On Uniform Convexity, Total Convexity and Convergence of the Proximal Point and Outer Bregman Projection Algorithms in Banach Spaces

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

Alfredo N. Iusem
The Institute of Pure and Applied Mathematics (IMPA),
Estrada Dona Castorina 110, Jardim Botânico,
Rio de Janeiro, RJ, CEP 22460-320, Brazil
e-mail: iusp@impa.br

Constantin Zălinescu
Faculty of Mathematics, "Al.I. Cuza" University,
6600 Iasi, Romania
e-mail: zalinesc@uaic.ro

April 28, 2002

Abstract

In this paper we study and compare the notions of uniform convexity of functions at a point and on bounded sets with the notions of total convexity at a point and sequential consistency of functions, respectively. We establish connections between these concepts of strict convexity in infinite dimensional settings and use the connections in order to obtain improved convergence results concerning the outer Bregman projection algorithm for solving convex feasibility problems and the generalized proximal point algorithm for optimization in Banach spaces.

Keywords: Uniform convexity at a point, total convexity at a point, uniform convexity on bounded sets, sequential consistency, generalized proximal point algorithm for optimization, outer Bregman projection algorithm for feasibility.

2000 Mathematics Subject Classification: Primary 90C25, 90C48, 26B25, 49J52; Secondary 46N10, 46N20, 90C30

1 Introduction

The aim of this paper is to study and compare several notions of convex analysis which proved to be useful in establishing convergence properties for fixed point and optimization algorithms in infinite dimensional Banach spaces. Uniform convexity of functions at a point and on bounded sets on one hand and total convexity at a point and sequential consistency on the other hand are, respectively, similar, although nonequivalent, concepts of strict convexity in the infinite dimensional setting. We establish connections between these concepts and use these connections in order to obtain improved convergence results concerning the outer Bregman projection algorithm for solving convex feasibility problems and the generalized proximal point algorithm for optimization in Banach spaces.

The idea of using uniform convexity properties of functions for convergence analysis of optimization algorithms goes back, as far as we know, to Polyak [30]. This idea was further developed by Levitin and Polyak [27] and was expanded by Zolezzi [38] into the area of stability analysis of optimization procedures in infinite dimensional spaces. The literature contains several slightly different definitions of uniform convexity of functions on a Banach space X (see, for instance, [3], [23] and [22]). The notion of uniform convexity of a function $f : X \to (-\infty, +\infty]$ at a point $x \in \mathrm{Dom}\,(f)$ we use in this paper (see Subsection 2.2) is that introduced and studied in [36] as a local counterpart of the definition of uniform convexity (on the whole domain) due to Vladimirov et al. [35]. Totally convex functions, although not under this name, were considered in [36] too but their usefulness for establishing convergence of fixed point, feasibility and optimization methods became apparent in [12] and [13]. The definition of total convexity of f at $x \in \mathrm{Dom}\,(f)$ we consider here (see Subsection 2.1) is that introduced in [13]. In the process of exploiting specific properties of totally convex functions in the convergence and stability analysis of fixed point algorithms several questions concerning the connections between total and uniform convexity naturally occurred:

1) If the function f is uniformly convex at $x \in \mathrm{Dom}\,(f)$, then it is also totally convex at x; when is the converse implication true?

2) If X is a locally uniformly convex space, then the powers of the norm $\|\cdot\|^r$, with $r \in (1, +\infty)$, are uniformly convex at any $x \in X$; can we find reasonably good estimates of the moduli of total convexity of these functions?

3) If f is a function which is totally convex at any point of X and Y is a dense subspace of X provided with a norm $\|\cdot\|_Y$ which is stronger than the norm induced from X, is the restriction of f to Y still totally convex at the points of Y (with respect to the norm $\|\cdot\|_Y$)?

4) (Robert T. Rockafellar) If X is finite dimensional, then the function f is totally convex at all points $x \in \mathrm{Int}\,(\mathrm{Dom}\,(f))$ if and only if f is strictly convex on $\mathrm{Int}\,(\mathrm{Dom}\,(f))$ and this happens if and only if for any $x \in \mathrm{Int}\,(\mathrm{Dom}\,(f))$ the conjugate function f^* is differentiable on $\partial f(x)$; do these implications hold true in spaces of infinite dimension?

5) (Simeon Reich) Functions f which are uniformly convex and have domains

with nonempty interior exist on reflexive spaces only; do sequentially consistent functions (i.e., uniformly totally convex functions) with solid domain exist on nonreflexive spaces?

6) (Yakov Alber) If f is uniformly convex on bounded sets, then it is sequentially consistent; does the converse implication hold?

The aim of this paper is to answer these questions and to use the resulting information in order to obtain improved convergence results for two algorithms: the outer Bregman projection method for solving feasibility problems proposed in [18] and the generalized proximal point method proposed by Censor and Zenios in [21] and whose convergence in infinite dimensional Banach spaces was studied in [14]. Questions 1), 2) and 3) are answered in Section 2. A partial answer to question 4) is given in Section 3. In Section 4 we answer questions 5) and 6). Also in Section 4 we show that the outer Bregman projection algorithm, previously known to converge in some special spaces only, is convergent in any uniformly convex and smooth Banach space and prove that the generalized proximal point algorithm is strongly convergent when applied to problems with totally convex objective functions. Claiming that the generalized proximal point algorithm produces strongly convergent sequences may raise some eyebrows: a well known example, due to Güler [25], shows that the classical proximal point algorithm for optimization [33], a particular version of the generalized proximal point method discussed here, may happen to produce sequences which are weakly, but not strongly, convergent. This aspect is clarified in Subsection 4.6.

2 Total Convexity Versus Uniform Convexity at a Point

2.1 In this paper X denotes a Banach space and $f : X \to (-\infty, +\infty]$ denotes a proper convex function whose domain, $\operatorname{Dom}(f)$, is not a singleton. The *right hand sided derivative of f at $x \in \operatorname{Dom}(f)$ in the direction d* is given by

$$f^\circ(x,d) := \lim_{t \searrow 0} \frac{f(x+td) - f(x)}{t}$$

and *the Bregman distance with respect to f between the points $x, y \in \operatorname{Dom}(f)$* (see [20]) is

$$D_f(y,x) := f(y) - f(x) - f^\circ(x, y-x). \tag{1}$$

The *modulus of total convexity of f at $x \in \operatorname{Dom}(f)$* is the function $\nu_f(x,\cdot) : [0,+\infty) \to [0,+\infty]$ defined by

$$\nu_f(x,t) = \inf \{ D_f(y,x) : y \in \operatorname{Dom}(f),\ \|y - x\| = t \}. \tag{2}$$

The function f is called *totally convex at $x \in \operatorname{Dom}(f)$* if $\nu_f(x,t) > 0$ whenever $t > 0$ (see [13]).

2.2 A notion strongly related to that of total convexity at $x \in \mathrm{Dom}(f)$ was introduced and studied in [36]: The function f is called *uniformly convex at* $x \in \mathrm{Dom}(f)$ if the function $\mu_f(x, \cdot) : [0, +\infty) \to [0, +\infty]$ defined by

$$\mu_f(x,t) = \inf_{\substack{y \in \mathrm{Dom}(f), \|y-x\|=t \\ \lambda \in (0,1)}} \left\{ \frac{\lambda f(x) + (1-\lambda) f(y) - f\left[\lambda x + (1-\lambda) y\right]}{\lambda(1-\lambda)} \right\}$$

is positive whenever $t > 0$. The function $\mu_f(x, \cdot)$ was termed *modulus of uniformly strict convexity* at x in [13] and *gage of uniform convexity* of f at x in [37]. According to [13, Proposition 1.2.5], we have that $\nu_f(x,t) \geq \mu_f(x,t)$ for all $t \geq 0$ and, therefore, if f is uniformly convex at $x \in \mathrm{Dom}(f)$, then it is totally convex at $x \in \mathrm{Dom}(f)$. The converse implication is not generally true, that is, a function f may be totally convex at the point $x \in \mathrm{Dom}(f)$ without being uniformly convex at that point. In order to show that, consider $x \in \mathrm{Dom}(f)$ and the function $\overline{\mu}_f(x, \cdot) : [0, +\infty) \to [0, +\infty]$ defined by

$$\overline{\mu}_f(x,t) = \inf_{\substack{y \in \mathrm{Dom}(f) \\ \|y-x\|=t}} \left\{ f(x) + f(y) - 2f\left(\tfrac{x+y}{2}\right) \right\}.$$

This is usually called the *modulus of locally uniform convexity* of f at $x \in \mathrm{Dom}(f)$ and f is said to be *locally uniformly convex at* x if $\overline{\mu}_f(x,t) > 0$ for all $t > 0$. The following result shows that locally uniform convexity of f at x and uniform convexity of f at x are equivalent notions and, thus, answers a question posed in [13, p. 25]. Consequently, the example given in [13, Section 2.3] pointing to a function on ℓ^1 which is everywhere Gâteaux differentiable on ℓ^1, totally convex at some points of ℓ^1, but not locally uniformly convex at any point, also shows that total convexity and uniform convexity at a point are not equivalent notions.

Lemma. *For any* $x \in \mathrm{Dom}(f)$ *and for any* $t \in [0, +\infty)$ *we have*

$$\mu_f(x,t) \geq \overline{\mu}_f(x,t) \geq \tfrac{1}{2}\mu_f(x,t). \tag{3}$$

Proof. It is clear that $\mu_f(x,t) \leq 2\overline{\mu}_f(x,t)$. The other inequality in (3) is proved now. Let $y \in \mathrm{Dom}(f)$ be such that $\|y - x\| = t > 0$ and suppose that $\lambda \in (0, \frac{1}{2}]$. Then, we have

$$\begin{aligned}
& f\left((1-\lambda)x + \lambda y\right) \\
& = f\left[(1-2\lambda)x + 2\lambda(\tfrac{1}{2}(y+x))\right] \\
& \leq (1-2\lambda)f(x) + 2\lambda f(\tfrac{1}{2}(y+x)) \\
& \leq (1-2\lambda)f(x) + 2\lambda\left[\tfrac{1}{2}\left(f(x)+f(y)\right) - \tfrac{1}{2}\overline{\mu}_f(x, \|y-x\|)\right] \\
& = (1-2\lambda)f(x) + \lambda\left(f(x)+f(y)\right) - \lambda\overline{\mu}_f(x,t) \\
& \leq (1-\lambda)f(x) + \lambda f(y) - \lambda(1-\lambda)\overline{\mu}_f(x,t),
\end{aligned}$$

showing that

$$\overline{\mu}_f(x,t) \le \frac{(1-\lambda)f(x) + \lambda f(y) - f\left((1-\lambda)x + \lambda y\right)}{\lambda(1-\lambda)},$$

for all $\lambda \in (0, \frac{1}{2}]$. A similar argument shows that the previous inequality also holds when $\lambda \in (\frac{1}{2}, 1]$. Consequently, $\overline{\mu}_f(x,t) \le \mu_f(x,t)$. ■

2.3 The example in [13, Section 2.3], quoted above, shows that, even for Gâteaux differentiable functions, uniform and total convexity at a point are not equivalent. However, according to Proposition 1.3.10 in [13], if f is everywhere finite and Fréchet differentiable, then f is totally convex at the point $x \in X$ if and only if f is uniformly convex at x. This equivalence holds under less restrictive conditions as follows from the next result.

