1}$$

We have

$$\Phi(\omega,x, y^k) \geq \inf\{g(z); z \in C\} + \omega \cdot \nu_f(x, \|y^k - x\|),$$

for all $k \in \mathbb{N}$. This implies that the sequence $\{y^k\}_{k\in\mathbb{N}}$ is bounded. Indeed, if C is bounded, then boundedness of $\{y^k\}_{k\in\mathbb{N}}$ is obvious. If C is unbounded, then the assumption that $\{y^k\}_{k\in\mathbb{N}}$ is unbounded implies

$$\lim_{k\to\infty} \nu_f(x, \|y^k - x\|) = +\infty$$

because of Proposition 2.1 in [11]. Hence,

$$\lim_{k\to\infty} \Phi(\omega, x, y^k) \geq \inf\{g(z); z \in C\} + \omega \cdot \lim_{k\to\infty} \nu_f(x, \|y^k - x\|) = +\infty,$$

and this contradicts the boundedness from below of the function $\Phi(\omega,x,\cdot)$ because of (3.1). Since B is reflexive, the bounded sequence $\{y^k\}_{k\in\mathbb{N}}$ has a weakly convergent subsequence $\{y^{j_k}\}_{k\in\mathbb{N}}$. Let y^* be the weak limit of this subsequence. The convex set C is closed and, consequently, weakly closed. This shows that $y^* \in C$. Taking into account that the function $\Phi(\omega, x, \cdot)$ is convex and, thus, lower semicontinuous on C, we deduce

$$\Phi(\omega, x, y^*) \leq \liminf_{k\to\infty} \Phi(\omega, x, y^{j_k}) = \inf\{\Phi(\omega, x, y);\ y \in C\}.$$

This implies that y^* is a minimizer of $\Phi(\omega, x, \cdot)$ and the proof is complete.□

The next result was repeatedly proven in slightly different contexts than ours (see [21], [9]). It establishes an essential relationship between the operators T_ω and the function g. Also, by explicitly writing the right hand side of the relation (3.2) below one deduces the firm nonexpansivity with respect to f [see (1.13)] of the operators T_ω.

Lemma 2. *If $z \in C$ is an optimal solution of the problem* (1.10)*, then, for any* $x \in C$,

$$D_f(z,x) - D_f(z,T_\omega(x)) - D_f(T_\omega(x),x) \geq \frac{1}{\omega}[g(T_\omega(x)) - g(z)] \geq 0, \qquad (3.2)$$

whenever $\omega > 0$.

Proof. Note that

$$D_f(z,x) - D_f(z,T_\omega(x)) - D_f(T_\omega(x),x) =$$

$$\langle f'(T_\omega(x)) - f'(x), z - T_\omega(x)\rangle.$$

Since the function $\Phi(\omega,x,\cdot)$ is convex and $T_\omega(x)$ is a minimizer of it we have that

$$0 \in \partial\Phi(\omega, T_\omega(x), \cdot) = \partial g(T_\omega(x)) + \omega \cdot [f'(T_\omega(x)) - f'(x)],$$

where the last equality follows from [26, Theorem 3.6] and Proposition 2.2(i) in [3]. In other words,

$$\omega \cdot [f'(x) - f'(T_\omega(x))] \in \partial g(T_\omega(x)),$$

that is, for some $u \in \partial g(T_\omega(x))$, we have

$$D_f(z,x) - D_f(z,T_\omega(x)) - D_f(T_\omega(x),x) = \frac{1}{\omega}u(T_\omega(x) - z).$$

Convexity of g implies that the right hand side of the last equation is at least equal to $(1/\omega)[g(T_\omega(x)) - g(z)]$. Hence,

$$D_f(z,x) - D_f(z,T_\omega(x)) - D_f(T_\omega(x),x) \geq \frac{1}{\omega}[g(T_\omega(x)) - g(z)] \geq 0,$$

where the last inequality results from the fact that z is a solution of (1.1).□

Now we are in position to prove Theorem 1(I). Precisely, we have the following result.

Lemma 3. *For any $x \in C$, the function $\omega \to T_\omega(x)$ is integrable with respect to any probability measure λ_k on any real interval $(0,b]$ and the integral*

$$\boldsymbol{T}_k(x) = \int_0^b T_\omega(x)d\lambda_k(\omega) \qquad (3.3)$$

belongs to C. Moreover, the sequence $\{x^k\}_{k\in\mathbb{N}}$ defined by (1.12) *exists, is contained in C, the sequence $\{g(x^k)\}_{k\in\mathbb{N}}$ converges and the inequality* (2.2) *holds.*

Proof. According to Lemma 2, we have

$$\nu_f(x, \|T_\omega(x) - x\|) \leq D_f(T_\omega(x), x) \leq D_f(z,x),$$

for all $\omega > 0$. Hence, the function $\omega \to \nu_f(x, \|T_\omega(x) - x\|)$ is bounded and, in virtue of

