Total convexity for powers of the norm in uniformly convex Banach spaces

Dan Butnariu* Alfredo N. Iusem† Elena Resmerita‡

February 7, 2000

Abstract

We show that in any uniformly convex Banach space the functions $f(x) = \|x\|^r$ with $r \in (1, \infty)$ are totally convex. Using this fact we establish a formula for determining Bregman projections on closed hyperplanes and half spaces. This leads to a method for solving linear operator equations (e.g., first kind Fredholm and Volterra equations) in spaces which are uniformly convex and smooth.

Keywords: Uniformly convex Banach space, Totally convex function, Duality mapping, Bregman projection.

1 Introduction

Let X be a uniformly convex Banach space, X^* its dual and $f : X \to \mathbb{R}$ a convex continuous function. For $x \in X$, let $\partial f(x)$ be the subdifferential of f at x, i.e.,

$$\partial f(x) = \{\xi \in X^* : f(y) - f(x) \geq \langle \xi, y - x\rangle, forally \in X\}.$$

Following [19], we define $D_f : X \times X \to \mathbb{R}$ as

$$D_f(x, y) = f(x) - f(y) - \inf\{\langle \xi, x - y\rangle : \xi \in \partial f(y)\}. \tag{1}$$

*Department of Mathematics, University of Haifa, 31905 Haifa, Israel

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

‡Department of Mathematics, University of Haifa, 31905 Haifa, Israel

The function D_f, called the *Bregman distance associated with* f, is always well defined because $\partial f(x)$ is nonempty and bounded, for all $x \in X$ (see, e.g., [22]), so that the infimum in (1) cannot be $-\infty$. It is easy to check that $D_f(x, y) \geq 0$ and that $D_f(x, x) = 0$, for all $x, y \in X$. If f is strictly convex then $D_f(x, y) = 0$ only when $x = y$.

For $t \in (0, \infty)$ and $z \in X$ let

$$U(z, t) = \{x \in X : \|x - z\| = t\}.$$

Following [5], we define $\nu_f : X \times [0, \infty) \to [0, \infty)$ as

$$\nu_f(z, t) = \inf \{D_f(x, z) : x \in U(z, t)\}. \tag{2}$$

and call it the *modulus of convexity of* f. The function f is said to be *totally convex* if $\nu_f(z, t) > 0$, for all $z \in X$ and all $t \in (0, \infty)$. Totally convex functions are strictly convex, but there exist strictly convex functions which are not totally convex (see [5]). Total convexity is a weaker condition than uniform convexity, i.e., uniformly convex functions are totally convex (see [5]). However, there are functions which are totally convex without being uniformly convex (see [7]).

Total convexity turns out to be a key property in the convergence analysis of the proximal point method with generalized distances for minimizing convex functions as well as in a large class of projection type algorithms for solving variational inequality problems. In these methods the role given to the Euclidean distance by Rockafellar [24], J. von Neumann [21] and Cimmino [11], respectively, is taken over by a Bregman distance associated with an auxiliary totally convex function f (cf. [8], [10], [15], [17], [20], [4], [6]). We should mention that in Banach spaces which are not Hilbertian the proximal point method with Bregman distances leads to simpler and easier to compute iterative formulae than the one with the metric distance induced by the norm of the Banach space (see [6]). Also, a pletora of examples shows that projection type algorithms with Bregman distances behave better from a computational point of view than their metric distance counterparts. Total convexity is also required in the convergence analysis of the methods for solving stochastic convex feasibility problems and for finding common fixed points of measurable families of operators in Banach spaces studied in [5] and [7]. It is therefore relevant to identify totally convex functions which may be used as auxiliary functions in the algorithms mentioned above. In finite dimensional spaces there is a large pool of known totally convex functions since, as shown in [7], any strictly convex function with closed domain is totally convex. In Banach spaces of infinite dimension identifying totally convex functions is a challenging problem. This happens because, in an infinite dimensional context, we need to find totally convex functions designed in such a way that specific algorithms like the proximal point method with Bregman distances and/or the projection type algorithms to be effectively and efficiently computable. It was already shown in [5] that the function $f(x) = \|x\|_p^p$ in L^p or ℓ^p is totally convex when $p \in (1, +\infty)$. This result was improved in [18] by showing total convexity of $f(x) = \|x\|_p^r$ with $r > 1$ in L^p or ℓ^p for $p \in (1, +\infty)$. In [6] the total convexity of the functions $f(x) = \|x\|^r$ with $r \geq 2$ was proved in any uniformly convex and uniformly smooth Banach space.

In the current work we show that, in any uniformly convex Banach space, the function $f(x) = \|x\|^r$ with $r > 1$ is totally convex and, necessarily, a Bregman function when X is smooth. This leads to implementability in uniformly convex and smooth Banach spaces not only of the proximal point method discussed in [6], but also to the possibility of numerically solving a large class of stochastic convex feasibility problems as those presented in Section 4.

2 Total convexity of the powers of the norm

In this sections we prove that the function $f(x) = \|x\|^r$ with $r > 1$ is totally convex in any uniformly convex Banach space X. We remark that f is always convex but, unless X is smooth, f is not differentiable and, therefore, ∂f is a point-to-set operator, i.e., it is not necessarily single valued (see, e.g., [14]). In that follows, we denote by $\delta_X : [0, 2] \to [0, 1]$ the modulus of uniform convexity of the space X, that is,

$$\delta_X(t) = \begin{cases} \inf \left\{ 1 - \frac{1}{2} \|x + y\| : \|x\| = 1 = \|y\|, \|x - y\| \geq t \right\}, & ift > 0, \\ \\ 0, & ift = 0. \end{cases}$$

Recall that, X is called *uniformly convex* if $\delta_X(t) > 0$, for all $t > 0$. Also, X is called *smooth* if the function $x \to \|x\|$ is differentiable at each point $x \neq 0$.