Proposition. *Suppose that f is lower semicontinuous. If $x \in \mathrm{Dom}(f)$ and f is Fréchet differentiable at x, then f is totally convex at x if and only if f is uniformly convex at x.*

Proof. Suppose that f is Fréchet differentiable and totally convex at x. Then $x \in \mathrm{Int}(\mathrm{Dom}(f))$ and, therefore, there exists a positive number t such that the closed ball of center x and radius t is contained in $\mathrm{Int}(\mathrm{Dom}(f))$. For any $y \in X$ with $\|y - x\| = t/2$ we have

$$f(y) \ge f(x) + \langle f'(x), y - x\rangle + \nu_f\left(x, \tfrac{t}{2}\right).$$

Applying Theorem 3.3.2 in [37], one deduce that, since f is Fréchet differentiable at x, there exists a selection $\gamma : \mathrm{Dom}\,(\partial f) \to X^*$ of ∂f which is continuous at x. Hence, $\gamma(x) = f'(x)$ and there exists a number $\tau > 0$ such that

$$\|z - x\| \le \tau \Rightarrow \|\gamma(z) - f'(x)\| \le \tfrac{1}{t}\nu_f\left(x, \tfrac{t}{2}\right). \tag{4}$$

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

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

Then,

$$\|w - x\| = \tfrac{1-\alpha}{2}\|u - x\| = \tfrac{1-\alpha}{2}t < \min(t, \tau),$$

showing that $w \in \mathrm{Int}(\mathrm{Dom}(f))$ and that

$$\|\gamma(w) - f'(x)\| \le \tfrac{1}{t}\nu_f\left(x, \tfrac{t}{2}\right) \tag{5}$$

because of (4). Obviously, we have

$$f(x) \ge f(w) + \langle \gamma(w), x - w\rangle = f(w) + \left\langle \gamma(w), \tfrac{1-\alpha}{2}(x - u)\right\rangle \tag{6}$$

and

$$\tfrac{1}{2}(x + u) = \tfrac{\alpha}{\alpha+1}u + \tfrac{1}{\alpha+1}w. \tag{7}$$

From (7) we deduce that

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

that is,

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

Summing up this inequality and (6) we get

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

Consequently,

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

where the last inequality results from (5). Hence, we have

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

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

$$\overline{\mu}_f(x,t) \geq \tfrac{1-\alpha}{2\alpha}\nu_f\left(x, \tfrac{t}{2}\right) > 0.$$

In view of Lemma 2.3, the proof is complete. ■

2.4 It was emphasized in [15] that efficient applicability of various iterative algorithms for solving feasibility and optimization problems in Banach spaces depends on the availability of large pools of totally convex functions and of good evaluations for their moduli of total convexity. Theorem 1 of Asplund [3] establishes locally uniform convexity of the function $\|\cdot\|^2$ at any $x \in X$ when X is locally uniformly convex. In [36, Theorem 4.1(i)] this result was extended to a larger class of functions including the functions $\|\cdot\|^r$ with $r \in (1, +\infty)$. These results are qualitative in the sense that they do not give estimates of the moduli of local uniform convexity of the functions $\|\cdot\|^r$ to which they apply. Having such estimates is of interest, among other things, for establishing convergence and/or determining error upper bounds for algorithms like those discussed in [2] and in Section 4 below. Adaptation of Asplund's proof in [3, Theorem 1] led to evaluations of the moduli of total convexity of the functions $\|\cdot\|^r$ based on the local modulus of convexity of the space $\delta_X : \{x \in X : \|x\| = 1\} \times [0, +\infty) \to [0, +\infty]$ defined by

$$\delta_X(x,t) := \inf\left\{1 - \tfrac{\|x+y\|}{2} : \|y\| = 1, \|x-y\| \geq t\right\},$$

with the usual convention that $\inf \emptyset = +\infty$. Clearly, $\delta_X(x,t) \in [0,1]$ when $t \in [0,2]$ and $\delta_X(x,t) = +\infty$, otherwise. Such an evaluation is given in [17] and presented in [13, Remark 1.4.15] for the functions $\|\cdot\|^r$ with $r \in [2, +\infty)$. The next result gives evaluations of the moduli of uniform convexity $\overline{\mu}_f(x,t)$ for the functions $f = \|\cdot\|^r$ with $r > 1$ in locally uniformly convex Banach spaces. In view of Lemma 2.3 and of the fact that $\nu_f(x,t) \geq \overline{\mu}_f(x,t)$ for all $t \geq 0$, this is also an evaluation of $\nu_f(x,t)$. For uniformly convex spaces sharper evaluations of $\nu_f(x,t)$ for $f = \|\cdot\|^r$ with $r > 1$ can be found in [16] and, in the particular case of $X = \mathcal{L}^p$, even better estimates are given in [26].

Proposition. *If X is a locally uniformly convex Banach space, $r \in (1, +\infty)$, and if the functions $f : X \to \mathbb{R}$ and $\phi : \mathbb{R} \to \mathbb{R}$ are defined by $f(x) := \|x\|^r$ and $\phi(t) := |t|^r$, respectively, then, for any $t \geq 0$, we have*

$$\overline{\mu}_f(0,t) = \left(1 - 2^{1-r}\right) t^r, \tag{8}$$

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

$$\overline{\mu}_f(x,t) \geq \|x\|^r \min\left\{\overline{\mu}_\phi\left(1, \tfrac{\alpha t}{\|x\|}\right), \zeta(a)\right\}, \tag{9}$$

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

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

$$\overline{\mu}_\phi(1,s) = \min\left\{1 + (1+s)^r - 2^{1-r}(2+s)^r, 1 + |1-s|^r - 2^{1-r}|2-s|^r\right\},$$

and

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

Proof. According to the definition of $\overline{\mu}_f$, we have

$$\overline{\mu}_f(0,t) = \inf\left\{\|y\|^r - 2\left\|\tfrac{y}{2}\right\|^r : \|y\| \geq t\right\} = \left(1 - 2^{1-r}\right) t^r,$$

and this proves (8). Now, suppose that $x \in X \setminus \{0\}$. Observe that, in this situation,

$$\overline{\mu}_f(x,t) = \|x\|^r \, \overline{\mu}_f\left(\tfrac{x}{\|x\|}, \tfrac{t}{\|x\|}\right)$$

and, therefore, it is sufficient to prove (9) under the assumption that $\|x\| = 1$ and $t > 0$. Assume that $\|x\| = 1$ and $t > 0$. Note that

$$\overline{\mu}_f(x,t) = \inf\left\{1 + \beta^r - 2\left\|\tfrac{x+\beta y}{2}\right\|^r : \|y\| = 1, \beta \geq 0, \|x - \beta y\| \geq t\right\}.$$

Fix $\alpha \in (0,1)$. Suppose that $\beta \geq 0$ and $y \in X$ are such that $\|y\| = 1$ and $\|x - \beta y\| \geq t$. We distinguish two complementary cases.

Case 1. Suppose that $|\beta - 1| \geq \alpha t$. Then, since $\left\|\frac{x+\beta y}{2}\right\| \leq \frac{1+\beta}{2}$, we deduce that

$$1 + \beta^r - 2\left\|\tfrac{x+\beta y}{2}\right\|^r \geq 1 + \beta^r - 2\left(\tfrac{1+\beta}{2}\right)^r \geq \overline{\mu}_\phi(1, \alpha t). \tag{10}$$

Case 2. *Suppose that* $|\beta - 1| < \alpha t$. Then, we have

$$\|x - y\| = \|x - \beta y + (\beta - 1)\, y\| \geq \|x - \beta y\| - |\beta - 1| > (1 - \alpha)t.$$

Denote $a = \delta_X\,(x, (1-\alpha)t)$. Now, we distinguish two possible situations. First, assume that $\beta \geq 1$. Since

$$a \leq 1 - \left\| \tfrac{x+y}{2} \right\| \tag{11}$$

and since

$$\|x + \beta y\| = \|x + y + (\beta - 1)\, y\| \leq \|x + y\| + \beta - 1$$

we get

$$\left\| \tfrac{x+\beta y}{2} \right\| \leq 1 - a + \tfrac{\beta - 1}{2}.$$

Therefore, we obtain

$$1 + \beta^r - 2 \left\| \tfrac{x+\beta y}{2} \right\|^r \geq 1 + \beta^r - 2 \left(\tfrac{\beta+1}{2} - a \right)^r . \tag{12}$$

The function $\psi : [1, +\infty) \to \mathbb{R}$ defined by

$$\psi(u) = 1 + u^r - 2 \left(\tfrac{u+1}{2} - a \right)^r ,$$

is differentiable and its derivative $\psi'(u)$ is nonnegative when $u \geq 1$. Hence, ψ is nondecreasing on $[1, +\infty)$ and, therefore, for any $u \geq 1$,

$$\psi(u) \geq \psi(1) = 2\left[1 - (1 - a)^r\right].$$

This and (12) imply that

$$1 + \beta^r - 2 \left\| \tfrac{x+\beta y}{2} \right\|^r \geq 2\left[1 - (1 - a)^r\right], \tag{13}$$

whenever $\beta \geq 1$.

Consider now $\beta \in [0, 1)$. Note that, according to (11),

$$\begin{aligned} \|x + \beta y\| &= \|\beta(x + y) + (1 - \beta)x\| \leq \beta\, \|x + y\| + 1 - \beta \\ &\leq 2\beta(1 - a) + 1 - \beta = 1 + \beta - 2a\beta. \end{aligned}$$

Consequently, we get

$$1 + \beta^r - 2 \left\| \tfrac{x+\beta y}{2} \right\|^r \geq 1 + \beta^r - 2 \left[\tfrac{1}{2} + \left(\tfrac{1}{2} - a \right) \beta \right]^r . \tag{14}$$

The function

$$\chi(u) = 1 + u^r - 2 \left[\tfrac{1}{2} + \left(\tfrac{1}{2} - a \right) u \right]^r$$

attains its minimum on the interval $[0, 1]$ at the point

$$u_0 = \begin{cases} \dfrac{(1-2a)^{\frac{1}{r-1}}}{2-(1-2a)^{\frac{r}{r-1}}} & \text{if } a < \frac{1}{2}, \\ 0 & \text{if } a \geq \frac{1}{2}. \end{cases}$$

Thus, according to (14), if $a \geq \frac{1}{2}$, then

$$1 + \beta^r - 2\left\|\frac{x+\beta y}{2}\right\|^r \geq \chi(0) = 1 - 2^{1-r}. \tag{15}$$

If $a < \frac{1}{2}$, then

$$1 + \beta^r - 2\left\|\frac{x+\beta y}{2}\right\|^r \geq \chi(u_0) = 1 - \left[2 - (1-2a)^{\frac{r}{r-1}}\right]^{1-r}. \tag{16}$$

