

BREGMAN DISTANCES, TOTALLY CONVEX FUNCTIONS AND A METHOD FOR SOLVING OPERATOR EQUATIONS IN BANACH SPACES

DAN BUTNARIU AND ELENA RESMERITA

January 18, 2005

Abstract

The aim of this paper is twofold. First, several basic mathematical concepts involved in the construction and study of Bregman type iterative algorithms are presented from a unified analytic perspective. Also, some gaps in the current knowledge about those concepts are filled in. Second, we employ existing results on total convexity, sequential consistency, uniform convexity and relative projections in order to define and study the convergence of a new Bregman type iterative method of solving operator equations.

2000 Mathematics Subject Classification: 46B99, 46T99, 47J25.

Keywords: Bregman distance, Bregman projection, relative projection, total convexity, essential strict convexity, E -space, Mosco convergence, uniform convexity, inverse monotone operator, inverse strongly-monotone operator.

1 Introduction

Let X be a Banach space and $f : X \rightarrow (-\infty, +\infty]$ be a proper convex function. For any $x \in \text{dom } f$ and $z \in X$ we denote by $f^\circ(x, z)$ the *right-hand side derivative* of f at x in the direction z , that is,

$$f^\circ(x, z) = \lim_{t \searrow 0} \frac{f(x + tz) - f(x)}{t}.$$

The function $D_f : \text{dom } f \times \text{dom } f \rightarrow [0, +\infty]$ given by

$$D_f(y, x) := f(y) - f(x) - f^\circ(x, y - x).$$

is called the *Bregman distance with respect to f* . In 1967 Lev Bregman [20] has discovered an elegant and effective technique for the use of the function D_f in the process of designing and analyzing feasibility and optimization algorithms. This opened a growing area of research in which Bregman's technique is applied in various ways in order to design and analyze iterative algorithms for solving not only feasibility and optimization problems, but also algorithms for solving variational inequalities, for approximating equilibria, for computing fixed points of nonlinear mappings and more (see [23], [24], [27], [40], [48], [49], [51], [64], [2], [5], [12], [15], [20], [26], [29], [28], [33], [36], [59], [36], [28] and the references therein). Although called a 'distance' by Censor and Lent [35], the function D_f is not a distance in the usual sense of the term: it is not always symmetric and may not satisfy the triangular inequality.

The main aim of this paper is to give an analytic perspective over the basic tools commonly used in the implementations of Bregman's technique and to show how these tools can be employed for building an algorithm of solving operator equations. An intrinsic feature of Bregman's technique of designing iterative algorithms consists of ensuring that the sequences $\{x^k\}_{k \in \mathbb{N}}$ these algorithms generate are contained in the domain of a convex function f and have the property that $\lim_{k \rightarrow \infty} D_f(x^k, x) = 0$ for points $x \in \text{dom } f$ which are necessarily solutions of the problem the algorithm is supposed to solve. If the function f is chosen such that, for any sequence $\{x^k\}_{k \in \mathbb{N}} \subset \text{dom } f$ and any vector $x \in \text{dom } f$, one has

$$\lim_{k \rightarrow \infty} D_f(x^k, x) = 0 \Rightarrow \lim_{k \rightarrow \infty} \|x^k - x\| = 0, \quad (1)$$

then the convergence of the algorithm to a solution of the problem is guaranteed. That brings up the question of whether, and under which conditions, a convex function f satisfies (1). Butnariu, Censor and Reich [25] were the first to point out that totally convex functions, introduced in their paper under a different name, have that property. The function f is called *totally convex at a point* $x \in \text{dom } f$ if its *modulus of total convexity at* x , i.e., the function $\nu_f(x, \cdot) : [0, +\infty) \rightarrow [0, +\infty]$ defined by

$$\nu_f(x, t) = \inf \{D_f(y, x) : y \in \text{dom } f, \|y - x\| = t\},$$

is positive on $(0, \infty)$. Resmerita [61] made clear that the totally convex functions are the only convex functions satisfying (1). The importance of the notion of total convexity in applications stems primarily from its equivalence to condition (1). In the process of using totally convex functions in the convergence analysis of various algorithms, additional useful properties of those functions were discovered and some questions concerning the connections of totally convex functions with other classes of functions and with the geometry of the underlying space were raised. The results in these respects are scattered in literature, are placed in the specific restrictive contexts of the algorithmic procedures they were supposed to serve and are presented in various terminologies. One of the aims of this paper is to present, in a unified approach and in full generality, the properties of totally convex functions which proved to be of interest in the theory of various algorithms and, at the same time, to fill in several gaps in the existing knowledge about those functions (see Section 2).

In finite dimensional Banach spaces there is no much difference between totally convex and strictly convex functions. Any strictly convex function with closed domain which is continuous on $\text{dom } f$ is totally convex. However, in infinite dimensional spaces, totally convex functions constitute a very special class of strictly convex functions and identifying totally convex functions with specific features required in applications is quite complicated. In an infinite dimensional setting, for reasons related to effective computability, the functions $\|\cdot\|^p$ with $p > 1$ are among the most likely candidates to be used in the build up of Bregman type algorithms. Therefore, it is interesting to know which are the Banach spaces in which the functions $\|\cdot\|^p$ with $p > 1$ are totally convex. Those spaces were termed locally totally convex spaces in [61]. The locally uniformly convex spaces are among the locally totally convex spaces. In Section 3 we give geometric characterizations of locally totally convex spaces. These characterizations lead us to the conjecture that there exist Banach spaces which are locally totally convex without being locally uniformly convex. However, we do not have any example to support this claim.

A notion strongly related to that of total convexity is that of projection onto a closed convex set relative to a totally convex function (see Section 4). This notion is a straightforward generalization of the concept of 'generalized projection' due to Alber [1, 2]. The projection relative to f onto a closed convex set K is a monotone operator from X to its dual X^* .

In some circumstances, projection operators relative to the totally convex function f can be seen as extensions of Bregman projection operators to larger sets. Bregman projections are the corner stones of a plethora of algorithms (see [36], [28] and the references therein). Relative projections were successfully used for designing and analyzing various algorithms by Alber et al. [3, 5, 6, 7] and by Alber and Reich [9]. The algorithm for solving equations in Banach spaces presented in Section 5, illustrates the way in which projections relative to totally convex functions can be used for that purpose. It is a procedure for finding solutions of the problem

$$Ax = 0 \text{ and } x \in \Omega,$$

where $A : X \rightarrow X^*$ is a given operator and Ω is a closed convex subset of X . This algorithm is a refinement of a procedure originating in Landweber's [52] method of finding minimum norm solutions of linear equations and in Polyak's [58] method of finding zeros for subgradients of convex functions. That procedure was subsequently extended by Bruck [21, 22] and by Alber [3] to a method of solving operatorial equations and variational inequalities in Banach spaces. The general iterative formula on which the algorithm is based is

$$y^{k+1} = P_{\Omega}^f[f'(y^k) - \alpha Ay^k],$$

where $\alpha \in (0, \infty)$, f' stands for the Gâteaux derivative of f and P_{Ω}^f stands for the projection operator onto Ω relative to f . We are proving convergence of this algorithm when applied to classes of operators A satisfying certain monotonicity type properties we call inverse monotony and strong inverse monotony relative to a totally convex function f . These properties require that the operator $A_{\alpha}^f x = f^{*'}[f'(x) - \alpha Ax]$, where f^* is the Fenchel conjugate of f , should satisfy some nonexpansivity type properties conditioned by the function f . In Hilbert spaces, such requirements are commonly satisfied by the operators A whose (set valued) inverses are strongly monotone because they are inverse (strongly) monotone relative to $f = \frac{1}{2} \|\cdot\|^2$. In this particular case P_{Ω}^f is exactly the metric projection operator P_{Ω} and the algorithmic procedure is $y^{k+1} = P_{\Omega}(y^k - \alpha Ay^k)$. In non-Hilbertian spaces, inverse strong monotonicity of A is less useful in order to ensure convergence of the algorithm $y^{k+1} = P_{\Omega}(y^k - \alpha Ay^k)$ mostly due to the lack of nonexpansivity of P_{Ω} . In this situation we take advantage of the nonexpansivity like properties of P_{Ω}^f when f is totally convex, and of the presumed inverse monotonicity of A relative to f , in order to ensure convergence of our algorithm.

The new results proven in this paper are as follows. Theorem 2.10 and Theorem 2.15 elucidate the connections between total convexity and uniform convexity in infinite and, respectively, in finite dimensional spaces. Theorem 3.3 is a geometric characterization of the Banach spaces in which the function $\|\cdot\|^2$ is totally convex at any point. If X is reflexive, then this result can be interpreted as another characterization of the E -spaces introduced by Fan and Glicksberg [44]. Theorem 4.6 establishes continuity of the projections relative to a totally convex functions with respect to the Mosco convergence. Theorem 4.8 shows a way of computing relative projections on hyperplanes and half-spaces. Theorems 5.5 and 5.8 are convergence results for the algorithm described above.

2 Totally Convex Functions

2.1 We start our presentation of the totally convex functions by summarizing some basic properties of the modulus of total convexity. They were stated in [28] and [48] under the

additional assumption that x is a point in the algebraic interior of $\text{dom } f$. A careful analysis of the proofs given there shows that this assumption is superfluous.

Proposition. *Let X be a Banach space and $f : X \rightarrow (-\infty, +\infty]$ be a lower semicontinuous convex function such that $\text{int}(\text{dom } f) \neq \emptyset$. If $x \in \text{dom } f$, then*

(i) *The domain of $\nu_f(x, \cdot)$ is an interval $[0, \tau_f(x))$ or $[0, \tau_f(x)]$ with $\tau_f(x) \in (0, +\infty]$, if $\partial f(x) \neq \emptyset$, and it is $\{0\}$, otherwise;*

(ii) *If $c \geq 1$ and $t \geq 0$, then $\nu_f(x, ct) \geq c\nu_f(x, t)$;*

(iii) *The function $\nu_f(x, \cdot)$ is superadditive, that is, for any $s, t \in [0, +\infty)$, one has $\nu_f(x, s+t) \geq \nu_f(x, s) + \nu_f(x, t)$;*

(iv) *The function $\nu_f(x, \cdot)$ is nondecreasing; it is strictly increasing on its domain if and only if f is totally convex at x ;*

(v) *The following equality holds*

$$\nu_f(x, t) = \inf \{D_f(y, x) : y \in \text{dom } f, \|y - x\| \geq t\}. \quad (2)$$

Proof: As mentioned at the beginning of this section, the statements (i)-(iv) are more general than the ones in [28, p. 26], but the proofs are the same. We now show that (v) also holds. To this end, denote by $\nu_f^\#(x, t)$ the right hand side of (2). Clearly,

$$\nu_f^\#(x, t) \leq \nu_f(x, t),$$

for all $t \geq 0$ and the equality holds for $t = 0$. Suppose that $t > 0$. We show that for any $y \in \text{dom } f$ having $\|y - x\| \geq t$ there exists $y' \in \text{dom } f$ such that $\|y' - x\| = t$ and

$$D_f(y, x) \geq D_f(y', x). \quad (3)$$

Since f is convex, the function $D_f(\cdot, x)$ is also convex. Then,

$$D_f(\alpha y + (1 - \alpha)x, x) \leq \alpha D_f(y, x) + (1 - \alpha)D_f(x, x) = \alpha D_f(y, x). \quad (4)$$

Now, take $\alpha = \frac{t}{\|y - x\|}$ in (4). This gives exactly (3) for $y' = y_\alpha$ and, clearly, one has $\|y_\alpha - x\| = t$. The proof is complete. \square

2.2 Total convexity is a strong form of strict convexity, as shown by the following result.

Proposition (cf. [28, p. 34]).

(i) *If f is totally convex at any $x \in \text{dom } f$, then it is strictly convex on this set;*

(ii) *If $\dim X < +\infty$, $\text{dom } f$ is closed and the function f restricted to $\text{dom } f$ is continuous and strictly convex, then f is totally convex at any $x \in \text{dom } f$.*

2.3 The following features of the moduli of total convexity give rules for generating totally convex functions by aggregating other totally convex functions.

Proposition (cf. [28, p. 35]).

(i) *Let $f_1, \dots, f_m : X \rightarrow (-\infty, +\infty]$ be totally convex functions with domains D_1, \dots, D_m , respectively, such that $D = \bigcap_{i=1}^m D_i \neq \emptyset$. Then, for any m nonnegative real numbers c_1, \dots, c_m with $\sum_{i=1}^m c_i > 0$, the function $h = \sum_{i=1}^m c_i f_i$ is totally convex and, for any $x \in D$ and for all $t \in [0, +\infty)$,*

$$\nu_h(x, t) \geq \sum_{i=1}^m c_i \nu_{f_i}(x, t);$$

(ii) If f is totally convex and lower semicontinuous with open domain D and ϕ is a real convex function, differentiable and strictly increasing on an open interval which contains $f(D)$, then the function $g : X \rightarrow (-\infty, +\infty]$ defined by $g = \phi \circ f$ (with $g(x) = +\infty$ for $x \notin D$) is totally convex and

$$\nu_g(x, t) \geq \phi'(f(x)) \cdot \nu_f(x, t)$$

for any $x \in D$ and for all $t \geq 0$.

2.4 Here are other total convexity criteria which allow building totally convex function on product spaces.

Proposition (cf. [48]). Let X_1, X_2 be real Banach spaces, $f_i : X_i \rightarrow (-\infty, +\infty]$, $i = 1, 2$, two proper convex lower semicontinuous functions whose domains have nonempty interiors. Define the function $f : X_1 \times X_2 \rightarrow (-\infty, +\infty]$ by $f(x_1, x_2) = f_1(x_1) + f_2(x_2)$. Then

(i) For any $x = (x_1, x_2) \in \text{dom } f$, such that $\partial f(x) \neq \emptyset$, the domain of $\nu_f(x, \cdot)$ is an interval $[0, \tau_f(x))$ or $[0, \tau_f(x)]$ with $0 \leq \tau_f(x) = \tau_{f_1}(x_1) + \tau_{f_2}(x_2) \leq \infty$, where

$$\tau_{f_i}(x_i) = \sup\{t : t \in \text{dom } \nu_{f_i}(x_i, \cdot)\}, i = 1, 2.$$

Moreover,

$$\nu_f(x, t) = \inf\{\nu_{f_1}(x_1, s) + \nu_{f_2}(x_2, t - s) : s \in [0, t]\}.$$

If $\partial f(x) = \emptyset$, then $\text{dom } \nu_f(x, \cdot) = \{0\}$.

(ii) For any $x = (x_1, x_2) \in \text{dom } f$ and $t \in \text{dom } \nu_f(x, \cdot)$,

$$\nu_f(x, t) \geq \min\{\nu_{f_i}(x_i, t/2)\}, i = 1, 2.$$

Therefore, if both functions f_i are totally convex, then f is totally convex.

2.5 The following characterization of total convexity makes clear why functions with this property are useful in convergence analysis of many algorithms: D_f -convergence of a sequence $\{y^n\}_{n \in \mathbb{N}}$ to x , that is, the fact that $\{D_f(y^n, x)\}_{n \in \mathbb{N}}$ converges to zero (see [14]) implies convergence in norm of $\{y^n\}_{n \in \mathbb{N}}$ to x .

