REDUNDANT AXIOMS IN THE DEFINITION OF BREGMAN FUNCTIONS

DAN BUTNARIU*, CHARLES BYRNE, AND YAIR CENSOR*

ABSTRACT. The definition of a *Bregman function*, given by Censor and Lent in 1981 on the basis of Bregman's seminal 1967 paper, was subsequently used in a plethora of research works as a tool for building sequential and inherently parallel feasibility and optimization algorithms. Solodov and Svaiter have recently shown that it is not "minimal". Some of its conditions can be derived from the others. In this note we illuminate this finding from a different perspective by presenting an alternative proof of the equivalence between the original and the simplified definitions of Bregman functions in which redundant conditions are eliminated. This implicitly shows that the seemingly different notion of Bregman functions recently introduced by Butnariu and Iusem, when transported to a proper setting in $\mathbb{R}^n$, is equivalent to the original concept. The results established in this context are also used to resolve a problem in proximity function minimization encountered by Byrne and Censor.

1. Introduction

The notion of a Bregman function in the n-dimensional Euclidean space $\mathbb{R}^n$ was introduced by Censor and Lent in [7, Definition 2.1] (see also Censor and Zenios [8, Definition 2.1.1]), on the basis of Bregman's work [3], in the following way.

Definition 1.1 A function $g : \Lambda \subseteq \mathbb{R}^n \to \mathbb{R}$ is called a *Bregman function* on the nonempty, open and convex set S, whose closure $\mathrm{cl}S \subseteq \Lambda$, if it satisfies the following conditions:

(i) g is continuous on $\mathrm{cl}S$;

(ii) g is strictly convex on $\mathrm{cl}S$;

(iii) g is continuously differentiable on S;

(iv) If $x \in \mathrm{cl}S$, $y \in S$ and $\alpha \in \mathbb{R}$, then the *partial level sets*

$$L_\alpha^g(y) := \{z \in \mathrm{cl}S \mid D_g(z,y) \leq \alpha\} \tag{1.1}$$

and

$$R_\alpha^g(x) := \{z \in S \mid D_g(x,z) \leq \alpha\}, \tag{1.2}$$

are bounded, where, for any $u \in S$ and $v \in \mathrm{cl}S$,

$$D_g(v,u) := g(v) - g(u) - \langle \nabla g(u), v-u \rangle, \tag{1.3}$$

and $\langle \cdot, \cdot \rangle$ is the standard inner product in $\mathbb{R}^n$.

Date: December 19, 2000. Revised: January 22, 2002.

Key words and phrases. Bregman function, Bregman distance, Kullback-Leibler distance, modulus of total convexity, sequential consistency, totally convex function.

*The work of this author was supported by research grants 293/97 and 592/00 from the Israel Science Foundation, founded by the Israel Academy of Sciences and Humanities.

(v) If $\{y^k\}_{k\in\mathbb{N}} \subset S$ ($\mathbb{N} := \{0, 1, 2, \ldots\}$) is a convergent sequence with $\lim_{k\to\infty} y^k = y$, then

$$\lim_{k\to\infty} D_g(y, y^k) = 0; \tag{1.4}$$

(vi) If $\{y^k\}_{k\in\mathbb{N}} \subset S$ is a convergent sequence with $\lim_{k\to\infty} y^k = y$, if the sequence $\{x^k\}_{k\in\mathbb{N}} \subset \mathrm{cl}S$ is bounded and if $\lim_{k\to\infty} D_g(x^k, y^k) = 0$, then $\lim_{k\to\infty} x^k = y$.

Our aim in this note is to show that a Bregman function in $\mathbb{R}^n$ can be equivalently defined as follows.