Taking into account that

$$2\left[1 - (1-a)^r\right] = \chi(1) \geq \chi(u_0),$$

and the inequalities (10), (13), (15) and (16) one deduces (9). ■

2.5 A question which naturally occurs in the applications of totally convex functions to studying convergence properties of feasibility algorithms like those presented in [13] is whether a function f which is totally convex at some points of the space X is still totally convex at those points when restricted to a linear subspace Y containing those points and provided with a stronger norm than the restriction of $\|\cdot\|$ to Y. The following example shows that this is not the case even if Y is dense in X (with respect to the norm of X). Consider $X = L^2[0,1]$ provided with its usual norm $\|\cdot\|_2$ and $Y = L^p[0,1]$ with its norm $\|\cdot\|_p$ for some $p \in (2, +\infty)$. Take $f : X \to \mathbb{R}$ given by $f(x) = \|x\|_2^r$ for a real number $r > 1$. The function f is totally convex at any point $x \in X$ as noted in Subsection 2.4. Let $a \in X \backslash Y$ and note that Y is dense in X. Therefore, there exists a sequence $\left\{x^k\right\}_{k\in\mathbb{N}}$ contained in Y such that $\lim_{k\to\infty}\left\|x^k - a\right\|_2 = 0$. We claim that $\lim_{k\to\infty}\left\|x^k\right\|_p = +\infty$. Indeed, if we assume otherwise, then there exists a subsequence $\left\{x^{i_k}\right\}_{k\in\mathbb{N}}$ which converges weakly in Y to some $\overline{x} \in Y$; thus, $\left\{x^{i_k}\right\}_{k\in\mathbb{N}}$ converges weakly in X to $\overline{x}$ (because X^* is a subspace of Y^*) and, therefore, we have

$$0 \leq \|\overline{x} - a\|_2 \leq \liminf_{k\to\infty} \left\|x^{i_k} - a\right\|_2 = \lim_{k\to\infty}\left\|x^k - a\right\|_2 = 0,$$

showing that $a = \overline{x} \in Y$, a contradiction. Since a cannot be zero we may assume without loss of generality that $x^k \neq 0$ for all $k \in \mathbb{N}$. Let $z^k = \frac{x^k}{\|x^k\|_p}$. Then,

$$\lim_{k\to\infty} D_f(z^k, 0) = \lim_{k\to\infty} \frac{\|x^k\|_2^r}{\|x^k\|_p^r} = 0,$$

because the sequence $\left\{\left\|x^k\right\|_2\right\}_{k\in\mathbb{N}}$ is bounded and $\lim_{k\to\infty}\left\|x^k\right\|_p = +\infty$. The restriction of f to Y, denoted h, has

$$0 \leq \nu_h(0,1) = \inf\left\{D_h(y,0) : \|y\|_p = 1\right\} \leq \inf_{k\in\mathbb{N}} D_f(z^k, 0) = \lim_{k\to\infty} D_f(z^k,0) = 0,$$

showing that, in contrast with f, the function h is not totally convex at zero.

3 Characteristic Properties of Totally Convex Functions

3.1 If X is a space of finite dimension and if f is a lower semicontinuous function, then f is totally convex at any point $x \in \mathrm{Int}(\mathrm{Dom}(f))$ if and only if f is strictly convex on $\mathrm{Int}(\mathrm{Dom}(f))$. This follows by an easy adaptation of the proof of Proposition 1.2.6 in [13]. On the other hand, Theorem 26.3 in [31] shows that, when $X = \mathbb{R}^n$, strict convexity of f on $\mathrm{Int}(\mathrm{Dom}(f))$ is equivalent to Fréchet differentiability of the conjugate function f^* at the points $x^* \in \mathrm{Int}(\mathrm{Dom}(f^*))$. In the light of these facts, Professor R. T. Rockafellar raised the question[1] whether a similar connection still holds for totally convex functions in general spaces (which may have infinite dimension). We are aiming now towards an answer to this question.

3.2 To this end, let $\mathcal{F}$ be the set of functions $\psi : [0, +\infty) \to [0, +\infty]$ which satisfy the following conditions:

(i) $\mathrm{Int}(\mathrm{Dom}(\psi)) \neq \emptyset$;
(ii) ψ is convex and lower semicontinuous;
(iii) $\psi(0) = 0$ and $\psi(t) > 0$ whenever $t > 0$.

Note that any $\psi \in \mathcal{F}$ has $\sup(\mathrm{Dom}(\psi)) > 0$ and it is continuous on the interval $[0, \sup(\mathrm{Dom}(\psi)))$ as follows from [37, Proposition 2.1.6].

Recall (see [37, Section 3.3]) that whenever a function $\varphi : [0, +\infty) \to [0, +\infty]$ has $\varphi(0) = 0$, its *pseudo-conjugate* $\varphi^{\#} : [0, +\infty) \to [0, +\infty]$, defined by

$$\varphi^{\#}(t) = \sup\{st - \varphi(s) : s \geq 0\},$$

is lower semicontinuous, convex, has

$$\varphi^{\#\#} := (\varphi^{\#})^{\#} = \overline{\mathrm{co}}\varphi \leq \varphi,$$

and $\varphi^{\#}(0) = 0$. If $\varphi \in \mathcal{F}$, then $\varphi^{\#\#} = \varphi$. Also, if φ is increasing and has $\liminf_{t\to\infty} \frac{\varphi(t)}{t} > 0$, then $\varphi^{\#\#} \in \mathcal{F}$ as follows from [36, Proposition A5].

3.3 The following characterization of the total convexity at a point will be of use in what follows.

Lemma. *Suppose that $x \in \mathrm{Dom}(f)$ 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 $y \in X$, we have*

$$f(y) - f(x) \geq f^{\circ}(x, y - x) + \varphi(\|y - x\|). \tag{17}$$

Proof. According to [36, Theorem 2.1], if there exists $\varphi \in \mathcal{F}$ such that (17) is satisfied for all $y \in X$, then f is totally convex at x. Conversely, suppose that

[1] This question was raised by Prof. R.T. Rockafellar during the Conference on Continuous Optimization held in Rio de Janeiro, Brazil, June 1999.

f is totally convex at $x \in \mathrm{Dom}(f)$. By [13, Proposition 1.2.2(ii)], we have that $\nu_f(x,\cdot)$ is nondecreasing and

$$\liminf_{t\to\infty} \frac{\nu_f(x,t)}{t} \geq \nu_f(x,1) > 0.$$

Therefore, Proposition A5 in [36] applies and shows that the (necessarily nonnegative) lower semicontinuous convex function $\varphi := \overline{\mathrm{co}}\nu_f(x,\cdot)$ is positive on $(0,+\infty)$. Clearly, $\varphi(t) \leq \nu_f(x,t)$, for all $t \geq 0$. Hence, $\varphi(0) = 0$, $\mathrm{Dom}(\varphi) \supseteq \mathrm{Dom}(\nu_f(x,\cdot))$ and we have

$$\begin{aligned} f(y) - f(x) &\geq f^\circ(x, y-x) + \nu_f(x, \|y-x\|) \\ &\geq f^\circ(x, y-x) + \varphi(\|y-x\|), \end{aligned} \tag{18}$$

for all $y \in X$. Since $\mathrm{Dom}(f)$ contains at least two points and $\partial f(x) \neq \emptyset$, it results that, for some $\overline{y} \in \mathrm{Dom}(f)\backslash\{x\}$, the right hand sided derivative $f^\circ(x, \overline{y}-x)$ is finite. Hence, by writing (18) for $y = \overline{y}$ we deduce that, for $t_0 = \|\overline{y}-x\| > 0$, we have $\nu_f(x,t_0) < +\infty$, that is, $[0,t_0) \subseteq \mathrm{Dom}(\nu_f(x,\cdot))$ which implies that $\mathrm{Int}[\mathrm{Dom}(\varphi)] \neq \emptyset$. Consequently, $\varphi \in \mathcal{F}$ and satisfies (17). ■

3.4 With these in mind we give a partial answer to Professor Rockafellar's question stated in Subsection 3.1 above. The next result, establishing basic properties of a function f which is totally convex and subdifferentiable at a point $x \in \mathrm{Dom}(f)$, was repeatedly proved under more restrictive conditions. It appears in [36, Theorem 2.1] under the assumptions that f should be lower semicontinuous and X reflexive. The reflexivity of X was dropped in [37, Corollary 3.4.4]. We show now that some of these properties are still satisfied even if f is not lower semicontinuous.

Proposition. *If the function f is totally convex at $x \in \mathrm{Dom}(f)$ and if $x^* \in \partial f(x)$, then $x^* \in \mathrm{Int}(\mathrm{Dom}(f^*))$ and any of the following equivalent conditions are satisfied:*

(i) There exists $\varphi \in \mathcal{F}$ such that, for any $y \in X$, we have

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

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

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

(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 $y^* \in \partial f(y)$, we have

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

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

$$\|y - x\| \le \theta\left(\|y^* - x^*\|\right).$$

Proof. Suppose that f is totally convex at $x \in \mathrm{Dom}(f)$ and $x^* \in \partial f(x)$. Since $f^\circ(x, y-x) \ge \langle x^*, y-x\rangle$, Lemma 3.3 shows that (i) holds. The equivalence $(ii)\Leftrightarrow(iii)$ and the equivalence $(iv)\Leftrightarrow(v)$ can be proved exactly as the corresponding implications in Theorem 2.1 of [36]. For proving $(i)\Rightarrow(ii)$ note that, whenever (i) holds and $y^* \in X^*$, we have that

$$\begin{aligned}
f^*(y^*) &= \sup_{y\in X}\{\langle y^*, y\rangle - f(y)\} \\
&\le \sup_{y\in X}\{\langle y^*, y\rangle - f(x) - \langle x^*, y-x\rangle - \varphi(\|y-x\|)\} \\
&= f^*(x^*) + \sup_{y\in X}\{\langle y^* - x^*, y\rangle - \varphi(\|y-x\|)\} \\
&= f^*(x^*) + \langle y^* - x^*, x\rangle + \sup_{y\in X}\{\langle y^* - x^*, y-x\rangle - \varphi(\|y-x\|)\} \\
&= f^*(x^*) + \langle y^* - x^*, x\rangle + \sup_{\substack{\|y-x\|=t\\ t\ge 0}}\{\langle y^* - x^*, y-x\rangle - \varphi(t)\} \\
&= f^*(x^*) + \langle y^* - x^*, x\rangle + \sup_{t\ge 0}\{t\,\|y^* - x^*\| - \varphi(t)\} \\
&= f^*(x^*) + \langle y^* - x^*, x\rangle + \varphi^{\#}(\|y^* - x^*\|).
\end{aligned}$$

Conversely, suppose that (ii) holds. Let $\overline{f}$ be the lower semicontinuous envelope of f. According to [8, Proposition 2.118] we have $\overline{f}^* = f^*$, $\overline{f}(x) = f(x)$ and $\partial\overline{f}(x) = \partial f(x)$. Hence, we have

$$\overline{f}^*(y^*) - \overline{f}^*(x^*) \le \langle y^* - x^*, x\rangle + \varphi^{\#}\left(\|y^* - x^*\|\right),$$

for all $y^* \in X^*$. This and [37, Corollary 3.4.4] imply that

$$\overline{f}(y) - \overline{f}(x) \ge \langle x^*, y-x\rangle + \varphi\left(\|y-x\|\right),$$

for all $y \in X$. Since $\overline{f}(y) \le f(y)$ and $\overline{f}(x) = f(x)$, the inequality in (i) results. The implication $(ii)\Rightarrow(iv)$ can be proved (without requiring that f is lower semicontinuous) in the same way in which the analogous implication was proven in [36, Theorem 2.1]. When f is lower semicontinuous the implication $(v)\Rightarrow(ii)$ results from [37, Corollary 3.4.4]. ■

3.5 Note that Proposition 3.4 does not require that x should be a point in the interior of the domain of f. When this happens, Proposition 3.4 can be converted into a characterization of total convexity at x.

