SIAM J. OPTIM., to appear

ITERATING BREGMAN RETRACTIONS*

HEINZ H. BAUSCHKE† AND PATRICK L. COMBETTES‡

Abstract. The notion of a Bregman retraction of a closed convex set in Euclidean space is introduced. Bregman retractions include backward Bregman projections, forward Bregman projections, as well as their convex combinations, and are thus quite flexible. The main result on iterating Bregman retractions unifies several convergence results on projection methods for solving convex feasibility problems. It is also used to construct new sequential and parallel algorithms.

Key words. backward Bregman projection, Bregman distance, Bregman function, Bregman projection, Bregman retraction, convex feasibility problem, forward Bregman projection, Legendre function, paracontraction, projection algorithm

AMS subject classifications. 90C25, 49M37, 65K05

PII. XXXXX

1. Standing assumptions, problem statement, and motivation.

We assume throughout this paper that

$$X \text{ is a Euclidean space with scalar product } \langle\cdot,\cdot\rangle \text{ and induced norm } \|\cdot\| \tag{1.1}$$

and that

$$\begin{aligned} &f\colon X \to \left]-\infty,+\infty\right] \text{ is a proper closed convex Legendre function}\\ &\text{such that } \operatorname{dom} f^* \text{ is open,}\end{aligned} \tag{1.2}$$

where f^* denotes the conjugate of f. Recall that a function is Legendre if it is both essentially smooth and essentially strictly convex (see, e.g., [31] for basic facts and notions from convex analysis). In addition, we assume that

$$\begin{aligned} &(C_i)_{i\in I} \text{ are finitely many closed convex sets in } X\\ &\text{such that } (\operatorname{int}\operatorname{dom} f)\cap \textstyle\bigcap_{i\in I} C_i \neq \varnothing.\end{aligned} \tag{1.3}$$

Our aim is to study algorithms for solving the fundamental *convex feasibility problem* (see [4], [14], [17], [20], and [27] for further information and references)

$$\text{find } x \in \bigcap_{i\in I} C_i. \tag{1.4}$$

Assumption (1.2) guarantees that we capture a large class of functions (see Example 2.1 below) for which the corresponding *Bregman distance*

$$D_f\colon X\times X \to [0,+\infty] : (x,y) \mapsto \begin{cases} f(x)-f(y)-\langle x-y,\nabla f(y)\rangle, & \text{if } y\in\operatorname{int}\operatorname{dom} f;\\ +\infty, & \text{otherwise,}\end{cases} \tag{1.5}$$

*Received by the editors July 2, 2002; accepted for publication XXX, XX 2002; published electronically DATE.

http://www.siam.org/journals/siopt/x-x/XXXX.html

†Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario N1G 2W1, Canada (hbauschk@uoguelph.ca). Research supported by the Natural Sciences and Engineering Research Council of Canada.

‡Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie – Paris 6, 75005 Paris, France (plc@math.jussieu.fr).

enjoys useful properties (Proposition 2.2). This type of directed distance was first introduced by Bregman in [8]; see [17] for a historical account. Now fix a closed convex set C in X such that $C \cap \operatorname{int} \operatorname{dom} f \neq \varnothing$ and a point $y \in \operatorname{int} \operatorname{dom} f$. Then there is a unique point in $C \cap \operatorname{int} \operatorname{dom} f$, called the *backward Bregman projection* (or simply the *Bregman projection*) of y onto C and denoted by $\overleftarrow{P}_C y$, which satisfies (Fact 2.3)

$$(\forall c \in C) \quad D_f(\overleftarrow{P}_C y, y) \leq D_f(c, y). \tag{1.6}$$

Moreover, if f *allows forward Bregman projections* (Definition 2.4), then there is analogously a unique point in $C \cap \operatorname{int} \operatorname{dom} f$, called the *forward Bregman projection* of y onto C and denoted by $\overrightarrow{P}_C y$, which satisfies (Fact 2.6)

$$(\forall c \in C) \quad D_f(y, \overrightarrow{P}_C y) \leq D_f(y, c). \tag{1.7}$$

If $f = \frac{1}{2}\|\cdot\|^2$, then both $\overleftarrow{P}_C y$ and $\overrightarrow{P}_C y$ coincide with the *orthogonal projection* of y onto C; however, the backward and forward Bregman projections differ generally, due to the asymmetry of D_f.