Definition 1.2 A function $g : \Lambda \subseteq \mathbb{R}^n \to \mathbb{R}$ is called a *Bregman function* on the nonempty, open and convex set S, where $\mathrm{cl}S \subseteq \Lambda$, if it satisfies the following conditions:

(i) g is continuous on $\mathrm{cl}S$;

(ii) g is strictly convex on $\mathrm{cl}S$;

(iii) g is differentiable on S;

(iv) If $x \in \mathrm{cl}S$ and $\alpha > 0$, then the partial level sets $R^g_\alpha(x)$ are bounded;

(v) If $\{x^k\}_{k\in\mathbb{N}} \subset S$ is a convergent sequence with the limit $x^* \in \mathrm{bd}S := (\mathrm{cl}S) \setminus S$, then the following limit exists and we have

$$\lim_{k\to\infty} \left\langle \nabla g(x^k), x^* - x^k \right\rangle = 0. \tag{1.5}$$

Comparing Definitions 1.1 and 1.2 we see that conditions (i) and (ii) are identical. Moreover, if condition (ii) holds then condition (iii) in Definition 1.1 is equivalent to condition (iii) in Definition 1.2, because if the convex function g is differentiable throughout the open set S then it is actually continuously differentiable on S by Rockafellar's Corollary 25.5.1 in [10]. This observation is due to Bauschke and Borwein [1, Remarks 4.2]. Conditions (v) in Definitions 1.1 and 1.2 are also equivalent, when conditions (i)–(iii) hold, because then D_g is continuous on $S \times S$. It has been repeatedly observed that the requirement that the level sets $L^g_\alpha(y)$, defined in (1.1), be bounded, for all $y \in S$, is redundant in Definition 1.1, see Kiwiel [9] and Bauschke and Borwein [1, Remarks 4.2]. We also reach the same conclusion from the more general result of Theorem 3.2, showing that the the functions g, which satisfy conditions *(ii)–(iii)* of Definition 1.2, have the property that $\cup_{y\in E} L^g_\alpha(y)$ is bounded whenever $E \subseteq S$ is nonempty and bounded.

We show below that condition (vi) of Definition 1.1 is also redundant. To this end we use a recently developed tool, the *modulus of total convexity* of a convex function, developed in [4]. That condition (vi) of Definition 1.1 can be derived from the other conditions in the definition has been recently proven, in a different way, by Solodov and Svaiter in [11, Theorem 2.4]. Their result shows that if g satisfies conditions (i)–(iii) then g is convergence consistent, and, therefore, satisfies condition (vi). As a matter of fact, as can be deduced from our Lemma 2.5(i), under these circumstances, g is convergence consistent if and only if it satisfies condition (vi).

These results affect the whole existing (and future) literature on Bregman functions, distances and projections in an obvious way. Definition 1.2 may also help identify additional useful members of the family of Bregman functions. In Section 2 we present various preliminary results that are used as tools, in Section 3, to prove, in a different way and by using different notions, the main result on the equivalence of the new and old definitions for Bregman functions. In Section 4 we describe a case in which the new definition helps to resolve a specific difficulty that was encountered in the course of a recent work on proximity function minimization.

The question (for which we thank an anonymous referee) whether condition (v) of Definition 1.1 or, equivalently, condition (v) of Definition 1.2, can also be derived from the other conditions remains open. It can be shown that if g is differentiable and convex on an open set Λ such that $\mathrm{cl}S \subseteq \Lambda$, then condition (v) of Definition 1.1 is satisfied (apply Lagrange's mean-value theorem combined with the uniform convexity of ∇g on compact subsets of Λ). However, strict convexity and differentiabilty of g on an open set which contains $\mathrm{cl}S$ is a stronger requirement than conditions (i)–(iii) and some useful Bregman functions like, for example, the negative entropy function, do not enjoy it.

2. Sequential consistency

2.1 Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a proper convex function whose (necessarily convex) domain is

$$\mathrm{dom} f := \{x \in \mathbb{R}^n \mid f(x) < +\infty\}. \tag{2.1}$$

With this function we associate the *generalized Bregman distances* $D_f^\flat, D_f^\#$: $\mathrm{dom} f \times \mathrm{dom} f \to [0, +\infty]$ given by

$$D_f^\flat(y, x) := f(y) - f(x) + f^\circ(x, x - y), \tag{2.2}$$

and

$$D_f^\#(y, x) := f(y) - f(x) - f^\circ(x, y - x), \tag{2.3}$$

respectively, where, for any $x \in \mathrm{dom} f$ and for each $d \in \mathbb{R}^n$,