Theorem 1. *If X is uniformly convex, then, for each $r \in (1, +\infty)$, the function $f(x) = \|x\|^r$ is totally convex.*

Proof. Define the function $\Phi : X \to \mathcal{P}(X^*)$ by

$$\Phi(x) = \frac{1}{r} \partial f(x).$$

Note that Φ is exactly the duality mapping of weight t^{r-1} on the space X. Therefore, there exists a positive real number K such that the following inequality, established in

[25, Theorem 1], holds for all $r > 1$:

$$\langle \xi - \eta, x - y \rangle \geq K \left\{ \max \left[\|x\|, \|y\| \right] \right\}^r \delta_X \left(\frac{\|x - y\|}{2 \max \left[\|x\|, \|y\| \right]} \right), \tag{3}$$

whenever $x, y \in X$ with $\|x\| + \|y\| \neq 0$, $\xi \in \Phi(x)$ and $\eta \in \Phi(y)$. Fix $z \in X$, $t \in (0, \infty)$, take $x \in U(z, t)$ and let $y = x - z$. Then $\|y\| = t$ and

$$D_f(x, z) = D_f(y + z, z) = \|z + y\|^r - \|z\|^r - r \cdot \inf \left\{ \langle \xi, y \rangle : \xi \in \Phi(z) \right\}, \tag{4}$$

in view of (1) and the definition of Φ. Define $\varphi : [0, \infty) \to [0, \infty)$ as

$$\varphi(\tau) = \frac{\|z + \tau y\|^r}{r}.$$

Then, we have

$$\|z + y\|^r - \|z\|^r = r[\varphi(1) - \varphi(0)]. \tag{5}$$

Our proof is based on the next result:

Claim 1: *If, for each* $\tau \in [0, 1]$, *we choose a point* $\xi(\tau) \in \Phi(z + \tau y)$, *then the following integral exists and*

$$\int_0^1 \langle \xi(\tau), y \rangle \, d\tau = \varphi(1) - \varphi(0). \tag{6}$$

We proceed to establish the claim. To this end, we observe that, if $g : X \to \mathbb{R}$ is a convex continuous function, then the function $\psi : \mathbb{R} \to \mathbb{R}$ defined by

$$\psi(\tau) = g(u + \tau v)$$

for arbitrarily fixed $u, v \in X$ is convex, continuous and has

$$\partial \psi(\tau) = \{ \langle \xi, v \rangle : \xi \in \partial g(u + \tau v) \}, \tag{7}$$

for any $\tau \in \mathbb{R}$ (cf. [13, Theorem 2.3.10]). Therefore, the operator $x \to \partial \psi(x)$ is maximal monotone and, for all $\tau \in \mathbb{R}$, $\partial \psi(\tau) = [\alpha(\tau), \beta(\tau)]$, for some $\alpha(\tau), \beta(\tau) \in \mathbb{R}$ with $\alpha(\tau) \leq \beta(\tau)$. From Rademacher's Theorem [22, p. 11], it follows that the set

$$T = \{ \tau \in \mathbb{R} : \alpha(\tau) < \beta(\tau) \}$$

is countable. Thus, if $\tau \notin T$, then the set $\partial\psi(\tau)$ reduces to a singleton $\{\alpha(\tau)\}$, ψ is differentiable at τ and $\alpha(\tau) = \psi'(\tau)$. Hence, ψ is differentiable outside a zero measure set. By consequence, for all $a, b \in \mathbb{R}$, with $a \leq b$, we have

$$\psi(b) - \psi(a) = \int_a^b \psi'(\tau)\, d\tau = \int_a^b \langle \xi(\tau), v\rangle\, d\tau, \tag{8}$$

for any choice of $\xi(\tau) \in \partial\psi(\tau)$. Now, we apply (8) to the case of $\psi = \varphi$, $g(x) = \|x\|^r/r$, $u = z$ and $v = y$, and we conclude that (6) holds, in view of the definition of Φ. The claim is established.

Now we use Claim 1 in order to complete the proof of the theorem. Fix a selector $\tau \to \xi(\tau)$ of the point-to-set mapping $\tau \to \Phi(z + \tau y)$. Application of Claim 1 combined with (5) implies

$$\|z + y\|^r - \|z\|^r = r \int_0^1 \langle \xi(\tau), y\rangle\, d\tau.$$

Therefore, for any $\eta \in \Phi(z)$, we have

$$\begin{aligned} \|z + y\|^r - \|z\|^r - r\langle \eta, y\rangle &= r\left[\int_0^1 \langle \xi(\tau), y\rangle\, d\tau - \langle \eta, y\rangle\right] \\ &= r\int_0^1 \langle \xi(\tau) - \eta, y\rangle\, d\tau = r\int_0^1 \frac{1}{\tau}\langle \xi(\tau) - \eta, \tau y\rangle\, d\tau. \end{aligned} \tag{9}$$