Proposition. *Suppose that* f *is continuous at the point* $x \in \mathrm{Int}\,(\mathrm{Dom}\,(f))$. *Then,* f *is totally convex at* x *if and only if any of the following three equivalent conditions is satisfied:*

(i') There exists $\varphi \in \mathcal{F}$ such that, for any $y \in X$ and for any $x^* \in \partial f(x)$ we have

$$f(y) - f(x) \geq \langle x^*, y - x \rangle + \varphi \left(\|y - x\| \right); \tag{19}$$

(ii') There exists $\varphi \in \mathcal{F}$ such that, for any $y^* \in X^*$ and for any $x^* \in \partial f(x)$ we have

$$f^*(y^*) - f^*(x^*) \leq \langle y^* - x^*, x \rangle + \varphi^{\#} \left(\|y^* - x^*\| \right); \tag{20}$$

(iii') $\partial f(x) \subseteq \operatorname{Int}(\operatorname{Dom}(f^*))$ 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:

(iv') 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 $y^* \in \partial f(y)$, we have

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

(v') There exists a nondecreasing function $\theta : [0, +\infty) \to [0, +\infty]$ with $\lim_{t \searrow 0} \theta(t) = 0$ such that, for any pair $(y, y^*) \in X \times X^*$ with $y^* \in \partial f(y)$, we have

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

Proof. The function f is continuous at $x \in \operatorname{Int}(\operatorname{Dom}(f))$ and, therefore, $\partial f(x) \neq \emptyset$. Also, we have

$$f^\circ(x, y) = \max \left\{ \langle x^*, y \rangle : x^* \in \partial f(x) \right\},$$

for all $y \in X$. Consequently, (i') holds if and only if (17) is satisfied, that is, f is totally convex at x if and only if (i') is true. The equivalence of (i') and (ii') follows from that of (i) and (ii) in Proposition 3.4 written for each $x^* \in \partial f(x)$. Suppose that (ii') holds. Then, for any $y^* \in X^*$ and for each $x^* \in \partial f(x)$, we have

$$0 \leq \frac{f^*(y^*) - f^*(x^*) - \langle y^* - x^*, x \rangle}{\|y^* - x^*\|} \leq \frac{\varphi^{\#} \left(\|y^* - x^*\| \right)}{\|y^* - x^*\|}. \tag{21}$$

Since $\varphi \in \mathcal{F}$ one can use Proposition A2(ii) from [36] in order to deduce that $\lim_{t \searrow 0} \frac{\varphi^{\#}(t)}{t} = 0$. Hence, letting $\|y^* - x^*\| \to 0$ in (21), we deduce that f^* is uniformly Fréchet differentiable on $\partial f(x)$, i.e., (iii') holds. Now, suppose that (iii') is satisfied. Define the function $\psi : [0, +\infty) \to [0, +\infty]$ by

$$\psi(t) = \sup \left\{ f^*(x^* + ty^*) - f^*(x^*) - t \langle y^*, x \rangle : \|y^*\| = 1, x^* \in \partial f(x) \right\}.$$

The convexity of f^* and the fact that $x \in \partial f^*(x^*)$ whenever $x^* \in \partial f(x)$, imply that ψ is nonnegative, lower semicontinuous, convex and has $\psi(0) = 0$. From (iii') it results that $\lim_{t \searrow 0} \frac{\psi(t)}{t} = 0$. Taking $\varphi := \psi^{\#}$ and applying Proposition A2(ii) from [36] to it one deduces that $\varphi \in \mathcal{F}$. From the definition of ψ we deduce that (ii') holds with $\varphi = \psi^{\#}$. Hence, (ii') and (iii') are equivalent.

Suppose that f is lower semicontinuous. Then, the equivalence of (iv') and (v'), as well as the implication $(v') \Rightarrow (ii')$ in the case that f is lower semicontinuous, result from Proposition 3.4. If (ii') holds, then the function $\theta : [0, +\infty) \to [0, +\infty]$ given by

$$\theta(t) = \begin{cases} \frac{\varphi^{\#}(t)}{t} & \text{if } t \in (0, +\infty) \backslash \left\{ \sup \left(\mathrm{Dom} \left(\varphi^{\#} \right) \right) \right\}, \\ 0 & \text{if } t = 0, \\ +\infty & \text{if } t = \sup \left(\mathrm{Dom} \left(\varphi^{\#} \right) \right) < +\infty, \end{cases}$$

for φ from (ii'), satisfies (v'), i.e., (ii') and (v') are also equivalent. ∎

3.6 The characterization of total convexity at interior points of the domain of f contained in Proposition 3.5 allows us to give an answer to a question which occurred in our discussions with Yakov Alber concerning the existence on an arbitrary Banach space X of functions f which, at any point $x \in \mathrm{Int}\,(\mathrm{Dom}\,(f))$, are totally convex and have $\nu_f(x, t) \geq c_f(x) t^2$ for all $t \in [0, 1]$, where $c_f(x)$ is a positive number. Our answer is contained in the following result.

Proposition. *Let $g : X^* \to (-\infty, +\infty]$ be a convex function which is weak* lower semicontinuous and strictly convex on its domain. Suppose that there exists an open ball $B(x^*, r)$, of center x^* and radius $r > 0$, such that $B(x^*, r) \subset \mathrm{Int}\,(\mathrm{Dom}\,(g))$ and such that*

(i) The function g is Gâteaux differentiable on $B(x^*, r)$;

(ii) There exists a real number $k > 0$ such that, for each $y^* \in B(x^*, r)$, the Gâteaux derivative g' satisfies

$$\|g'(y^*) - g'(x^*)\| \leq k \|y^* - x^*\|.$$

Then the vector $x := g'(x^*)$ belongs to X and the function $f := g^* : X \to (-\infty, +\infty]$ is totally convex at x and has

$$\nu_f(x, t) \geq \begin{cases} \frac{t^2}{2k} & \text{if } t \in [0, kr], \\ \frac{r}{2}\left(t - \frac{kr}{2}\right) & \text{if } t > kr. \end{cases} \tag{22}$$

Proof. Using [6, Lemma 5.1] one obtains that the mapping ∂f is single valued on its domain. The Asplund-Rockafellar Theorem (see [4] or [37, Corollary 3.3.4]) ensures that, in the current circumstances, $x = g'(x^*) = (f^*)'(x^*) \in X$. Define the function $\theta : [0, +\infty) \to [0, +\infty]$ by

$$\theta(t) = \begin{cases} kt & \text{if } t \in [0, r), \\ +\infty & \text{if } t \geq r. \end{cases} \tag{23}$$

Observe that, if $y^* \in B(x^*, r)$, then, according to hypothesis (ii), we have

$$\|g'(y^*) - g'(x^*)\| \leq \theta\left(\|y^* - x^*\|\right). \tag{24}$$

If $y \in X$ and $y^* \in \partial f(y)$, then

$$\|y - x\| \leq \theta\left(\|y^* - x^*\|\right), \tag{25}$$

since this is exactly (24) when $y^* \in B(x^*, r)$ and, otherwise, $\theta\left(\|y^* - x^*\|\right) = +\infty$. In other words, the function f satisfies condition (v') of Proposition 3.5.

This implies that the function f also satisfies (20) for some $\varphi \in \mathcal{F}$. Analyzing the proof of Proposition 3.5 (see the part concerning the implication $(v')\Rightarrow(ii')$ in the proof of Theorem 2.1 from [36]) it follows that, if (25) holds, then the inequality (20) is satisfied for some $\varphi \in \mathcal{F}$ such that

$$\varphi^{\#}(t) = \int_0^t \theta(s)ds$$

Hence, taking into account (23), we get that $\varphi^{\#}(t) = \frac{1}{2}kt^2$ for $t \in [0,r]$, and $\varphi^{\#}(t) = +\infty$, otherwise. Since $\varphi = \varphi^{\#\#}$, we deduce that

$$\varphi(s) = \begin{cases} \frac{s^2}{2k} & \text{if } s \in [0,kr], \\ \frac{r}{2}\left(s - \frac{kr}{2}\right) & \text{if } s > kr. \end{cases}$$

Observing that (19) holds with φ given here and that this implies $\nu_f(x,s) \geq \varphi(s)$ for all $s \geq 0$, we deduce (22). ■

3.7 In order to guarantee convergence of some infinite dimensional optimization algorithms in a Banach space X (see [28, Section 4], [2], [1] and [13, Section 3.5.2]) one has to ensure existence on X of a convex function f with the following property:

(A) Whenever the set $E \subset \operatorname{Int}(\operatorname{Dom}(f))$ is bounded, there exists a positive constant $c_f(E)$ such that

$$\inf_{x\in E} \nu_f(x,t) \geq c_f(E)t^2,$$

for all $t \in [0,1]$.

Obviously, condition (A) is a stronger requirement than this occurring in Yakov Alber's question (see Subsection 3.6). For spaces $L^q = L^q(\Omega)$, where Ω is a space of finite measure and $q \in (1,2]$, validity of condition (A) for the function $f = \|\cdot\|^2$ was established in the proof of [13, Corollary 3.5.9]. We will show next, using Proposition 3.6, that in the space $X = \ell^q$ with $q \in (1,2]$ the q-th power of the norm has this property. Precisely, we have the next result:

Corollary. *For any $q \in (1,2]$ the function $f(x) = \frac{1}{q}\|x\|_q^q$ on $X = \ell^q$ satisfies condition (A).*

Proof. We first prove the following result, which is probably known but we do not have a reference for it.

Claim 1. Let $p \in [2,+\infty)$ be such that $q = p/(p-1)$ and $h : \mathbb{R}^n \to \mathbb{R}$ be defined by $h(x) := \frac{1}{p}\|x\|_p^p$. Then

$$\|h'(x) - h'(y)\|_q \leq (p-1)\left(\|x\|_p + \|y\|_p\right)^{\frac{p-2}{p}} \|x-y\|_p \quad \forall\, x,y \in \mathbb{R}^n. \tag{26}$$

In order to prove this claim, consider the function $\varphi : \mathbb{R} \to \mathbb{R}$ defined by $\varphi(t) = \frac{1}{p}|t|^p$. We have that

$$\varphi'(t) = |t|^{p-1}\operatorname{sgn} t \text{ and } \varphi''(t) = (p-1)|t|^{p-2},$$

for every $t \in \mathbb{R}$. Note that for any two real numbers t and s, there exists a real number θ between t and s such that

$$\varphi'(t) - \varphi'(s) = \varphi''(\theta)(t-s).$$

Let $x = (x_1, ..., x_n) \in \mathbb{R}^n$ and $y = (y_1, ..., y_n) \in \mathbb{R}^n$ be fixed. Then, for each integer k between 1 and n, there exists ξ_k between x_k and y_k such that

$$\varphi'(x_k) - \varphi'(y_k) = (p-1)\left|\xi_k\right|^{p-2}(x_k - y_k).$$

Since

$$h'(x) = \left(|x_1|^{p-1}\operatorname{sgn} x_1, \ldots, |x_n|^{p-1}\operatorname{sgn} x_n\right),$$

we have that

$$\begin{aligned}
\|h'(x) - h'(y)\|_q^q &= |\varphi'(x_1) - \varphi'(y_1)|^q + \ldots + |\varphi'(x_n) - \varphi'(y_n)|^q && (27)\\
&= (p-1)^q \left(|\xi_1|^{q(p-2)}|x_1 - y_1|^q + \ldots + |\xi_n|^{q(p-2)}|x_n - y_n|^q\right)\\
&\leq (p-1)^q \left(|\xi_1|^p + \ldots + |\xi_n|^p\right)^{\frac{p-2}{p-1}} \left(|x_1 - y_1|^p + \ldots + |x_n - y_n|^p\right)^{\frac{q}{p}},
\end{aligned}$$

where the last inequality results from Hölder's inequality written for the conjugate numbers $\frac{p-1}{p-2} = \frac{p}{q(p-2)}$ and $\frac{p}{q}$. Taking into account that $|\xi_k| \leq |x_k| + |y_k|$, we get (26) from (27) and Claim 1 is proved.