Proposition 2.1(iv) in [11], this cannot happen unless the set $\{\|T_\omega(x) - x\|;\ \omega > 0)\}$ is bounded. Consequently, the function $\omega \to T_\omega(x)$ is bounded on $(0, b]$. This function is measurable too because $(\omega, y) \to \Phi(\omega, x, y)$ is a Carathèodory function and, therefore, Theorem 8.2.11 in [5] applies. Being bounded and measurable, the function $\omega \to T_\omega(x)$ is integrable on $(0, b]$. According to the definition of the (Bochner) integral, the vector $\boldsymbol{T}_k(x)$ is the limit of a sequence of integrals of step functions which converge almost everywhere to the integrand and such that each step function of the sequence has the form

$$s(\omega) = \sum_{i=1}^{m} 1_{A_i}(\omega) \cdot T_{\omega_i}(x),$$

where $A_1, ..., A_m$ is an $\mathcal{A}_k$-measurable partition of $(0, b]$ and $\omega_i \in A_i$, for each $i \in \{1, ..., m\}$. The integral with respect to λ_k of the step function s is a convex combination of the elements $T_{\omega_i}(x)$. In other words, $\boldsymbol{T}_k(x)$ is the limit of a sequence of convex combinations of elements of C. Since C is convex and closed, it follows that $\boldsymbol{T}_k(x) \in C$. Notice that, for each nonnegative integer k, we have

$$x^{k+1} = \boldsymbol{T}_k(x^k).$$

Therefore, the sequence $\{x^k\}_{k\in\mathbb{N}}$ exists and is contained in C. According to (1.11), we have

$$g(x) - g(T_\omega(x)) \geq \omega D_f(T_\omega(x), x),$$

for all $\omega > 0$. Since the functions g and $D_f(\cdot, x)$ are convex and lower semicontinuous, integrating the last relation and taking into account Jensen's inequality, we obtain

$$g(x) - g(\boldsymbol{T}_k(x)) \geq \int_0^b \omega D_f(T_\omega(x), x) d\lambda_k(\omega) \geq 0,$$

which implies that the sequence $\{g(x^k)\}_{k\in\mathbb{N}}$ is nonincreasing and (2.2) holds. Combining these facts with the boundedness from below of the function g it results that $\{g(x^k)\}_{k\in\mathbb{N}}$ is convergent.□

The next result establishes an inequality which is essential for the analysis of the proximal point method and also shows the firm nonexpansivity of the operators $\boldsymbol{T}_k$ with respect to f.

Lemma 4. *If $z \in C$ is an optimal solution of* (1.10) *and if the function $D_f(z, \cdot)$ is convex, then, for any $k \in \mathbb{N}$,*

$$D_f(z, x^k) - D_f(z, x^{k+1}) - D_f(x^{k+1}, x^k) \geq \frac{g(x^{k+1}) - g(z)}{b} \geq 0. \tag{3.4}$$

Proof. According to (3.2), for each nonnegative integer k, we have

$$D_f(z, x^k) - D_f(z, T_\omega(x^k)) - D_f(T_\omega(x^k), x^k) \geq \frac{g(T_\omega(x^k)) - g(z)}{\omega} \geq 0, \tag{3.5}$$

for all $\omega > 0$. Note that each of the functions $\omega \to D_f(T_\omega(x^k), x^k)$, $\omega \to D_f(z, T_\omega(x^k))$

is measurable, because of Lemma 3 and of the continuity of the functions $D_f(\cdot, x^k)$ and $D_f(z, \cdot)$. The function $\omega \to g(T_\omega(x^k))$ is measurable because of Lemma 3 and of the lower semicontinuity of g (which implies that the level sets of g are open, hence, measurable sets with respect to the Borel σ-algebra induced by the metric topology of C). Since the functions $\omega \to D_f(T_\omega(x^k), x^k)$, $\omega \to D_f(z, T_\omega(x^k))$ and $\omega \to g(T_\omega(x^k))$ are also bounded as follows from (3.5), it results that they are integrable on $(0, b]$ with respect to any probability measure λ_k. Integrating the inequality (3.5) with respect to λ_k and taking into account Jensen's inequality, we get

$$D_f(z, x^k) - D_f(z, x^{k+1}) - D_f(x^{k+1}, x^k) =$$

$$D_f(z, x^k) - D_f\left[z, \boldsymbol{T}_k(x^k)\right] - D_f\left[\boldsymbol{T}_k(x^k), x^k\right] \geq$$

$$D_f(z, x^k) - \int_0^b D_f(z, T_\omega(x^k)) d\lambda_k(\omega) - \int_0^b D_f(T_\omega(x^k), x^k) d\lambda_k(\omega) \geq \tag{3.6}$$

$$\frac{1}{b}\left[\int_0^b g(T_\omega(x)) d\lambda_k(\omega) - g(x^k)\right] \geq \frac{g(\boldsymbol{T}_k(x^k)) - g(z)}{b} =$$

$$\frac{g(x^{k+1}) - g(z)}{b} \geq 0,$$

for all $k \in \mathbb{N}$, where $\boldsymbol{T}_k$ represents the operator defined by (3.3).□

Now we are in position to prove Theorem 1(II).