Proposition (cf. [61]). Let X be a Banach space and $f : X \rightarrow (-\infty, +\infty]$ be a convex function. If $x \in \text{dom } f$, then the following statements are equivalent:

(i) The function f is totally convex at x ;

(ii) For any sequence $\{y^n\}_{n \in \mathbb{N}} \subseteq \text{dom } f$,

$$\lim_{n \rightarrow \infty} D_f(y^n, x) = 0 \Rightarrow \lim_{n \rightarrow \infty} \|y^n - x\| = 0;$$

(iii) For any sequence $\{y^n\}_{n \in \mathbb{N}} \subseteq \text{dom } f$,

$$\liminf_{n \rightarrow \infty} D_f(y^n, x) = 0 \Rightarrow \liminf_{n \rightarrow \infty} \|y^n - x\| = 0;$$

(iv) For any sequence $\{y^n\}_{n \in \mathbb{N}} \subseteq \text{dom } f$,

$$\lim_{n \rightarrow \infty} D_f(y^n, x) = 0 \Rightarrow \liminf_{n \rightarrow \infty} \|y^n - x\| = 0;$$

(v) For any $\varepsilon > 0$, there exists $\delta > 0$ such that if $y \in \text{dom } f$ and $D_f(y, x) \leq \delta$, then $\|x - y\| \leq \varepsilon$.

2.6 Characteristic properties of totally convex functions were given in [31], [61], [32]. Denote by \mathcal{F} the set of functions $\psi : [0, +\infty) \rightarrow [0, +\infty]$ which satisfy the conditions below:

- (i) $\text{int dom } \psi \neq \emptyset$;
- (ii) ψ is convex and lower semicontinuous;
- (iii) $\psi(0) = 0$ and $\psi(t) > 0$ whenever $t > 0$.

The following characterization appears in [31]:

Proposition. *Suppose that $x \in \text{dom } f$ and $\partial f(x) \neq \emptyset$. The function f is totally convex at x if and only if there exists a function $\varphi \in \mathcal{F}$ such that, for any $y \in X$, one has*

$$f(y) - f(x) \geq f^\circ(x, y - x) + \varphi(\|y - x\|).$$

2.7 Some features of totally convex functions which make such functions interesting in applications are listed below.

Proposition (cf. [31]). *If the function f is totally convex at $x \in \text{dom } f$ and if $x^* \in \partial f(x)$, then $x^* \in \text{int}(\text{dom } f^*)$ and any of the following equivalent conditions are satisfied:*

- (i) *There exists $\varphi \in \mathcal{F}$ such that, for any $y \in X$, one has*

$$f(y) - f(x) \geq \langle x^*, y - x \rangle + \varphi(\|y - x\|);$$

- (ii) *There exists $\varphi \in \mathcal{F}$ such that, for any $y^* \in X^*$, one has*

$$f^*(y^*) - f^*(x^*) \leq \langle y^* - x^*, x \rangle + \varphi^\#(\|y^* - x^*\|);$$

- (iii) *The function f^* is Fréchet differentiable at x^* .*

If, in addition, f is lower semicontinuous, then these conditions are also equivalent to each of the following requirements:

- (iv) *There exists $\varphi \in \mathcal{F}$ such that, for any pair $(y, y^*) \in X \times X^*$ with $y^* \in \partial f(y)$, one has*

$$\langle y^* - x^*, y - x \rangle \geq \varphi(\|y - x\|);$$

- (v) *There exists a nondecreasing function $\theta : [0, +\infty) \rightarrow [0, +\infty]$ with $\lim_{t \searrow 0} \theta(t) = 0$ such that, for any pair $(y, y^*) \in X \times X^*$ with $y^* \in \partial f(y)$, one has*

$$\|y - x\| \leq \theta(\|y^* - x^*\|).$$

2.8 Under additional assumptions on the function f , one obtains the following characterization of the total convexity of f at a point:

Proposition (cf. [31]). *Suppose that f is continuous at the point $x \in \text{int}(\text{dom } f)$. Then, the following conditions are equivalent:*

- (i) *The function f is totally convex at x ;*
- (ii) *There exists $\varphi \in \mathcal{F}$ such that, for any $y \in X$ and for any $x^* \in \partial f(x)$ one has*

$$f(y) - f(x) \geq \langle x^*, y - x \rangle + \varphi(\|y - x\|);$$

- (iii) *There exists $\varphi \in \mathcal{F}$ such that, for any $y^* \in X^*$ and for any $x^* \in \partial f(x)$ one has*

$$f^*(y^*) - f^*(x^*) \leq \langle y^* - x^*, x \rangle + \varphi^\#(\|y^* - x^*\|);$$

(iv) $\partial f(x) \subseteq \text{int}(\text{dom}(f^*))$ and the function f^* is uniformly Fréchet differentiable on $\partial f(x)$.

If, in addition, the function f is lower semicontinuous, then these conditions are equivalent to each of the following requirements:

(v) There exists $\varphi \in F$ such that, for any $x^* \in \partial f(x)$ and for any pair $(y, y^*) \in X \times X^*$ with $y^* \in \partial f(y)$, one has

$$\langle y^* - x^*, y - x \rangle \geq \varphi(\|y - x\|);$$

(vi) There exists a nondecreasing function $\theta : [0, +\infty) \rightarrow [0, +\infty]$ with $\lim_{t \searrow 0} \theta(t) = 0$ such that, for any pair $(y, y^*) \in X \times X^*$ with $y^* \in \partial f(y)$, one has

$$\|y - x\| \leq \theta(\|y^* - x^*\|).$$

2.9 It follows from Proposition 2.2 that in finite dimensional Banach spaces any strictly convex function which is continuous on its domain is totally convex at any point of its domain. In infinite dimensional Banach spaces the distinction between strict convexity and total convexity is more drastic as we explain below. We start by reviewing the relationships between the modulus of total convexity and several moduli of uniform convexity. Recall that the modulus of uniform convexity of f at $x \in \text{dom } f$ is the function $\mu_f(x, \cdot) : [0, +\infty) \rightarrow [0, +\infty]$ defined by

$$\mu_f(x, t) = \inf\{f(x) + f(y) - 2f\left(\frac{x+y}{2}\right) : y \in \text{dom } f, \|y - x\| = t\}.$$

The function f is called *uniformly convex at x* if $\mu_f(x, t) > 0$, whenever $t > 0$.

The modulus of uniform convexity of f is the function $\mu_f : [0, +\infty) \rightarrow [0, +\infty]$ defined by

$$\mu_f(t) = \inf\{f(x) + f(y) - 2f\left(\frac{x+y}{2}\right) : x, y \in \text{dom } f, \|y - x\| = t\}.$$

The function f is called *uniformly convex* if $\mu_f(t) > 0$, whenever $t > 0$.

It was shown in [28] that uniform convexity of a function at a point implies total convexity at that point, while the converse is not true. The last fact is illustrated by the function $g : \ell^1 \rightarrow \mathbb{R}$ given by

$$g(x) = \sum_{k=1}^{\infty} |x_k|^{1+\frac{1}{k}}.$$

This function is totally convex at any point of the convex set

$$C = \{x \in \ell^1 : \limsup_{k \rightarrow \infty} |x_k|^{\frac{1}{k}} < 1\},$$

but it is not uniformly convex at any point of ℓ^1 . However, under additional hypothesis, one has the following result:

Proposition (cf. [31]). *Suppose that f is lower semicontinuous. If f is Fréchet differentiable at $x \in \text{dom } f$, then f is totally convex at x if and only if f is uniformly convex at x .*

2.10 In order to have a better understanding of the connections between the notion of total convexity and that of uniform convexity, recall that the modulus of uniformly strict convexity of f on the set $E \subseteq X$ is the function $\mu_f(E, \cdot) : [0, +\infty) \rightarrow [0, +\infty]$ given by

$$\mu_f(E, t) = \inf\{\mu_f(x, t) : x \in E \cap \text{dom } f\}.$$

The function f is said to be *uniformly strictly convex on bounded sets* (cf. [10]) if for each nonempty bounded subset E of X , $\mu_f(E, t) > 0$ for all $t \in (0, +\infty)$. Also, recall (cf. [28]) that the *modulus of total convexity of the function f on the set E* is the function $\nu_f(E, \cdot) : [0, +\infty) \rightarrow [0, +\infty]$ given by

$$\nu_f(E, t) = \inf \{ \nu_f(x, t) : x \in E \cap \text{dom } f \}.$$

The function f is called *totally convex on bounded sets* if $\nu_f(E, t)$ is positive for any nonempty bounded set E and any $t > 0$. It immediately results from Proposition 2.1 that, if f is totally convex on bounded sets and the set E is nonempty and bounded, then $\nu_f(E, ct) \geq c\nu_f(E, t)$ for any $c \geq 1$ and $t \geq 0$. Consequently, in this case, the function $\nu_f(E, \cdot)$ is strictly increasing on $[0, +\infty)$. The relevance of the concept of total convexity on bounded sets in convergence analysis of various algorithms (see [31], [33], [36], [28], [14]) steams from the fact that functions which are totally convex on bounded sets are sequentially consistent. Recall that the function f is called *sequentially consistent* if 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,

$$\lim_{k \rightarrow \infty} D_f(y^k, x^k) = 0 \implies \lim_{k \rightarrow \infty} \|x^k - y^k\| = 0.$$

We follow [68] and call the function f *uniformly convex on bounded sets* if for each positive real numbers r and t the function $f_r = f + \iota_{B(0,r)}$ has $\mu_{f_r}(X, t) > 0$, where $\iota_{B(0,r)}$ stands for the indicator function of the closed ball of center 0 and radius r in X .

The following result shows the connections between sequential consistency, uniform convexity on bounded sets, total convexity on bounded sets and uniformly strict convexity on bounded sets. One should note that, for lower semicontinuous functions, the sequential consistency, the total convexity on bounded sets and the uniform convexity on bounded sets are equivalent. Uniform strict convexity seems to be a stronger condition than sequential consistency, although we do not have examples of functions which are sequentially consistent without being uniformly strictly convex.

Theorem. *Let $f : X \rightarrow (-\infty, +\infty]$ be a convex function whose domain contains at least two points. The following statements hold true:*

- (i) *The function f is sequentially consistent if and only if it is totally convex on bounded sets;*
- (ii) *If f is also lower semicontinuous, then f is sequentially consistent if and only if f is uniformly convex on bounded sets;*
- (iii) *If f is uniformly strictly convex on bounded sets, then it is sequentially consistent and the converse implication holds when f is lower semicontinuous, Fréchet differentiable on its domain and the Fréchet derivative f' is uniformly continuous on bounded sets.*

Proof: The statements (i) and (ii) were proved in [31, Proposition 4.2]. In analyzing the proof of that result one should take into account that the definition of uniform convexity on bounded sets given in [31, Subsection 4.1] is incorrectly formulated. It is congruent with what we call here uniformly strict convexity on bounded sets. However, the notion of uniform convexity de facto used in the paper [31] is that occurring here. The fact that uniformly strict convexity on bounded sets implies total convexity on bounded sets follows immediately from [28, Proposition 1.2.5]. We show next that if f is lower semicontinuous, totally convex on bounded sets, Fréchet differentiable on $\text{dom } f$ and its Fréchet derivative f' is uniformly continuous on bounded sets, then f is also uniformly strictly convex on bounded sets. Note that $\text{dom } f$ is open because f is Fréchet differentiable on it. Let E be a bounded subset of X such that $E \cap \text{dom } f \neq \emptyset$. Suppose that $t > 0$ is a number such that $\nu_f(E, \frac{t}{2})$

is finite. Then, by the total convexity of f on bounded sets, we deduce that $\nu_f(E, \frac{t}{2}) > 0$. Since f' is uniformly continuous on the bounded set $E \cap \text{dom } f$, there exists a number $\tau > 0$ such that, for any $x, z \in E \cap \text{dom } f$, one has

$$\|z - x\| \leq \tau \Rightarrow \|f'(z) - f'(x)\| \leq \frac{1}{t} \nu_f(E, \frac{t}{2}). \quad (5)$$

Fix $x \in E \cap \text{dom } f$ arbitrarily. For any $y \in X$ with $\|y - x\| = t/2$ one has

$$f(y) \geq f(x) + \langle f'(x), y - x \rangle + \nu_f(E, \frac{t}{2}).$$

Let $\alpha \in (0, 1)$ be such that $1 - \alpha < \frac{2\tau}{t}$ and assume that $u \in \text{dom } f$ is such that $\|u - x\| = t$. Define

$$w = x + \frac{1-\alpha}{2}(u - x).$$

Clearly, $w \in \text{dom } f$ because it is a convex combination of $x, u \in \text{dom } f$. Note that

$$\|w - x\| = \frac{1-\alpha}{2} \|u - x\| = \frac{1-\alpha}{2} t \leq \tau,$$

and this implies that

$$\|f'(w) - f'(x)\| \leq \frac{1}{t} \nu_f(E, \frac{t}{2}) \quad (6)$$

because of (5). Obviously, one has

$$\begin{aligned} f(x) &\geq f(w) + \langle f'(w), x - w \rangle \\ &= f(w) + \frac{1-\alpha}{2} \langle f'(w), x - u \rangle \end{aligned} \quad (7)$$

and

$$\frac{1}{2}(x + u) = \frac{\alpha}{\alpha+1} u + \frac{1}{\alpha+1} w. \quad (8)$$

From (8) one deduces that

$$\frac{\alpha}{\alpha+1} f(u) + \frac{1}{\alpha+1} f(w) \geq f\left(\frac{1}{2}(x + u)\right),$$

that is,

$$\alpha f(u) + f(w) \geq (1 + \alpha) f\left(\frac{1}{2}(x + u)\right).$$

Summing up this inequality and (7) one gets

$$f(x) + \alpha f(u) \geq (1 + \alpha) f\left(\frac{1}{2}(x + u)\right) + \langle f'(w), \frac{1-\alpha}{2}(x - u) \rangle.$$

Consequently,

$$\begin{aligned} &f(x) + f(u) - 2f\left(\frac{1}{2}(x + u)\right) \\ &\geq \frac{1-\alpha}{\alpha} \left[f\left(\frac{1}{2}(x + u)\right) - f(x) + \frac{1}{2} \langle f'(w), x - u \rangle \right] \\ &\geq \frac{1-\alpha}{\alpha} \left[\nu_f(E, \frac{t}{2}) + \frac{1}{2} \langle f'(x), u - x \rangle + \frac{1}{2} \langle f'(w), x - u \rangle \right] \\ &\geq \frac{1-\alpha}{\alpha} \left[\nu_f(E, \frac{t}{2}) - \frac{1}{2} \|f'(x) - f'(w)\| \|u - x\| \right] \\ &= \frac{1-\alpha}{\alpha} \left[\nu_f(E, \frac{t}{2}) - \frac{t}{2} \|f'(x) - f'(w)\| \right] \\ &\geq \frac{1-\alpha}{\alpha} \left[\nu_f(E, \frac{t}{2}) - \frac{t}{2} \frac{\nu_f(E, \frac{t}{2})}{t} \right], \end{aligned}$$

where the last inequality results from (6). Hence, one has

$$f(x) + f(u) - 2f\left(\frac{1}{2}(x + u)\right) \geq \frac{1-\alpha}{2\alpha} \nu_f\left(x, \frac{t}{2}\right),$$

for all $u \in X$ with $\|u - x\| = t$ and this implies that

$$\mu_f(x, t) \geq \frac{1 - \alpha}{2\alpha} \nu_f \left(E, \frac{t}{2} \right) > 0.$$

Since this inequality holds for any $x \in E \cap \text{dom } f$, the proof is complete. \square

2.11 The following property of totally convex functions on bounded sets was used in the study of generic fixed point properties of some classes of operators done in [23]. Although stated in [32, Lemma 2.1] for uniformly convex functions, the proof is done in fact for functions which are totally convex on bounded sets.