$$f^\circ(x, d) := \lim_{t \searrow 0} \frac{f(x + td) - f(x)}{t}, \tag{2.4}$$

is the *one-sided directional derivative* of f at x in the direction d. Recall (see, for instance, [5, Proposition 1.1.2]) that the limit in (2.4) exists and is finite whenever there exists a real number $t_0 := t_0(x, d) > 0$ such that $x + td \in \mathrm{dom} f$, for all $t \in (0, t_0]$ (and, in particular, when $x \in \mathrm{int}(\mathrm{dom} f)$, the interior of $\mathrm{dom} f$). Otherwise, we necessarily have that $f^\circ(x, d) = +\infty$. This ensures that $D_f^\#(y, x)$ is finite on $\mathrm{dom} f \times \mathrm{int}(\mathrm{dom}) f$. If $x, y \in \mathrm{dom} f$, then the convexity of f ensures that

$$f(y) - f(x) \geq f^\circ(x, y - x) \geq -f^\circ(x, x - y), \tag{2.5}$$

showing that the functions $D_f^\flat$ and $D_f^\#$ are well–defined, that is, they are nonnegative. Obviously,

$$D_f^\flat(y, x) \geq D_f^\#(y, x) \tag{2.6}$$

and $D_f^\flat(x, x) = D_f^\#(x, x) = 0$. If $y \neq x$, then we still may have that $D_f^\flat(y, x) = 0$ (e.g., when f is a linear functional). However, if f is strictly convex, then $D_f^\#(y, x) > 0$ whenever $y \neq x$ (cf. [5, Proposition 1.1.4]).

The generalized Bregman distances, given above, introduced and first studied by Kiwiel in [9], are natural extensions of the original concept of Bregman distance defined in [7] because the original notion is the restriction of $D_f^\flat(y, x)$ and of $D_f^\#(y, x)$ to the set of pairs $(y, x) \in \mathrm{dom} f \times \mathrm{int}(\mathrm{dom} f)$ in which x is a differentiability point of f.

2.2 Next we establish several facts which will be used in the sequel. Recall (see [5, Section 1.2]) that the *modulus of total convexity* of the function f at the point $x \in \mathrm{dom} f$ is the function $\nu_f(x, \cdot) : [0, +\infty) \to [0, +\infty]$ defined by

$$\nu_f(x,t) := \inf\left\{ D_f^{\#}(y,x) \mid y \in \mathrm{dom} f,\ \|y-x\| = t \right\}, \tag{2.7}$$

where $\|\cdot\|$ is the Euclidean norm. It follows from (2.6) that

$$\nu_f(x, \|y-x\|) \leq D_f^{\flat}(y,x), \tag{2.8}$$

for any $x, y \in \mathrm{dom} f$. Observe that $\nu_f(x,t) = +\infty$ when there is no point $y \in \mathrm{dom} f$ such that $\|y - x\| = t$. For any nonempty set $E \subseteq \mathrm{dom} f$ we define the function $\nu_f(E, \cdot) : [0, +\infty) \to [0, +\infty]$ by

$$\nu_f(E,t) := \inf\left\{ \nu_f(x,t) \mid x \in E \right\}. \tag{2.9}$$

Several properties of this function, which are used below, are summarized in the following lemma.

Lemma 2.2 *If E is a nonempty subset of* $\mathrm{int}(\mathrm{dom} f)$ *then the following hold:*
(i) If $c \in [1, +\infty)$, then $\nu_f(E, ct) \geq c\nu_f(E,t)$, for any $t \in [0, +\infty)$;
(ii) The function $\nu_f(E, \cdot)$ is nondecreasing on $[0, +\infty)$;
(iii) The domain of $\nu_f(E, \cdot)$ is an interval containing 0.

Proof. (i) This follows from [5, Proposition 1.2.2] which implies that, whenever $x \in E$ and $c \geq 1$, we have $\nu_f(x, ct) \geq c\nu_f(x,t)$.

(ii) If $0 < t_1 < t_2 < +\infty$, then

$$\nu_f(E, t_2) = \nu_f(E, \frac{t_2}{t_1} t_1) \geq \frac{t_2}{t_1} \nu_f(E, t_1), \tag{2.10}$$

because of (i).