Lemma 5. *If the function $D_f(z, \cdot)$ is convex whenever z is an optimal solution of the problem* (1.10)*, then the sequence $\{g(x^k)\}_{k\in\mathbb{N}}$ converges nonincreasingly to the minimal value of g over C. In this case, the sequence $\{x^k\}_{k\in\mathbb{N}}$ is bounded, has weak accumulation points and all these points are optimal solutions of the problem* (1.10)*. Moreover, the equation* (2.4) *holds and, if f is uniformly convex, then* (2.5) *is also satisfied.*

Proof. Let z be an optimal solution of the problem (1.10). According to (3.4), the sequence $\{D_f(z, x^k)\}_{k\in\mathbb{N}}$ is nonincreasing, hence, convergent. This implies that the entire sequence $\{x^k\}_{k\in\mathbb{N}}$ is included in the bounded set $R_\alpha^f(z)$, where $\alpha = D_f(z, x^0)$. Consequently, $\{x^k\}_{k\in\mathbb{N}}$ is bounded (cf. condition (C)). The sequence $\{x^k\}_{k\in\mathbb{N}}$ has weak accumulation points because B is reflexive. All weak accumulation points of $\{x^k\}_{k\in\mathbb{N}}$ are contained in C because closed convex sets are weakly closed. Observe that (3.4) also implies

$$D_f(z, x^k) - D_f(z, x^{k+1}) \geq D_f(x^{k+1}, x^k) \geq 0, \tag{3.7}$$

where the left hand side converges to zero as $k \to \infty$ because $\{D_f(z, x^k)\}_{k\in\mathbb{N}}$ is convergent. This shows that the left hand side of (3.4) converges to zero as $k \to \infty$.

Again according to (3.4), we have

$$D_f(z, x^k) - D_f(z, x^{k+1}) - D_f(x^{k+1}, x^k) \geq \frac{g(x^{k+1}) - g(z)}{b} \geq 0. \tag{3.8}$$

By letting $k \to \infty$ in this relation, we deduce (2.3).

Let x^* be a weak accumulation point of the sequence $\{x^k\}_{k\in\mathbb{N}}$. Assume that $\{x^{j_k}\}_{k\in\mathbb{N}}$ is a subsequence of $\{x^k\}_{k\in\mathbb{N}}$ which converges weakly to x^*. Since g is weakly lower semicontinuous we have

$$g(x^*) \leq \lim_{k\to\infty} g(x^{j_k}) = g(z).$$

This implies $g(z) = g(x^*)$, because z is an optimal solution of (1.10) and $x^* \in C$. Thus, any weak accumulation point of the sequence $\{x^k\}_{k\in\mathbb{N}}$ is an optimal solution of (1.10).

Observe that, since $\{D_f(z, x^k)\}_{k\in\mathbb{N}}$ is convergent, (3.7) implies (2.4). According to [11], for any $k \in \mathbb{N}$,

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

Using this and the already proven formula (2.4), we deduce that

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

If the function f is uniformly convex, then the modulus of convexity μ_f is strictly increasing and continuous from the right at zero (cf. [3]). Therefore, the last equality cannot hold unless (2.5) is satisfied.□

From now and till the end of this section we work under the assumption of the hypothesis of Theorem 2 which essentially says that the operators $\boldsymbol{T}_k$ are firmly nonexpansive with respect to f. It was shown in [12, Section 6.1] that the condition (b) of Theorem 2 implies condition (a) in the same theorem. Thus, by proving the next result we implicitly show that when the condition (b) holds the sequence $\{x^k\}_{k\in\mathbb{N}}$ converges weakly to an optimal solution of the problem (1.10). If condition (d) or condition (e) of Theorem 2 is satisfied, then weak convergence and strong convergence of sequences in C are equivalent. Thus, the fact that the Fréchet derivative f' is continuous on $\mathcal{D}^\circ$ (cf. [26, Corollary, p.20]), combined with condition (C), shows that, if (d) or (e) holds, then (c) is also satisfied. Hence, the following two results also prove that each of the conditions (d) and (e) implies strong convergence of the sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by the proximal point method (1.12).

Lemma 6. *If the function f satisfies the condition (a) of Theorem 2, then any sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by the proximal point method* (1.12) *converges weakly to an optimal solution of the problem* (1.10).