Since, for any $\tau \in [0, 1]$, $\xi(\tau) \in \Phi(z + \tau y)$, and $\tau y = (z + \tau y) - z$, we conclude from (9) and (3) that

$$\begin{aligned} &\|z + y\|^r - \|z\|^r - r\langle \eta, y\rangle \\ \geq\ & rK \int_0^1 \frac{\{\max[\|z + \tau y\|, \|z\|]\}^r}{\tau} \delta_X\left(\frac{\|\tau y\|}{2\max[\|z + \tau y\|, \|z\|]}\right) d\tau, \end{aligned}$$

for all $\eta \in \Phi(z)$, where the last integral exists because δ_X is nondecreasing on [0,2]. Taking infimum over $\eta \in \Phi(z)$ in the left hand side of the last inequality and using (4), we get

$$D_f(x, z) \geq rK \int_0^1 \frac{\{\max[\|z + \tau y\|, \|z\|]\}^r}{\tau} \delta_X\left(\frac{\|\tau y\|}{2\max[\|z + \tau y\|, \|z\|]}\right) d\tau. \tag{10}$$

Clearly, $\max[\|z+\tau y\|, \|z\|] \le \|z\| + \tau\|y\|$. Since δ_X is nondecreasing on $[0,2]$, we have

$$\delta_X\left(\frac{\|\tau y\|}{2\max[\|z+\tau y\|, \|z\|]}\right) \ge \delta_X\left(\frac{\|\tau y\|}{2(\|z\| + \tau\|y\|)}\right) = \delta_X\left(\frac{\tau t}{2(\|z\| + \tau t)}\right), \tag{11}$$

because $\|y\| = t$. Now, we claim that

$$\max[\|z+\tau y\|, \|z\|] \ge \frac{\tau}{2}\|y\|.$$

This is clearly the case if $\|z\| \ge \frac{\tau}{2}\|y\|$. Otherwise, we have $\|z\| < \frac{\tau}{2}\|y\|$ and, therefore,

$$\begin{aligned} \|z+\tau y\| &\ge |\tau| \cdot \|y\| - \|z\| \\ &\ge \tau\|y\| - \frac{\tau}{2}\|y\| = \frac{\tau}{2}\|y\|, \end{aligned} \tag{12}$$

and the result also holds. Replacing (11) and (12) in (10) we have that, for all $x \in U(z,t)$,

$$D_f(x,z) \ge rK\left(\frac{t}{2}\right)^r \int_0^1 \tau^{r-1}\delta_X\left(\frac{\tau t}{2(\|z\| + \tau t)}\right) d\tau.$$

Taking infimum over $x \in U(z,t)$ in the left hand side of the last inequality and using (2) we obtain

$$\nu_f(z,t) \ge rK\left(\frac{t}{2}\right)^r \int_0^1 \tau^{r-1}\delta_X\left(\frac{\tau t}{2(\|z\| + \tau t)}\right) d\tau > 0, \tag{13}$$

where the last inequality holds because X is uniformly convex. This completes the proof.

Remark. The total convexity of the function $f(x) = \|x\|^r$ is guaranteed by Theorem 1 for $r > 1$ in uniformly convex Banach spaces. For $r \ge 2$, the function $f(x) = \|x\|^r$ is totally convex even if X is only *locally uniformly convex*, that is, if for each $x \in U(0,1)$, the function $\mu_X(x,\cdot) : [0,2] \to [0,1]$ defined by

$$\mu_X(x,t) = \begin{cases} \inf\left\{1 - \frac{1}{2}\|x+y\| : \|y\| = 1, \|x-y\| \ge t\right\}, & if t > 0, \\ 0, & if t = 0, \end{cases}$$

is positive for $t > 0$. Recall that X is locally uniformly convex if and only if the function $x \to \|x\|^2$ is locally uniformly strictly convex (cf. [12, Proposition 2.11, p. 50]). Observe that, for $r \ge 2$, we have $f(x) = \phi(h(x))$, where $\phi : [0,\infty) \to [0,\infty)$ is the convex nondecreasing differentiable function $\phi(t) = t^{r/2}$ and $h : X \to \mathbb{R}$ is given by $h(x) = \|x\|^2$. Using [19, Proposition 3], we deduce that, in this case,

$$D_f(x, y) = D_\phi(h(x), h(y)) + \phi'(h(y))D_h(x, y).$$

Hence,

$$D_f(x, y) \geq \phi'(h(y))D_h(x, y).$$

This implies

$$\nu_f(y, t) \geq \frac{r}{2} \|y\|^{r-2} \nu_h(y, t) \geq \frac{r}{2} \|y\|^{r-2} \Delta_h(y, t),$$

where $\Delta_h(y, t)$ stands for the modulus of uniformly strict convexity of h as defined in [12, p. 50]. Since, as noted above, when X is locally uniformly convex the function h is locally uniformly strictly convex, it results that $\Delta_h(y, t) > 0$, for all $t > 0$. Therefore, whenever $y \in X$ and $y \neq 0$, we have $\nu_f(y, t) > 0$ for all $t > 0$. If $y = 0$, then $D_f(x, y) = \|x\|^r$ and, thus, $\nu_f(y, t) = t^r$ which is positive whenever $t > 0$.

Theorem 1 leads to the following result which is of special interest in the sequel.