Now, let $q \in (1, 2]$, $p = q/(q-1)$ and define the function $g : \ell^p \to \mathbb{R}$ by $g(x) := \frac{1}{p}\|x\|_p^p$. Then g is Fréchet differentiable on ℓ^p and we have

$$g'(x) = \left(|x_1|^{p-1}\operatorname{sgn} x_1, |x_2|^{p-1}\operatorname{sgn} x_2, \ldots\right) \in \ell^q.$$

Moreover

$$\|g'(x)\|_q = \|x\|_p^{p-1} \quad \forall\, x \in \ell^p. \tag{28}$$

Let $x = (x_1, x_2, \ldots)$ and $y = (y_1, y_2, \ldots)$ be sequences in ℓ^p. Applying (26) to the vectors $x^{[n]} = (x_1, ..., x_n)$ and $y^{[n]} = (y_1, ..., y_n)$ and then passing to the limit for $n \to \infty$ we get

$$\|g'(x) - g'(y)\|_q \leq (p-1)\left(\|x\|_p + \|y\|_p\right)^{\frac{p-2}{p}} \|x - y\|_p \quad \forall\, x, y \in \ell^p.$$

This inequality shows that g' is Lipschitz on bounded subsets of ℓ^p. Also, it shows that if r is a positive real number, $E \subset \ell^p$ is a nonempty bounded set and $x \in E$, then for any $y \in B(x, r)$ we have

$$\|g'(x) - g'(y)\|_q \leq (p-1)\left(2\|x\|_p + r\right)^{\frac{p-2}{p}} \|x - y\|_p.$$

As a consequence, for any $x \in E$ and for any $y \in \ell^p$ with $\|x - y\|_p < r$, we have

$$\|g'(x) - g'(y)\|_q \leq k(E, r)\|x - y\|_p, \tag{29}$$

where

$$k(E,r) := (p-1)\left(2\sup_{x\in E}\|x\|_p + r\right)^{\frac{p-2}{p}}.$$

Recall (see, for instance, [8]) that $g^* = f$ and, therefore,

$$f' = g^{*\prime} = (g')^{-1}. \tag{30}$$

Take a nonempty bounded set $F \subset \ell^q$ and denote $E = f'(F)$. Clearly, E is a bounded subset of ℓ^p (cf. (28)). Fix $r > 0$ such that $k := k(E,r)$ has $kr \geq 1$. If $x \in F$, then $x = g'(x^*)$ for some $x^* \in E$ because f' and g' are inverse to each other (cf. (30)). In view of (29), application of Proposition 3.6 gives

$$\nu_f(x,t) \geq \frac{1}{2k}t^2,$$

for all $x \in F$ and for all $t \in [0, kr] \supset [0,1]$. ■

4 Sequentially Consistent Functions and Convergence of Some Iterative Algorithms

4.1 The aim of this section is to establish connections among the notions of (locally) uniform convexity on bounded sets, total convexity on bounded sets and sequential consistency of convex functions. These connections are used in order to improve upon existing convergence results concerning algorithms for feasibility and optimization in infinite dimensional Banach spaces. To precise the terminology, recall that the function $f : X \to (-\infty, +\infty]$ is called *uniformly convex on bounded sets* if for each bounded nonempty subset E of X the function $\overline{\mu}_f(E,\cdot) : X \to [0,+\infty]$ given by

$$\overline{\mu}_f(E,t) = \inf\left\{\overline{\mu}_f(x,t) : x \in E \cap \mathrm{Dom}\,(f)\right\},$$

is positive whenever $t > 0$. It can be easily seen that f is uniformly convex on bounded sets if and only if f is *uniformly strictly convex* in the sense given to this term by Asplund in [3, p. 231]. Also, Lemma 2.3 shows that f is uniformly convex on bounded sets if and only if for any bounded set $E \subset X$ we have

$$\mu_f(E,t) := \inf\left\{\mu_f(x,t) : x \in E \cap \mathrm{Dom}\,(f)\right\} > 0,$$

for all $t > 0$. The function f is called *totally convex on bounded sets* if for each bounded nonempty subset E of X the function $\nu_f(E,\cdot) : X \to [0,+\infty)$ defined by

$$\nu_f(E,t) = \inf\left\{\nu_f(x,t) : x \in E \cap \mathrm{Dom}\,(f)\right\},$$

is positive on $(0,+\infty)$. The notion termed in [13] sequential consistency occurred first in [20] was also studied in [10] and [11]. The function f is called *sequentially consistent* (*with the norm topology of the space* X) if for any two sequences

$\{x^k\}_{k\in\mathbb{N}}$ and $\{y^k\}_{k\in\mathbb{N}}$ contained in $\mathrm{Dom}\,(f)$ such that $\{x^k\}_{k\in\mathbb{N}}$ is bounded and $\lim_{k\to\infty} D_f(y^k,x^k)=0$ one has $\lim_{k\to\infty}\|x^k-y^k\|=0$.

4.2 We have noted in Section 2 that, in general, total convexity of f at a point $x\in\mathrm{Dom}\,(f)$ does not imply uniform convexity of f at x. The next result shows that, by contrast, for lower semicontinuous convex functions, uniform and total convexity on bounded sets are equivalent notions and, in this way, we answer a question raised by Yakov Alber. Even if f is not lower semicontinuous, total convexity of f on bounded sets is still equivalent to sequential consistency of f as well as to the *property (*)* considered in [18] in the context of the theory of outer Bregman projection algorithms for solving convex feasibility problems.

Proposition. *If $f:X\to(-\infty,+\infty]$ is a convex function whose domain contains at least two different points, then the following conditions are equivalent:*

(i) The function f is totally convex on bounded sets;

(ii) The function f is sequentially consistent;

(iii) The function f has the following property:

(*) For any nonempty bounded set $E\subset X$, there exists a function $\eta_E:[0,+\infty)\times[0,+\infty)\to[0,+\infty)$ such that, for each real number $a>0$, the function $\eta_E(a,\cdot)$ is strictly increasing, continuous, convex, and $\nu_f(E,t)\geq\eta_E(a,t)$ whenever $t\in[0,a)$.

Moreover, if f is lower semicontinuous, then these conditions hold if and only if the function f is uniformly convex on bounded sets.

Proof. The equivalence of conditions (i) and (ii) follows from [13, Lemma 2.1.2] by taking there $C=\mathrm{Dom}\,(f)$. In order to show that $(i)\Leftrightarrow(iii)$, let $E\subset X$ be a bounded set such that $E\cap\mathrm{Dom}\,(f)\neq\emptyset$. According to Lemma 2.4 in [18], the function $\nu_f(E,\cdot)$ has $\nu_f(E,ct)\geq c\nu_f(E,t)$ whenever $c\geq1$ and $t\geq0$, it is nondecreasing and it is increasing on its domain if and only if it is positive for all $t>0$. Suppose that (i) holds. Then, $\nu_f(E,\cdot)$ is increasing on its domain and

$$\liminf_{t\to\infty}\frac{\nu_f\,(E,t)}{t}\geq\nu_f(E,1)>0.$$

Hence, applying Proposition A5 from [36] to $\nu_f(E,\cdot)$, one deduces that $\overline{co}\nu_f(E,\cdot)$ is lower semicontinuous, convex and positive on $(0,+\infty)$. Define $\eta_E:[0,+\infty)\times[0,+\infty)\to[0,+\infty]$ by $\eta_E(a,t)=\overline{co}\nu_f(E,t)$ whenever $a\geq0$ and, then, property (*) holds. Conversely, suppose that condition (iii) holds. Then, we have that $\nu_f(E,t)\geq\eta_E(1,t)>0$ whenever $t\in(0,1]$. For $t>1$ we have that $\nu_f(E,t)\geq\nu_f(E,1)>0$ because $\nu_f(E,\cdot)$ is nondecreasing. Hence, f is totally convex on bounded sets.

Now, let f be lower semicontinuous. Suppose that f is uniformly convex on bounded sets. Take E a bounded set such that $E\cap\mathrm{Dom}\,(f)\neq\emptyset$. Denote by B the closed convex hull of E and $C:=\{x\in X:d(x,B)\leq1\}$. Obviously, C is closed, convex and bounded and $B\subset\mathrm{Int}\,(C)$. Let ι_C be the indicator function of C and define $g=f+\iota_C$. The function g is uniformly convex (i.e., $\mu_g(X,t)>0$ for all $t>0$) because f is uniformly convex on bounded sets. Therefore, one can apply [36, Theorem 2.2] in order to deduce that there exists a function $\psi \in \mathcal{F}$ such that

$$g(y) - g(x) \geq g^\circ(x, y - x) + \psi\left(\|y - x\|\right),$$

for all $x, y \in \operatorname{Dom}(g)$. Therefore, we have

$$f(y) - f(x) \geq g^\circ(x, y - x) + \psi\left(\|y - x\|\right),$$

for all $x \in B \cap \operatorname{Dom}(f)$ and all $y \in C \cap \operatorname{Dom}(f)$. Note that, by the definition of g, whenever $x \in B \cap \operatorname{Dom}(f)$ and $y \in C \cap \operatorname{Dom}(f)$ we have $g^\circ(x, y - x) = f^\circ(x, y - x)$ and, thus,

$$f(y) - f(x) \geq f^\circ(x, y - x) + \psi\left(\|y - x\|\right).$$

Hence, if $x \in B \cap \operatorname{Dom}(f)$, $y \in \operatorname{Dom}(f)$ and $\|y - x\| = t \in (0, 1]$, then

$$f(y) - f(x) \geq f^\circ(x, y - x) + \psi(t),$$

showing that $\nu_f(E, t) \geq \nu_f(B, t) \geq \psi(t) > 0$, for all $t \in (0, 1]$. Since $\nu_f(E, \cdot)$ is nondecreasing, it results that $\nu_f(E, t) > 0$ for all $t > 0$, i.e., f is totally convex on bounded sets. Conversely, assume that f is totally convex on bounded sets. Then, for all $x \in B \cap \operatorname{Dom}(f)$ and all $y \in \operatorname{Dom}(f)$ with $\|y - x\| = t$, we have

$$f(y) - f(x) \geq f^\circ(x, y - x) + \nu_f(B, t).$$

Therefore, the function $h = f + \iota_B$ has

$$\begin{aligned} h(y) - h(x) &\geq h^\circ(x, y - x) + \nu_f(B, \|y - x\|) \\ &\geq h^\circ(x, y - x) + \overline{co}\nu_f(B, \|y - x\|), \end{aligned}$$

whenever $x, y \in \operatorname{Dom}(h)$. According to [37, Theorem 3.5.10(ii)], this implies that h is uniformly convex (i.e., $\mu_h(X, t) > 0$ for all $t > 0$) and this implies that $\mu_f(E, t) \geq \mu_f(B, t) \geq \mu_h(X, t) > 0$, for all $t > 0$. ■