Proof. According to Lemma 5, it is sufficient to show that the sequence $\{x^k\}_{k\in\mathbb{N}}$ has an unique weak accumulation point. Observe that, according to (3.4), for any optimal solution z of the problem (1.10), the sequence $\{D_f(z, x^k)\}_{k\in\mathbb{N}}$ is nonincreasing, hence,

convergent. Assume, by contradiction, that x' and x'' are two different weak accumulation points of the sequence $\{x^k\}_{k\in\mathbb{N}}$. Using Lemma 5 again, we deduce that x' and x'' are optimal solutions of (1.10) and, thus, the sequences $\{D_f(x', x^k)\}_{k\in\mathbb{N}}$ and $\{D_f(x'', x^k)\}_{k\in\mathbb{N}}$ converge. Let $\{x^{i_k}\}_{k\in\mathbb{N}}$ and $\{x^{j_k}\}_{k\in\mathbb{N}}$ be subsequences of $\{x^k\}_{k\in\mathbb{N}}$ which converge weakly to x' and x'', respectively. Observe that

$$\left|\langle f'(x^{i_k}) - f'(x^{j_k}), x' - x''\rangle\right| =$$

$$\left|(D_f(x', x^{i_k}) - D_f(x', x^{j_k})) + (D_f(x'', x^{j_k}) - D_f(x'', x^{i_k}))\right| \leq$$

$$\left|D_f(x', x^{i_k}) - D_f(x', x^{j_k})\right| + \left|D_f(x'', x^{j_k}) - D_f(x'', x^{i_k})\right|.$$

Letting $k \to \infty$ in this inequality one obtains that

$$\lim_{k\to\infty}\left[\langle f'(x^{i_k}) - f'(x^{j_k}), x' - x''\rangle\right] = 0.$$

According to (2.6), this implies $x' = x''$, a contradiction.□

Now we are going to show that condition (c) of Theorem 2 is also sufficient for ensuring weak convergence of the sequence $\{x^k\}_{k\in\mathbb{N}}$ to optimal solutions of the problem (1.10). Also, this condition together with the uniform convexity of f (i.e. the requirement (f) of Theorem 2), implies strong convergence. To this end, define the function $r : \mathcal{D} \to \mathbb{R}_+$ by

$$r(z) = \limsup_{k\to\infty} D_f(z, x^k). \tag{3.9}$$

This function, introduced in [27], is convex and lower semicontinuous on $\mathcal{D}^\circ$ because $D_f(\cdot, x^k)$ is convex and continuous on that set, for each $k \in \mathbb{N}$. Also, for any $z \in C$ which is an optimal solution of the problem (1.10), the sequence $\{D(z, x^k)\}_{k\in\mathbb{N}}$ is nonincreasing (see (3.4)) and, thus,

$$r(z) = \lim_{k\to\infty} D_f(z, x^k). \tag{3.10}$$

With these facts in mind we prove the following result.

Lemma 7. *If the condition (c) of Theorem 2 is satisfied, then the sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by the proximal point method* (1.12) *converges weakly to an optimal solution of the problem* (1.10). *In this case, if the function f is uniformly convex, then the sequence $\{x^k\}_{k\in\mathbb{N}}$ converges strongly.*

Proof. According to (3.9), if $x \in \mathcal{D}$, then

$$r(x) = \limsup_{k\to\infty}\left[f(x) - f(x^k) - \langle f'(x^k), x - x^k\rangle\right] =$$

$$f(x) - \liminf_{k\to\infty}\left[f(x^k) - \langle f'(x^k), x - x^k\rangle\right],$$

where the function

$$x \to \liminf_{k\to\infty}\left[f(x^k) - \langle f'(x^k), x - x^k\rangle\right]$$

is concave. This implies that r is strictly convex on $\mathcal{D}^\circ$, because f is strictly convex on $\mathcal{D}^\circ$ as being totally convex (cf. [11, Proposition 3.1]). Any weak accumulation point of $\{x^k\}_{k\in\mathbb{N}}$ is a minimizer of the function r over the set $\mathcal{D}$. Indeed, if x^* is a weak accumulation point of $\{x^k\}_{k\in\mathbb{N}}$ and if $\{x^{i_k}\}_{k\in\mathbb{N}}$ is a subsequence of $\{x^k\}_{k\in\mathbb{N}}$ which converges weakly to x^*, then, according to (3.10),

$$0 \le r(x^*) = \lim_{k\to\infty} D_f(x^*, x^{i_k}) \le$$

$$f(x^*) - \liminf_{k\to\infty} f(x^{i_k}) + \limsup_{k\to\infty}\langle f'(x^{i_k}), x^{i_k} - x^*\rangle \le$$

$$\limsup_{k\to\infty}\langle f'(x^{i_k}), x^{i_k} - x^*\rangle,$$

where the last inequality holds because f is weakly lower semicontinuous. Condition (c) of Theorem 2 implies that the last limit is nonpositive. Hence, $r(x^*) = 0$, that is, x^* is a minimizer of r over $\mathcal{D}$. The function r cannot have more than one minimizer over the set $\mathcal{D}$ because it is strictly convex. Therefore, the sequence $\{x^k\}_{k\in\mathbb{N}}$ has exactly one weak accumulation point, i.e., the sequence $\{x^k\}_{k\in\mathbb{N}}$ converges weakly to x^*. Now, suppose that the function f is uniformly convex. Then, for any nonnegative integer k,

$$0 \le \mu_f(\|x^* - x^k\|) \le D_f(x^*, x^k).$$

Hence,

$$\lim_{k\to\infty} \mu_f(\|x^* - x^k\|) = 0.$$

Since f is uniformly convex, its modulus of convexity μ_f is strictly increasing and continuous from the right at zero (cf. [3]). Hence, the last equality cannot hold unless