Corollary 1. *If X is uniformly convex then, for each $r \in (1, +\infty)$, the function $f(x) = \|x\|^r$ has the following properties:*

(i) For any $\alpha \geq 0$ and for any $y \in X$, the set

$$R_\alpha^f(y) = \{x \in X : D_f(y, x) \leq \alpha\}, \tag{14}$$

is bounded;

(ii) [Sequentialconsistency] For any two sequences $\{x^k\}_{k\in\mathbb{N}}$ and $\{y^k\}_{k\in\mathbb{N}}$ in X such that the first is bounded,

$$\lim_{k\to\infty} D_f(y^k, x^k) = 0 \Longrightarrow \lim_{k\to\infty} \left\|x^k - y^k\right\| = 0. \tag{15}$$

Proof. (i) Suppose, by contradiction, that for some $\alpha \geq 0$ and for some $y \in X$, there exists a unbounded sequence $\{x^k\}_{k\in\mathbb{N}}$ in $R_\alpha^f(y)$. Note that, for each nonnegative integer k, there exists some $\xi^k \in \partial f(x^k)$, such that

$$\begin{aligned}
\alpha &\geq D_f(y, x^k) = f(y) - f\left(x^k\right) - \left\langle \xi^k, y - x^k \right\rangle \\
&= f(y) - f\left(x^k\right) + \left\langle \xi^k, x^k \right\rangle - \left\langle \xi^k, y \right\rangle \\
&= f(y) - f\left(x^k\right) + r\left\|x^k\right\|^r - \left\langle \xi^k, y \right\rangle \\
&\geq \|y\|^r - \left\|x^k\right\|^r + r\left\|x^k\right\|^r - r\|y\| \cdot \left\|x^k\right\|^{r-1} \\
&= \|y\|^r + \left\|x^k\right\|^{r-1}\left[(r-1)\left\|x^k\right\| - r\|y\|\right],
\end{aligned}$$

where the second equality holds because of Asplund's Theorem (see [12, p. 25]) which ensures that $\partial f(x)$ is exactly the duality mapping of weight $\theta(t) = rt^{r-1}$. Letting here $k \to \infty$ leads to a contradiction.

(ii) We start by observing that, in our circumstances, the inequality (13) still holds. Thus, if $z \in C$, then for any real number $t > 0$ and for any nonempty bounded set $C \subset X$, we have

$$\begin{aligned} \nu_f(z,t) &\geq rK\left(\frac{t}{2}\right)^r \int_0^1 \tau^{r-1}\delta_X\left(\frac{\tau t}{2\left(\|z\| + \tau t\right)}\right) d\tau \\ &\geq rK\left(\frac{t}{2}\right)^r \int_0^1 \tau^{r-1}\delta_X\left(\frac{\tau t}{2\left(M + \tau t\right)}\right) d\tau > 0, \end{aligned}$$

whenever M is a upper bound of the set C. By taking here the infimum with respect to $z \in C$, we get

$$\inf\left\{\nu_f(z,t) : z \in C\right\} > 0, \tag{16}$$

whenever $t > 0$ and C is a bounded subset of X.

Now, suppose by contradiction that the norm-compatibility condition does not hold. Then, there exist a bounded sequence $\left\{x^k\right\}_{k\in\mathbb{N}}$ and a sequence $\left\{y^k\right\}_{k\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} D_f(y^k, x^k) = 0,$$

but $\left\{\left\|x^k - y^k\right\|\right\}_{k\in\mathbb{N}}$ does not converge to zero. Hence, there exists a number $\alpha > 0$, a subsequence $\{x^{j_k}\}_{k\in\mathbb{N}}$ of $\left\{x^k\right\}_{k\in\mathbb{N}}$ and a subsequence $\{y^{j_k}\}_{k\in\mathbb{N}}$ of $\left\{y^k\right\}_{k\in\mathbb{N}}$ such that, for each positive integer k, we have $\|x^{j_k} - y^{j_k}\| \geq \alpha$. Consider the bounded set C of all terms of $\left\{x^k\right\}_{k\in\mathbb{N}}$. Thus, according to (16), we get

$$\begin{aligned} D_f(y^{j_k}, x^{j_k}) &\geq \nu_f(x^{j_k}, \left\|x^{j_k} - y^{j_k}\right\|) \\ &\geq \nu_f(x^{j_k}, \alpha) \geq \inf\left\{\nu_f(z,t) : z \in C\right\} > 0, \end{aligned}$$

for all $k \in \mathbb{N}$. Letting $k \to \infty$ in this inequality leads to a contradiction.

3 Computing Bregman projections

In this section X denotes a uniformly convex and smooth Banach space. Our aim is to show how Bregman projections with respect to the function $f(x) = \|x\|^r$ with $r \in (1, +\infty)$

onto closed hyperplanes and half spaces in X can be effectively computed. Recall that, according to [2] and [9], when K is a closed convex subset of X, the *Bregman projection with respect to* f *onto* K is defined by

$$\Pi_K^f(x) = \arg\min \{D_f(y,x) : y \in K\}.$$

Note that the minimizer $\Pi_K^f(x)$ of $D_f(\cdot,x)$ over K, provided that it exists, is unique because $D_f(\cdot,x)$ is strictly convex as f is so. In our circumstances, existence of $\Pi_K^f(x)$ follows from Corollary 1 combined with [1, Proposition 2.1]. The relevance of computing $\Pi_K^f(x)$ when K is a closed hyperplane or half space will be made clear in Section 4 where these results will be used for solving applied mathematical problems.