Proposition. *If the function f is totally convex on bounded sets, then for each $x \in \text{dom } f$ and each $\epsilon > 0$, there exists $\delta > 0$ such that if $y, z \in \text{dom } f$, $\|y - x\| \leq \delta$ and $D_f(z, y) \leq \delta$, then $\|x - z\| \leq \epsilon$.*

2.12 In finite dimensional spaces, one of the most useful totally convex functions is the negentropy, that is, the negative of the Boltzmann-Shannon entropy (see [36]). Its extension to $\mathcal{L}^1(\Omega)$, when Ω is a nonempty bounded and measurable subset of \mathbb{R}^d , is the function $f : \mathcal{L}^1(\Omega) \rightarrow (-\infty, +\infty]$ defined by

$$f(x) = \begin{cases} \int_{\Omega} x(t) \ln x(t) dt & \text{if } x \geq 0 \text{ a.e. and } x \ln x \in \mathcal{L}^1(\Omega), \\ +\infty & \text{otherwise.} \end{cases} \quad (9)$$

This function is used in methods which aim at reconstructing a function from a finite number of moments (see [18], [19]) as well as in regularization methods for ill-posed inverse problems (see, e.g., [41], [43]). Note that the domain of the negentropy f is not equal to $\mathcal{L}_+^1(\Omega)$, but it is a much smaller set. For instance, if $\Omega = [0, 1]$, then the nonnegative function x defined by $x(t) = \frac{1}{t(\ln t - 1)^2}$ for $t \in (0, 1]$ and $x(0) = 0$, belongs to $\mathcal{L}^1(\Omega)$, while $x \ln x$ does not, that is, x is not contained in $\text{dom } f$. This implies that, contrary to what is claimed in [13, Remark 5.7], the domain of the function defined by

$$I_{\psi}(x) := f(x) - \|x\|_{\mathcal{L}^1(\Omega)}$$

for any $x \in \mathcal{L}_+^1(\Omega)$, which obviously equals that of the negentropy, is not $\mathcal{L}_+^1(\Omega)$, but a smaller set.

The interior of the domain of f is empty since the interior of $\mathcal{L}_+^1(\Omega)$ is so. Thus, the function is not differentiable, while its directional derivative $f^\circ(x, y - x)$ exists and belongs to $[-\infty, +\infty)$, for all $x, y \in \text{dom } f$. The following result shows that the extended negentropy preserves the most basic properties which make its finite counterpart of much interest in the convergence analysis of algorithms in \mathbb{R}^n .

Proposition. *The function f defined by (9) has the following properties:*

(i) *For any $x, y \in \text{dom } f$, one has*

$$\|y - x\|^2 \leq \left(\frac{2}{3} \|y\| + \frac{4}{3} \|x\| \right) D_f(y, x);$$

(ii) *For any $x \in \text{dom } f$ and $\alpha > 0$, the following level sets are weakly compact in $\mathcal{L}^1(\Omega)$:*

$$\{y \in \text{dom } f : D_f(y, x) \leq \alpha\}.$$

(iii) The function f is totally convex on bounded sets and, for any nonempty bounded set $E \subseteq \text{dom } f$, there exist two positive real numbers a and b such that, for any number $t \geq 0$, we have

$$\nu_f(E, t) \geq \frac{t^2}{at + b};$$

(iv) The function f is uniformly convex on bounded sets.

Proof: (i) See [18]. (ii) See Theorem 1.3 in [42, Chapter 8]. (iii) It follows immediately from (i). (iv) results from Theorem 2.10. \square

2.13 Total convexity is a somewhat stronger form of essentially strict convexity. Recall, cf. [13], that the function f is *essentially strictly convex* if $(\partial f)^{-1}$ is locally bounded on its domain and f is strictly convex on every convex subset of $\text{dom}(\partial f)$. This notion is of interest in the theory of some classes of fixed point algorithms ([12], [13]); it was introduced by Rockafellar [62] in \mathbb{R}^n and was further generalized and studied by Bauschke, Borwein and Combettes [13] in the context of infinite dimensional Banach spaces. The following result, establishing the connection between totally convex and essentially strictly convex functions, is a reformulation of a similar one from [61] where it was proved under the additional, but superfluous, hypotheses that $\text{dom}(\partial f)$ is convex. Since total convexity of a function is a local notion, convexity of the set on which total convexity is discussed is not necessary.

Proposition. *Suppose that X is reflexive and that f is a proper convex lower semi-continuous function which is totally convex at any point of $\text{dom}(\partial f)$. Then f is essentially strictly convex.*

2.14 The converse implication in Proposition 2.13 is not true. To see that, let $f = \frac{1}{2}\|\cdot\|^2$ and suppose that the space X is strictly convex and reflexive. Then, according to [13, Definition 5.2], the function f is essentially strictly convex. However, the function f may not be totally convex as happens in the case when X is the space contained in Vanderwerff's example (see [14, p. 3]). That space is strictly convex and reflexive but does not have the Kadec-Klee property; in such a situation f is not totally convex, as follows from Proposition 3.2 below. In finite dimensional Banach spaces, essential strict convexity seems to be slightly different from total convexity. In this respect, it is interesting to note that the most useful essentially strictly convex functions in \mathbb{R}^n (see, for instance, [12]) are also totally convex. Moreover, any function which has closed domain and which is strictly convex and continuous on its domain, as well as any strictly convex function whose domain is the entire space is totally convex (cf. [28, Proposition 1.2.6]) and essentially strictly convex (cf. Proposition 2.13). On one hand, we do not have any example of a function which simultaneously satisfies the assumptions of Proposition 2.13 and is essentially strictly convex without being totally convex. On the other hand, we do not have a proof for the equivalence of the two notions even in finite dimensional Banach spaces except for the case $\text{dom } f = \mathbb{R}^n$ discussed below.

2.15 It was noted in Subsection 2.9 that, in general, strict convexity and total convexity are not equivalent notions. However, if $\text{dom } f = \mathbb{R}^n$, then f is totally convex at any point of \mathbb{R}^n if and only if it is strictly convex as follows from Proposition 2.2. Moreover, for functions f having $\text{dom } f = \mathbb{R}^n$, the strict convexity is a property equivalent to the uniform convexity on bounded sets. This may be already known, but we do not have any reference for it. In the case of functions f defined and differentiable on the whole space \mathbb{R}^n , the equivalence between strict convexity and uniform convexity on bounded sets was implicitly shown in [45, Theorem 1.2]. Without requiring differentiability of f , but in the case of functions f with

$\text{dom } f = \mathbb{R}$, the same equivalence is proved in [68, Proposition 3.6.5]. Our next statement shows that for functions f which are finite everywhere on \mathbb{R}^n , uniform convexity on bounded sets is equivalent to strict convexity.

Theorem. *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex function. Then the following conditions are equivalent:*

- (i) f is strictly convex;
- (ii) f is essentially strictly convex;
- (iii) f is totally convex at any $x \in \mathbb{R}^n$;
- (iv) f is uniformly convex at any $x \in \mathbb{R}^n$;
- (v) f is totally convex on bounded sets;
- (vi) f is sequentially consistent;
- (vii) f is uniformly convex on bounded sets.

Proof: The equivalences (v) \Leftrightarrow (vi) \Leftrightarrow (vii) result from Theorem 2.10. The implications (v) \Rightarrow (iv) \Rightarrow (iii) \Rightarrow (ii) follow from Theorem 2.10, [28, Proposition 1.2.5] and Proposition 2.13, while the implication (ii) \Rightarrow (i) follows from [13, Theorem 5.11]. We prove here that (i) \Rightarrow (vii). To this end, one applies Theorem 3.5.10 in [68]. According to this theorem, it is sufficient to show that if E is a nonempty bounded subset of \mathbb{R}^n , then there exists a function $\phi : [0, +\infty) \rightarrow [0, +\infty]$ with $\phi(0) = 0$ which is strictly increasing and such that

$$\langle \xi - \zeta, u - v \rangle \geq \phi(\|u - v\|), \quad (10)$$

whenever $u, v \in E$ and $\xi \in \partial f(u)$, $\zeta \in \partial f(v)$. Observe that it is enough to prove that this condition is satisfied on sets E which are closed balls of center zero. So, let E be a closed ball of center zero and radius $M > 0$. Denote

$$E_t := \{(u, v, \xi, \zeta) \in \mathbb{R}^{4n} : u, v \in E, \|u - v\| = t, \xi \in \partial f(u), \zeta \in \partial f(v)\}.$$

Define $\theta : [0, +\infty) \rightarrow [0, +\infty]$ by

$$\theta(t) = \inf \{\langle \xi - \zeta, u - v \rangle : (u, v, \xi, \zeta) \in E_t\}.$$

Clearly, $\theta(t) = \infty$ for $t > 2M$, $\text{dom } \theta = [0, 2M]$ and $\theta(0) = 0$. Since f is finite everywhere and convex, it is also continuous on \mathbb{R}^n and, thus, it is bounded on bounded sets. Consequently, according to [28, Proposition 1.1.11], the multi-valued mapping $\partial f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ transforms bounded sets into bounded sets. So, the set E_t is contained in the bounded set $E \times E \times \partial f(E) \times \partial f(E)$. In order to show closedness of E_t , let $\{(u^k, v^k, \xi^k, \zeta^k)\}_{k \in \mathbb{N}}$ be a sequence contained in E_t and converging to (u, v, ξ, ζ) in \mathbb{R}^{4n} . Then, for each $k \in \mathbb{N}$, one has $\|u^k - v^k\| = t$, $\xi^k \in \partial f(u^k)$ and $\zeta^k \in \partial f(v^k)$. Since the mapping ∂f is upper semicontinuous (cf. [57, Proposition 2.5]), it follows that $\xi \in \partial f(u)$ and $\zeta \in \partial f(v)$. Obviously, one also has that $\|u - v\| = t$ and these show that $(u, v, \xi, \zeta) \in E_t$, that is, E_t is closed.

Due to the compactness of E_t , for any $t \in (0, 2M]$, there exists $u_t, v_t \in E$ and $\xi_t \in \partial f(u_t), \zeta_t \in \partial f(v_t)$ such that

$$\|u_t - v_t\| = t \text{ and } \theta(t) = \langle \xi_t - \zeta_t, u_t - v_t \rangle > 0,$$

where the last inequality follows from the strict monotonicity of ∂f because f is strictly convex. We claim that θ is strictly increasing on $[0, 2M]$. To prove that, let $0 < t_1 < t_2 \leq 2M$. Denote $\bar{u} = u_{t_2}$, $\bar{v} = v_{t_2}$, $\bar{\xi} = \xi_{t_2}$ and $\bar{\zeta} = \zeta_{t_2}$. Let $\bar{t} = t_1/t_2$ and $\bar{w} := \bar{u} + \bar{t}(\bar{v} - \bar{u})$. Clearly, $\bar{w} \in E$ and

$$\|\bar{w} - \bar{u}\| = \bar{t} \|\bar{v} - \bar{u}\| = t_1.$$

Therefore, using the definition of θ , one deduces that, for any $\eta \in \partial f(\bar{w})$, one has that

$$\theta(t_1) \leq \langle \eta - \bar{\xi}, \bar{w} - \bar{u} \rangle = \bar{t} \langle \eta - \bar{\xi}, \bar{v} - \bar{u} \rangle \leq \langle \eta - \bar{\xi}, \bar{v} - \bar{u} \rangle. \quad (11)$$

On the other hand, one has that

$$\theta(t_2) = \langle \bar{\zeta} - \bar{\xi}, \bar{v} - \bar{u} \rangle = \langle \bar{\zeta} - \eta, \bar{v} - \bar{u} \rangle + \langle \eta - \bar{\xi}, \bar{v} - \bar{u} \rangle, \quad (12)$$

where

$$\langle \bar{\zeta} - \eta, \bar{v} - \bar{u} \rangle = \bar{t}^{-1} \langle \bar{\zeta} - \eta, \bar{w} - \bar{u} \rangle > 0$$

because of the strict monotonicity of ∂f . This, (11) and (12) combined imply

$$\theta(t_2) > \langle \eta - \bar{\xi}, \bar{v} - \bar{u} \rangle \geq \theta(t_1),$$

proving that θ is strictly increasing on $[0, 2M]$.

Define the function $\phi : [0, +\infty) \rightarrow [0, +\infty]$ by

$$\phi(s) = \begin{cases} 2 \int_0^s \tau^{-1} \theta(\tau) d\tau & \text{if } s > 0, \\ 0 & \text{if } s = 0. \end{cases}$$

We are going to show that this function is well defined and satisfies condition (10). Let $t \in (0, 2M]$, let $u, v \in E$ be such that $\|u - v\| = t$ and let $\xi \in \partial f(u)$, $\zeta \in \partial f(v)$. From the definition of θ one has that

$$\langle \zeta - \xi, v - u \rangle \geq \theta(\|u - v\|).$$

For any $s \in (0, 1]$, denote $u_s := u + s(v - u)$ and take $\eta_s \in \partial f(u_s)$. Then one gets

$$\langle \eta_s - \xi, u_s - u \rangle \geq \theta(\|u_s - u\|),$$

that is,

$$\langle \eta_s - \xi, v - u \rangle \geq s^{-1} \theta(s \|v - u\|). \quad (13)$$

Letting $\tau = s \|v - u\|$ one obtains

$$\frac{\langle \eta_s - \xi, v - u \rangle}{\|v - u\|} \geq \frac{\theta(\tau)}{\tau}. \quad (14)$$

Let $\{s_k\}_{k \in \mathbb{N}} \subset (0, 1]$ be a sequence converging to zero. Then the sequence $\{u_{s_k}\}_{k \in \mathbb{N}}$ converges to u and, due to the local boundedness of the operator ∂f , the sequence $\{\eta_{s_k}\}_{k \in \mathbb{N}}$ is bounded. So, by eventually taking a subsequence, we may assume that $\{\eta_{s_k}\}_{k \in \mathbb{N}}$ converges to some vector η . By the upper semicontinuity of the operator ∂f we deduce that $\eta \in \partial f(u)$. Since in (14) the vector ξ is an arbitrary element of $\partial f(u)$, it results that

$$\frac{\langle \eta_{s_k} - \eta, v - u \rangle}{\|v - u\|} \geq \frac{\theta(\tau_k)}{\tau_k}, \quad (15)$$

where $\tau_k = s_k \|v - u\|$. So, letting $k \rightarrow \infty$ in (15), one deduces that $\lim_{\tau \rightarrow 0^+} \frac{\theta(\tau)}{\tau} = 0$, and this implies that the function ϕ is well defined. Since θ is strictly increasing, it results that $\phi(s) = 0$ if and only if $s = 0$. From the inequality (13) one also has that

$$\langle \eta_s, v - u \rangle \geq \langle \xi, v - u \rangle + s^{-1} \theta(s \|v - u\|). \quad (16)$$

If one chooses a point $\eta_s \in \partial f(u + s(v - u))$ for each $s \in [0, 1]$, then the following integral exists and

$$\int_0^1 \langle \eta_s, v - u \rangle ds = f(v) - f(u).$$

This fact was proved in a particular case in [30] and the proof can be extended without much change. Hence, integrating in (16), one gets

$$f(v) - f(u) \geq \langle \xi, v - u \rangle + \int_0^1 s^{-1} \theta(s \|v - u\|) ds.$$

Interchanging u and v in the previous arguments one deduces that

$$f(u) - f(v) \geq \langle \zeta, u - v \rangle + \int_0^1 s^{-1} \theta(s \|v - u\|) ds.$$

Adding the last two inequalities one obtains

$$\langle \zeta - \xi, v - u \rangle \geq 2 \int_0^1 s^{-1} \theta(s \|v - u\|) ds = 2 \int_0^{\|v-u\|} \frac{\theta(\tau)}{\tau} d\tau = \phi(\|v - u\|).$$

This completes the proof. \square

3 Totally Convex Banach Spaces

3.1 Identifying totally convex functions in Banach spaces of infinite dimension is a problem of practical interest in optimization and in other areas. If they happen to be totally convex, then the functions $h_r = \|\cdot\|^r$, $r > 1$, are among the most convenient such functions to work with (see [28, Chapter 3], [6]). Thus, it is natural to ask whether, and under which conditions, the function h_r is totally convex as to ensure that D -convergence and norm convergence are equivalent. Recall that, in uniformly convex Banach spaces, uniform convexity of the function h_r and then, implicitly, total convexity, was established by Asplund [10], Clarkson [38], Vladimirov et al. [66] for $r \geq 2$. Zălinescu [67] proved that for $r > 1$, the functions h_r are uniformly convex at any $x \in X$ in locally uniformly convex spaces, and thus, totally convex at any $x \in X$. This also follows from the next result which gives evaluations of the moduli of total convexity of h_r in (locally) uniformly convex Banach spaces. These evaluations were established in [30] and [31] and were used in the design and convergence analysis of the algorithms presented in [28, Chapter 3], [7] and [33].