With backward and forward Bregman projections in place, we now describe three projection methods for solving (1.4). To this end, fix an index selector map $\mathrm{i}\colon \mathbb{N} = \{0, 1, 2, \dots\} \to I$ that takes on each value in I infinitely often, and a starting point $y_0 \in \operatorname{int} \operatorname{dom} f$. The *method of backward Bregman projections* generates a sequence $(y_n)_{n\in\mathbb{N}}$ by

$$(\forall n \in \mathbb{N}) \quad y_{n+1} = \overleftarrow{P}_{C_{\mathrm{i}(n+1)}} y_n. \tag{1.8}$$

Analogously, if f allows forward Bregman projections, then the update rule for the *method of forward Bregman projections* is

$$(\forall n \in \mathbb{N}) \quad y_{n+1} = \overrightarrow{P}_{C_{\mathrm{i}(n+1)}} y_n. \tag{1.9}$$

Well-known *cyclic* versions arise if $I = \{1, \dots, N\}$ and $\mathrm{i}(n) = n \bmod N$, where the range of the mod function is assumed to be $\{1, \dots, N\}$. The sequence $(y_n)_{n\in\mathbb{N}}$ generated by (1.8) (or by (1.9), if f allows forward Bregman projections) is known to solve (1.4) asymptotically: indeed, $(y_n)_{n\in\mathbb{N}}$ converges to some point in $\bigcap_{i\in I} C_i$, see [5] and [16] (or [7], respectively).

The third algorithm is due to Byrne and Censor [12], who adapted Csiszár and Tusnády's classical alternating minimization procedure [22] to a product space setting (see also Section 5). Their algorithm assumes two constraints, $I = \{1, 2\}$, and a sequence $(y_n)_{n\in\mathbb{N}}$ is generated using *alternating backward-forward Bregman projections*:

$$(\forall n \in \mathbb{N}) \quad y_{n+1} = \big(\overleftarrow{P}_{C_2} \circ \overrightarrow{P}_{C_1}\big) y_n. \tag{1.10}$$

They show that, under appropriate conditions, $(y_n)_{n\in\mathbb{N}}$ converges to some point in $C_1 \cap C_2$, see [12, Theorem 1].

The striking resemblance in the update rules of the three preceding algorithms motivates this paper. Our objective is to provide a unified convergence analysis of these algorithms using the notion of a *Bregman retraction*, which encompasses both backward and forward Bregman projections. The main theorem not only recovers known convergence results but also provides a theoretical basis for the application of new sequential and parallel methods.

It is instructive to contrast our Bregman retraction-based framework with Censor and Reich's [16] framework, which is built on *paracontractions* (Definition 3.11). While backward Bregman projections are both Bregman retractions and paracontractions, the two notions differ in general; actually, Examples 3.12 and 3.13 show that neither framework contains the other.

The key advantage of the Bregman retraction-based framework presented here is its applicability: the conditions on f are mild and easy to check. Moreover, simple constraint qualifications guarantee that Bregman retractions — in the form of backward Bregman projections (and forward Bregman projections, if f allows them) — always exist.

The paper is organized as follows. Background material on Bregman distances and associated projections is included in Section 2. In Section 3, Bregman retractions are introduced, analyzed, and illustrated by examples. The main result is proved in Section 4 and applications are presented in Section 5.

2. Preliminary results. Below is a selection of functions satisfying our assumptions (see [5] for additional examples).

EXAMPLE 2.1. [5] Suppose $X = \mathbb{R}^J$ and, for every $x \in X$, write $x = (\xi_j)_{j=1}^J$. Then the following functions satisfy (1.2) (here and elsewhere, we use the convention $0 \cdot \ln(0) = 0$):

(i) $f : x \mapsto \frac{1}{2}\|x\|^2 = \frac{1}{2}\sum_{j=1}^J |\xi_j|^2$, with $\operatorname{dom} f = \mathbb{R}^J$ (*energy*);

(ii) $f : x \mapsto \sum_{j=1}^J \xi_j \ln(\xi_j) - \xi_j$, with $\operatorname{dom} f = [0, +\infty[^J$ (*negative entropy*);

(iii) $f : x \mapsto \sum_{j=1}^J \xi_j \ln(\xi_j) + (1-\xi_j)\ln(1-\xi_j)$, with $\operatorname{dom} f = [0,1]^J$ (*Fermi-Dirac entropy*);

(iv) $f : x \mapsto -\sum_{j=1}^J \ln(\xi_j)$, with $\operatorname{dom} f = \,]0, +\infty[^J$ (*Burg entropy*);

(v) $f : x \mapsto -\sum_{j=1}^J \sqrt{\xi_j}$, with $\operatorname{dom} f = [0, +\infty[^J$.