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

This completes the proof.□

4. COMMENTS AND APPLICATIONS

In this section we show that the class of Banach spaces in which the proximal point method (1.12) and the results guaranteeing its convergence, Theorems 1, 2 and Corollary 1 above, can be applied includes all smooth and uniformly convex Banach spaces. Also, we show that, in some particular instances, application of the general proximal point method described in this paper is more convenient from a computational point of view than the application of other versions of the proximal point method discussed in literature.

The practical applicability of Theorems 1 and 2 and of Corollary 1 essentially depends on the possibility of identifying totally convex functions on the Banach space B in which the optimization problem (1.10) is formulated. As noted in [11] uniformly convex functions are totally convex. This already provides a significant pool of functions with which the results mentioned above are implementable. The archetypal example in this sense is the function $f(x) = \|x\|^2$ when B is a Hilbert spaces. Examples of applications of the proximal point algorithm involving this function in $B = \mathbb{R}^n$ can be found in [21] and [15]. The classical proximal point method, invented for the resolution of optimization problems in Hilbert spaces (see [28]) is, in fact, the particular version of (1.12) in which $f(x) = \|x\|^2$. Totally convex functions which are not uniformly convex were also used before in some applications of the proximal point method in $B = \mathbb{R}^n$ (see [16], [15] and [21]) although the connection between the total convexity and the behavior of the proximal point method remained unobserved till now. This is typically the case of the negentropy defined by

$$f(x) = \begin{cases} \sum_{i=1}^{n} x_i \cdot log x_i \quad if \ x_1, ..., x_n \geq 0, \\ \\ + \infty, \qquad otherwise, \end{cases}$$

with the usual convention that $0 \cdot \log 0 = 0$. It has been shown in [10] that this function is totally (but not uniformly) convex.

Our results extend the applicability of the proximal point method in spaces which are neither Hilbertian nor finite dimensional. This is typically the case of spaces like $\mathcal{L}^p$ and l^p with $p \in (1, + \infty)$, $p \neq 2$. It has been proved in [11] (and using rather different arguments in [23]) that, in these spaces, the function $f(x) = \|x\|^p$ is totally convex (although it is uniformly convex only if $p \geq 2$). This result was later extended in [22] where it was shown that in the spaces $\mathcal{L}^p$ and l^p the function $f(x) = \|x\|^s$ with $s > 1$ is totally convex. We show next that this result holds in a more general context when $s \geq 2$.

Proposition 1. *If B is a smooth uniformly convex Banach space, then, for each $s \in [2, + \infty)$, the function $f(x) = \|x\|^s$ is totally convex.*

Proof. First we consider the case $s = 2$. For $f(x) = \|x\|^2$ we have

$$D_f(x, z) = \|x\|^2 - \|z\|^2 - 2\langle J(z), x - z \rangle,$$

where $J : B \to B^*$ is the duality mapping of B. Since $\langle J(z), z \rangle = \|z\|^2$, we have

$$D_f(x, z) = \|x\|^2 + \|z\|^2 - 2\langle J(z), x \rangle. \tag{4.1}$$

From [1, Theorem 7.5] we have that

$$\|x\|^2 + \|z\|^2 - 2\langle J(z), x \rangle \geq 8[\eta(x, z)]^2 \delta_B \left(\frac{\|x - z\|}{4\eta(x, z)} \right), \tag{4.2}$$

where δ_B is the modulus of convexity of the space B (see [8]) and

$$\eta(x,z) := \left[\frac{1}{2}\left(\|x\|^2 + \|z\|^2\right)\right]^{1/2}.$$

Letting $\|x - z\| = t$ and combining (4.1) and (4.2) we get

$$D_f(x,z) \geq 4\left(\|x\|^2 + \|z\|^2\right)\delta_B\left(\frac{t}{4\eta(x,z)}\right).$$

Taking the infimum with respect to all $x \in B$ such that $\|x - z\| = t$ on both sides of this inequality we obtain

$$\nu_f(z,t) \geq 4\|z\|^2\delta_B\left[\frac{\sqrt{2}t}{4(t^2 + 2\|z\|^2 + 2t\|z\|)^{1/2}}\right],$$

because $t = \|x - z\| \geq \|x\| - \|z\|$ and because the function δ_B is strictly increasing on (0,1). Since B is uniformly convex the function δ_B is positive on (0,2] and, thus, the quantity on the right hand side in the last inequality is positive whenever $t > 0$. This shows that $\nu_f(z,t) > 0$ for all $t > 0$. Hence, the function $f(x) = \|x\|^2$ is totally convex.