Proposition. (i) *If X is a locally uniformly convex Banach space and $\phi(t) := |t|^r$, then, for any $t \geq 0$, one has*

$$\mu_{h_r}(0, t) = (1 - 2^{1-r}) t^r,$$

and, when $x \neq 0$ and α is any number in $(0, 1)$, one has

$$\nu_{h_r}(x, t) \geq \mu_{h_r}(x, t) \geq \|x\|^r \min \left\{ \mu_\phi \left(1, \frac{\alpha t}{\|x\|} \right), \zeta(\alpha) \right\},$$

where $\zeta : [0, 1] \rightarrow [0, \infty)$ is the function

$$\zeta(s) = \begin{cases} 1 - [2 - (1 - 2s)^{\frac{r}{r-1}}]^{1-r} & \text{if } s \in [0, \frac{1}{2}), \\ 1 - 2^{1-r} & \text{if } s \geq \frac{1}{2}, \end{cases}$$

$$\mu_\phi(1, s) = \min \{1 + (1 + s)^r - 2^{1-r}(2 + s)^r, 1 + |1 - s|^r - 2^{1-r}|2 - s|^r\},$$

and

$$a := \delta_X \left(\frac{x}{\|x\|}, (1 - \alpha) \frac{t}{\|x\|} \right);$$

(ii) If X is uniformly convex and $r \in (1, \infty)$, then the function h_r is totally convex, and, for any $z \in X$ and $t > 0$, one has:

$$\nu_{h_r}(z, t) \geq rM \left(\frac{t}{2} \right)^r \int_0^1 \tau^{r-1} \delta_X \left(\frac{\tau t}{2(\|z\| + \tau t)} \right) d\tau > 0,$$

where M is some positive constant and δ_X is the modulus of convexity of X .

3.2 It is known that geometric properties of Banach spaces, such as strict convexity, locally uniform convexity, uniform convexity can be characterized analytically by properties of the square of the norm (see [37]). In [61], a Banach space X is called *locally totally convex* if the function $h = \frac{1}{2}\|\cdot\|^2$ is totally convex at each $x \in X$.

Locally uniformly convex spaces are locally totally convex; separable or reflexive Banach spaces can be equivalently renormed for becoming locally uniformly convex spaces (see [50] and [65], respectively).

It follows from [68, Section 3.7] that (locally) uniform convexity of a Banach space is equivalent to (locally) uniform convexity of a large class of functions Φ of the norm, where

$$\Phi(x) = \int_0^{\|x\|} \varphi(s) ds \tag{17}$$

and φ is a weight function. An analogous result is true in the case of total convexity, as shown by the result below. Recall that the Banach space X is said to be an *E-space* if it is reflexive, strictly convex and has the Kadec-Klee property. According to [37, p. 47], a Banach space is an *E-space* if and only if it is *weakly uniformly convex*, that is, for any two sequences $\{x^n\}_{n \in \mathbb{N}}, \{y^n\}_{n \in \mathbb{N}} \subseteq X$ with $\|x^n\| = \|y^n\| = 1, n \in \mathbb{N}$, one has

$$\lim_{n \rightarrow \infty} \langle x^*, x^n + y^n \rangle = 2 \quad \Rightarrow \quad \lim_{n \rightarrow \infty} \|x^n - y^n\| = 0,$$

whenever $x^* \in X^*, \|x^*\| = 1$. The *E-spaces*, introduced by Fan and Glicksberg [44], were studied since they are the natural setting for the research of strong Tykhonov and Hadamard well-posedness of convex best approximation problems (see, e.g., [39], [46]), and variational inequalities in Banach spaces (see, e.g., [6]).

Proposition (cf. [61]). *The following conditions are equivalent:*

- (i) X is locally totally convex;
 - (ii) There exists a weight function φ such that the function Φ defined at (17), is totally convex at any $x \in X$;
 - (iii) For any weight function φ , the function Φ is totally convex at any $x \in X$.
 - (iv) The function $\frac{1}{2}\|\cdot\|^2$ is totally convex at each point of the unit sphere of X .
- Moreover, if X is reflexive, then these conditions are also equivalent to the following:
- (v) X is an *E-space*;
 - (vi) X has the Kadec-Klee property.

3.3 The next result shows another characterization of locally totally convex spaces and points out some geometric properties of the norm in this context. It was suggested to us by Constantin Zălinescu. In view of Proposition 3.2, if X is reflexive, then this result provides also a characterization of the E -spaces. In what follows, we denote by S_X the unit sphere of X .

Theorem. *Let X be a Banach space. Then the following conditions are equivalent:*

- (i) *The space X is locally totally convex;*
- (ii) *For any $x \in S_X$ and for any real number $\varepsilon > 0$, there exists $\delta = \delta(x, \varepsilon) > 0$ such that, for all $y \in S_X$ with $\|y - x\| \geq \varepsilon$, there exists $\lambda_0 \in (0, 1)$ such that*

$$\|(1 - \lambda_0)x + \lambda_0 y\| < 1 - \lambda_0 \delta; \quad (18)$$

- (iii) *For any $x \in S_X$ and for any real number $\varepsilon > 0$, there exists $\delta = \delta(x, \varepsilon) > 0$ such that, for all $y \in S_X$ with $\|y - x\| = \varepsilon$, there exists $\lambda_0 \in (0, 1)$ such that the inequality (18) holds.*

Proof: (i) \Rightarrow (ii): Let $h = \frac{1}{2}\|\cdot\|^2$. Suppose, by contradiction, that for some $\|x^0\| = 1$, for some $\varepsilon_0 > 0$ and for any $n \in \mathbb{N}$, there is $y^n \in X$ with $\|y^n\| = 1$ and $\|y^n - x^0\| \geq \varepsilon_0$ such that

$$\|x^0 + \lambda(y^n - x^0)\| \geq 1 - \lambda/n = \|x^0\| - \lambda/n,$$

whenever $\lambda \in (0, 1)$. Subtracting $\|x^0\|$, dividing the resulting inequality by λ and letting $\lambda \rightarrow 0^+$, one gets $-g^\circ(x^0, y^n - x^0) \leq 1/n$, where $g = \|\cdot\|$. Consequently, $D_h(y^n, x^0) \leq 1/n$ for any $n \in \mathbb{N}$, since $h^\circ(x^0, y^n - x^0) = \|x^0\| \cdot g^\circ(x^0, y^n - x^0)$. By Proposition 2.5 and the total convexity of h at x^0 , it follows that $\{y^n\}_{n \in \mathbb{N}}$ converges to x^0 , which is a contradiction.

(ii) \Rightarrow (i): According to Proposition 2.1(v) and Theorem 3.1(iv), it is sufficient to show that

$$\inf\{D_h(y, x) : \|y - x\| \geq \varepsilon\} > 0, \quad (19)$$

for any $x \in S_X$ and $\varepsilon > 0$. Take $y \in X$ with $\|y - x\| \geq \varepsilon$. Denote $v = y/\|y\|$ and $\varepsilon_0 = \varepsilon/2$. Then, for some $\delta = \delta(x, \varepsilon) > 0$ and for any $w \in S_X$ with $\|x - w\| \geq \varepsilon_0$, there is $\lambda_0 \in (0, 1)$ such that

$$1 - \|x + \lambda_0(w - x)\| > \lambda_0 \delta. \quad (20)$$

Observe that

$$D_h(y, x) \geq \frac{1}{2}\|y\|^2 - \frac{1}{2} - \frac{\|x + \lambda(y - x)\|^2 - 1}{2\lambda} := \zeta_\lambda(x, y) \quad (21)$$

for any $\lambda \in (0, 1)$, because

$$h^\circ(x, y - x) \leq \frac{h(x + \lambda(y - x)) - h(x)}{\lambda}.$$

We distinguish the following possible situations:

Case 1: If $\|x\| - \|y\| \geq \varepsilon/2$, then

$$\zeta_{1/2}(x, y) \geq \frac{1}{4}(\|x\| - \|y\|)^2 \geq \frac{1}{16}\varepsilon^2.$$

By using (21) and taking infimum over $y \in X$ such that $\|y - x\| \geq \varepsilon$, inequality (19) follows.

Case 2: Suppose that $0 \leq \|x\| - \|y\| < \varepsilon/2$. Denote $\alpha = 1/\|y\|$, $\beta = (\lambda_0 + \alpha^2(1 - \lambda_0))^{1/2}$, $z = x + \lambda_0(v - x)$, and $\gamma = \|z\| + (1 - \lambda_0)(\alpha - 1)$. Then one has

$$x + \lambda_0(y - x) = \|y\|[z + (1 - \lambda_0)(\alpha - 1)x].$$

Note that

$$\|x + \lambda_0(y - x)\| \leq \|y\|[\|z\| + (1 - \lambda_0)(\alpha - 1)].$$

Then one gets

$$\begin{aligned} D_h(y, x) \geq \zeta_{\lambda_0}(x, y) &\geq \frac{\|y\|^2}{2} \left(1 - \alpha^2 - \frac{\gamma^2 - \alpha^2}{\lambda_0}\right) \\ &= \frac{\|y\|^2}{2\lambda_0} (\beta + \gamma)(\beta - \gamma) \geq \frac{\|y\|^2}{2\lambda_0} (1 - \|z\|), \end{aligned}$$

where the last inequality holds because $\beta + \gamma \geq 1$ and $\beta \geq 1 + (1 - \lambda_0)(\alpha - 1)$. Observe that

$$\|v - x\| = \left\|x - y + \frac{\|y\| - 1}{\|y\|}y\right\| \geq [\|x - y\| - (1 - \|y\|)] \geq \varepsilon/2 = \varepsilon_0.$$

If $\|y\| \geq \varepsilon/2$ then, by using (20), one obtains that $\zeta_{\lambda_0}(x, y) \geq \frac{\delta\varepsilon^2}{8}$. Otherwise, one has $\|x\| > \varepsilon/2$ because $\|y - x\| \geq \varepsilon$. Consequently, $\|y\| > \|x\| - \varepsilon/2 > 0$ and, thus,

$$\zeta_{\lambda_0}(x, y) \geq \frac{\delta}{2} \left(\|x\| - \frac{\varepsilon}{2}\right)^2 > 0.$$

Therefore, in both situations, inequality (19) holds.

Case 3: Suppose that $0 \leq \|y\| - \|x\| < \varepsilon/2$. If one denotes $\alpha = \|y\|$ and preserves the notations for β and γ , then the argument in this case is similar to that in Case 2.

Case 4: If $\|y\| - \|x\| \geq \varepsilon/2$, then one proceeds as in Case 1 with x and y interchanged.

The proof of $(i) \Leftrightarrow (iii)$ is identical to that of $(i) \Leftrightarrow (ii)$, due to Proposition 2.1(v). \square

3.4 As pointed out in Subsection 2.9, there are totally convex functions which are not locally uniformly convex. However, we do not know any example of a Banach space which is locally totally convex without being locally uniformly convex. The following proposition seems to support our conjecture that locally totally convex spaces are not necessarily locally uniformly convex. The key fact is that λ_0 in the statement below does not depend on y , while in the previous theorem it does.

Proposition. *Let X be a Banach space. Then the following conditions are equivalent:*

- (i) *The space X is locally uniformly convex;*
- (ii) *For any $x \in S_X$ and for any real number $\varepsilon > 0$, there exists $\delta = \delta(x, \varepsilon) > 0$ and $\lambda_0 \in (0, 1)$ such that, for all $y \in S_X$ with $\|y - x\| \geq \varepsilon$, one has*

$$\|(1 - \lambda_0)x + \lambda_0y\| < 1 - \lambda_0\delta. \quad (22)$$

Proof: $(i) \Rightarrow (ii)$: Let $x \in S_X$ and $\varepsilon > 0$. From the definition of locally uniformly convex spaces, there exists $\delta > 0$ and $\lambda_0 = 1/2$ such that (22) holds.

$(ii) \Rightarrow (i)$ It follows immediately. \square

4 Relative Projections onto Closed Convex Sets

4.1 In this section we present the basic properties of a class of operators we call relative projections. They are natural generalizations of the Bregman projections introduced in [20] and of the generalized projections defined and studied by Alber [2], [4]. In the sequel we

assume that the Banach space X is reflexive and $f : X \rightarrow (-\infty, +\infty]$ represents a proper lower semicontinuous function which is strictly convex on its domain $\text{dom } f$. As usual, we denote by $f^* : X^* \rightarrow (-\infty, +\infty]$ the *Fenchel conjugate of f* , that is, the function defined by

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

We follow Alber [4] and, with the function f we associate the function $W^f : X^* \times X \rightarrow (-\infty, +\infty]$ defined by

$$W^f(\xi, x) = f(x) - \langle \xi, x \rangle + f^*(\xi).$$

Clearly, W^f is nonnegative and, for any $\xi \in \text{dom } f^*$, the function $W^f(\xi, \cdot)$ is strictly convex.

4.2 For any $\xi \in X^*$ and for any subset K of X denote

$$W^f(\xi, K) = \inf \{ W^f(\xi, x) : x \in K \}.$$

An element $y \in K$ such that

$$W^f(\xi, y) = W^f(\xi, K) \tag{23}$$

is called *projection of ξ onto K relative to f* . The following result shows sufficient conditions for existence of such projections. In the case of a coercive function f , this result, as well as Proposition 4.4 below occur in [4].

Proposition. *Suppose that $\xi \in \text{dom } f^*$ and that the set $K \subseteq \text{dom } f$ is nonempty, closed and convex. If the set*

$$R_\alpha^f(\xi) = \{ z \in \text{dom } f : W^f(\xi, z) \leq \alpha \}$$

is bounded for any $\alpha \in [0, +\infty)$, then there exists a unique projection of ξ onto K relative to f . In particular, this happens if any of the following conditions holds:

- (i) $\xi \in \text{range}(\partial f)$ and f is totally convex;
- (ii) f is coercive (in the sense that $\lim_{\|x\| \rightarrow \infty} \frac{f(x)}{\|x\|} = +\infty$).