The assumptions imposed on f in (1.2) guarantee the following very useful properties of D_f.

PROPOSITION 2.2. *Let D_f be defined as in* (1.5). *Then*

(i) *D_f is continuous on* $(\operatorname{int}\operatorname{dom} f)^2$.

(ii) *If $x \in \operatorname{dom} f$ and $y \in \operatorname{int}\operatorname{dom} f$, then $D_f(x,y) \geq 0$, and $D_f(x,y) = 0 \Leftrightarrow x = y$.*

(iii) *If $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are two sequences in $\operatorname{int}\operatorname{dom} f$ converging to $x \in \operatorname{int}\operatorname{dom} f$ and $y \in \operatorname{int}\operatorname{dom} f$, respectively, then $D_f(x_n, y_n) \to 0 \Leftrightarrow x = y$.*

(iv) *If $x \in \operatorname{int}\operatorname{dom} f$ and $(y_n)_{n\in\mathbb{N}}$ is a sequence in $\operatorname{int}\operatorname{dom} f$ such that the sequence $\big(D_f(x, y_n)\big)_{n\in\mathbb{N}}$ is bounded, then $(y_n)_{n\in\mathbb{N}}$ is bounded and all its cluster points belong to $\operatorname{int}\operatorname{dom} f$.*

(v) *If $x \in \operatorname{int}\operatorname{dom} f$ and $(y_n)_{n\in\mathbb{N}}$ is a sequence in $\operatorname{int}\operatorname{dom} f$ such that $D_f(x, y_n) \to 0$, then $y_n \to x$.*

Proof. (i): This follows from the definition of D_f and the continuity of f (respectively ∇f) on $\operatorname{int}\operatorname{dom} f$; see [31, Theorem 10.1] (respectively [31, Theorem 25.5]). (ii): [5, Theorem 3.7.(iv)]. (iii): This is a consequence of (i) and (ii). (iv): [5, Theorem 3.7.(vi) and Theorem 3.8.(ii)]. (v): (See also [7, Fact 2.18].) By (iv), $(y_n)_{n\in\mathbb{N}}$ is bounded and has all its cluster points in $\operatorname{int}\operatorname{dom} f$. Pick an arbitrary cluster point of $(y_n)_{n\in\mathbb{N}}$, say $y_{k_n} \to y \in \operatorname{int}\operatorname{dom} f$. Then $D_f(x, y_{k_n}) \to 0$ and thus $x = y$ by (iii). □

We now turn to backward and forward Bregman projections.

FACT 2.3 (backward Bregman projection). *Suppose C is a closed convex set in X such that $C \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$. Then, for every $y \in \operatorname{int}\operatorname{dom} f$, there exists a unique point $\overleftarrow{P}_C y \in C \cap \operatorname{dom} f$ such that $D_f(\overleftarrow{P}_C y, y) \leq D_f(c, y)$, for all $c \in C$. The point*

$\overleftarrow{P}_C y$ *is called the* backward Bregman projection *(or simply the* Bregman projection*) of* y *onto* C*, and it is characterized by*

$$\overleftarrow{P}_C \in C \cap \operatorname{int}\operatorname{dom} f \quad \textit{and} \quad (\forall c \in C)\ \big\langle c - \overleftarrow{P}_C y, \nabla f(y) - \nabla f(\overleftarrow{P}_C y)\big\rangle \leq 0; \tag{2.1}$$

equivalently, by

$$\overleftarrow{P}_C \in C \cap \operatorname{int}\operatorname{dom} f \quad \textit{and} \quad (\forall c \in C)\ D_f(c, y) \geq D_f(c, \overleftarrow{P}_C y) + D_f(\overleftarrow{P}_C y, y). \tag{2.2}$$

Finally, the operator $\overleftarrow{P}_C$ *is continuous on* $\operatorname{int}\operatorname{dom} f$*.*

Proof. Under the present assumptions on f, the claims follow from [5, Theorem 3.14 and Proposition 3.16], except for the continuity of $\overleftarrow{P}_C$, which we derive now. Suppose $(x_n)_{n\in\mathbb{N}}$ is a sequence in $\operatorname{int}\operatorname{dom} f$ converging to $\bar{x} \in \operatorname{int}\operatorname{dom} f$. Set $(c_n)_{n\in\mathbb{N}} = (\overleftarrow{P}_C x_n)_{n\in\mathbb{N}}$ and $\bar{c} = \overleftarrow{P}_C \bar{x}$. We must show that $(c_n)_{n\in\mathbb{N}}$ converges to $\bar{c}$. Using Proposition 2.2.(i) and (2.2), we have

$$D_f(\bar{c}, \bar{x}) \leftarrow D_f(\bar{c}, x_n) \geq D_f(\bar{c}, c_n) + D_f(c_n, x_n) \geq D_f(\bar{c}, c_n). \tag{2.3}$$