(iii) Clearly, $\nu_f(E, 0) = 0$. According to (ii), if $t \in \mathrm{dom}(\nu_f(E, \cdot))$, then $[0, t) \subseteq \mathrm{dom}(\nu_f(E, \cdot))$.

2.3 A feature of Bregman functions, as defined in Definition 1.1, essential in their applications, is the following property which was termed *sequential consistency* in [5]: If the sequence $\{x^k\}_{k\in\mathbb{N}} \subset \mathrm{int}(\mathrm{dom} f)$ is bounded and the sequence $\{y^k\}_{k\in\mathbb{N}}$ is contained in $\mathrm{dom} f$, then

$$\lim_{k\to\infty} D_f^{\flat}(y^k, x^k) = 0 \quad \text{implies} \quad \lim_{k\to\infty} \left\|y^k - x^k\right\| = 0. \tag{2.11}$$

We show now that sequential consistency is a common feature of a large class of convex functions in $\mathbb{R}^n$. Note that the converse implication in (2.11) may not hold even if the function f does satisfy all the requirements of the Theorem 2.3 below. Take $n = 1$ and

$$f(x) = \begin{cases} x \ln x, & \text{if } x > 0, \\ 0, & \text{if } x = 0, \\ +\infty, & \text{if } x < 0. \end{cases} \tag{2.12}$$

Then, for any sequence $\{x^k\}_{k\in\mathbb{N}} \subset (0, e^{-1}]$ which converges to zero the sequence $\{y^k\}_{k\in\mathbb{N}}$ defined by

$$y^k = -\frac{1}{1 + \ln x^k} \tag{2.13}$$

has the property that $\lim_{k\to\infty} \left\|y^k - x^k\right\| = 0$, but $\lim_{k\to\infty} D_f^{\flat}(y^k, x^k) = 1$.

Theorem 2.3 *If the convex function* $f : \mathbb{R}^n \to (-\infty, +\infty]$ *has the following properties:*

(i) *The set* $\operatorname{dom} f$ *is closed and* $\operatorname{int}(\operatorname{dom} f)$ *is nonempty;*

(ii) *The function* f *is continuous and strictly convex on* $\operatorname{dom} f$;

(iii) *The function* f *is differentiable on* $\operatorname{int}(\operatorname{dom} f)$;

then, the function f *is sequentially consistent.*

The proof of this result is based on a pair of lemmas of intrinsic interest. To present them we will assume, for the reminder of this section, that f is a function satisfying conditions (i)–(iii) of the theorem.

2.4 Observe that, if $x \in \operatorname{int}(\operatorname{dom} f)$, then the function $D_f^\flat(\cdot, x) = D_f^\#(\cdot, x)$ is convex and continuous on $\operatorname{dom} f$ as follows from [5, Corollary 1.1.6]. The following result shows that, in our circumstances, the function $D_f^\flat$ has a more general continuity property.

Lemma 2.4 *Under the assumptions of Theorem 2.3 the function* $(x, y) \to f^\circ(x, y - x)$ *is upper semicontinuous on* $\operatorname{dom} f \times \operatorname{dom} f$ *and the function* $(y, x) \to D_f^\#(y, x)$ *is lower semicontinuous on* $\operatorname{dom} f \times \operatorname{dom} f$.

Proof. Taking into account (2.3), it is sufficient to show that the function $(x, y) \to f^\circ(x, y - x)$ is upper semicontinuous on $\operatorname{dom} f \times \operatorname{dom} f$. To this end, for any real number $t > 0$, define the function $\psi_t : \operatorname{dom} f \times \operatorname{dom} f \to [0, +\infty]$ by

$$\psi_t(x, y) := \frac{f(x + t(y - x)) - f(x)}{t}. \tag{2.14}$$

It follows from [5, Proposition 1.1.2(i)] that the family of functions $\{\psi_t \,|\, t > 0\}$, is point-wise nondecreasing, that is,

$$0 < s < t < +\infty \quad \text{implies} \quad \psi_s(x, y) \le \psi_t(x, y), \tag{2.15}$$