4.3 In the analysis of various iterative algorithms for feasibility and optimization in Banach spaces (see, for instance, [13, Chapters 2 and 3]) two conditions are simultaneously imposed upon the space X for ensuring that the sequences these algorithms produce exist and converge weakly: (a) reflexivity of X and (b) existence on X of a lower semicontinuous, sequentially consistent, convex function f with $\operatorname{Int}(\operatorname{Dom}(f)) \neq \emptyset$. The reflexivity of X is required because it guarantees weak sequential compactness of the unit ball and this is used in the convergence proofs. In some cases, one can prove that the sequences generated by such algorithms converge weakly without involving weak sequential compactness of the unit ball of the space. Instead, one may take advantage of the properties of a sequentially consistent function f in the interior of the domain of which the sequences are necessarily contained. This suggests that it may be possible to ensure convergence of those algorithms and, hence, their applicability in non reflexive Banach spaces under condition (b) only. The following consequence of Proposition 4.2 shows that this is not the case and, in this way, answers a question posed by Simeon Reich: requiring existence of a sequentially consistent, lower semicontinuous, convex function f on X such that $\mathrm{Int}\,(\mathrm{Dom}\,(f)) \neq \emptyset$, one implicitly requires reflexivity of X.

Corollary. *If there exists a lower semicontinuous convex function* $f : X \to (-\infty, +\infty]$ *which is sequentially consistent and has* $\mathrm{Int}\,(\mathrm{Dom}\,(f)) \neq \emptyset$*, then the Banach space* X *is reflexive.*

Proof. Observe that there exists a closed ball $B \subseteq \mathrm{Int}\,(\mathrm{Dom}\,(f))$ and that, according to Proposition 4.2, the function f is uniformly convex on bounded sets. Consequently, the function $g = f + \iota_B$ is uniformly convex, lower semicontinuous and proper and, therefore, Theorem 3.5.13 in [37] applies showing that X is reflexive. ■

4.4 Proposition 4.2 allows us to show that the so called *Outer Bregman Projection Method (OBPM)* for finding solutions of consistent stochastic convex feasibility problems in smooth, uniformly convex Banach spaces converges in more general conditions than previously known. Recall that a *stochastic convex feasibility problem* requires finding $x \in X$ such that

$$g(\omega, x) \leq 0, \qquad \rho\text{-a.e.}(\Omega), \tag{31}$$

where $(\Omega, \mathcal{A}, \rho)$ is a complete probability space and $g : \Omega \times X \to \mathbb{R}$ is a *convex Charathéodory function*, that is, $g(\cdot, x)$ is measurable for each $x \in X$ and $g(\omega, \cdot)$ is convex and continuous for each $\omega \in \Omega$. Some significant problems of applied mathematics can be equivalently rewritten in the form (31) and solved as such. This is the case of constrained convex optimization problems as well as of problems of finding Nash equilibria as shown in [19] where the convergence of particular versions of OBPM was proved and their computational behavior was discussed and illustrated. The fact that any linear Fredholm (and, implicitly, Volterra) equation in L^p can be equivalently rewritten in the form (31) was shown in [16, Section 3]. In [18], where the OPBM was first proposed and studied in its general form, it is also shown how OBPM typically behaves when applied to finding subsolutions of nonlinear integral equations in $L^{3/2}$.

The OBPM is the following iterative procedure of generating sequences in X, provided that X is separable, smooth and uniformly convex and $r \in (1, +\infty)$:

$$x^{k+1} := \int_\Omega J_r^{-1}\left[s_k(\omega)\Gamma_k(\omega) + J_r x^k\right] d\rho(\omega), \tag{32}$$

where $J_r : X \to X^*$ is the duality mapping of weight $t \to rt^{r-1}$, J_r^{-1} stands for the inverse of J_r (which exists because X is smooth and uniformly convex), $\Gamma_k : \Omega \to X$ is a (necessarily existing) measurable selector of the point-to-set mapping $\omega \to \partial g(\omega, \cdot)(x^k) : \Omega \to X$ and $s_k(\omega)$ is a (necessarily existing) solution of the equation

$$\left\langle \Gamma_k(\omega), J_r^{-1}\left[s\Gamma_k(\omega) + J_r x^k\right] - x^k \right\rangle = -g(\omega, x^k) \tag{33}$$

when $g(\omega, x^k) > 0$, and $s_k(\omega) = 0$, otherwise. Theorem 2.1(I) in [18] gives sufficient conditions ensuring that, no matter how the point x^0 is chosen in X, the sequence $\{x^k\}_{k\in\mathbb{N}}$, recursively defined by (32) and (33) with x^0 as initial point, is well defined, bounded, has weak accumulation points, any weak accumulation point of it is a solution of (31) and $\lim_{k\to\infty}\|x^{k+1}-x^k\|=0$. Also, Theorem 2.1$(II)$ in [18] shows that, if the function $f(x)=\|x\|^r$ has the property (*) given in Proposition 4.2(iii), then the size of the constraint violations at each iterative step converges in mean to zero, that is,

$$\lim_{k\to\infty}\int_\Omega \max\left[0, g(\omega, x^k)\right] d\rho(\omega) = 0. \tag{34}$$

Combining these facts and the results in [18, Section 3] one deduces that (34) holds in Hilbert spaces, Lebesgue spaces and Sobolev spaces. We are in position to complete this result and prove that these properties of the sequences $\{x^k\}_{k\in\mathbb{N}}$ generated by the OBPM are valid in any separable, smooth and uniformly convex Banach space.

Theorem. *Suppose that X is a separable, smooth and uniformly convex Banach space, (Ω, A, ρ) is a complete probability space, $r\in(1,+\infty)$ and $g: \Omega\times X\to\mathbb{R}$ is a convex Charathéodory function such that the problem* (31) *has at least one solution x^* for which the function $\varphi(x^*,\cdot): X\to\mathbb{R}$ defined by*

$$\varphi(x^*, y) := (r-1)\|y\|^r - \langle J_r(y), x^*\rangle$$

is convex. Then, for any initial point x^0, the sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by the OBPM is well defined, bounded, has weak accumulation points, each weak accumulation point of $\{x^k\}_{k\in\mathbb{N}}$ is a solution of the problem (31), $\lim_{k\to\infty}\|x^{k+1}-x^k\|=0$, *and the constraint violations converge in mean to zero, i.e.,* (34) *holds.*

In particular, this happens in each of the following situations:

(i) The problem (31) *has $x^*=0$ among its solutions;*

(ii) X is a Hilbert space and $r=2$;

(iii) X is one of the Lebesgue spaces L^p or one of the Sobolev spaces $W^{m,p}$ with $p\in(1,2]$ and $m>1$, $r=p$ and the problem (31) *has a solution x^* which is almost everywhere nonnegative.*

Proof. According to [18, Theorem 2.1] it is sufficient to show that, in the given circumstances, the function $f(x)=\|x\|^r$ has the property (*). This follows from Proposition 4.2 because, when X is uniformly convex, the function $f(x)=\|x\|^r$ is lower semicontinuous and uniformly convex on bounded sets. Note that $\varphi(0,\cdot)$ is convex and this shows that the conclusion holds when (i) is satisfied. If X is a Hilbert space and $r=2$, then $\varphi(x^*,y)=\|y\|^2-2\langle x^*,y\rangle$ is convex for any $x^*\in X$. Under condition (iii) convexity of $\varphi(x^*,\cdot)$ when x^* is almost everywhere nonnegative follows from [15, Lemma 6.3]. ∎

4.5 Let X be a reflexive Banach space and consider a function $g: X\to(-\infty,+\infty]$ which is convex, lower semicontinuous and whose domain, $\mathrm{Dom}\,(g)$, is closed. Suppose that g has a minimizer and that there exists a convex function $f: X\to(-\infty,+\infty]$ such that

(a) $\mathrm{Dom}\,(g)\subseteq\mathrm{Int}\,(\mathrm{Dom}\,(f))$;

(b) For any $\alpha \in (0, +\infty)$ and $x \in \mathrm{Dom}\,(g)$, the set

$$R_\alpha^f(x) := \{y \in \mathrm{Dom}\,(g) : D_f(x, y) \leq \alpha\}$$

is bounded;

(c) f is Fréchet differentiable and totally convex at any point $x \in \mathrm{Dom}\,(g)$.

The following procedure, which we call in what follows *the generalized proximal point algorithm,* was first studied by Censor and Zenios [21] in $\mathbb{R}^n$; its convergence in infinite dimensional Banach spaces was discussed in [14]:

$$x^{k+1} = \arg\min \left\{ g(x) + \omega_k D_f(x, x^k) \right\}, \tag{35}$$

where $x^0 \in \mathrm{Dom}\,(g)$ and $\{\omega_k\}_{k\in\mathbb{N}}$ is a bounded sequence in $(0, +\infty)$. Rockafellar's [33] proximal point algorithm is the particular version of (35) in which X is a Hilbert space and $f(x) = \frac{1}{2}\|x\|^2$. However, if X is not Hilbertian but still uniformly smooth and uniformly convex, then the generalized proximal point method even with $f(x) = \frac{1}{2}\|x\|^2$ essentially differs from Rockafellar's algorithm due to the fact that the equality $D_f(x, y) = \frac{1}{2}\|x - y\|^2$ is equivalent to the parallelogram law. It was shown in [14, Corollary 1] that, in spite of this fact, the generalized proximal point method preserves most of the well known properties of the proximal point algorithm: the sequence $\left\{x^k\right\}_{k\in\mathbb{N}}$ is well defined, bounded, has weak accumulation points, all its weak accumulation points are minimizers of g, the sequence $\left\{g(x^k)\right\}_{k\in\mathbb{N}}$ converges nonincreasingly,

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

and

$$\lim_{k\to\infty} g(x^k) = \inf g(x).$$

An easy adaptation of the proof of Lemma 5 in [14] shows that, if f is uniformly convex on bounded sets, then we also have $\lim_{k\to\infty} \left\|x^{k+1} - x^k\right\| = 0$. All these do not necessarily mean that the sequences generated by the generalized proximal point method converge (even weakly) to a minimizer of g (even if a minimizer of g is presumed to exist). Theorem 2 in [14] shows that if it is possible to choose the function f such that, in addition to the conditions (a), (b) and (c) above, will satisfy some quite restrictive requirements, then the sequences generated by the generalized proximal point method converge weakly, and sometimes even strongly, to minimizers of g. To the best of our knowledge, existence of functions like those required in [14, Theorem 2] can be ensured in Hilbert spaces and in very special non Hilbertian spaces like, for instance, the Lebesgue space ℓ^p with $p \in (1, +\infty)$. Using the properties of totally convex functions established above, we are in position to prove the next result which is a strong convergence criterion for the generalized proximal point method. It can be applied in any uniformly convex and uniformly smooth Banach space because, in such spaces, the functions $f(x) = \|x\|^r$ with $r \in (1, +\infty)$ not only satisfy the conditions (a), (b) and (c), but they also are uniformly convex and uniformly smooth (in the sense that their Gâteaux derivatives are uniformly continuous) on bounded sets (cf. [36, Corollary 4.2]). In order to state our result, recall (see [29]) that *the greatest quasi-inverse* of the function $t \to \nu_g(x^*, t)/t$ is the function $\xi_g(x^*, \cdot) : [0, +\infty) \to [0, +\infty]$ given by

$$\xi_g(x^*, t) = \sup \{s \geq 0 : \nu_g(x^*, s) \leq st\}. \tag{36}$$

Also, observe that, although reflexivity of X is not explicitly required in the next theorem, functions f which satisfy the hypothesis exist on reflexive spaces only (see Proposition 4.2 and Corollary 4.3).