Hence $\big(D_f(\bar{c}, c_n)\big)_{n\in\mathbb{N}}$ is bounded. By Proposition 2.2.(iv), $(c_n)_{n\in\mathbb{N}}$ is bounded and all its cluster points belong to $C \cap \operatorname{int}\operatorname{dom} f$. Let $\hat{c}$ be such a cluster point, say $c_{k_n} \to \hat{c} \in \operatorname{int}\operatorname{dom} f$. Using the definition of $\bar{c}$, Proposition 2.2.(i), and (2.2), we deduce $D_f(\hat{c}, \bar{x}) \geq D_f(\bar{c}, \bar{x}) \leftarrow D_f(\bar{c}, x_{k_n}) \geq D_f(\bar{c}, c_{k_n}) + D_f(c_{k_n}, x_{k_n}) \to D_f(\bar{c}, \hat{c}) + D_f(\hat{c}, \bar{x}) \geq D_f(\hat{c}, \bar{x})$; thus $D_f(\bar{c}, \hat{c}) = 0$ and hence, by Proposition 2.2.(ii), $\bar{c} = \hat{c}$. ☐

Definition 2.4. *The function* f allows forward Bregman projections *if it satisfies the following additional properties:*

(i) $\nabla^2 f$ *exists and is continuous on* $\operatorname{int}\operatorname{dom} f$*;*
(ii) D_f *is convex on* $(\operatorname{int}\operatorname{dom} f)^2$*;*
(iii) *For every* $x \in \operatorname{int}\operatorname{dom} f$*,* $D_f(x, \cdot)$ *is strictly convex on* $\operatorname{int}\operatorname{dom} f$*.*

Remark 2.5. The function f allows forward Bregman projections if and only if it satisfies the standing assumptions of [7], which allows us to apply the results of [7]. This equivalence follows from [7, Remark 2.1] and

$$D_f \text{ is convex on } (\operatorname{int}\operatorname{dom} f)^2 \quad \Leftrightarrow \quad D_f \text{ is convex on } X^2. \tag{2.4}$$

We now verify (2.4). The implication "$\Leftarrow$" is clear. To establish "$\Rightarrow$", let us fix $(y_1, y_2) \in (\operatorname{int}\operatorname{dom} f)^2$, $(x_1, x_2) \in (\operatorname{dom} f)^2$, and $(\lambda_1, \lambda_2) \in \,]0,1[^2$ such that $\lambda_1 + \lambda_2 = 1$. For $\varepsilon \in \,]0,1[$ and $i \in \{1,2\}$, set $x_{i,\varepsilon} = (1-\varepsilon)x_i + \varepsilon y_i \in \operatorname{int}\operatorname{dom} f$. Then $D_f(\lambda_1 x_{1,\varepsilon} + \lambda_2 x_{2,\varepsilon}, \lambda_1 y_1 + \lambda_2 y_2) \leq \lambda_1 D_f(x_{1,\varepsilon}, y_1) + \lambda_2 D_f(x_{2,\varepsilon}, y_2)$. Now take $y \in \operatorname{int}\operatorname{dom} f$. Since f is closed and convex, so is $D_f(\cdot, y)$. Hence, as $\varepsilon \downarrow 0^+$, the line segment continuity property of $D_f(\cdot, y)$ [31, Corollary 7.5.1] results in $D_f(\lambda_1 x_1 + \lambda_2 x_2, \lambda_1 y_1 + \lambda_2 y_2) \leq \lambda_1 D_f(x_1, y_1) + \lambda_2 D_f(x_2, y_2)$. Thus D_f is convex on $\operatorname{dom} f \times \operatorname{int}\operatorname{dom} f = \operatorname{dom} D_f$ and, thereby, on X^2.