for all pairs $(x, y) \in \operatorname{dom} f \times \operatorname{dom} f$. For any fixed real number $t \in (0, 1]$, the function ψ_t is continuous on $\operatorname{dom} f \times \operatorname{dom} f$. Indeed, observe that, for $x, y \in \operatorname{dom} f$, we have $x + t(y - x) = ty + (1 - t)x \in \operatorname{dom} f$ and that the function f is continuous on $\operatorname{dom} f$. According to (2.14) and (2.15), we have that

$$f^\circ(x, y - x) = \lim_{t \searrow 0} \psi_t(x, y), \tag{2.16}$$

whenever $(x, y) \in \operatorname{dom} f$. Thus, $(x, y) \to f^\circ(x, y - x)$ is upper semicontinuous on $\operatorname{dom} f$ because it is the point-wise limit of a nonincreasing family of continuous functions.

2.5 Under the assumptions of Theorem 2.3 the function $\nu_f(E, \cdot)$ has some special properties shown in the following lemma.

Lemma 2.5 *Suppose that* E *is a nonempty bounded subset of* $\operatorname{int}(\operatorname{dom} f)$. *Then, the following statements hold:*

(i) *For any* $t > 0$, *we have* $\nu_f(E, t) > 0$;

(ii) *If the right end-point of the domain of* $\nu_f(E, \cdot)$,

$$\tau_f(E) := \sup \{t \,|\, \nu_f(E, t) < +\infty\}, \tag{2.17}$$

is positive, then the function $\nu_f(E, \cdot)$ *is strictly increasing on the interval* $[0, \tau_f(E))$.

Proof. (i) Let t be a positive real number and suppose, by way of negation, that $\nu_f(E, t) = 0$. Then, according to (2.9), there exists a sequence $\{x^k\}_{k \in \mathbb{N}} \subset E$

such that

$$\lim_{k\to\infty} \nu_f(x^k, t) = 0. \tag{2.18}$$

For each $k \in \mathbb{N}$, the set $\{y \in \mathrm{dom}f \mid \|y - x^k\| = t\}$ is compact because $\mathrm{dom}f$ is closed. Therefore, for each $k \in \mathbb{N}$, there exists a vector $y^k \in \mathrm{dom}f$ at which the lower semicontinuous function $D_f^{\#}(\cdot, x^k)$ (see Lemma 2.4) attains its minimum on this set, that is, such that $\|y^k - x^k\| = t$ and $\nu_f(x^k, t) = D_f^{\#}(y^k, x^k)$. The sequence $\{y^k\}_{k\in\mathbb{N}}$ is bounded because

$$\|y^k\| \leq \|y^k - x^k\| + \|x^k\| = t + \|x^k\|, \tag{2.19}$$

and $\{x^k\}_{k\in\mathbb{N}}$ is bounded (because it is contained in E). Let $\{x^{i_k}\}_{k\in\mathbb{N}}$ and $\{y^{i_k}\}_{k\in\mathbb{N}}$ be convergent subsequences of $\{x^k\}_{k\in\mathbb{N}}$ and $\{y^k\}_{k\in\mathbb{N}}$, respectively, and denote by x^* and y^* their respective limits. Clearly, x^* and y^* are contained in the closed set $\mathrm{dom}f$. According to Lemma 2.4, the function $D_f^{\#}$ is lower semicontinuous on $\mathrm{dom}f \times \mathrm{dom}f$ and, applying (2.18), we have

$$0 \leq D_f^{\#}(y^*, x^*) \leq \liminf_{k\to\infty} D_f^{\#}(y^{i_k}, x^{i_k}) = 0, \tag{2.20}$$

showing that $D_f^{\#}(y^*, x^*) = 0$. This can not hold unless $x^* = y^*$ because f is strictly convex. However, we also have

$$\|x^* - y^*\| = \lim_{k\to\infty} \|x^{i_k} - y^{i_k}\| = t > 0, \tag{2.21}$$

and this is a contradiction.

(ii) Suppose that $0 < t_1 < t_2 < \tau_f(E)$. Then the inequalities in (2.10) still hold and the last of these inequalities is strict because, according to (i), $\nu_f(E, t_1) > 0$.