The effective computability of $\Pi_K^f(x)$ essentially depends on the computability of the duality mapping $J_r : X \to \mathcal{P}(X^*)$ with the weight function $\theta(t) = rt^{r-1}$, which is given by $J_r(x) = \partial f(x)$ – see Asplund's Theorem [12, p. 18]. Since X is smooth, f is differentiable and, therefore, $J_r(x) = \{f'(x)\}$, for all $x \in X$. Observe that the function $\theta : [0,\infty) \to [0,\infty)$ is invertible and

$$\theta^{-1}(t) = \left(\frac{t}{r}\right)^{\frac{1}{r-1}}.$$

The function θ^{-1} is a weight function too and the duality mapping associated with it, $J_r^* : X^* \to \mathcal{P}(X)$, is given by

$$J_r^*(\xi) = \partial\chi(\|\cdot\|_*)(\xi),$$

where

$$\chi(t) := \int_0^t \theta^{-1}(u)\,du = (r-1)\cdot r^{\frac{r}{1-r}} \cdot t^{\frac{r}{r-1}}.$$

Taking into account that X is uniformly convex, we deduce that the function $\|\cdot\|_*$ is continuously differentiable (cf. [12, Theorem 2.13, p. 52]) and, thus, J_r^* is single valued, continuous and

$$J_r^*(\xi) = r^{\frac{1}{1-r}} \|\xi\|_*^{\frac{1}{r-1}} (\|\cdot\|_*)'(\xi),$$

for all $\xi \in X^*$. Recall (cf. [12, Proposition 4.7, p. 27]) that:

$$J_r^*\left(J_r(x)\right) = x, \tag{17}$$

for all $x \in X$ and

$$J_r\left(J_r^*(\xi)\right) = \xi, \tag{18}$$

for all $\xi \in X^*$.

The next result shows that computing $\Pi_K^f(x)$ when X is uniformly convex and smooth and K is a closed hyperplane or half space is, practically speaking, equivalent to solving a usually nonlinear equation of the form $\Phi(s) = 0$, where $\Phi : [0, \infty) \to \mathbb{R}$ is a known continuous function.

Theorem 2. *Let X be a uniformly convex and smooth Banach space and let*

$$K = \{z \in X : \langle a, z\rangle = b\}, \tag{19}$$

where $a \in X^* \backslash \{0\}$ and $b \in \mathbb{R}$. For any $x \in X$ and for all $r \in (1, +\infty)$, the following statements hold:

(i) The equation

$$\langle a, J_r^* (sa + J_r(x))\rangle = b \tag{20}$$

has solutions s such that

$$sign(s) = sign(b - \langle a, x\rangle); \tag{21}$$

(ii) The Bregman projection $\Pi_K^f(x)$ with respect to the function $f(x) = \|x\|^r$ is given by

$$\Pi_K^f(x) = J_r^* (sa + J_r(x)), \tag{22}$$

with $s \in \mathbb{R}$ being a solution of the equation (20);

(iii) Formula (22) *remains true when K is the half space $\{z \in X : \langle a, z\rangle \geq b\}$ and s is a nonnegative solution of* (20).

Proof. (i) Denote

$$u(s) = J_r^* (sa + J_r(x)). \tag{23}$$

We distinguish three cases which we discuss separately.

Case 1. Suppose that $\langle a, x\rangle = b$. Observe that $u(0) = J_r^* (J_r(x)) = x$ because of (17). Thus, for $s = 0$,

$$\langle a, u(s)\rangle = \langle a, x\rangle = b,$$

that is, $s = 0$ is a solution of (20) in this case.

Case 2. Suppose that $\langle a, x\rangle < b$. Consider the function $\Phi : [0, \infty) \to \mathbb{R}$ defined by

$$\Phi(s) = \langle a, u(s)\rangle - b.$$

with $u(s)$ given by (23) and note that

$$\Phi(0) = \langle a, x\rangle - b < 0. \tag{24}$$

The function Φ is continuous on $[0,\infty)$, because J_r^* is continuous on X^* (as noted above, if X is uniformly convex, then J_r^* is continuous on X^*). According to [12, Proposition 4.7, p. 27], we obtain

$$\begin{aligned}
u(s) &= J_r^*\left(sa + J_r(x)\right) \\
&= J_r^*\left[s\left(a + \frac{1}{s}J_r(x)\right)\right] \\
&= \frac{\chi\left(s\left\|a+\frac{1}{s}J_r(x)\right\|_*\right)}{\chi\left(\left\|a+\frac{1}{s}J_r(x)\right\|_*\right)} \cdot J_r^*\left(a + \frac{1}{s}J_r(x)\right) \\
&= s^{\frac{r}{r-1}} J_r^*\left(a + \frac{1}{s}J_r(x)\right).
\end{aligned}$$

Taking into account the continuity of J_r^*, we deduce that the following limit exists and we have

$$\lim_{s\to\infty} J_r^*\left(a + \frac{1}{s}J_r(x)\right) = J_r^*(a).$$

Hence,

$$\lim_{s\to\infty}\Phi(s) = \lim_{s\to\infty}\left[s^{\frac{r}{r-1}}\left\langle a, J_r^*\left(a+\frac{1}{s}J_r(x)\right)\right\rangle - b\right] = +\infty.$$

This, together with (24), shows that the continuous function Φ vanishes at some point s in $[0,\infty)$, i.e. for some $s \in [0,\infty)$, the equation (20) is satisfied.

Case 3. Suppose that $\langle a, x\rangle > b$. This is equivalent to $\langle -a, x\rangle < -b$ and we can apply the reasoning done in Case 2 with $-a$ instead of a and $-b$ instead of b.

In any of the possible cases the equation (20) has solutions. This proves (i).