Define $g(x) = \|x\|^s$ for some $s > 2$. Let $\varphi(t) = t^{s/2}$ and note that $g = \varphi \circ f$, where $f(x) = \|x\|^2$. From [23, Proposition 3] we have

$$D_g(x,z) - D_\varphi(f(x), f(z)) + \varphi'(f(x))D_f(x,z) \geq \varphi'(f(z))D_f(x,z).$$

This implies

$$D_g(x,z) \geq (s/2)\|z\|^{s-2}D_f(x,z).$$

Taking on both sides of this inequality the infimum over the set of all $x \in B$ such that $\|x - z\| = t$ we obtain

$$\nu_g(z,t) \geq (s/2)\|z\|^{s-2}\nu_f(z,t).$$

Since, as shown above, the function f is totally convex, it follows that $\nu_f(z,t) > 0$ and, therefore, $\nu_g(z,t) > 0$, whenever $t > 0$. This completes the proof.□

We claimed above that our convergence analysis enhances the applicability of the proximal point method not only by extrapolating it to spaces in which it was not known before that it may work, but also by allowing much freedom for the choice of the function f. This freedom is exploited by fitting the function f to the nature of the given problem (1.10) in such a way that computation of the values $T_\omega(x^k)$ (i.e., solving the inclusion (2.10)) may be quite easy. In order to illustrate this idea consider the following example. Let $B = l^p$ with $p \in (1, +\infty)$, $p \neq 2$, and assume that in (1.10) we have $C = B$ and that g is a function with separated variables, that is, for any $x = (x_1, x_2, ...) \in l^p$,

$$g(x) = \sum_{i=1}^{\infty} g_j(x_j), \tag{4.3}$$

where each $g_j : \mathbb{R} \to \mathbb{R}$ is convex and differentiable. In these circumstances, the inclusion (2.10) which should be solved in order to determine x^{k+1} has the particular form (1.7), no matter which function f is chosen for running the proximal point method (1.12). Letting $f(x) = \|x\|^p$ the equation (1.7) is reduced to the system of equations

$$g_j'(x_j) + \omega_k p|x_j|^{p-2}x_j = \omega_k p|x_j^k|^{p-2}x_j^k, \tag{4.4}$$

for $j = 1, 2, \ldots\ldots$. The system of equations is *uncoupled*, that is, each equation can be solved separately and has as unique solution the number x_j^{k+1}. This contrasts to the situation which occurs when one tries to apply the proximal point method with the function $f(x) = \|x\|^2$ as suggested in [2]. The algorithm will be weakly convergent in this case too. However, the equation (1.7) corresponding to this new function f is equivalent to the system of *coupled* equations

$$g_j'(x_j) + 2\omega_k\|x\|^{2-p}|x_j|^{p-2}x_j = 2\omega_k\|x^k\|^{2-p}|x_j^k|^{p-2}x_j^k, \tag{4.5}$$

for $j = 1, 2, \ldots$. The right hand side terms in both equations (4.4) and (4.5) are computable. Nevertheless, the presence in (4.5) of the quantity $\|x\|^{2-p}$ which contains infinitely many variables makes the advantage of choosing $f(x) = \|x\|^p$ instead of $f(x) = \|x\|^2$ self-evident.

Theorem 1 and Corollary 1 in combination with Proposition 1 show that convergence of the sequence $\{g(x^k)\}_{k\in\mathbb{N}}$ towards the minimal value of the optimization problem (1.10) when $\{x^k\}_{k\in\mathbb{N}}$ is generated by the proximal point algorithm can be guaranteed in nonrestrictive conditions. Comparing the hypothesis of these results with the hypothesis of Theorem 2 one can see that (weak or strong) convergence of the sequence $\{x^k\}_{k\in\mathbb{N}}$ towards an optimal solution of (1.10) is a much more demanding property. Condition (a) of Theorem 2 is the least restrictive requirement in this sense. This condition, first studied in [9] in the case of Hilbert spaces, seems to be less restrictive than conditions (b) and (c) of the same Theorem 2 although we have no examples showing that these conditions are not equivalent. Condition (b) is satisfied whenever B is a Hilbert space and $f(x) = \|x\|^2$. Also, it holds when $B = l^p$ and $f(x) = \|x\|^p$ with $p > 1$ (cf. [8, Proposition 8.2, p. 111]). It is an interesting open question whether the conditions in Theorem 2 can be relaxed. In this case, the next naturally occurring question will be whether the convergence of the sequence $\{x^k\}_{k\in\mathbb{N}}$ is still superlinear as happens in some particular instances of the proximal point method surveyed in [21] and [24].