2.6 The next result completes the proof of Theorem 2.3.

Lemma 2.6 *Under the assumptions of Theorem 2.3 the function f is sequentially consistent.*

Proof. Let $\{x^k\}_{k\in\mathbb{N}}$ be a bounded sequence in $\mathrm{int}(\mathrm{dom}f)$ and suppose that $\{y^k\}_{k\in\mathbb{N}} \subset \mathrm{dom}f$ is a sequence such that $\lim_{k\to\infty} D_f^\flat(y^k, x^k) = 0$. Denoting $E := \{x^k \mid k \in \mathbb{N}\}$, defining $t_k := \|x^k - y^k\|$ and using (2.8), we get that the following limit exists and we have

$$0 \leq \nu_f(E, t_k) \leq \nu_f(x^k, t_k) \leq D_f^\flat(y^k, x^k), \tag{2.22}$$

for all $k \in \mathbb{N}$ and, thus, we deduce that

$$\lim_{k\to\infty} \nu_f(E, t_k) = 0. \tag{2.23}$$

Suppose, by way of negation, that the sequence $\{t_k\}_{k\in\mathbb{N}}$ does not converge to zero. Then, it has a subsequence $\{t_{i_k}\}_{k\in\mathbb{N}}$ such that, for some real number $\varepsilon_0 > 0$, we have $t_{i_k} \geq \varepsilon_0$, whenever $k \in \mathbb{N}$. The inequality (2.22) ensures that all t_k excepting, eventually, finitely many of them, are in the interval $[0, \tau_f(E))$. Hence, ε_0 is in this interval too. Applying Lemma 2.5(ii) we deduce that

$$\nu_f(E, t_{i_k}) \geq \nu_f(E, \varepsilon_0) > 0, \tag{2.24}$$

and this contradicts (2.23).

3. Back to Bregman functions

3.1 In this section we use the results presented above to deduce the simplification of Definition 1.1, i.e., its equivalence with Definition 1.2. To this end, throughout this section, we consider a function $g : \Lambda \subseteq \mathbb{R}^n \to \mathbb{R}$ which satisfies the requirements of Definition 1.2. With this function we associate the function $f : \mathbb{R}^n \to (-\infty, +\infty]$ defined by

$$f(x) := \begin{cases} g(x), & \text{if } x \in \text{cl}S, \\ +\infty, & \text{otherwise.} \end{cases} \tag{3.1}$$

Clearly, the function f satisfies the conditions (i)–(iii) of Theorem 2.3. Also, for any pair $(x, y) \in \text{cl}S \times S$, we have $D_f^\flat(x, y) = D_g(x, y)$. Using these facts we prove the following result.

Proposition 3.1 *If* $g : \Lambda \subseteq \mathbb{R}^n \to \mathbb{R}$ *is a function satisfying the conditions (i)–(iii) of Definition 1.2 for the nonempty, open and convex subset S of Λ, then g also satisfies over S the condition (vi) of Definition 1.1.*

Proof. Observe that the function f defined by (3.1) is, by Lemma 2.6, sequentially consistent because it satisfies the conditions of Theorem 2.3. Also, $D_g(y, x) = D_f^\flat(y, x)$, for any pair $(y, x) \in \text{cl}S \times S$. Hence, if $\{x^k\}_{k\in\mathbb{N}}$ is a bounded sequence in $\text{cl}S$ and $\{y^k\}_{k\in\mathbb{N}} \subset S$ is a sequence such that $\lim_{k\to\infty} D_g(x^k, y^k) = 0$ and $\lim_{k\to\infty} y^k = y^*$, then it follows from Theorem 2.3 that $\lim_{k\to\infty} \|y^k - x^k\| = 0$ and, thus, $\lim_{k\to\infty} x^k = \lim_{k\to\infty} y^k = y^*$. This shows that condition (vi) of Definition 1.1 holds.

3.2 Now we show that the boundedness of the sets $L_\alpha^g(y)$, with $y \in S$, defined in (1.1), involved in Definition 1.1, can be derived from conditions (i)–(iii) of Definition 1.2. In fact, we prove the next, more general, result.