(ii) Suppose that K is given by (19). According to [1, Proposition 2.2], it is sufficient to show that when s is a solution of (20), we have $u(s) \in K$ and

$$\langle f'(x) - f'(u(s)), z - u(s)\rangle \le 0, \tag{25}$$

for any $z \in K$. Note that $u(s) \in K$ because of the way in which s is chosen. Also, observe that (25) can be rewritten as

$$\langle J_r(x) - J_r(u(s)), z - u(s)\rangle \le 0$$

for all $z \in K$. In turn, the last inequality is equivalent to

$$\langle J_r(x) - J_r\left(J_r^*(sa + J_r(x))\right), z - J_r^*\left(sa + J_r(x)\right)\rangle \le 0.$$

According to (18), this amounts to

$$\langle J_r(x) - sa - J_r(x), z - J_r^* (sa + J_r(x))\rangle \leq 0$$

which is exactly

$$-s \langle a, z - J_r^* (sa + J_r(x))\rangle \leq 0$$

for all $z \in K$. Since s is a solution of (20), for any $z \in K$, we get

$$\langle a, z - J_r^* (sa + J_r(x))\rangle = \langle a, z\rangle - \langle a, J_r^* (sa + J_r(x))\rangle = b - b = 0,$$

and so (25) holds.

(iii) Assume that K is the half-space $\{z \in X : \langle a, z\rangle \geq b\}$. Using again Proposition 2.2 in [1], it is sufficient to prove that (25) holds for any $z \in K$, when $x \notin K$, i.e., when $\langle a, x\rangle < b$. Observe that

$$\begin{aligned} & \langle f'(x) - f'(u(s)), z - u(s)\rangle \\ = \; & \langle J_r(x) - J_r(J_r^* sa + J_r(x)), z - J_r^* (sa + J_r(x))\rangle \\ = \; & -s \langle a, z - J_r^* (sa + J_r(x))\rangle , \end{aligned}$$

for all $z \in K$. Since s is a nonnegative solution of equation (20) (which exists according to (i)), we deduce that (25) is equivalent to

$$\langle a, z - J_r^* (sa + J_r(x))\rangle \geq 0,$$

which is exactly

$$\langle a, z\rangle - \langle a, u(s)\rangle \geq 0.$$

The last inequality holds because $\langle a, z\rangle \geq b$ and $\langle a, u(s)\rangle = b$.

4 Applications

In this section we show several applications of the results presented above. We start with an application to solving a class of linear equations in a smooth uniformly convex Banach space X. Let $(\Omega, \mathcal{A}, \mu)$ be a complete probability space. *Let* $K : \Omega \to X^*$ *and* $h : \Omega \to \mathbb{R}$ *be* μ*-integrable functions*[1]*. Find* $x \in X$ *such that*

$$\langle K(\omega), x\rangle = h(\omega), \quad \mu\text{−}a.e., \tag{26}$$

[1] *In what follows, measurability and integrability of functions with values in Banach spaces are in the sense of Bochner.*

presuming that such a x exists. This problem appears in practice in various particular forms. Among them we recall the following:

(a) *The Fredholm integral equation of the first type* (cf. [16], [23]) is the particular instance of the problem above in which Ω is an interval, $X = \mathcal{L}^p(\Omega)$ and $h \in \mathcal{L}^1(\Omega)$;

(b) *The best approximation problem of the function* $h \in \mathcal{L}^1[a,b]$ is the problem of finding $x = \{x^n\}_{n\in\mathbb{N}} \in X := \ell^p$ such that, for almost all $\omega \in \Omega := [a,b]$,

$$\sum_{n=0}^{\infty} K_n(\omega)x^n = h(\omega),$$

where, for each $t \in [a,b]$, we denote $K(t) := \{K_n(t)\}_{n\in\mathbb{N}} \in \ell^q$. If $h \in \mathcal{L}^p[a,b]$ and $\{K_n\}_{n\in\mathbb{N}}$ is a base of $\mathcal{L}^p[a,b]$, then this is exactly the problem of finding the Fourier coefficients of h with respect to the given base.

Denote $\Omega_i = \Omega \times \{i\}$, $i = 1,2$. Let the set $\Omega_* = \Omega_1 \cup \Omega_2$ be provided with the σ-algebra

$$\mathcal{A}_* = \{(A_1 \times \{1\}) \cup (A_2 \times \{2\}) : A_i \in \mathcal{A}, i = 1,2\}$$

and with the measure

$$\mu_* \left[(A_1 \times \{1\}) \cup (A_2 \times \{2\})\right] = \frac{1}{2}\left[\mu(A_1) + \mu(A_2)\right].$$

Clearly, $(\Omega_*, \mathcal{A}_*, \mu_*)$ is a complete probability space. We define the function $g : \Omega_* \times X \to \mathbb{R}$ by

$$g((\omega,i),x) = \begin{cases} \langle K(\omega), x\rangle - h(\omega), & if\, i = 1, \\ \\ -\left[\langle K(\omega), x\rangle - h(\omega)\right], & if\, i = 2. \end{cases} \tag{27}$$

Note that g is a convex Carathéodory function, that is, for any $x \in X$, the function $g(\cdot, x)$ is measurable and, for any $(\omega, i) \in \Omega_*$, the function $g((\omega,i),\cdot)$ is convex and continuous on X. Observe that finding a solution of (26) is equivalent to finding $x \in X$ such that

$$g((\omega,i),x) \leq 0, \quad \mu_*{-}a.e. \tag{28}$$

Define

$$A(x,y) = \frac{1}{2}\int_\Omega \langle K(\omega), y\rangle \cdot sign\left[\langle K(\omega), x\rangle - h(\omega)\right] d\mu(\omega), \tag{29}$$

for any $x, y \in X$. Note that $A(x, \cdot)$ belongs to X^*. Let

$$C(x) = \frac{1}{2} \int_\Omega h(\omega) \cdot sign\left[\langle K(\omega), x\rangle - h(\omega)\right] d\mu(\omega). \tag{30}$$