We noted in Section 1 that there are problems of practical interest for the resolution of which the sequential proximal point method with appropriately chosen numbers ω_k behaves better than its parallel counterpart. This is the case of unrestricted optimization problems (1.10) in which g is differentiable but the equation $g'(x) = 0$ is numerically ill-conditioned. Then, the corresponding optimization problems (1.11), when

rewritten in their equivalent form (1.7), may be also ill-conditioned for ω situated in the neighborhood of 0. In order to avoid solving problems (1.7) with small numbers ω, one has to use in (1.12) probability measures λ_k having their support contained in some intervals $[a_k, b]$ with $0 < a_k < b$. The proposition below implies that, in such a case, the sequential proximal point method with $\omega_k = a_k$ ensures faster convergence of $\{g(x^k)\}_{k\in\mathbb{N}}$. Its proof involves some standard arguments concerning the monotonicity of the functions $\omega \to g(T_\omega(x^k))$ and $\omega \to D_f(T_\omega(x^k), x^k)$.

Proposition 2. *If $C = B$, if the function g is Gâteaux differentiable on B, and if there exists a number $a \in (0, b]$ such that $\lambda_k([a, b]) = 1$, then, for any $\tau \in (0, a)$, we have*

$$g(T_\tau(x^k)) \le g(T_a(x^k)) \le g\Big[\int_0^b T_\omega(x^k) d\lambda_k(\omega)\Big]. \tag{4.6}$$

Proof: For any $x \in \mathcal{D}$ denote $h(x) = D_f(x, x^k)$. Let α and β be positive real numbers such that $\alpha < \beta$. According to (1.11), we have

$$g(T_\alpha(x^k)) + \alpha h(T_\alpha(x^k)) \le g(T_\beta(x^k)) + \alpha h(T_\beta(x^k)) \tag{4.7}$$

and

$$g(T_\beta(x^k)) + \alpha h(T_\beta(x^k)) \le g(T_\alpha(x^k)) + \beta h(T_\alpha(x^k)). \tag{4.8}$$

Adding these two inequalities we get

$$(\beta - \alpha)\Big[h(T_\alpha(x^k)) - h(T_\beta(x^k))\Big] \ge 0,$$

which shows that the function $\omega \to h(T_\omega(x^k))$ is nonincreasing on $(0, +\infty)$. Multiplying the inequality (4.7) by $1/\alpha$, and multiplying the inequality (4.8) by $1/\beta$, and adding the resulting relations we get

$$(\frac{1}{\alpha} - \frac{1}{\beta})\Big[g(T_\beta(x^k)) - g(T_\alpha(x^k))\Big] \ge 0,$$

which shows that the function $\omega \to g(T_\omega(x^k))$ is nondecreasing on $(0, +\infty)$. This proves the first inequality in (4.6). For proving the second inequality in (4.6), define the function φ:$[0, 1] \to \mathbb{R}$ by

$$\varphi(t) = g((1 - t)x' + tx^*),$$

where $x^* := \int_0^b T_\omega(x^k) d\lambda_k(\omega)$ and $x' := T_a(x^k)$. Since g is convex, the function φ is convex too and we have

$$g(x^*) - g(x') = \varphi(1) - \varphi(0) \ge \varphi'(0) = \langle g'(x'), x^* - x'\rangle.$$

Taking into account (1.11) we deduce that $g'(x') = -ah'(x')$. Hence,

$$g(x^*) - g(x') \ge -a\langle h'(x'), x^* - x'\rangle$$

$$= -a\int_0^b \langle h'(x'),\, T_\omega(x^k) - x'\rangle d\lambda_k(\omega).$$

because λ_k is a probability measure and because $h'(x')$ is linear and continuous. Since the function h is convex on B and the function $\omega \to h(T_\omega(x^k))$ is nonincrasing, we also have

$$\langle h'(x'), T_\omega(x^k) - x'\rangle \leq h(T_\omega(x^k)) - h(x') = h(T_\omega(x^k)) - h(T_a(x^k)) \leq 0,$$

whenever $\omega \geq a$. Using the fact that $\lambda_k((0,a)) = 1 - \lambda_k([a,b]) = 0$ we get

$$g(x^*) - g(x') \geq -a\int_0^b \langle h'(x'),\, T_\omega(x^k) - x'\rangle d\lambda_k(\omega)$$