Theorem 3.2 *If $E \subseteq S$ is nonempty and bounded, then the set*

$$L_\alpha^g(E) := \cup_{y\in E} L_\alpha^g(y) \tag{3.2}$$

is bounded, for any $\alpha \in [0, +\infty)$.

Proof. Suppose, by way of negation, that the set $L_\alpha^g(E)$ is unbounded. Then, there exist two sequences $\{y^k\}_{k\in\mathbb{N}} \subseteq E$ and $\{z^k\}_{k\in\mathbb{N}} \subseteq \text{cl}S$, such that $\{z^k\}_{k\in\mathbb{N}}$ is unbounded and, for any $k \in \mathbb{N}$, we have $z^k \in L_\alpha^g(y^k)$. Consequently,

$$\nu_f(E, \|z^k - y^k\|) \le D_f^\flat(z^k, y^k) \le \alpha < +\infty, \tag{3.3}$$

for all $k \in \mathbb{N}$. This shows that $\|z^k - y^k\| \in \text{dom}(\nu_f(E, \cdot))$, for all $k \in \mathbb{N}$. Since $\{z^k\}_{k\in\mathbb{N}}$ is unbounded and $\{y^k\}_{k\in\mathbb{N}}$ is bounded (because it is contained in the bounded set E), there exist two subsequences $\{y^{i_k}\}_{k\in\mathbb{N}}$ and $\{z^{i_k}\}_{k\in\mathbb{N}}$, of $\{y^k\}_{k\in\mathbb{N}}$ and $\{z^k\}_{k\in\mathbb{N}}$, respectively, such that the sequence $\{\|z^{i_k} - y^{i_k}\|\}_{k\in\mathbb{N}}$ is strictly increasing and

$$\lim_{k\to\infty} \|z^{i_k} - y^{i_k}\| = +\infty. \tag{3.4}$$

According to Lemma 2.5, the function $\nu_f(E, \cdot)$ is strictly increasing on $\text{dom}(\nu_f(E, \cdot))$. Also, according to (3.3), $\|z^{i_k} - y^{i_k}\| \in \text{dom}(\nu_f(E, \cdot))$. Hence, the following limit exists and, according to Lemmas 2.2(i) and 2.5(i), we have

$$\alpha \ge \lim_{k\to\infty} \nu_f(E, \|z^{i_k} - y^{i_k}\|) \ge \lim_{k\to\infty} \|z^{i_k} - y^{i_k}\| \, \nu_f(E, 1) = +\infty. \tag{3.5}$$

This contradicts the finiteness of α.

3.3 Proposition 3.1 and Theorem 3.2 imply that Definitions 1.1 and 1.2 are equivalent. But, is Definition 1.2 "minimal" ? That conditions (i), (ii) and (iii) of Definition 1.2 are independent from each other is obvious. We show next that condition (iv) of Definition 1.2 can not be deduced from its first three conditions. The following example shows that there are functions which satisfy conditions (i)–(iii) of Definition 1.2 without satisfying condition (iv). Take $n = 1$ and $g(x) = e^{-x}$, for all $x \in \mathbb{R}^1$. Clearly, this function satisfies (i)–(iii), but $D_g(1, z) = e^{-1} - ze^{-z}$ and this shows that $R^g_\alpha(1)$ is not bounded when $\alpha = e^{-1}$.

4. An application

Theorem 2.3 not only helps to simplify the definition of a Bregman function but also resolves a problem encountered by Byrne and Censor in [6]. That paper is concerned with the minimization of proximity functions of the form

$$F(x) := \sum_{i=1}^{I} D_{f_i}\left(P^{f_i}_{C_i}(x), x\right), \tag{4.1}$$

defined for $x \in U := \cap_{i=1}^{I} S_i \neq \emptyset$, where, for all $i = 1, 2, \ldots, I$, the f_i are Bregman functions with zones S_i, the C_i are nonempty closed convex sets in $\mathbb{R}^n$ and each $P^{f_i}_{C_i}(x) \in S_i$ is the Bregman projection, with respect to f_i, of x onto the set C_i, see, e.g., Censor and Zenios [8, Definition 2.1.2]). It was assumed in [6] that the Bregman distances D_{f_i} are jointly convex, so that the function F is a proper convex function on U and can be extended to $\mathbb{R}^n$ by letting $F(x) := +\infty$, for $x \in \mathbb{R}^n \backslash U$. The closure of F, denoted by cl$F$, is defined, for all $x \in \mathbb{R}^n$, by (see [10, p. 52]):