Then, for each $x \in X$, the set

$$H(x) = \{z \in X; \langle A(x, \cdot), z\rangle \leq C(x)\} \tag{31}$$

is a well defined half space of X. With these notations we have the following result:

Theorem 3. *Suppose that the equation* (26) *has solutions in X and that $r \in (1, \infty)$. Then, for any initial point $x^0 \in X$, the sequence $\left\{x^k\right\}_{k\in\mathbb{N}}$ recursively generated in X by*

$$x^{k+1} = J_r^*\left(s_k A(x^k, \cdot) + J_r(x^k)\right), \tag{32}$$

with s_k a solution of the equation

$$\left\langle A(x^k, \cdot), J_r^*\left(sA(x^k, \cdot) + J_r(x^k)\right)\right\rangle = C(x^k),$$

has the following properties:
(i) $\left\{x^k\right\}_{k\in\mathbb{N}}$ is bounded and has weak accumulation points;
(ii) The following limit exists and we have

$$\lim_{k\to\infty} \left\|x^{k+1} - x^k\right\| = 0; \tag{33}$$

(iii) The weak accumulation points of $\left\{x^k\right\}_{k\in\mathbb{N}}$ are solutions of the equation (26);
(iv) If the duality mapping J_r is sequentially weakly-to-weak* continuous, then $\left\{x^k\right\}_{k\in\mathbb{N}}$ converges weakly to a solution of (26).

Proof. According to [7, Theorem 4.3] and Corollary 1 we deduce that, for any initial point $x^0 \in X$, the sequence $\left\{x^k\right\}_{k\in\mathbb{N}}$ recursively generated by

$$x^{k+1} = \Pi^f_{K\left(x^k\right)}\left(x^k\right), \tag{34}$$

with

$$K(x) := \left\{z \in X : \int_{\Omega_*(x)} g\left(\left(\omega, i\right), z\right) d\mu_*\left(\omega, i\right) \leq 0\right\}$$

and

$$\Omega_*(x) = \left\{\left(\omega, i\right) \in \Omega_* : g\left(\left(\omega, i\right), x\right) > 0\right\}$$

is bounded, has weak accumulation points and each such point is a solution of (28). As noted above, solutions of (28) are also solutions of (26) and conversely. It remains to show the following claim which implies that the sequences generated according to the recursive rule (32) are exactly the sequences (34).

Claim 2. The function $\psi : X \times X \to \mathbb{R}$ given by

$$\psi(z,x) = \int_{\Omega_*(x)} g((\omega,i),z)\, d\mu_*(\omega,i)$$

is well defined and for the half space $H(x)$ described in (31) *we have*

$$H(x) = \{z \in X : \psi(z,x) \leq 0\} \tag{35}$$

In order to prove the claim observe that well definedness of ψ follows from [7, Lemma 4.2]. For showing (35) note that

$$\begin{aligned}
\psi(z,x) &= \int_{\Omega_*(x)} g((\omega,i),z)\, d\mu_*(\omega,i) \\
&= \int_{\Omega_*} g((\omega,i),z) \cdot sgg((\omega,i),x)\, d\mu_*(\omega,i) \\
&= \int_{\Omega_1} g((\omega,1),z) \cdot sgg((\omega,1),x)\, d\mu_*(\omega,i) \\
&\quad + \int_{\Omega_2} g((\omega,2),z) \cdot sgg((\omega,2),x)\, d\mu_*(\omega,i),
\end{aligned}$$

where $sg(t) = 1$, if $t > 0$, and $sg(t) = 0$, otherwise. Hence,

$$\begin{aligned}
\psi(z,x) &= \frac{1}{2}\int_\Omega [\langle K(\omega),z\rangle - h(\omega)] \cdot sg[\langle K(\omega),x\rangle - h(\omega)]\, d\mu(\omega) \\
&\quad -\frac{1}{2}\int_\Omega [\langle K(\omega),z\rangle - h(\omega)] \cdot sg[-\langle K(\omega),x\rangle + h(\omega)]\, d\mu(\omega) \\
&= \frac{1}{2}\int_\Omega [\langle K(\omega),z\rangle - h(\omega)] \cdot sign[\langle K(\omega),x\rangle - h(\omega)]\, d\mu(\omega) \\
&= A(x,z) - C(x)
\end{aligned}$$

and this implies that $\psi(z,x) \leq 0$ if and only if $A(x,z) - C(x) \leq 0$, i.e. $z \in K(x)$ if and only if $z \in H(x)$.

fhFU6.7135in3.1825in0ptA numerical experimentelena2.bmp

Remarks. (i) If X is a finite dimensional space, then Theorem 3 guarantees (strong) convergence of the sequence $\{x^k\}_{k\in\mathbb{N}}$ defined by (32), no matter how r is chosen in $(1,+\infty)$.

(ii) If X is one of the spaces ℓ^p with $1<p<+\infty$, then the equation (26) is exactly the best approximation problem and, for any $r\in(1,+\infty)$, the sequence generated according to (32) converges weakly to a solution of the problem (26) because, in this case, the duality mapping J_r is sequentially weakly-to-weak* continuous (cf. [12, Proposition 4.14, p. 73]). For the same reason, weak convergence of $\{x^k\}_{k\in\mathbb{N}}$ holds when X is a Hilbert space and $r=2$ in which case the Bregman projections are the metric projections.