Theorem. *Let $g : X \to (-\infty, +\infty]$ be a lower semicontinuous function with closed domain which is totally convex at any point $x \in \operatorname{Dom}(g)$. Suppose that the function g has a minimizer and that $f : X \to (-\infty, +\infty]$ is a lower semicontinuous convex function which satisfies the conditions (a) and (b) above and is uniformly convex and uniformly smooth on bounded subsets of $\operatorname{Dom}(g)$. Then, for any initial point $x^0 \in \operatorname{Dom}(g)$, the sequence $\{x^k\}_{k\in\mathbb{N}}$ recursively generated by the generalized proximal point algorithm (35) is well defined, contained in $\operatorname{Dom}(g)$, and has the following properties provided that the sequence $\{\omega_k\}_{k\in\mathbb{N}}$ is bounded:*

(i) $\{x^k\}_{k\in\mathbb{N}}$ converges strongly to a minimizer x^* of g and, for any nonnegative integer k, the distance of x^{k+1} to x^* can be estimated by

$$\left\|x^{k+1} - x^*\right\| \leq \xi_g\left(x^*, b\left\|f'(x^{k+1}) - f'(x^k)\right\|\right), \tag{37}$$

where $b = \sup_{k\in\mathbf{N}} \omega_k$ and $\xi_g(x^*, \cdot)$ is the greatest quasi-inverse of the function $t \to \nu_g(x^*, t)/t$;

(ii) The sequence $\{g(x^k)\}_{k\in\mathbb{N}}$ converges, $\lim_{k\to\infty} g(x^k) = g(x^*)$ and

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

(iii) If there exists a real number $\alpha > 0$ such that $\nu_g(x^*, t) \geq \alpha t$ for all $t \in (0, 1)$, then the generalized proximal point algorithms converges in finitely many steps;

(iv) If $X = \mathbb{R}^n$ and f is twice continuously differentiable on $\operatorname{Int}(\operatorname{Dom}(f))$, or if X is a Hilbert space and $f = \|\cdot\|^2$, then there exists a constant $Q > 0$ such that

$$\left\|x^{k+1} - x^*\right\| \leq \xi_g\left(x^*, Q\left\|x^{k+1} - x^k\right\|\right),$$

for all $k \in \mathbb{N}$.

Proof. Well definedness of the sequence $\{x^k\}_{k\in\mathbb{N}}$ and the fact that it is contained in $\operatorname{Dom}(g)$ result from [14, Lemma 1]. Let x^* be a minimizer of g. Then, according to [14, Lemma 2], for each nonnegative integer k, we have

$$\begin{aligned} D_f(x^*, x^k) - D_f(x^*, x^{k+1}) - D_f(x^{k+1}, x^k) &\geq \frac{1}{\omega_k}\left[g(x^{k+1}) - g(x^*)\right] \\ &\geq \frac{1}{b}\left[g(x^{k+1}) - g(x^*)\right] \geq 0. \end{aligned}$$

Taking into account (1) and (2), this implies

$$\begin{aligned}\left\langle f'(x^{k+1}) - f'(x^k), x^* - x^{k+1}\right\rangle &\geq \frac{1}{b}\left[g(x^{k+1}) - g(x^*)\right] \\ &\geq \frac{1}{b}\left[g^\circ(x^*, x^{k+1} - x^*) + \nu_g(x^*, \left\|x^{k+1} - x^*\right\|)\right].\end{aligned}$$

Since x^* is a minimizer of g, we also have $g^\circ(x^*, x^{k+1} - x^*) \geq 0$ and, thus, we obtain

$$b\left\|f'(x^{k+1}) - f'(x^k)\right\| \left\|x^* - x^{k+1}\right\| \geq \nu_g(x^*, \left\|x^{k+1} - x^*\right\|)$$

which, according to (36), implies (37). Corollary 1 in [14] guarantees that

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

and that the sequence $\left\{x^k\right\}_{k\in\mathbb{N}}$ is bounded. Hence, for the bounded set E consisting of all the terms of the sequence we have

$$\begin{aligned}0 = \lim_{k\to\infty} D_f(x^{k+1}, x^k) &\geq \limsup_{k\to\infty} \nu_f\left(E, \left\|x^{k+1} - x^k\right\|\right) \\ &\geq \liminf_{k\to\infty} \nu_f\left(E, \left\|x^{k+1} - x^k\right\|\right) \geq 0,\end{aligned}$$

that is,

$$\lim_{k\to\infty} \nu_f\left(E, \left\|x^{k+1} - x^k\right\|\right) = 0. \tag{38}$$

The function f is uniformly convex on bounded sets. Therefore, $\nu_f(E, \cdot)$ is positive and increasing on its domain and is continuous from the right at 0 (cf. [18, Lemma 2.4]). Hence, the equality (38) can not hold unless $\lim_{k\to\infty}\left\|x^{k+1} - x^k\right\| = 0$. The function f being uniformly smooth on the bounded set E, one obtains that

$$\lim_{k\to\infty} \left\|f'(x^{k+1}) - f'(x^k)\right\| = 0. \tag{39}$$

It follows from Lemma 3.3.1 in [37] that $\lim_{t\searrow 0}\xi_g(x^*, t) = 0$ because $\nu_g(x^*, t) > 0$ and $\nu_g(x^*, ct) \geq c\nu_g(x^*, t)$ when $t > 0$ and $c \geq 1$. Consequently, using (37) and (39), we deduce that

$$\lim_{k\to\infty} \left\|x^{k+1} - x^*\right\| = 0,$$

i.e., the sequence $\left\{x^k\right\}_{k\in\mathbb{N}}$ converges strongly to x^*. This completes the proof of (i). The proof of (ii) follows directly from [14, Corollary 1]. For proving (iii) note that, in view of (36), if $\nu_g(x^*, t) \geq \alpha t$ for all $t \in (0, 1)$, then $\xi_g(x^*, t) = 0$ for all $t \in [0, \alpha) \cap (0, 1)$. Since the sequence $\left\{x^k\right\}_{k\in\mathbb{N}}$ converges strongly to x^*, there exists a $k_0 \in \mathbb{N}$ such that $\left\|x^k - x^*\right\| < \alpha$ for all $k \geq k_0$ and, thus, $x^k = x^*$ for all $k \geq k_0$ because of (37). Combining (37) with [7, Proposition 2.10] one obtains (iv). ■

4.6 In connection with Theorem 4.5, it is worthwhile to mention that the issue of strong convergence of the sequences generated by the proximal point method was already raised by Rockafellar in [33] and has been treated by several authors (see, for instance, [9]). Among the results in this direction, a well known example due to Güler [25] seems to disagree with Theorem 4.5. This example shows that, even if X is a Hilbert space and $f(x) = \|x\|^2$, it may happen that the generalized proximal point algorithm for optimization (35) produces sequences $\{x^k\}_{k\in\mathbb{N}}$ which converge weakly without converging strongly. Precisely, in Güler's example $X = \ell^2$, $f(x) = \|x\|_2^2$ and the function to minimize $g : X \to (-\infty, +\infty]$ is given by

$$g(x) = \begin{cases} \sum_{k=1}^{\infty} \alpha_k \left[\arctan\left(\frac{x_k}{x_{k+1}}\right)\right]^{\beta_k} \sqrt{x_k^2 + x_{k+1}^2} & \text{if } x \in \ell_+^2, \\ \\ +\infty & \text{otherwise,} \end{cases} \tag{40}$$

with the conventions that $\arctan(t/0) = \pi/2$ for all $t \geq 0$; here, ℓ_+^2 denotes the subset of ℓ^2 consisting of sequences with nonnegative terms only, $\beta_k = \pi^2 2^{k-2}$, and the α_k's are positive real numbers recursively defined so as to ensure, e.g., that $g(x)$ is finite for all $x \in \ell_+^2$. Using previous results due to Baillon [5], Güler [25] shows that, for some $x^0 \in \ell^2$, the sequence generated by (35) with g defined in (40) is weakly but not strongly convergent, independently of how the bounded sequence of regularization coefficients $\{\omega_k\}_{k\in\mathbb{N}}$ is chosen. Clearly, the space X and the function f involved in Güler's argument satisfy the corresponding requirements of Theorem 4.5. In this example it is the function g which does not satisfy the conditions of Theorem 4.5. One should observe that the function g defined at (40) is positively homogeneous of degree 1, and henceforth linear on half lines through the origin. Thus, g is neither strictly convex, nor, "a fortiori", totally convex at all the points of its domain (as required by Theorem 4.5).

4.7 A problem more general than this of minimizing a convex function is that of finding zeroes of maximal monotone operators: Let $T : X \to \mathcal{P}(X^*)$ be a maximal monotone operator; find $\bar{x} \in X$ such that $0 \in T(\bar{x})$. The generalized proximal point algorithm for this problem recursively generates sequences $\{x^k\}_{k\in\mathbb{N}}$ in X through the iteration

$$\omega_k \left[f'(x^k) - f'(x^{k+1})\right] \in T(x^{k+1}), \tag{41}$$

where the initial point $x^0 \in \mathrm{Dom}\,(T)$ is arbitrarily chosen, the sequence of positive real numbers $\{\omega_k\}_{k\in\mathbb{N}}$ is bounded and $f : X \to (-\infty, +\infty]$ is such that

(a') $\mathrm{Dom}\,(T) \subseteq \mathrm{Int}\,(\mathrm{Dom}\,(f))$;

(b') For any $\alpha \in (0, +\infty)$ and $x \in \mathrm{Dom}\,(T)$, the set

$$R_\alpha^f(x) := \{y \in \mathrm{Dom}\,(T) : D_f(x, y) \leq \alpha\}$$

is bounded;

(c') f is uniformly convex and uniformly smooth on bounded subsets of $\mathrm{Dom}\,(T)$.

Note that, if $T = \partial g$ for some lower semicontinuous convex function $g : X \to (-\infty, +\infty]$, then (35) and (41) are equivalent, because (41) is just the necessary and sufficient condition for the minimization subproblem (35).