Proof: It was noted in Subsection 4.1 that the function $W^f(\xi, \cdot)$ is strictly convex. This guarantees that there is no more than one vector y satisfying (23). Since the function $W^f(\xi, \cdot)$ is also convex, lower semicontinuous and $R_\alpha^f(\xi)$ is bounded for each $\alpha \in [0, +\infty)$, it results that $W^f(\xi, \cdot)$ has at least one minimizer in the convex set K , that is, the projection of ξ onto K relative to f exists.

Suppose that $\xi \in \text{range}(\partial f)$, f is totally convex and that, for some $\alpha \geq 0$, the set $R_\alpha^f(\xi)$ is unbounded. Then there exists a sequence $\{y^k\}_{k \in \mathbb{N}}$ contained in $R_\alpha^f(\xi)$ such that $\lim_{k \rightarrow \infty} \|y^k\| = +\infty$. According to Proposition 2.1, for any $x \in \text{dom } f$ and for any natural number k such that $\|y^k - x\| \geq 1$, one has

$$\begin{aligned} \alpha &\geq W^f(\xi, y^k) = D_f(y^k, x) + f^\circ(x, y^k - x) - \langle \xi, y^k - x \rangle + W^f(\xi, x) \\ &\geq \nu_f(x, 1) \|y^k - x\| + f^\circ(x, y^k - x) - \langle \xi, y^k - x \rangle. \end{aligned} \tag{24}$$

Since $\xi \in \text{range}(\partial f)$, there exists a vector $x^0 \in \text{dom } \partial f$ such that $\xi \in \partial f(x^0)$. One can check (or see, e.g., [57, p. 41]) that

$$f^\circ(x^0, y^k - x^0) \geq \langle \xi, y^k - x^0 \rangle.$$

This, combined with (24), implies that

$$\alpha \geq \nu_f(x^0, 1) \|y^k - x^0\|. \tag{25}$$

Since f is totally convex at x^0 , it results that $\nu_f(x^0, 1) > 0$. Therefore, by letting $k \rightarrow \infty$ in (25), one gets a contradiction. Hence, the set $R_\alpha^f(\xi)$ is bounded for all $\alpha \geq 0$.

Now, suppose that f is coercive and that, for some $\alpha \geq 0$, there exists a sequence $\{x^k\}_{k \in \mathbb{N}}$ contained in $R_\alpha^f(\xi)$ such that $\lim_{k \rightarrow \infty} \|x^k\| = +\infty$. Then

$$\alpha \geq W^f(\xi, x^k) \geq f^*(\xi) + \|x^k\| \left(\frac{f(x^k)}{\|x^k\|} - \|\xi\|_* \right)$$

and letting here $k \rightarrow +\infty$ one gets a contradiction. Hence, in this case too, the set $R_\alpha^f(\xi)$ is bounded for all $\alpha \geq 0$. \square

4.3 Whenever it exists, we denote by $P_K^f(\xi)$ the projection of ξ onto $K \subseteq \text{dom } f$ with respect to f . The function $\xi \rightarrow P_K^f(\xi) : \text{dom } P_K^f \rightarrow K$ is called the *projection operator onto K relative to f* . Proposition 4.2 shows that if f is totally convex, then $\text{range}(\partial f) \subseteq \text{dom } P_K^f$. Also, if f is coercive, then $\text{dom } P_K^f = \text{dom } f^*$. One can ask whether coercivity implies total convexity or vice-versa. The examples below show that this is not the case. First, consider the situation when $X = \mathbb{R}$ and $f : [0, +\infty) \rightarrow [0, +\infty)$ is given by $f(x) = \sqrt{x}$. This function is not coercive, while, by Proposition 2.2, it is totally convex at any point of its domain. Also, it was noted in Subsection 2.14 that there are Banach spaces on which the coercive function $\frac{1}{2}\|\cdot\|^2$ is strictly convex (even Gâteaux differentiable - see [14]) without being totally convex at all points of the space.

Note that the existence of relative projections onto closed convex sets does not involve differentiability requirements on f . However, in applications, the case when f is Gâteaux differentiable is of most interest. From now on, and until the end of this section, we presume that the following condition holds:

Assumption 1. *The function f is Gâteaux differentiable on $\text{int}(\text{dom } f)$.*

In this situation, one has that

$$f^*(f'(x)) = \langle f'(x), x \rangle - f(x), \quad (26)$$

for any $x \in \text{int}(\text{dom } f)$ and, consequently,

$$D_f(y, x) = W^f(f'(x), y).$$

Recall that the *Bregman projection of x onto the nonempty closed convex set $K \subseteq \text{dom } f$* (as defined in [20]), is the necessarily unique vector $\Pi_K^f(x) \in K$, satisfying

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

According to Proposition 4.2, since $f'(x) \in \text{int}(\text{dom } f^*)$ for all $x \in \text{int}(\text{dom } f)$, one has that

$$\Pi_K^f(x) = P_K^f(f'(x)) = (P_K^f \circ f')(x). \quad (27)$$

4.4 Similarly to the metric projection in Hilbert spaces, Bregman projections and generalized projections (i.e., the projections relative to $h = \frac{1}{2}\|\cdot\|^2$) have variational characterizations - see [2] and [5]. This property extends to relative projections with respect to totally convex differentiable functions.

Proposition. *Let f be totally convex at any point of $\text{int}(\text{dom } f)$ and let $K \subseteq \text{int}(\text{dom } f)$ be a nonempty, closed and convex set. If $\xi \in X^*$ and $\hat{x} \in K$, then the following conditions are equivalent:*

- (a) The vector \hat{x} is the projection of ξ onto K relative to f ;
(b) The vector \hat{x} is the unique solution in K of the following variational inequality:

$$\langle \xi - f'(z), z - y \rangle \geq 0, \quad \forall y \in K; \quad (28)$$

- (c) The vector \hat{x} is the unique solution contained in K of the following variational inequality:

$$W^f(\xi, z) + W^f(f'(z), y) \leq W^f(\xi, y), \quad \forall y \in K.$$

Proof: We prove first that (a) and (b) are equivalent. Suppose that (a) holds. Then, for any $x \in K$ one has $W^f(\xi, \hat{x}) \leq W^f(\xi, x)$. In particular, this holds for $x = (1 - \tau)\hat{x} + \tau y$ for all $y \in K$ and for all $\tau \in [0, 1]$. Hence, one obtains

$$0 \geq W^f(\xi, \hat{x}) - W^f(\xi, (1 - \tau)\hat{x} + \tau y) \geq \left\langle [W^f(\xi, \cdot)]'((1 - \tau)\hat{x} + \tau y), \tau(\hat{x} - y) \right\rangle,$$

where $[W^f(\xi, \cdot)]' = f' - \xi$. Therefore, for any $\tau \in (0, 1]$, one has that

$$0 \geq \langle f'((1 - \tau)\hat{x} + \tau y) - \xi, \hat{x} - y \rangle$$

and, letting here $\tau \rightarrow 0^+$, one obtains (28) because the function $\langle f'(\cdot) - \xi, \hat{x} - y \rangle$ is continuous due to the norm-to-weak* continuity of f' . Now, suppose that $\hat{x} \in K$ satisfies (28). Then, for any $y \in K$, one has

$$W^f(\xi, y) - W^f(\xi, \hat{x}) \geq \langle [W^f(\xi, \cdot)]'(\hat{x}), y - \hat{x} \rangle = \langle f'(\hat{x}) - \xi, y - \hat{x} \rangle \geq 0,$$

showing that \hat{x} minimizes $W^f(\xi, \cdot)$ over K , i.e., $\hat{x} = P_K^f(\xi)$. In order to show that (b) and (c) are equivalent, it is sufficient to observe that

$$W^f(\xi, \hat{x}) + W^f(f'(\hat{x}), y) - W^f(\xi, y) = \langle f'(\hat{x}) - \xi, \hat{x} - y \rangle$$

for any $y \in K$. \square

4.5 Combining Proposition 4.4 with (27), one re-finds the variational characterizations of the Bregman projections given in [28].

Corollary. *Suppose that f is totally convex at any point of $\text{int}(\text{dom } f)$. Let $x \in \text{int}(\text{dom } f)$ and let $K \subseteq \text{int}(\text{dom } f)$ be a nonempty, closed and convex set. If $\hat{x} \in K$, then the following conditions are equivalent:*

- (i) The vector \hat{x} is the Bregman projection of x onto K with respect to f ;
(ii) The vector \hat{x} is the unique solution of the variational inequality

$$\langle f'(x) - f'(z), z - y \rangle \geq 0, \quad \forall y \in K;$$

- (iii) The vector \hat{x} is the unique solution of the variational inequality

$$D_f(y, z) + D_f(z, x) \leq D_f(y, x), \quad \forall y \in K.$$

4.6 If X is a Hilbert space and $f = \frac{1}{2} \|\cdot\|^2$, then the operators P_K^f and Π_K^f coincide and are equal to the metric projection operator onto the set K . It is known (see, for instance, [39]) that the metric projection operators in Hilbert spaces have valuable continuity and stability properties. Such properties were also proved in [61] for Bregman projections with respect to totally convex functions. As shown by the next result, which extends [61, Theorem 2.2.12] and [47, Theorem 3.1], some of these properties remain true for the relative projection operators P_K^f in any reflexive Banach space. Recall (cf. [54]) that a sequence $\{K_n\}_{n \in \mathbb{N}}$ of subsets of X is called *convergent (in Mosco's sense) to the set K* (we write *M -convergent* for short) if

$$K = w - \overline{\lim}_{n \rightarrow \infty} K_n = s - \underline{\lim}_{n \rightarrow \infty} K_n,$$

where $s - \underline{\lim}_{n \rightarrow \infty} K_n$ represents the collection of all $y \in X$ which are limits (in the strong convergence sense) of sequences with the property that $x^n \in K_n$ for all $n \in \mathbb{N}$ and $w - \overline{\lim}_{n \rightarrow \infty} K_n$ denotes the collection of all $x \in X$ such that there exists a sequence $\{y^n\}_{n \in \mathbb{N}}$ in X converging weakly to x and with the property that there exists a subsequence $\{K_{i_n}\}_{n \in \mathbb{N}}$ of $\{K_n\}_{n \in \mathbb{N}}$ such that $y^n \in K_{i_n}$ for all $n \in \mathbb{N}$.

Theorem. *Suppose that f is lower semicontinuous and at least one of the following conditions is satisfied:*

- (i) f is coercive;
- (ii) f is totally convex on bounded sets;

If K and K_n , ($n \in \mathbb{N}$), are nonempty closed convex subsets of $\text{int}(\text{dom } f)$ and if the sequence $\{K_n\}_{n \in \mathbb{N}}$ is M -convergent to K , then for any sequence $\{\xi^n\}_{n \in \mathbb{N}}$ contained in $\text{int}(\text{dom } f^)$ which converges to some $\xi \in \text{int}(\text{dom } f^*)$, the following limits exist and we have*

$$\lim_{n \rightarrow \infty} W^f(\xi^n, K_n) = W^f(\xi, K) \quad (29)$$

and

$$w - \lim_{n \rightarrow \infty} P_{K_n}^f(\xi^n) = P_K^f(\xi). \quad (30)$$

Proof: Denote $\hat{x}^n = P_{K_n}^f(\xi^n)$ and $\hat{x} = P_K^f(\xi)$. Suppose that $\{z^n\}_{n \in \mathbb{N}}$ is a sequence in X such that $z^n \in K_n$ for all $n \in \mathbb{N}$ and $\lim_{n \rightarrow \infty} z^n = z$. Since $\{K_n\}_{n \in \mathbb{N}}$ is M -convergent to K it results that $z \in K$. By Proposition 4.4 one deduces that

$$W^f(\xi^n, \hat{x}^n) + W^f(f'(\hat{x}^n), z^n) \leq W^f(\xi^n, z^n), \quad \forall n \in \mathbb{N}. \quad (31)$$

Since f and f^* are lower semicontinuous they are continuous on the interior of their respective domains. Consequently, W^f is continuous on $\text{int}(\text{dom } f^*) \times \text{int}(\text{dom } f)$ and, therefore, the sequence $\{W^f(\xi^n, z^n)\}_{n \in \mathbb{N}}$ is convergent. By (31), this implies that the sequence $\{W^f(\xi^n, \hat{x}^n)\}_{n \in \mathbb{N}}$ is bounded. We claim that $\{\hat{x}^n\}_{n \in \mathbb{N}}$ is bounded too. In order to show that, suppose first that f is coercive. Then the following inequality, which results immediately from the definition of $W^f(\xi^n, \hat{x}^n)$, can not hold unless $\{\hat{x}^n\}_{n \in \mathbb{N}}$ is bounded:

$$W^f(\xi^n, \hat{x}^n) \geq f^*(\xi^n) + \|\hat{x}^n\| \left(\frac{f(\hat{x}^n)}{\|\hat{x}^n\|} - \|\xi^n\| \right)$$

Now, assume that f is totally convex on bounded sets. By Theorem 2.10, the function f is also uniformly convex on bounded sets. Thus, according to [68, Proposition 2.4.13], for each $n \in \mathbb{N}$, one has that

$$\xi^n \in \text{int}(\text{dom } f^*) = \text{int}(\text{dom } \partial f^*) = \text{int}(\text{range}(\partial f)) = \text{int}(\text{range}(f')).$$

Thus, for each $n \in \mathbb{N}$, there exists $u^n \in \text{int}(\text{dom } f)$ such that $f'(u^n) = \xi^n$. The sequence $\{u^n\}_{n \in \mathbb{N}}$ is contained in the set $\bigcup_{n \in \mathbb{N}} \partial f^*(\xi^n)$. Since the operator ∂f^* is monotone, it is also locally bounded and, so, there exists a neighborhood V of ξ such that $\bigcup_{\zeta \in V} \partial f^*(\zeta)$ is bounded. The sequence $\{\xi^n\}_{n \in \mathbb{N}}$ being convergent to ξ , there exists a positive integer n_0 such that $\xi^n \in V$ for all $n \geq n_0$. Hence, for any $n \geq n_0$ one has

$$u^n \in \partial f^*(\xi^n) \subseteq \bigcup_{\zeta \in V} \partial f^*(\zeta),$$

and this implies that the sequence $\{u^n\}_{n \in \mathbb{N}}$ is bounded. Thus, the function $\nu_f(\{u^n\}_{n \in \mathbb{N}}, \cdot)$ is strictly increasing (see Subsection 2.10). Observe that

$$\nu_f(\{u^n\}_{n \in \mathbb{N}}, \|\hat{x}^n - u^n\|) \leq D_f(\hat{x}^n, u^n) = W^f(\xi^n, \hat{x}^n),$$

where the sequence $\{W^f(\xi^n, \hat{x}^n)\}_{n \in \mathbb{N}}$ is bounded. Since the sequence $\{u^n\}_{n \in \mathbb{N}}$ is also bounded, this inequality can not hold unless the sequence $\{\hat{x}^n\}_{n \in \mathbb{N}}$ is bounded too.