Fact 2.6 (forward Bregman projection). *Suppose* f *allows forward Bregman projections and* C *is a closed convex set in* X *such that* $C \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$*. Then, for every* $y \in \operatorname{int}\operatorname{dom} f$*, there exists a unique point* $\overrightarrow{P}_C y \in C \cap \operatorname{dom} f$ *such that* $D_f(y, \overrightarrow{P}_C y) \leq D_f(y, c)$*, for all* $c \in C$*. The point* $\overrightarrow{P}_C y$ *is called the* forward Bregman projection *of* y *onto* C *and it is characterized by*

$$\overleftarrow{P}_C y \in C \cap \operatorname{int}\operatorname{dom} f \quad \textit{and} \quad (\forall c \in C)\ \big\langle c - \overrightarrow{P}_C y, \nabla^2 f(\overrightarrow{P}_C y)(y - \overrightarrow{P}_C y)\big\rangle \leq 0; \tag{2.5}$$

equivalently, by

(2.6)
$$\overleftarrow{P}_C y \in C \cap \operatorname{int}\operatorname{dom} f \;\; \textit{and} \;\; (\forall c \in C) \;\; D_f(c,y) \geq D_f(c, \overrightarrow{P}_C y) + D_{D_f}\big((c,c),(y,\overrightarrow{P}_C y)\big).$$

Finally, the operator $\overrightarrow{P}_C$ *is continuous on* $\operatorname{int}\operatorname{dom} f$.

Proof. This follows from [7, Lemma 2.9, Lemma 3.5, Lemma 3.6, and Corollary 3.7]. ☐

The key requirement in Definition 2.4 is the convexity of D_f, which is studied separately in [6]. Not every Legendre function allows forward Bregman projections, but the most important ones from Example 2.1 do:

Example 2.7 (functions allowing forward Bregman projections). [7, Example 2.16] Let $X = \mathbb{R}^J$. Then the energy, the negative entropy, and the Fermi-Dirac entropy allow forward Bregman projections.

The following example shows that backward and forward Bregman projections are different notions.

Example 2.8 (entropic averaging in $\mathbb{R}^2$). Let $f\colon \mathbb{R}^2 \to \left]-\infty,+\infty\right] : (\xi_1,\xi_2) \mapsto \sum_{i=1}^2 \xi_i \ln(\xi_i) - \xi_i$ be the negative entropy on $\mathbb{R}^2$, and let $\Delta = \{(\xi_1,\xi_2) \in \mathbb{R}^2 : \xi_1 = \xi_2\}$. Then $\operatorname{dom} f = \left[0,+\infty\right[^2$ and clearly $\Delta \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$. Using (2.1) and (2.5), it is straightforward to verify that, for every $(\xi_1,\xi_2) \in \operatorname{int}\operatorname{dom} f = \left]0,+\infty\right[^2$,

$$\overleftarrow{P}_\Delta(\xi_1,\xi_2) = \big(\sqrt{\xi_1\xi_2}, \sqrt{\xi_1\xi_2}\big) \quad \text{and} \quad \overrightarrow{P}_\Delta(\xi_1,\xi_2) = \big(\tfrac{1}{2}(\xi_1+\xi_2), \tfrac{1}{2}(\xi_1+\xi_2)\big). \tag{2.7}$$

These formulae can also be deduced from Example 3.16 below.

We close this section with a characterization of convergence for Bregman monotone sequences. Note that when $f = \frac{1}{2}\|\cdot\|^2$, Bregman monotonicity reverts to the standard notion of Fejér monotonicity, which is discussed in detail in [4] and [21].

Proposition 2.9 (Bregman monotonicity). *Suppose C is a closed convex set in X such that $C \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$. Suppose further $(y_n)_{n\in\mathbb{N}}$ is a sequence which is* Bregman monotone *with respect to $C \cap \operatorname{int}\operatorname{dom} f$, i.e., it lies in $\operatorname{int}\operatorname{dom} f$ and*

$$(\forall c \in C \cap \operatorname{int}\operatorname{dom} f)(\forall n \in \mathbb{N}) \quad D_f(c, y_{n+1}) \leq D_f(c, y_n). \tag{2.8}$$

Then: $(y_n)_{n\in\mathbb{N}}$ converges to some point in $C\cap\operatorname{int}\operatorname{dom} f \Leftrightarrow$ all cluster points of $(y_n)_{n\in\mathbb{N}}$ are in C.