(iii) A question whose answer we do not know is whether weak convergence of the sequence $\{x^k\}_{k\in\mathbb{N}}$ defined by (32) can be ensured under requirements less demanding than the weak-to-weak* continuity of J_r.

fhFU5.0004in3.1825in0ptThe measure of the set of violated restrictionselena3.bmp

An example. A detailed analysis of the proof of convergence of the algorithm given by (32) shows that this is essentially a procedure of reducing at each iterative step the size of the set

$$\Gamma_k=\left\{(\omega,i)\in\Omega_*:\max\left[g((\omega,i),x^k),0\right]>0\right\},$$

that is the size of the set of points ω at which the original equation is violated. In order to show how the algorithm works we consider the following Fredholm equation (of the first kind) which has the function

$$\bar{x}(t)=\begin{cases} t^{-1/2} & if t>0,\\ 0 & if t=0,\end{cases}$$

as a solution in $L^{3/2}([0,1])$:

$$\int_0^1 \exp(s\sqrt{t})x(t)dt=\frac{2}{s}(\exp(s)-1), a.e.$$

We compute[2] the 20-th iterate x^k starting from $x^0(t)\equiv 0.3$ and taking $f(x)=\|x\|^{3/2}$ in the space $L^{3/2}([0,1])$. Figure 1 shows the successive approximations produced by the algorithm. Observe that they accumulate to the curve denoted by $x^{20}(t)$. In fact, this curve is a reasonably good approximation of a solution of the given equation. The size of Γ_{20}, measured by the L^1-norm of $\max[g(\cdot,x^{20}),0]$ is approximately 0.004. Figure 2

[2]The computations were done using a package of programs written by Mr. Eyal Masad.

presents the way in which the Bregman distances $D_f(x^{k+1}, x^k)$ (the tall bars) and the L^1-norms of $\max\left[g(\cdot, x^k), 0\right]$ (the short bars) are varying along the computational process. Observe that they are decreasing and that after 20 steps we have $D_f(x^{20}, x^{19}) \approx 0.0001$. The metric distances between successive iterates decrease too as indicated in Figure 2 by the black bars. The behavior of the algorithm in this particular case is typical for all consistent equations we solved numerically in spaces L^p with $p > 1$: After a small number of iterations the iterates accumulate to a function with a rather small residual (i.e., an almost feasible function) from which further iterates very little differ for as long as the computational process is continued. This suggests that the algorithm converges strongly but we do not have a proof of that.

Acknowledgments

The authors gratefully acknowledge the financial support of the Israel Science Foundation. Also, the authors wish to thank Mr. Eyal Masad from Technion-The Israel Institute of Technology for allowing them to use a package of programs he has created.

References

[1] Alber, Ya. and Butnariu, D., Convergence of Bregman projection methods for solving convex feasibility problems in reflexive Banach spaces, *Journal of Optimization Theory and Applications*, **92** (1997) 33-61.

[2] Bregman, L.M., The relaxation method for finding common points of convex sets and its application to the solution of convex programming, *USSR Computational Mathematics and Mathematical Physics*, **7** (1967) 200-217.

[3] Browder, F.E., Nonlinear operators and nonlinear equations of evolution in Banach spaces. Proceeding of Symposia in Pure Mathematics, *American Mathematical Society,* **18** 2 (1976).

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

[5] Butnariu, D. and Iusem, A.N., Local moduli of convexity and their application to finding almost common fixed points of measurable families of operators, in: Y. Censor and S. Reich (Eds.) "Recent Developments in Optimization Theory and Nonlinear Analysis", *Contemporary Mathematics*, vol. 204, (1997) pp. 61-92.

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

[7] Butnariu, D., Iusem, A.N. and Burachik, R., Iterative methods of solving stochastic convex feasibility problems and applications, to be published in *J. of Computational Optimization and Applications*, **15** 3 (2000).

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

[9] Censor, Y. and Lent, A., An iterative row action method for intervakl convex programming, *Journal of Optimization Theory and Applications,* **34** (1981) 321-353.

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

[11] Cimmino, G., Calcolo approsimato per le soluzione di sistemi di equazioni lineari, *La Riserca Scientifica,* Roma, XVI, Anno IX, **2** (1938) 326-333.

[12] Ciorãnescu, I., Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems, *Kluwer Academic Publishers*, 1990.

[13] Clarke, F.H., Optimization and Nonsomooth Analysis, *John Willey and Sons*, New York, 1983.

[14] Diestel, J., Geometry of Banach Spaces — Selected Topics, *Springer Verlag*, Berlin (1975).

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

[16] Guo, D., Lakshmikantham, V. and Liu, X., Nonlinear Integral Equations in Abstract Spaces, *Kluwer Academic Publishers*, 1996.

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

[18] Iusem, A.N. and Isnard, C.A., A mixed Hölder-Minkowsky inequality and total convexity in Banach spaces, (to be published).

[19] Iusem, A.N., Isnard, C.A. and Butnariu, D., A mixed Hölder-Minkowsky inequality, *Proceedings of the American Mathematical Society*, **127** 3 (1999) 2405-2415.

[20] Kiwiel, K., Proximal minimization methods with generalized Bregman functions, *SIAM J. on Control and Optimization*, **35** (1997) 1142-1168.

[21] Neumann, J. (von), Functional Operators - Vol. II, The Geometry of Orthogonal Spaces, in: "Annals of Mathematics Studies" #22, Princeton University Press, 1950.

[22] Phelps, R.R., Convex Functions, Monotone Operators and Differentiability, *Springer Verlag*, Berlin, 1993.

[23] Pipkin, A.C., A Course on Integral Equations, Springer Verlag, 1991.

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

[25] Xu, Zong-Ben and Roach, G.F., Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces, *Journal of Mathematical Analysis and Applications,* **157** (1991) 189–210.