We show next that the only weak accumulation point of $\{\hat{x}^n\}_{n \in \mathbb{N}}$ is \hat{x} . Clearly, this implies (30). To this end, note that, if $\{\zeta^n\}_{n \in \mathbb{N}}$ is strongly convergent to ζ in X^* and if $\{u^n\}_{n \in \mathbb{N}}$ is weakly convergent to u in X , then

$$W^f(\zeta, u) \leq \liminf_{n \rightarrow \infty} W^f(\zeta^n, u^n) \quad (32)$$

because of the weak lower semicontinuity of f and the lower semicontinuity of f^* . Since $\hat{x} \in K$ and K is the Mosco limit of the sequence of sets $\{K_n\}_{n \in \mathbb{N}}$, there exists a sequence $\{v^n\}_{n \in \mathbb{N}}$ in X such that $v^n \in K_n$ for all $n \in \mathbb{N}$ and $\lim_{n \rightarrow \infty} v^n = \hat{x}$. Let \bar{x} be an arbitrary weak accumulation point of $\{\hat{x}^n\}_{n \in \mathbb{N}}$ and let $\{\hat{x}^{i_n}\}_{n \in \mathbb{N}}$ be a subsequence of $\{\hat{x}^n\}_{n \in \mathbb{N}}$ which converges weakly to \bar{x} . Clearly, $\bar{x} \in K$ and, by (32), one has

$$\begin{aligned} W^f(\xi, \hat{x}) &\leq W^f(\xi, \bar{x}) \leq \liminf_{n \rightarrow \infty} W^f(\xi^{i_n}, \hat{x}^{i_n}) \\ &\leq \liminf_{n \rightarrow \infty} W^f(\xi^{i_n}, v^{i_n}) = \lim_{n \rightarrow \infty} W^f(\xi^n, v^n) = W^f(\xi, \hat{x}). \end{aligned}$$

This shows that $W^f(\xi, \hat{x}) = W^f(\xi, \bar{x})$ where \hat{x} is the only minimizer of $W^f(\xi, \cdot)$ in K . Consequently, one gets $\hat{x} = \bar{x}$. Since \bar{x} is an arbitrary weak accumulation point of $\{\hat{x}^n\}_{n \in \mathbb{N}}$, one obtains (30).

In order to prove (29) observe that, since $w - \lim_{n \rightarrow \infty} \hat{x}^n = \hat{x}$, one can use (32) again in order to deduce that

$$\begin{aligned} W^f(\xi, \hat{x}) &\leq \liminf_{n \rightarrow \infty} W^f(\xi^n, \hat{x}^n) \leq \limsup_{n \rightarrow \infty} W^f(\xi^n, \hat{x}^n) \\ &\leq \lim_{n \rightarrow \infty} W^f(\xi^n, v^n) = W^f(\xi, \hat{x}). \end{aligned}$$

This shows that $\lim_{n \rightarrow \infty} W^f(\xi^n, \hat{x}^n) = W^f(\xi, \hat{x})$, i.e., (29) holds. \square

4.7 A direct consequence of Theorem 4.6 is the monotonicity of the operator P_K^f .

Proposition. *Suppose that f is either coercive or totally convex at each point of its domain. Then, for any nonempty closed convex subset K of $\text{dom } f$, the operator P_K^f is monotone. Moreover, if $\text{dom } P_K^f = X^*$ and $K \subseteq \text{int}(\text{dom } f)$, then P_K^f is maximal monotone.*

Proof: According to the definition of P_K^f , for any $\xi, \zeta \in \text{dom } P_K^f$, one has that

$$W^f(\xi, P_K^f(\xi)) \leq W^f(\xi, P_K^f(\zeta))$$

and

$$W^f(\zeta, P_K^f(\zeta)) \leq W^f(\zeta, P_K^f(\xi)).$$

Adding these inequalities one obtains

$$\langle P_K^f(\xi) - P_K^f(\zeta), \xi - \zeta \rangle \geq 0,$$

showing the operator P_K^f is monotone. According to Theorem 4.6, the monotone operator P_K^f is hemicontinuous. Now, suppose that $\text{dom } P_K^f = X^*$ and $K \subseteq \text{int}(\text{dom } f)$. Applying [56, Corollary, p. 106], maximal monotonicity of P_K^f results. \square

4.8 Computing relative projections may not be an easy task. Clearly, for any $\xi \in \text{int}(\text{dom } f^*)$, we have

$$P_X^f(\xi) = f^{*'}(\xi)$$

because

$$W^f(\xi, f^{*'}(\xi)) = f^*(\xi) - \langle \xi, f^{*'}(\xi) \rangle + f(f^{*'}(\xi)).$$

It is easy to see that (26) applies with f and f^* interchanged, because $f = f^{**}$. If K is a hyperplane or a half space, then formulas for computing the Bregman projection $P_K^f(f'(x)) = \Pi_K^f(x)$ are given in [5] and [30] in the case when $f = \|\cdot\|^p$ for some $p \in (1, \infty)$ and X is smooth and strictly convex. Iterative algorithms for computing Bregman projections on more general sets are presented in [14]. The following result gives a formula for computing relative projections onto closed hyperplanes and half spaces. It applies, for instance, when X is smooth and $f = \|\cdot\|^p$ with $p \in (1, \infty)$.

Theorem. *Suppose that $\text{dom } f = X$, f is totally convex at any $x \in X$, $\text{dom } f^* = X^*$, $f^{*'}$ is positively homogeneous of degree $q > 0$, $a \in X^* \setminus \{0\}$ and $b \in \mathbb{R}$.*

(i) *If $K = \{x \in X : \langle a, x \rangle = b\}$, then*

$$P_K^f(\xi) = f^{*'}(sa + \xi)$$

where s is a necessarily existing solution of the equation

$$\langle a, f^{*'}(sa + \xi) \rangle = b. \quad (33)$$

(ii) *If $K = \{x \in X : \langle a, x \rangle \geq b\}$, then*

$$P_K^f \xi = \begin{cases} f^{*'}(sa + \xi) & \text{if } f^{*'}(\xi) \notin K \\ f^{*'}(\xi) & \text{if } f^{*'}(\xi) \in K, \end{cases}$$

where s is a necessarily existing positive solution of the equation (33).

Proof: We first justify that $\text{dom } f^{*'} = X^*$. Indeed, this is so because the function f is essentially strictly convex (cf. Proposition 2.13) and, according to Theorem 5.9 in [13], it satisfies $\text{dom } f^{*'} = \text{int}(\text{dom } f^*)$ which is exactly X^* in our context.

Now we show that there exists $s \in \mathbb{R}$ such that $\langle a, f^{*'}(sa + \xi) \rangle = b$. To this end, consider the function

$$\psi(s) = \langle a, f^{*'}(sa + \xi) \rangle - b.$$

Clearly, $\psi(0) = \langle a, f^{*'}(\xi) \rangle - b$. If $f^{*'}(\xi) \in K$, then $s = 0$ is the required solution. Otherwise, suppose that $\psi(0) < 0$ and observe that the function ψ is continuous. By the homogeneity of $f^{*'}$ one has $f^{*'}(0) = 0^q f^{*'}(0) = 0$ and, thus, one obtains that

$$\langle a, f^{*'}(a) \rangle = \langle a - 0, f^{*'}(a) - f^{*'}(0) \rangle > 0,$$

due to the strict monotonicity of $f^{*'}$. Consequently, for any $s > 0$, one has

$$\psi(s) = s^q \langle a, f^{*'}(a + s^{-1}\xi) \rangle - b.$$

Taking into account that $f^{*'}$ is norm-to-weak* continuous, one deduces that

$$\lim_{s \rightarrow \infty} \psi(s) = \langle a, f^{*'}(a) \rangle \lim_{s \rightarrow \infty} s^q - b = \infty.$$

This, the fact that $\psi(0) < 0$, and the continuity of ψ imply that ψ has a root s in the interval $(0, \infty)$. Now, observe that for the root s of ψ and for all $y \in K$ one has

$$\begin{aligned} \langle \xi - f'(f^{*'}(sa + \xi)), f^{*'}(sa + \xi) - y \rangle &= \langle \xi - (sa + \xi), f^{*'}(sa + \xi) - y \rangle \\ &= -s \langle a, f^{*'}(sa + \xi) - y \rangle = 0, \end{aligned}$$

showing that $f^{*'}(sa + \xi)$ satisfies the variational inequality (28). The case $\psi(0) > 0$ can be treated similarly by noting that it is exactly the previous case with a and b replaced by $-a$ and $-b$, respectively. This proves (i). The proof of (ii) is similar. \square

5 A Method of Solving Operator Equations

5.1 In this section we assume that X is a reflexive Banach space and $f : X \rightarrow \mathbb{R}$ is a function satisfying the following conditions:

Assumption 2.

- (i) f is totally convex on bounded sets;
- (ii) f , as well as its Fenchel conjugate f^* are defined and (Gâteaux) differentiable on X and X^* , respectively;
- (iii) f' is uniformly continuous on bounded sets.

Note that, if f is lower semicontinuous, then the differentiability of f^* required by Assumption 2(ii) can be deduced from the strict convexity of f which is guaranteed by Assumption 2(i) (cf. [63, Theorem 11.13]). In particular, this happens when X is finite dimensional because, in this case, f is continuous on X as being convex.

We consider the problem of finding a vector x such that

$$Ax = 0 \text{ and } x \in \Omega, \tag{34}$$

where $A : X \rightarrow X^*$ is a given operator and Ω is a nonempty closed convex subset of X . We are going to present an iterative method for solving the equation (34) for a class of operators which, when X is a Hilbert space, contains all inverse strongly monotone operators (see [53] and Lemma 5.4 below) and, implicitly, the gradients of many differentiable functions of interest in optimization (see [11, Corollary 10]).

For each $\alpha \in (0, \infty)$ we define the operator $A_\alpha^f : X \rightarrow X$ by

$$A_\alpha^f x = f^{*'}(f'(x) - \alpha Ax). \tag{35}$$

Note that $Ax = 0$ if and only if x is a fixed point of A_α^f . We say that the operator A is *inverse-monotone relative to f on the set Ω* if there exists a real number $\alpha > 0$ and a vector $z \in \Omega$ such that

$$\langle Ay, A_\alpha^f y - z \rangle \geq 0, \quad \forall y \in \Omega. \quad (36)$$

In this case, the vector z involved in (36) is called *monotonicity pole* of A .

5.2 The most typical example of an operator which is inverse-monotone relative to f is the operator $A = f'$ in the case when the function f has a global minimizer $z \in X$. In this case, for any $\alpha \in (0, 1)$, we have that

$$A_\alpha^f x = f^{*'}((1 - \alpha)f'(x)) = ((1 - \alpha)^{-1}f)^{*'}(f'(x)).$$

Also, according to [63, Theorem 11.8(b)], we have that $((1 - \alpha)^{-1}f)^{*'}(0) = z$ because z also minimizes the function $(1 - \alpha)^{-1}f$ on X . Hence,

$$\begin{aligned} \langle Ax, A_\alpha^f x - z \rangle &= \left\langle f'(x), \left(\frac{1}{1 - \alpha} f \right)^{*'}(f'(x)) - z \right\rangle \\ &= \left\langle f'(x) - 0, \left(\frac{1}{1 - \alpha} f \right)^{*'}(f'(x)) - \left(\frac{1}{1 - \alpha} f \right)^{*'}(0) \right\rangle \geq 0, \end{aligned}$$

where the last inequality is due to the monotonicity of the operator $((1 - \alpha)^{-1}f)^{*'}$.

5.3 The following result establishes connections between the properties of relative inverse-monotonicity defined here and the property of total nonexpansivity introduced in [28]. Recall that an operator $B : X \rightarrow X$ is called *totally nonexpansive with respect to the function f on the set Ω* if there exists a vector $z \in \Omega$ such that

$$D_f(z, Bx) + D_f(Bx, x) \leq D_f(z, x), \quad \forall x \in \Omega. \quad (37)$$

A vector z for which the condition (37) is satisfied is called *nonexpansivity pole of B with respect to f* and it is necessarily a fixed point of B .

Lemma. *The operator A is inverse-monotone relative to f with constant α if and only if the operator A_α^f is totally nonexpansive with respect to f , that is, for some $z \in \Omega$ the following inequality holds:*

$$D_f(z, A_\alpha^f x) + D_f(A_\alpha^f x, x) \leq D_f(z, x), \quad \forall x \in \Omega.$$

In this case, $z \in \Omega$ is a monotonicity pole of A if and only if it is a nonexpansivity pole of A_α^f .

Proof: Observe that for any $x \in \Omega$ one has

$$D_f(z, x) - D_f(z, A_\alpha^f x) = f(A_\alpha^f x) - f(x) - \langle f'(x), z - x \rangle + \langle f'(A_\alpha^f x), z - A_\alpha^f x \rangle,$$

where $f'(A_\alpha^f x) = f'(x) - \alpha Ax$. Hence,

$$\begin{aligned} D_f(z, x) - D_f(z, A_\alpha^f x) &= f(A_\alpha^f x) - f(x) - \langle f'(x), A_\alpha^f x - x \rangle + \alpha \langle Ax, A_\alpha^f x - z \rangle \\ &= D_f(A_\alpha^f x, x) + \alpha \langle Ax, A_\alpha^f x - z \rangle. \end{aligned}$$

This implies that A is inverse-monotone relative to f on Ω and z is a monotonicity pole of it if and only if A_α^f is totally nonexpansive with respect to f and z is a nonexpansivity pole of it. \square

5.4 It is interesting to observe that if X is a Hilbert space and if $f(x) = \frac{1}{2} \|x\|^2$, then the inverse monotonicity condition (36) can be rewritten in a more familiar form. Operators A on a Hilbert space X which satisfy the condition

$$\langle Ax - Az, x - z \rangle \geq \alpha \|Ax - Az\|^2, \quad \forall x, z \in X \quad (38)$$

are called inverse-strongly-monotone operators (see, for instance, [53, p. 320]). These are nonexpansive operators whose (set valued) inverses are strongly monotone. Note that, according to the lemma proved below, an inverse-strongly-monotone operator A which has a zero at some point $z \in X$ is necessarily inverse-monotone relative to $f = \frac{1}{2} \|\cdot\|^2$ on X and has z as a monotonicity pole. In other words, the notion of inverse monotonicity relative to a function f can be viewed as a generalization of that of inverse-strong-monotonicity for operators in Hilbert spaces.

Lemma. *If X is a Hilbert space and $f = \frac{1}{2} \|\cdot\|^2$, then the operator $A : X \rightarrow X$ is inverse-monotone relative to f on Ω if and only if, for some $z \in \Omega$, one has that $Az = 0$ and*

$$\langle Ax, x - z \rangle \geq \alpha \|Ax\|^2, \quad \forall x \in \Omega. \quad (39)$$

Proof: If A is inverse-monotone relative to f on Ω and $z \in \Omega$ is a monotonicity pole of it then, by Lemma 3.1, z is a nonexpansivity pole of A_α^f and, thus, z is a fixed point of A_α^f . Since, in these circumstances, f' is the identity, it follows that $z = A_\alpha^f z = z - \alpha Az$ and, therefore, z is a solution of (34). Hence,

$$\begin{aligned} \langle Ax, x - z \rangle - \alpha \|Ax\|^2 &= \langle Ax - Az, x - z \rangle - \alpha \|Ax - Az\|^2 \\ &= \langle Ax - Az, x - z - \alpha(Ax - Az) \rangle \\ &= \langle Ax, A_\alpha^f x - z \rangle \geq 0. \end{aligned}$$

Conversely, if A satisfies (39) for some $z \in \Omega$ with $Az = 0$, then the previous chain of equalities shows that (36) holds. \square

5.5 Now we are going to describe an iterative procedure for solving the equation (34). To this end, for each real number $\alpha > 0$, we define the operator $S_\alpha^f : X \rightarrow X$ by

$$S_\alpha^f x = P_\Omega^f(f'(x) - \alpha Ax), \quad (40)$$

where P_Ω^f is the projection operator relative to f onto the set Ω . According to (27) we also have

$$S_\alpha^f = \Pi_\Omega^f \circ A_\alpha^f. \quad (41)$$

We are interested in the behavior of the following iterative process:

$$\text{Choose } y^0 \in \Omega \text{ and define } y^{k+1} = S_\alpha^f y^k, \quad \forall k \in \mathbb{N}. \quad (42)$$

The next result shows that if the operator A is inverse-monotone relative to a function f having the properties required at the beginning of this section, then the equation (34) has solutions and the sequences $\{y^k\}_{k \in \mathbb{N}}$ generated according to the rule (42) produces weak approximations of solutions of (34). A careful analysis of the proof shows that (42) is essentially a method of approximating fixed points of the operator S_α^f .