Proof. The implication "$\Rightarrow$" is clear. "$\Leftarrow$": pick $c \in C \cap \operatorname{int}\operatorname{dom} f$. Then the sequence $\big(D_f(c,y_n)\big)_{n\in\mathbb{N}}$ is decreasing and nonnegative, hence bounded. By Proposition 2.2.(iv), $(y_n)_{n\in\mathbb{N}}$ is bounded and all its cluster points lie in $\operatorname{int}\operatorname{dom} f$. Let $\{c,\hat{c}\} \subset C \cap \operatorname{int}\operatorname{dom} f$ be two cluster points of $(y_n)_{n\in\mathbb{N}}$, say $y_{k_n} \to c$ and $y_{l_n} \to \hat{c}$. By Proposition 2.2.(iii), $D_f(c, y_{k_n}) \to 0$. Since $(y_n)_{n\in\mathbb{N}}$ is Bregman monotone, we have $D_f(c,y_n) \to 0$ and, in particular, $D_f(c, y_{l_n}) \to 0$. Using Proposition 2.2.(v), we conclude $c = \hat{c}$. ☐

3. Bregman retractions.

3.1. Properties and examples.

Definition 3.1 (Bregman retraction). *Suppose C is a closed convex set in X such that $C \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$ and μ is a function from $\operatorname{dom}\mu = (C \cap \operatorname{int}\operatorname{dom} f) \times \operatorname{int}\operatorname{dom} f$ to $\left[0,+\infty\right[$. Then $R\colon \operatorname{dom} R = \operatorname{int}\operatorname{dom} f \to C \cap \operatorname{int}\operatorname{dom} f$ is a* Bregman retraction of C with modulus μ, *if the following two properties hold for every $c \in C \cap \operatorname{int}\operatorname{dom} f$ and every $x \in \operatorname{int}\operatorname{dom} f$:*

(i) $D_f(c,x) \geq D_f(c,Rx) + \mu(c,x)$.

(ii) *If* $(x_n)_{n\in\mathbb{N}}$ *is a sequence in* $\operatorname{int}\operatorname{dom} f$ *and* y *is a point in* $\operatorname{int}\operatorname{dom} f$ *such that* $x_n \to x$, $Rx_n \to y$, *and* $\mu(c, x_n) \to 0$, *then* $x = y$.

PROPOSITION 3.2 (basic properties of Bregman retractions). *Suppose* C *is a closed convex set in* X *such that* $C \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$. *Suppose further* R *is a Bregman retraction of* C *with modulus* μ. *Then the following holds true for every* $c \in C \cap \operatorname{int}\operatorname{dom} f$ *and every* $x \in \operatorname{int}\operatorname{dom} f$:

(i) *Suppose* $(x_n)_{n\in\mathbb{N}}$ *is a sequence in* $\operatorname{int}\operatorname{dom} f$ *and* y *is a point in* $\operatorname{int}\operatorname{dom} f$ *such that* $x_n \to x$ *and* $Rx_n \to y$. *Then* $\mu(c, x_n) \to 0 \Leftrightarrow x = y$.

(ii) $\mu(c, x) = 0 \Leftrightarrow x = Rx \Leftrightarrow x \in C$.

Proof. (i): the implication "$\Rightarrow$" is clear by Definition 3.1.(ii). If $x = y$, then (use Proposition 2.2.(i) and Definition 3.1.(i)) $0 = D_f(c, x) - D_f(c, x) \leftarrow D_f(c, x_n) - D_f(c, Rx_n) \geq \mu(c, x_n) \geq 0$. Hence $\mu(c, x_n) \to 0$ and "$\Leftarrow$" is verified. (ii): The first equivalence is a special case of (i), while the implication $x = Rx \Rightarrow x \in C$ is clear. Now assume $x \in C$. Then Definition 3.1.(i) with $c = x$ yields $0 = D_f(x, x) \geq D_f(x, Rx) + \mu(x, x) \geq D_f(x, Rx) \geq 0$. Hence $D_f(x, Rx) = 0$ and thus $x = Rx$ by Proposition 2.2.(ii). ☐

Every nonempty closed convex set in X possesses a Bregman retraction with respect to the energy:

EXAMPLE 3.3 (orthogonal projection). Suppose $f = \frac{1}{2}\|\cdot\|^2$ and C is a nonempty closed convex set in X. Then its orthogonal projection P_C is a Bregman retraction with modulus $\mu\colon (c, x) \mapsto \frac{1}{2}\|x - P_C x\|^2$.

Proof. This will turn out to be a special case of Example 3.6 or 3.7. ☐

However, the next example shows that there exist Bregman retractions that are not projections.