$$= -a\int_a^b \langle h'(x'),\, T_\omega(x^k) - x'\rangle d\lambda_k(\omega)$$

$$\geq -a\int_a^b (h(T_\omega(x^k)) - h(T_a(x^k)))d\lambda_k(\omega) \geq 0,$$

and this completes the proof.□

In practical applications of the proximal point method the iterates are determined numerically and it often happens that instead of the precise vectors $T_\omega(x^k)$ one has to use approximations a_ω of them. It may also happen that the vectors $T_\omega(x^k)$ are close to the boundary of C and the approximations a_ω are unfeasible. Accumulated errors may lead to approximate iterates lying outside the interior of $\mathcal{D}$ or to points a_ω at which the function g is not defined. In some circumstances it is possible to avoid such difficulties by using a parallel proximal point algorithm with probability measures λ_k concentrated at two points provided that the chosen function f satisfies the requirements of Theorem 1. This is the case when in (1.10) the function g is Lipschitz of some constant K and $C = \{x \in B;\ h(x) \leq 0\}$, where $h : B \to \mathbb{R}$ is uniformly convex and continuous. Note that

$$h(\frac{1}{2}(a_\alpha + a_\beta)) \leq \frac{1}{2}(h(a_\alpha) + h(a_\beta)) - \mu_h(\|a_\alpha - a_\beta\|), \tag{4.9}$$

where μ_h is the modulus of convexity of h. Since h is continuous, if the maximal error ϵ involved in computing vectors $T_\omega(x^k)$ is small, then the sum $\frac{1}{2}(h(a_\alpha) + h(a_\beta))$ is close to zero even if one or both of its terms are positive. The function $\omega \to g(T_\omega(x^k))$ is monotonically increasing (see the proof of Proposition 3), and since

$$\frac{1}{K}|g(T_\beta(x^k)) - g(T_\alpha(x^k))| \leq \|T_\beta(x^k) - T_\alpha(x^k)\|, \tag{4.10}$$

the value of $\|a_\alpha - a_\beta\|$ grows with the difference $\beta - \alpha$ because $\|T_\beta(x^k) - T_\alpha(x^k)\|$ does so. This indicates that, due to the strict monotonicity of μ_h, it may be possible to chose the numbers α and β in such a way that, for a sufficiently small error ϵ, the right hand side of (4.9) is negative. This will ensure that the new iterate $a^{k+1} := \frac{1}{2}(a_\alpha + a_\beta)$

is feasible (even if a_α or a_β or both are not). Clearly, $\|\frac{1}{2}(T_\beta(x^k) + T_\alpha(x^k)) - a^{k+1}\| \leq \epsilon$ and this implies that, in the procedure (1.12) with $\lambda_k =$ the probability measure concentrated at α and β, we have

$$|g(x^{k+1}) - g(a^{k+1})| \leq K\epsilon$$

showing that $\{g(a^{k+1})\}_{k\in\mathbb{N}}$ is a sequence approximating the minimal value of the problem (1.10) because Theorem 1 guarantees convergence of $\{g(x^{k+1})\}_{k\in\mathbb{N}}$ towards this value.

ACKNOWLEDGMENTS

This work was done during Dan Butnariu's visit to the Department of Mathematics of the University of Texas at Arlington. The authors are grateful to Professors Y. Censor, I. Dragan and S. Reich for comments which helped us improve an earlier version of this material.

REFERENCES

[1] Alber, Ya., Metric and generalized projections in Banach spaces: properties and applications, in: *Theory and Applications of Nonlinear Operators of Monotone and Accretive Type*, Edited by A. Kartsatos, *Marcel Dekker*, New York, 1996, 15-50.

[2] Alber, Ya., Burachik, R.S. and Iusem, A.N., A proximal point method for nonsmooth convex optimization problems in Banach spaces, *to be published.*

[3] Alber, Ya. and Butnariu, D., Convergence of Bregman-projection methods for solving convex feasibility problems in reflexive Banach spaces, *J. Optim. Theory Appl.*, Vol. 92, 1, 1997, 33-61.

[4] Araujo, A., The non-existence of smooth demands in general Banach spaces, *J. Math. Economics,* 17, 1988, 309-319.

[5] Aubin, J.-P. and Frankowska, H., Set Valued Analysis, *Birkhäuser*, Basel, 1990.

[6] Bregman, L.M., The relaxation method for finding common points of convex sets and its application to the solution of convex programming, *USSR Comp. Math. and Math. Phys.,* 7, 1967, 200-217.

[7] Brezis, H., Analyse Fonctionelle -- Théorie et Applications, *Masson*, Paris, 1987.

[8] Browder, F.E., Nonlinear Operators and Nonlinear Equations of Evolution in Banach Spaces, *Proceedings of Symposia in Pure Mathematics, American Mathematical Society,* 18, 2, 1976.

[9] Burachik, R.S., Generalized proximal point methods for the variational inequality problem, Ph.D. Thesis, *Instituto de Mathematica Pura e Aplicada,* Rio de Janeiro, 1995.

[28] Rockafellar, R.T., Monotone operators and the proximal point algorithm, *SIAM Journal Contr. Optim.*, 14, 1976, 877-898.

[29] Vladimirov, A.A., Nesterov, Y.E. and Chekanov, Y.N., Uniformly convex functions (Russian), *Vestnik Moskovkaya Universiteta, Series Mathematika i Kybernetika*, 3, 1978, 12-23.