Theorem. *If the operator A is inverse-monotone relative to f on Ω with constant $\alpha > 0$, then the equation (34) has solutions and any sequence $\{y^k\}_{k \in \mathbb{N}}$ generated according*

to the rule (42) is bounded, has weak accumulation points and $\lim_{k \rightarrow \infty} Ay^k = 0$. If, in addition, the operator A is sequentially weakly-weakly* continuous on Ω , then any weak accumulation point of $\{y^k\}_{k \in \mathbb{N}}$ is a solution of (34). In this case, if the equation (34) has unique solution, then the sequence $\{y^k\}_{k \in \mathbb{N}}$ converges weakly to that solution.

Proof: According to Lemma 5.3, the operator A_α^f is totally nonexpansive with respect to f on Ω . This implies that there exists a vector $z \in \Omega$ such that

$$D_f(z, A_\alpha^f x) \leq D_f(z, x), \quad \forall x \in \Omega. \quad (43)$$

Taking here $x = z$, one obtains that $z = A_\alpha^f z$. Consequently, one has $f'(z) = f'(A_\alpha^f z)$ which, together with (35), implies $Az = 0$. Hence, the equation (34) has solutions. According to (43) combined with Corollary 4.5 and (41) one has that

$$D_f(z, S_\alpha^f x) + D_f(S_\alpha^f x, A_\alpha^f x) \leq D_f(z, A_\alpha^f x) \leq D_f(z, x), \quad \forall x \in \Omega.$$

Writing these inequalities for $x = y^k$ (which is contained in Ω) one deduces that

$$D_f(z, y^{k+1}) + D_f(y^{k+1}, A_\alpha^f y^k) \leq D_f(z, A_\alpha^f y^k) \leq D_f(z, y^k), \quad \forall k \in \mathbb{N}. \quad (44)$$

This implies that the sequence $\{D_f(z, y^k)\}_{k \in \mathbb{N}}$ is nonincreasing and, therefore, bounded. Thus, the sequence $\{y^k\}_{k \in \mathbb{N}}$ is contained in the set $R_\beta^f(f'(z))$, where β is a positive upper bound of the sequence $\{D_f(z, y^k)\}_{k \in \mathbb{N}}$. The set $R_\beta^f(f'(z))$ is bounded because f is totally convex (see the argument in the proof of Proposition 4.2). Hence, the sequence $\{y^k\}_{k \in \mathbb{N}}$ is bounded and, since the space X is reflexive, there exists a subsequence $\{y^{i_k}\}_{k \in \mathbb{N}}$ of $\{y^k\}_{k \in \mathbb{N}}$ which converges weakly to some point \bar{y} . Since all y^{i_k} 's are contained in Ω , the vector \bar{y} is contained in Ω too (Ω is convex and closed and, thus, weakly closed). We claim that the sequences $\{y^{i_k}\}_{k \in \mathbb{N}}$ and $\{A_\alpha^f y^{i_k}\}_{k \in \mathbb{N}}$ have the same weak limit \bar{y} . For proving this it is sufficient to show that

$$\lim_{k \rightarrow \infty} \|A_\alpha^f y^k - y^k\| = 0. \quad (45)$$

Observe that (44) implies

$$0 \leq D_f(z, A_\alpha^f y^k) - D_f(z, y^{k+1}) \leq D_f(z, y^k) - D_f(z, y^{k+1}), \quad \forall k \in \mathbb{N}.$$

As noted above, the sequence $\{D_f(z, y^k)\}_{k \in \mathbb{N}}$ converges. Therefore, the last inequality shows that the sequence $\{D_f(z, A_\alpha^f y^k)\}_{k \in \mathbb{N}}$ converges and has the same limit as the sequence $\{D_f(z, y^k)\}_{k \in \mathbb{N}}$. By (38) one also has

$$D_f(z, y^k) - D_f(z, A_\alpha^f y^k) = D_f(A_\alpha^f y^k, y^k) + \alpha \langle Ay^k, A_\alpha^f y^k - z \rangle$$

where the last term on the right hand side is nonnegative because of the inverse monotonicity of A . Thus, one obtains that

$$0 \leq \lim_{k \rightarrow \infty} D_f(A_\alpha^f y^k, y^k) \leq \lim_{k \rightarrow \infty} [D_f(z, y^k) - D_f(z, A_\alpha^f y^k)] = 0, \quad (46)$$

which shows that $\lim_{k \rightarrow \infty} D_f(A_\alpha^f y^k, y^k) = 0$. The function f is totally convex on bounded sets and the sequence $\{y^k\}_{k \in \mathbb{N}}$ is bounded. Therefore, the function $\nu_f(E, \cdot)$ associated to the set $E := \{y^k\}_{k \in \mathbb{N}}$, is strictly increasing on $[0, \infty)$ (see Subsection 2.10), vanishes at 0 and satisfies

$$\nu_f(E, \|A_\alpha^f y^k - y^k\|) \leq D_f(A_\alpha^f y^k, y^k), \quad \forall k \in \mathbb{N}.$$

This and (46) imply that

$$\lim_{k \rightarrow \infty} \nu_f(E, \|A_\alpha^f y^k - y^k\|) = 0.$$

Due to the strict monotonicity of $\nu_f(E, \cdot)$, this cannot happen unless (45) holds. Since f' is uniformly continuous on bounded sets, by (45) one deduces that

$$\lim_{k \rightarrow \infty} \|f'(A_\alpha^f y^k) - f'(y^k)\|_* = 0.$$

Note that, according to (35), one also has that $f'(A_\alpha^f y^k) - f'(y^k) = -\alpha A y^k$. The last two equalities imply that $\lim_{k \rightarrow \infty} \|A y^k\|_* = 0$.

Now, suppose that A is sequentially weakly-weakly* continuous. Since the sequence $\{y^{i_k}\}_{k \in \mathbb{N}}$ converges weakly to \bar{y} , we obtain that $\|A \bar{y}\|_* \leq \lim_{k \rightarrow \infty} \|A y^{i_k}\|_* = 0$, showing that \bar{y} is a solution of (34). If the solution of (34) is unique, then the sequence $\{y^k\}_{k \in \mathbb{N}}$ has a unique weak accumulation point and this completes the proof. \square

5.6 The sequential weak-weak* continuity condition imposed upon the operator A in Theorem 5.5 is restrictive. However, this condition always holds when A is linear and continuous or, alternatively, when X is finite dimensional and A is continuous. An immediate application of Theorem 5.5 to convex unconstrained optimization in \mathbb{R}^n is emphasized in the result below:

Corollary. *The following statements are true:*

(a) *Suppose that $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a differentiable strictly convex function. If f' is uniformly continuous on bounded sets and if f has a global minimizer, then for each $\alpha \in (0, 1)$, the sequence $\{y^k\}_{k \in \mathbb{N}}$ generated according to the rule*

$$y^0 \in \mathbb{R}^n \text{ and } y^{k+1} = f^*((1 - \alpha)f'(y^k)), \quad \forall k \in \mathbb{N},$$

converges to the minimizer of f .

(b) *If $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ is a differentiable convex function whose derivative φ' is Lipschitzian with Lipschitz constant L and if φ has at least one minimizer in \mathbb{R}^n , then the sequence generated according to the rule*

$$y^0 \in \mathbb{R}^n \text{ and } y^{k+1} = y^k - \frac{1}{L} \varphi'(y^k), \quad \forall k \in \mathbb{N},$$

is bounded and its accumulation points are minimizers of φ .

Proof: (a) Theorem 2.15 ensures that f satisfies Assumption 2. As noted in Subsection 5.2, in these circumstances, the operator $A = f'$ is inverse monotone relative to f . Applying Theorem 5.5 and taking into account that f has a single minimizer (i.e., the equation $f'(x) = 0$ has a single solution) yields the conclusion.

(b) According to [11, Corollary 10], the operator $A = \varphi'$ is inverse strongly monotone with constant $\alpha = L^{-1}$. By Lemma 5.4, this implies that φ' is inverse-monotone relative to the function $f = \frac{1}{2} \|\cdot\|^2$ on the whole space X with the same constant α . Consequently, Theorem 5.5 applies to $A = \varphi'$ and, keeping in mind that f' is the identity, the result follows. \square

5.7 The following technical result, which will be used later, is a generalization of a result known as Opial's Lemma [55, Lemma 2]. It says that if X is a Hilbert space, if $g = \frac{1}{2} \|\cdot\|^2$ and if $T : \Omega \rightarrow X$ is a nonexpansive mapping on the nonempty closed convex subset Ω of X , then

for any sequence $\{z^k\}_{k \in \mathbb{N}} \subseteq \Omega$ which is weakly convergent and has $\lim_{k \rightarrow \infty} \|Tz^k - z^k\| = 0$, the vector $z = w - \lim_{k \rightarrow \infty} z^k$ is necessarily a fixed point of T . We extend this result to not necessarily Hilbertian Banach spaces and to nonexpansive operators in a generalized sense.

Lemma. *Suppose that the function $g : X \rightarrow \mathbb{R}$ is totally convex on bounded sets and has bounded on bounded sets Gâteaux derivative g' . If $T : \Omega \rightarrow X$ is an operator such that*

$$D_g(Ty, Tx) \leq D_g(y, x), \quad \forall x, y \in \Omega, \quad (47)$$

then for any weakly convergent sequence $\{z^k\}_{k \in \mathbb{N}} \subseteq \Omega$ which has $\lim_{k \rightarrow \infty} D_g(Tz^k, z^k) = 0$, the vector $z = w - \lim_{k \rightarrow \infty} z^k$ is a fixed point of T .

Proof: Observe that, for any $x \in X$, one has

$$\begin{aligned} D_g(z^k, x) - D_g(z^k, z) &= -g(x) - \langle g'(x), z^k - x \rangle + g(z) + \langle g'(z), z^k - z \rangle \\ &= D_g(z, x) + \langle g'(x) - g'(z), z - z^k \rangle, \end{aligned}$$

where $\lim_{k \rightarrow \infty} \langle g'(x) - g'(z), z - z^k \rangle = 0$. Therefore,

$$\liminf_{k \rightarrow \infty} D_g(z^k, x) \geq D_g(z, x) + \liminf_{k \rightarrow \infty} D_g(z^k, z),$$

which shows that, when $x \neq z$, one has

$$\liminf_{k \rightarrow \infty} D_g(z^k, x) > \liminf_{k \rightarrow \infty} D_g(z^k, z). \quad (48)$$

We claim that

$$\liminf_{k \rightarrow \infty} D_g(Tz^k, Tz) = \liminf_{k \rightarrow \infty} D_g(z^k, Tz). \quad (49)$$

In order to show that, observe that

$$\begin{aligned} D_g(Tz^k, Tz) &= g(Tz^k) - g(Tz) - \langle g'(Tz), Tz^k - Tz \rangle \\ &= D_g(z^k, Tz) + [g(Tz^k) - g(z^k)] + \langle g'(Tz), z^k - Tz^k \rangle. \end{aligned} \quad (50)$$

The function g is sequentially consistent (cf. Theorem 2.10). Consequently, since

$$\lim_{k \rightarrow \infty} D_g(Tz^k, z^k) = 0,$$

one also has that

$$\lim_{k \rightarrow \infty} \|Tz^k - z^k\| = 0. \quad (51)$$

This implies that

$$\lim_{k \rightarrow \infty} \langle g'(Tz), z^k - Tz^k \rangle = 0. \quad (52)$$

From the convexity of g one deduces that

$$\langle g'(Tz^k), Tz^k - z^k \rangle \geq g(Tz^k) - g(z^k) \geq \langle g'(z^k), Tz^k - z^k \rangle, \quad (53)$$

where the sequence $\{z^k\}_{k \in \mathbb{N}}$ is bounded as being weakly convergent and $\{Tz^k\}_{k \in \mathbb{N}}$ is bounded because of (51). The function g' being bounded on bounded sets the sequences $\{g'(z^k)\}_{k \in \mathbb{N}}$ and $\{g'(Tz^k)\}_{k \in \mathbb{N}}$ are bounded too. Thus, letting $k \rightarrow \infty$ in (53) we deduce that

$$\lim_{k \rightarrow \infty} [g(Tz^k) - g(z^k)] = 0.$$

This, combined with (52) and (53), proves (49). Now, suppose by contradiction that $Tz \neq z$. Then, according to (48), one has that

$$\liminf_{k \rightarrow \infty} D_g(z^k, Tz) > \liminf_{k \rightarrow \infty} D_g(z^k, z). \quad (54)$$

By (47), one deduces that

$$\liminf_{k \rightarrow \infty} D_g(z^k, z) \geq \liminf_{k \rightarrow \infty} D_g(Tz^k, Tz) = \liminf_{k \rightarrow \infty} D_g(z^k, Tz), \quad (55)$$

where the last equality results from (49). Since (54) and (55) contradict each other, our assumption that $Tz \neq z$ is false, that is, z is a fixed point of T . \square

5.8 It was noted above that applicability of Theorem 5.5 is restricted because of the requirement that the operator A should be sequentially weakly-weakly* continuous. Using the generalization of Opial's lemma proved above we can give another variant of Theorem 5.5 in which that requirement is dropped. The price to pay for dropping that condition is the need to strengthen our demands on the function f and on the nonexpansivity of A . In order to present the alternative version of Theorem 5.5 we will call the operator A *inverse strongly-monotone on Ω relative to f* if, for some real number $\alpha > 0$, the operator A_α^f is nonexpansive on Ω with respect to f in the sense given to this term in [53], that is, if

$$D_f(A_\alpha^f x, A_\alpha^f y) \leq D_f(x, y), \quad \forall x, y \in \Omega. \quad (56)$$

Now we are in position to prove the new version of Theorem 5.5 which applies to operators which are inverse strongly-monotone with respect to f .