EXAMPLE 3.4. Let $f = \frac{1}{2}\|\cdot\|^2$ and $C = \{x \in X \colon \|x\| \leq 1\}$. Fix $\varepsilon \in \left]0, 1\right[$, define $\lambda\colon X \to \left[0, +\infty\right[\colon x \mapsto 1 + \min\big\{\varepsilon, \|x - P_C x\|\big\}$, and let $R\colon X \to C \colon x \mapsto \big(1 - \lambda(x)\big)x + \lambda(x) P_C x$, where P_C is the orthogonal projection onto C. Then R is a Bregman retraction of C with modulus $\mu\colon (c, x) \mapsto \frac{1}{2}\big(2 - \lambda(x)\big)\lambda(x)\|x - P_C x\|^2$.

Proof. Fix $x \in X$ and $c \in C$. It follows from standard properties of orthogonal projections (see, e.g., [4, Corollary 2.5]) that

$$\|x - c\|^2 - \|x - Rx\|^2 \geq \big(2 - \lambda(x)\big)\lambda(x)\|x - P_C x\|^2, \tag{3.1}$$

which corresponds to Definition 3.1.(i). Now assume $(x_n)_{n\in\mathbb{N}}$ converges to x. Since P_C, and hence λ, is continuous, we have $P_C x_n \to P_C x$ and $Rx_n \to Rx$. Assume further $\mu(c, x_n) \to 0$. Then $x_n - P_C x_n \to 0$ and thus $Rx_n = x_n + \lambda(x_n)(P_C x_n - x_n) \to x$. Hence $Rx = x$ and therefore R is a Bregman retraction. ☐

REMARK 3.5. In passing, note that if C is a closed convex set in X such that $(\operatorname{int} C) \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$ and $y \in \operatorname{int}\operatorname{dom} f \smallsetminus C$, then both points $\overleftarrow{P}_C y$ and $\overrightarrow{P}_C y$ belong to $(\operatorname{bdry} C) \cap \operatorname{int}\operatorname{dom} f$. Now let R and C as in Example 3.4. Since R maps points outside C to the interior of C, there is no function f such that R is the backward or forward Bregman projection onto C with respect to D_f.

The following two examples contain Example 3.3 if we let $f = \frac{1}{2}\|\cdot\|^2$.

EXAMPLE 3.6 (backward Bregman projection). Suppose C is a closed convex set in X such that $C \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$. Then the backward Bregman projection $\overleftarrow{P}_C$ is a continuous Bregman retraction with modulus $\mu\colon (c, x) \mapsto D_f(\overleftarrow{P}_C x, x)$.

Proof. This follows from Fact 2.3 and Proposition 2.2.(iii). ☐

EXAMPLE 3.7 (forward Bregman projection). Suppose f allows forward Bregman projections and C is a closed convex set in X such that $C \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$. Then the forward Bregman projection $\overrightarrow{P}_C$ is a continuous Bregman retraction with modulus

$$\mu\colon (c,x) \mapsto D_{D_f}\big((c,c),(x,\overrightarrow{P}_C x)\big) = D_f(c,x) - D_f(c,\overrightarrow{P}_C x) + \big\langle c - \overrightarrow{P}_C x, \nabla^2 f(\overrightarrow{P}_C x)(x - \overrightarrow{P}_C x)\big\rangle. \tag{3.2}$$

Proof. See [7, Lemma 2.9] for the nonnegativity of D_{D_f} and for the expression of D_{D_f}. Fact 2.6 states that $\overrightarrow{P}_C$ is continuous and (2.6) verifies Definition 3.1.(i). It remains to establish condition (ii) of Definition 3.1. So pick $c \in C \cap \operatorname{int}\operatorname{dom} f$ and $(x_n)_{n\in\mathbb{N}}$ in $\operatorname{int}\operatorname{dom} f$ such that $x_n \to x \in \operatorname{int}\operatorname{dom} f$, $\overrightarrow{P}_C x_n \to y \in \operatorname{int}\operatorname{dom} f$, and $\mu(c,x_n) \to 0$. By [7, Lemma 2.9], D_{D_f} is continuous on $(\operatorname{int}\operatorname{dom} f)^4$ and therefore $\mu(c,x_n) \to D_{D_f}\big((c,c),(x,y)\big)$. Altogether, $D_{D_f}\big((c,c),(x,y)\big) = 0$ and [7, Lemma 2.10] implies $x = y$. ☐

The following example is motivated by [30, Section 4.7].