$$(\mathrm{cl}F)(x) = \liminf_{y \to x} F(y). \tag{4.2}$$

The question that arose in [6] was whether all points $x \in \mathbb{R}^n$ for which $(\mathrm{cl}F)(x) = 0$ are contained in $\cap_{i=1}^{I} C_i$. The affirmative answer, which we were not able to furnish in [6], follows from Theorem 2.3. Indeed, if $(\mathrm{cl}F)(x) = 0$, then there exists a convergent sequence $\left\{x^k\right\}_{k \in \mathbb{N}} \subset U$ such that $\lim_{k \to \infty} x^k = x$ and $\lim_{k \to \infty} F(x^k) = 0$. For each $i = 1, 2, \ldots, I$, we then have

$$\lim_{k \to \infty} D_{f_i}\left(P^{f_i}_{C_i}(x^k), x^k\right) = 0. \tag{4.3}$$

According to Theorem 2.3, this implies that, for each $i = 1, 2, \ldots, I$, we have

$$\lim_{k \to \infty} \left\| P^{f_i}_{C_i}(x^k) - x^k \right\| = 0, \tag{4.4}$$

that is, $\lim_{k \to \infty} P^{f_i}_{C_i}(x^k) = x$. Since the sets C_i are closed and $P^{f_i}_{C_i}(x^k) \in C_i$, for all $k \in \mathbb{N}$ and $i = 1, 2, \ldots, I$, it follows that $x \in \cap_{i=1}^{I} C_i$.

This analysis enables us to strengthen Theorem 4.1 of Byrne and Censor's [6] by removing from it the assumption called there Assumption **A3**. It should be noted that the joint convexity of D_{f_i}, required above, is a restrictive condition whose realizability was recently studied by Bauschke and Borwein [2].

References

[1] H.H. Bauschke and J.M. Borwein, Legendre functions and the method of random Bregman projections, *Journal of Convex Analysis*, **4**:27–67, 1997.

[2] H.H. Bauschke and J.M. Borwein, Joint and separate convexity of the Bregman distance, In: D. Butnariu, Y. Censor and S. Reich (Editors), *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications*, Elesevier Science, Amsterdam, The Netherlands, 2001, pp. 23–36.

[3] L.M. Bregman, The relaxation method for finding the common point of convex sets and its application to the solution of problems in convex programming, *USSR Computational Mathematics and mathematical Physics*, **7**:200–217, 1967.

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

[5] D. Butnariu and A.N. Iusem, *Totally Convex Function for Fixed Points Computation and Infinite Dimensional Optimization*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.

[6] C. Byrne and Y. Censor, Proximity function minimization using multiple Bregman projections, with applications to split feasibility and Kullback-Leibler distance minimization, *Annals of Operations Research*, **105**:77–98, 2001.

[7] Y. Censor and A. Lent, An iterative row-action method for interval convex programming, *Journal of Optimization Theory and Applications*, **34**:322–353, 1981.

[8] Y. Censor and S.A. Zenios, *Parallel Optimization: Theory, Algorithms and Applications*, Oxford University Press, New York, 1997.

[9] K. Kiwiel, Free-steering relaxation methods for problems with strictly convex costs and linear constraints, *Mathematics of Operations Research,* **22**:326–349, 1997.

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

[11] M.V. Solodov and B.F. Svaiter, An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions, *Mathematics of Operations Research,* **25**:214–230, 2000.

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

Charles Byrne, Department of Mathematical Sciences, University of Massachusetts at Lowell, Lowell, MA 01854, USA
E-mail address: `byrnec@cs.uml.edu`

Yair Censor, Department of Mathematics, University of Haifa, 31905 Haifa, Israel
E-mail address: `yair@math.haifa.ac.il`