Theorem. *If the operator A is inverse strongly-monotone relative to f on Ω with constant $\alpha > 0$ and if the equation (34) has at least one solution, then any sequence $\{y^k\}_{k \in \mathbb{N}}$ generated according to the rule (42) has the following properties:*

- (i) *It is bounded, has weak accumulation points and any such point is a solution of (34);*
- (ii) *If the equation (34) has unique solution, then the sequence $\{y^k\}_{k \in \mathbb{N}}$ converges weakly to that solution;*
- (iii) *If the space X has finite dimension, then $\{y^k\}_{k \in \mathbb{N}}$ converges to a solution of (34).*

Proof: Let $z \in \Omega$ be a solution of (34). Then z is a fixed point of A_α^f and one has

$$D_f(z, A_\alpha^f x) = D_f(A_\alpha^f z, A_\alpha^f x) \leq D_f(z, x), \quad \forall x \in \Omega. \quad (57)$$

Consequently, (43) holds and all the considerations in the proof of Theorem 5.5 up to, and including, formula (46) can be reproduced without any modification. Formula (46) implies that

$$\lim_{k \rightarrow \infty} D_f(A_\alpha^f y^k, y^k) = 0.$$

Also, according to [48, Proposition 4], the derivative f' is bounded on bounded sets because it is uniformly continuous on bounded sets (cf. Assumption 2). Therefore, one can apply Lemma 5.7 to the weakly convergent subsequence $\{y^{i_k}\}_{k \in \mathbb{N}}$. It implies that \bar{y} is a fixed point of the operator A_α^f , i.e., a solution of the problem (34). This proves (i). Clearly, (ii) results from (i). Now, suppose that X has finite dimension. By Corollary 4.5 and (57), one has that

$$D_f(\bar{y}, y^{k+1}) + D_f(y^{k+1}, A_\alpha^f y^k) \leq D_f(\bar{y}, y^k), \quad \forall k \in \mathbb{N}.$$

This implies that $\{D_f(\bar{y}, y^k)\}_{k \in \mathbb{N}}$ is convergent. Since it has a subsequence convergent to zero, then the whole sequence converges to zero. By Assumption 2 and Theorem 2.10, f is sequentially consistent. Hence, $\|y^k - \bar{y}\| \rightarrow 0$ as $k \rightarrow \infty$ and this completes the proof. \square

5.9 If X is a Hilbert space and $f = \frac{1}{2} \|\cdot\|^2$, then the operator A is inverse strongly-monotone relative to f with constant α if and only if A is inverse strongly-monotone (see (38)) with constant $\alpha/2$. Therefore, Theorem 5.8 leads to the following partially known result (see [16, Subsection 2.3] and [34, Corollary 4.1]):

Corollary. *Suppose that X is a Hilbert space and $\varphi : X \rightarrow (-\infty, +\infty]$ is a convex function which is Gâteaux differentiable on the interior of its domain and has Lipschitzian derivative φ' . If Ω is a nonempty closed convex subset of $\text{int}(\text{dom } \varphi)$ such that $\inf_{z \in \Omega} \varphi(z)$ is achieved, then any sequence generated according to the rule*

$$y^{k+1} = P_{\Omega}(y^k - 2L^{-1}\varphi'(y^k)), \quad \forall k \in \mathbb{N},$$

where $P_{\Omega}(x)$ denotes the metric projection onto Ω and L is the Lipschitz constant of φ' , has the following properties, no matter how the initial point $y^0 \in \Omega$ is chosen;

(i) $\{y^k\}_{k \in \mathbb{N}}$ is bounded, has weak accumulation points, and any such point y^* belongs to Ω and satisfies $\varphi(y^*) = \inf_{z \in \Omega} \varphi(z)$;

(ii) If φ has a unique minimum on Ω , then the sequence $\{y^k\}_{k \in \mathbb{N}}$ converges weakly to a minimizer of φ on Ω ;

(iii) If X is finite dimensional, then the sequence $\{y^k\}_{k \in \mathbb{N}}$ converges to a minimizer of φ on Ω .

Proof: Apply Theorem 5.8 by taking into account that, in the given circumstances, the operator φ' is inverse strongly monotone relative to $f = \frac{1}{2} \|\cdot\|^2$ with constant $\alpha = 2L^{-1}$ as follows from [11]. \square

6 Acknowledgments

The authors wish to thank Anca Croitoru and Constantin Zălinescu for valuable suggestions concerning some topics discussed in this work. Also, the second author gratefully acknowledges the support of the Institute for Pure and Applied Mathematics, UCLA, California (September - December 2003), where part of this work was completed.

References

- [1] Alber, Y., Generalized projection operators in Banach spaces: Properties and applications, in: "Functional Differential Equations" edited by M.E. Draklin and E. Litsyn, Vol. 1, The research Institute of Judea and Samaria, Kedumim-Ariel, 1993.
- [2] Alber, Y., Metric and generalized projection operators in Banach spaces: Properties and applications, in: Theory and Applications of Nonlinear Operators of Accretive and Monotone Type, Dekker, New York, (1996) 15-50.
- [3] Alber, Y., Generalized projections, decompositions, and the Pythagorea-type Theorem in Banach spaces, App. Math. Lett., 11 (1998) 115-121.
- [4] Alber, Y., Some new characteristics of Banach spaces, preprint 2004.
- [5] Alber, Y. and Butnariu, D., Convergence of Bregman-projection methods for solving consistent convex feasibility problems in reflexive Banach spaces, J. Optimization Theory Appl. 92(1) (1997) 33-61.

- [6] Alber, Y., Butnariu, D. and Ryazantseva, I., Regularization and resolution of monotone variational inequalities with operators given by hypomonotone approximations, *J. Nonlin. Convex Anal.*, to appear, 2005.
- [7] Alber, Y., Iusem, A.N. and Solodov, M.V., Minimization of nonsmooth convex functionals in Banach spaces, *J. Convex Anal.* 4 (1997), 235-255.
- [8] Alber, Y. and Notik, A.I., Geometric properties of Banach spaces and approximate methods for solving nonlinear operator equations, *Soviet Math. Dokl.* 29 (1984), 611-615.
- [9] Alber, Y. and Reich, S., An iterative method for solving a class of nonlinear operator equations in Banach spaces, *Panamer. Math. J.* 4(2) (1994) 3954.
- [10] Asplund, E., Averaged norms, *Israel J. Math.* 5 (1967) 227-233.
- [11] Baillon, J.-B. and Haddad, G., Quelques propriétés des opérateurs angle-bornés et n -cyclément monotones, *Israel J. Math.* 26 (1977) 137-150.
- [12] Bauschke, H.H. and Borwein, J.M., Legendre functions and the method of random Bregman projections, *J. Convex Anal.* 4 1 (1997) 27-67.
- [13] Bauschke, H.H., Borwein, J.M. and Combettes, P.L., Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces, *Commun. Contemp. Math.* 3(4) (2001) 615-647.
- [14] Bauschke, H.H. and Combettes, P.L., Construction of best Bregman approximations in reflexive Banach spaces, *Proc. Amer. Math. Soc.* 131(12) (2003) 3757-3766.
- [15] Bauschke, H.H. and Lewis, A.S., Dykstra's algorithm with Bregman projections: a convergence proof, *Optimization* 48 (2000) 409-427.
- [16] Bertsekas, D., *Nonlinear Programming: 2nd Edition*, Athena Scientific, Belmont, 1999.
- [17] Bessaga, C. and Pełczyński, A., *Selected Topics in Infinite-Dimensional Topology*, PWN-Polish Scientific Publishers, Warszawa, 1975.
- [18] Borwein, J.M. and Lewis, A.S., Convergence of best entropy estimates, *SIAM J. Optim.* 1 (1991) 191-205.
- [19] Borwein, J.M. and Limber, M.A., On entropy maximization via convex programming, preprint available at <http://citeseer.ist.psu.edu/71881.html>, 1996.
- [20] Bregman, L.M., The relaxation method for finding common points of convex sets and its application to the solution of problems in convex programming, *USSR Computational Mathematics and Mathematical Physics* 7 (1967) 200-217.
- [21] Bruck, R.E., A strongly convergent iterative solution of $0 \in U(x)$ for a maximal monotone operator U in Hilbert space, *J. Math. Anal. Appl.* 48 (1974), 114-126.
- [22] Bruck, R.E., An iterative solution of a variational inequality for certain monotone operators in Hilbert space, *Bull. Amer. Math. Soc.* 81 (1975) 890-892.
- [23] Burachik, R.S., Generalized proximal point methods for the variational inequality problem, PhD Thesis, Instituto de Mathematica Pura e Aplicada (IMPA), Rio de Janeiro, RJ, Brazil, 1995.

- [24] Burachik, R.S. and Schemberg, S., A proximal point method for the variational inequality problem in Banach spaces, *SIAM J. Control Optim.* 39(5) (2001) 1633-1649.
- [25] Butnariu, D., Censor, Y. and Reich, S., Iterative averaging of entropic projections for solving stochastic convex feasibility problems, *Comput. Optim. Appl.* 8 (1997) 21-39.
- [26] 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) 61-92.
- [27] Butnariu, D. and Iusem, A.N., On a proximal point method for convex optimization in Banach spaces, *Numer. Funct. Anal. Optim.* 18 (1998) 723-744.
- [28] Butnariu, D. and Iusem, A.N., *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization*, Kluwer Academic Publishers, Dordrecht, 2000.
- [29] Butnariu, D., Iusem, A.N. and Burachik, R.S., Iterative methods for solving stochastic convex feasibility problems and applications, *Comput. Optim. Appl.* 15 (2000) 269-307.
- [30] Butnariu, D., Iusem, A.N. and Resmerita, E., Total convexity for powers of the norm in uniformly convex Banach spaces, *J. Convex Anal.* 7 (2000) 319-334.
- [31] Butnariu, D., Iusem, A.N. and Zălinescu, C., On uniform convexity, total convexity and convergence of the proximal point and outer Bregman projection algorithms in Banach spaces, *J. Convex Anal.* 10 (2003) 35-61.
- [32] Butnariu, D., Reich, S. and Zaslavski, A., Weak convergence for orbits of nonlinear operators in reflexive Banach spaces, *Numer. Funct. Anal. Optim.* 24 (2003) 489-508.
- [33] Butnariu, D. and Resmerita, E., The outer Bregman projection method for stochastic feasibility problems in Banach spaces, in: *Inherently Parallel Algorithms for Feasibility and Optimization and their Applications*, (D. Butnariu, Y. Censor and S. Reich, eds.), Elsevier, Amsterdam, 2001.
- [34] Byrne, C., A unified treatment of some iterative algorithms in signal processing and image reconstruction, *Inverse Problems* 20 (2004) 103-120.
- [35] Censor, Y. and Lent, A., An iterative row action method for interval convex programming, *J. Optimization Theory Appl.* 34 (1981) 321-353.
- [36] Censor, Y. and Zenios, S., *Parallel Optimization: Theory, Algorithms and Applications*, Oxford University Press, New York, 1997.
- [37] Ciorănescu, I., *Geometry of Banach Spaces, Duality mappings, and Nonlinear Problems*, Kluwer Academic Publishers, Dordrecht, 1990.
- [38] Clarkson, J.A., Uniformly convex spaces, *Trans. Amer. Math. Soc.* 40 (1936) 396-414.
- [39] Dontchev, A.L. and Zolezzi, T., *Well-Posed Optimization Problems*, Springer-Verlag, Berlin, 1993.
- [40] Eckstein, J., Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming, *Math. Oper. Res.* 18(1) (1993) 202-226.

- [41] Eggermont, P.P.B., Maximum entropy regularization for Fredholm integral equations of the first kind, *SIAM J. Math. Anal.* 24 (1993) 1557-1576.
- [42] Ekeland, I. and Temam, R., *Convex Analysis and Variational Problems*, North-Holland, Amsterdam, 1976.
- [43] Engl, H. and Landl, G., Convergence rates for maximum entropy regularization, *SIAM J. Num. Anal.* 30 (1993) 1509-1536.
- [44] Fan, K. and Glicksberg, I., Some geometric properties of the spheres in a normed linear space, *Duke Math. J.* 25 (1958) 553-568.
- [45] Glowinski, R., Lions, J.L. and Tremolieres, R., *Numerical Analysis of Variational Inequalities*, North-Holland Publisher Co., Amsterdam, 1981.
- [46] Holmes, R., *A Course on Optimization and Best Approximation*, Lecture Notes in Mathematics 257, Springer, 1972.
- [47] Ibaraki, T., Kimura, Y., and Takahashi, W., Convergence theorems for generalized projections and maximal monotone operators in Banach spaces, *Abstract and Applied Analysis* 10 (2003) 621-629.
- [48] Iusem, A.N. and Gárciga-Otero, R., Inexact versions of proximal point and augmented Lagrangian algorithms in Banach spaces, *Numer. Funct. Anal. Optim.* 22 (2001) 609-640.
- [49] Iusem, A.N. and Gárciga-Otero, R., Augmented Lagrangian methods for cone-constrained optimization in Banach spaces, *J. Nonlinear Convex Anal.* 3(2) (2002) 155-176.
- [50] Kadec, M.I., Spaces isomorphic to a locally uniformly convex space. (Russian) *Izv. Vys. U cebn. Zaved. Matematika* 6 (13) (1959) 51-57.
- [51] Kiwiel, K.C., Proximal minimization methods with generalized Bregman functions, *SIAM J. Control Optim.* 35(4) (1997) 1142-1168.
- [52] Landweber, L., An iterative formula for Fredholm integral equations of the first kind, *Amer. J. Math.* 73 (1951) 615-624.
- [53] Liu, F. and Nashed, M.Z., Regularization of nonlinear ill-posed variational inequalities and convergence rates, *Set-Valued Anal.* 6 (1998), 313-344.
- [54] Mosco, U., Approximation of the solutions of some variational inequalities, *Ann. Sc. Norm. Sup. Pisa* 21 (1967) 373-394.
- [55] Opial, Z., Weak convergence of the sequence of successive approximations for nonexpansive mappings, *Bull. Amer. Math. Soc.* 73 (1967) 591-597.
- [56] Pascali, D. and Sburlan, S., *Nonlinear Mappings of Monotone Type*, Stijthoff&Noordhoff International Publishers, Amsterdam, 1978.
- [57] Phelps, R.R., *Convex Functions, Monotone Operators and Differentiability*, Springer Verlag, Berlin, 1993.
- [58] Polyak, B., A general method for solving extremum problems, *Dokl. Akad. nauk SSSR* 174 (1967), 593-597.

- [59] Reich, S., A weak convergence theorem for the alternating method with Bregman distances, in: Theory and Applications of Nonlinear Operators of Accretive and Monotone Type, Dekker, New York, (1996) 313-318.
- [60] Resmerita, E., Fixed point and optimization methods in infinite dimensional Banach spaces, PhD Thesis, University of Haifa, Israel, 2003.
- [61] Resmerita, E., On total convexity, Bregman projections and stability in Banach spaces, J. Convex Anal. 11 (2004) 1-16.
- [62] Rockafellar, T.R., Convex Analysis, Princeton University Press, Princeton, 1970.
- [63] Rockafellar, T.R. and Wets, R. J.-B. Variational Analysis, Springer Verlag, Berlin, 1998.
- [64] Solodov, M.V. and Svaiter, B.F., An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions, Math. Oper. Res. 25(2) (2000) 214-230.
- [65] Troyanski, S.L., On locally uniformly convex and differentiable norms in certain non-separable Banach spaces, Studia Math. 37 (1971) 173-180.
- [66] Vladimirov, A.A., Nesterov, Y.E. and Chekanov, Y.N., Uniformly convex functionals, Vestnik Moscov. Univ. Ser.15, Vychisl. Mat. Kibernet. 3 (1978) 12-23.
- [67] Zălinescu, C., On uniformly convex functions, J. Math. Anal. Appl. 95 (1983) 344-374.
- [68] Zălinescu, C., Convex Analysis in General Vector Spaces, World Scientific, River Edge, N.J., 2002.

Dan Butnariu: Department of Mathematics, University of Haifa, 31905 Haifa, Israel
E-mail address: dbutnaru@math.haifa.ac.il

Elena Resmerita: Johann Radon Institute for Computational and Applied Mathematics,
 Austrian Academy of Sciences, Altenbergerstrasse 69, Linz 4040, Austria
E-mail address: elena.resmerita@ricam.oeaw.ac.at