For X a Hilbert space provided with $f(x) = \|x\|^2$, it was proved by Rockafellar in [33] that the sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by (41) is weakly convergent to a zero of T if T has zeroes, and it is unbounded otherwise. Also in [33], it is shown that, in these circumstances, the algorithm (41) produces sequences which are necessarily strongly convergent to zeroes of T provided that the inverse operator T^{-1} is *Lipschitz continuous near* 0, that is, there exist two positive constants M and ϵ such that

$$\|x - y\| \leq M \|u - v\|$$

for all $x, y \in \operatorname{Dom}(T)$ and all $u \in T(x)$ and all $v \in T(y)$ such that $\|u\|, \|v\| \leq \epsilon$. This happens because, in such a case, the sequence $\{x^k\}_{k\in\mathbb{N}}$ is linearly convergent to its limit x^* which is the unique zero of T, and henceforth $\lim_{k\to\infty} \|x^k - x^*\| = 0$.

There exists yet another option to force strong convergence upon the sequences generated by the proximal point algorithm (41) when T is a maximal monotone operator with zeros in the particular case that X is a Hilbert space and $f(x) = \frac{1}{2}\|x\|^2$. It was proposed by Solodov and Svaiter in [34] and it consists of modifying the algorithm (41) by introducing an additional step: after finding the vector $y^k \in X$ such that

$$\omega_k \left(x^k - y^k\right) \in T(y^k),$$

one defines x^{k+1} as being the projection of y^k onto the intersection of the sets $\{x \in X : \langle x, x^0 - x^k\rangle \leq 0\}$ and $\{x \in X : \langle x, x^k - y^k\rangle \leq 0\}$. The sequence $\{x^k\}_{k\in\mathbb{N}}$ produced in this way converges strongly to a zero of T even if T^{-1} is not Lipschitz continuous near 0 as required in Rockafellar's previously quoted result (and, for that matter, even if T does not satisfy condition (i) of the proposition proved below).

The algorithm (41) in Banach spaces which are not necessarily Hilbertian was considered in [11], and an analysis for the case of inexact solutions of (41) appears in [24]. In these two references, Rockafellar's weak convergence result is extended to Banach spaces, under assumptions on f similar to (a'), (b') and (c') above.

Let us now look at the result of Theorem 4.5, and some consequences for the method applied to finding zeroes of maximal monotone operators (41). Theorem 4.5 ensures strong convergence for the optimization case when g is totally convex at any point of its domain. In view of Proposition 3.4, this assumption implies differentiability of the conjugate g^* at the points of the range of ∂g. Since $\partial g^*(x^*) = (\partial g)^{-1}(x^*)$ whenever x^* is in the range of the operator ∂g, differentiability of g^* at the points of the range is equivalent to $(\partial g)^{-1}$ being point-to-point. Thus, application of Theorem 4.5 suggests that strong convergence of the sequence generated by (41) might hold when the inverse operator T^{-1} of the maximal monotone operator T is point-to-point near zero. This is of course a weaker condition than the local Lipschitz continuity of T^{-1} required in the strong convergence criterion, due to Rockafellar and mentioned above, but it still implies uniqueness of the zero of T. We prove next a result related to this conjecture. We mention that Corollary 1.1 of [32] establishes that a maximal monotone operator which is point-to-point in a open set U is norm-to-weak continuous in U; our next proposition requires strong continuity of T^{-1}, which does not follow, as far as we know, from being point-to-point. Also, one should observe that condition (ii) below is satisfied quite often in practical situations as follows from [11, Theorem 3.2(b)].

Proposition. *Let $T : X \to \mathcal{P}(X^*)$ be a maximal monotone operator with zeroes. Suppose that $\{x^k\}_{k\in\mathbb{N}}$ is a sequence generated by the algorithm* (41) *with a bounded sequence of positive regularization coefficients $\{\omega_k\}_{k\in\mathbb{N}}$ and a function f satisfying the conditions (a'), (b') and (c'). If*

(i) *T^{-1} is point-to-point and norm-to-norm continuous in a neighborhood of $0 \in X^*$ and*

(ii) $\lim_{k\to\infty} \left\| f'(x^k) - f'(x^{k+1}) \right\| = 0$,

then $\{x^k\}_{k\in\mathbb{N}}$ is strongly convergent to $T^{-1}(0)$.

Proof. Since $\{\omega_k\}_{k\in\mathbb{N}}$ is bounded, we get

$$\lim_{k\to\infty} \omega_k[f'(x^k) - f'(x^{k+1})] = 0.$$

Since T^{-1} is point-to-point near 0, (41) implies that

$$x^{k+1} = T^{-1}\left(\omega_k[f'(x^k) - f'(x^{k+1})]\right),$$

and this and the norm-to-norm continuity of T^{-1} imply that $\{x^k\}_{k\in\mathbb{N}}$ is strongly convergent to $T^{-1}(0)$. ■

Acknowledgments

The authors are grateful to Yakov Alber, Simeon Reich and Robert T. Rockafellar for the interesting and productive mathematical discussions which led to some of the questions this work is intended to answer. Also, the authors wish to thank the referees and to Patrick L. Combettes for comments which helped improve an earlier version of this work.

D. Butnariu gratefully acknowledges the support of the Israel Science Foundation founded by the Israeli Academy for Sciences and Humanities.

References

[1] Y. Alber and A.N. Iusem, Extensions of subgradient techniques for nonsmooth optimization in Banach spaces, *Set-Valued Analysis*, **9** (2001), 315-335.

[2] Y. Alber, A.N. Iusem and M.V. Solodov, Minimization of nonsmooth convex functionals in Banach spaces, *Journal of Convex Analysis*, **4** (1997), 235-255.

[3] E. Asplund, Averaged norms, *Israel Journal of Mathematics* **5** (1967), 227-233.

[4] E. Asplund and R.T. Rockafellar, Gradients of convex functions, *Transactions of the American Mathematical Society* **139** (1969), 443-467.

[5] J.B. Baillon, Un example concernant le comportement asymptotique de la solution du problème $0 \in du/dt + \partial\varphi(u)$, *Journal of Functional Analysis* **28** (1978), 369-376.

[6] H.H. Bauschke, J.M. Borwein and P.L. Combettes, Essential smoothness, essential strict convexity and Legendre functions in Banach spaces, *Communications in Contemporary Mathematics* **3** (2001), 615-647.

[7] H.H. Bauschke and A.S. Lewis, Dykstra's algorithm with Bregman projections: A convergence proof, *Optimization* **48** (2000), 409-428.

[8] J.F. Bonnans and A. Shapiro, Perturbation Analysis of Optimization Problems, *Springer Verlag*, New York, 2000.

[9] R.E. Bruck and S. Reich, Nonexpansive projections and resolvents of acretive operators in Banach spaces, *Huston Journal of Mathematics* **3** (1977), 459-470.

[10] R. Burachik, Generalized proximal point methods for the variational inequality problem, Ph.D. Thesis, *The Institute of Pure and Applied Mathematics (IMPA)*, Rio de Janeiro, Brazil, 1995.

[11] R. Burachik and S. Scheimberg, A proximal point algorithm for the variational inequality problem, *SIAM Journal of Control and Optimization* **39** (2000), 1633-1649.

[12] D. Butnariu, Y. Censor and S. Reich, Iterative averaging of entropic projections for solving stochastic convex feasibility problems, *Computational Optimization and Applications*, **8** (1997), 21-39.

[13] D. Butnariu and A.N. Iusem, Totally Convex Functions for Fixed Point Computation and Infinite Dimensional Optimization, *Kluwer Academic Publishers*, 2000.

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

[15] D. Butnariu, A.N. Iusem and R. Burachik, Iterative methods for solving stochastic convex feasibility problems and applications, *Computational Optimization and Applications* **15** (2000), 269-307.

[16] D. Butnariu, A.N. Iusem and E. Resmerita, Total convexity for the powers of the norm in uniformly convex Banach spaces, *Journal of Convex Analysis* **7** (2000), 319-334.

[17] D. Butnariu and I. Markowicz, An evaluation of the moduli of total convexity for the powers of the norms in locally uniformly convex Banach spaces, *Technical Report*, Department of Mathematics, University of Haifa, March 1997.

[18] D. Butnariu and E. Resmerita, The outer Bregman projection method for stochastic feasibility problems in Banach spaces, in: D. Butnariu, Y. Censor and S. Reich (Eds.): "Inherently Parallel Algorithms in Feasibility and Optimization and their Applications", pp. 69-87, *Elsevier*, Amsterdam, 2001.

[19] D. Butnariu and E. Resmerita, Averaged subgradient methods for constrained convex optimization and Nash equilibria computation, *Optimization*, to appear.

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

[21] Y. Censor and S.A. Zenios, Proximal minimization algorithm with D-functions, *Journal of Optimization Theory and Applications* **73** (1992), 455-468.

[22] I. Ciorănescu, Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems, *Kluwer Academic Publishers*, Dordrecht, 1990.

[23] J. Diestel, Geometry of Banach Spaces - Selected Topics, *Springer Verlag*, Berlin, 1975.

[24] R. Gárciga Otero and A.N. Iusem, Inexact versions of proximal point and augmented Lagrangian algorithms in Banach spaces, *Numerical Functional Analysis and Optimization,* to appear.

[25] O. Güler, On the convergence of the proximal point algorithm for convex minimization, *SIAM Journal on Control and Optimization* **29** (1991), 403-419.

[26] A.N. Iusem and C. Isnard, On mixed Hölder-Minkowski inequalities and total convexity of certain functions in $L^p(\Omega)$, *Mathematical Inequalities and Applications* **3** (2000), 519-537.

[27] E.S. Levitin and B.T. Polyak, Convergence of minimizing sequences in conditional extremum problems, *Soviet Mathematics-Doklady* **7** (1966), 764-767.

[28] O. Nevalinna and S. Reich, Strong convergence of contraction semigroups and of iterative methods for acretive operators in Banach spaces, *Israel Journal of Mathematics*, **32** (1979), 44-58.

[29] J.-P. Penot and M. Volle, Inversion of real valued functions and applications, *Zeitschrift für Operations Research* **34** (1990), 117-141.

[30] B.T. Polyak, Existence theorems and convergence of minimizing sequences in extremum problems with restrictions, *Soviet Mathematics-Doklady* **7** (1966), 73-75.

[31] R.T. Rockafellar, Convex Analysis, *Princeton University Press*, Princeton, New Jersey, 1970.

[32] R.T. Rockafellar, Local boundedness of nonlinear monotone operators, *Michigan Mathematical Journal,* **16** (1969), 397-407.

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

[34] M.V. Solodov and B.F. Svaiter, Forcing strong convergence of proximal point iterations in a Hilbert space. *Mathematical Programming,* **87** (2000), 189-202.

[35] A.A. Vladimirov, Y.E. Nesterov and Y.N. Chekanov, On uniformly convex functionals (Russian), *Vestnik Moskovskaya Universiteta, Seria Matematika y Kibernetika* **3** (1978), 12-23.

[36] C. Zălinescu, On uniformly convex functions, *Journal of Mathematical Analysis and Applications* **95** (1983), 344-374.

[37] C. Zălinescu, Convex Analysis in General Vector Spaces, *to appear.*

[38] T. Zolezzi, On equiwellset minimum problems, *Applied Mathematics and Optimization* **4** (1978), 209-223.