EXAMPLE 3.8. Let $X = \mathbb{R}^J$, f be the negative entropy, and

$$C = \big\{x \in X\colon x \geq 0 \text{ and } \langle x, \mathbf{1}\rangle \leq 1\big\}, \quad \text{where} \quad \mathbf{1} = (1,\ldots,1) \in X. \tag{3.3}$$

Let

$$R\colon \operatorname{int}\operatorname{dom} f \to C \cap \operatorname{int}\operatorname{dom} f\colon x \mapsto \begin{cases} x, & \text{if } x \in C;\\ x/\langle x,\mathbf{1}\rangle, & \text{otherwise.}\end{cases} \tag{3.4}$$

Then $R = \overleftarrow{P}_C = \overrightarrow{P}_C$. Consequently, R is a continuous Bregman retraction of C.

Proof. Fix $c \in C$ and $x \in \operatorname{int}\operatorname{dom} f \smallsetminus C$. Then, abusing notation slightly,

$$\begin{aligned}\langle c - Rx, \nabla f(x) - \nabla f(Rx)\rangle &= \big\langle c - x/\langle x,\mathbf{1}\rangle, \ln(x) - \ln\big(x/\langle x,\mathbf{1}\rangle\big)\big\rangle\\ &= \big\langle c - x/\langle x,\mathbf{1}\rangle, \ln\big(\langle x,\mathbf{1}\rangle\big)\mathbf{1}\big\rangle\\ &= \ln\big(\langle x,\mathbf{1}\rangle\big)\big(\langle c,\mathbf{1}\rangle - 1\big)\\ &\leq 0.\end{aligned}$$

By (2.1), we see that $Rx = \overleftarrow{P}_C x$. Similarly,

$$\begin{aligned}\big\langle c - Rx, \nabla^2 f(Rx)(x - Rx)\big\rangle &= \big\langle c - x/\langle x,\mathbf{1}\rangle, \big(\langle x,\mathbf{1}\rangle - 1\big)\cdot\mathbf{1}\big\rangle\\ &= \big(\langle x,\mathbf{1}\rangle - 1\big)\big(\langle c,\mathbf{1}\rangle - 1\big)\\ &\leq 0.\end{aligned}$$

Thus, using (2.5), $Rx = \overrightarrow{P}_C x$. ☐

REMARK 3.9. In [30, Section 4.7], it is observed that the *orthogonal* projection of an arbitrary point in $\mathbb{R}^J$ onto C is hard to compute explicitly, and hence the use of the following extension $\widetilde{R}$ of R is suggested. Denoting the nonnegative part of a vector $x \in \mathbb{R}^J$ by x^+ (i.e., x^+ is the orthogonal projection of x onto the nonnegative orthant), the extension $\widetilde{R}$ is defined by

$$\widetilde{R}\colon X \to C\colon x \mapsto \begin{cases} x^+/\langle x^+,\mathbf{1}\rangle, & \text{if } \langle x^+,\mathbf{1}\rangle > 1;\\ x^+, & \text{otherwise.}\end{cases} \tag{3.5}$$

It is important to note that for certain points $x \in \operatorname{dom} f \smallsetminus C$ and $c \in C$, the inequality $\|\widetilde{R}x - c\| = \|Rx - c\| \leq \|x - c\|$ does *not* hold. Indeed, take $X = \mathbb{R}^2$, let $c = (1,0)$, consider the ray emanating from 0 that makes an angle of $\pi/6$ with $[0,+\infty[\cdot c$, and let x be the orthogonal projection of c onto this ray. Then $\|\widetilde{R}x-c\| = \|Rx-c\| > \|x-c\|$ (and this example can be lifted to $\mathbb{R}^J$, where $J \geq 3$). Therefore, Example 3.8 shows that an operator which is not a Bregman retraction with respect to the energy may turn out to be a Bregman retraction with respect to some other function.

3.2. Comparison with Censor and Reich's paracontractions. Let us first recall the concept of a *Bregman function*, as defined in [15] or [17] (see also [5], [9], [25], and [32] for more concise definitions).

DEFINITION 3.10. *Let S be a nonempty open convex subset of $\mathbb{R}^J$, let $g\colon \overline{S} \to \mathbb{R}$ be a continuous and strictly convex function, and let D_g be the corresponding Bregman distance. Then g is a* Bregman function *with* zone S *if the following conditions hold:*

(i) *g is continuously differentiable on S;*
(ii) *For every $x \in \overline{S}$ the sets $\big(\{y \in S\colon D_g(x,y) \leq \eta\}\big)_{\eta\in\mathbb{R}}$ are bounded;*
(iii) *For every $y \in S$ the sets $\big(\{x \in \overline{S}\colon D_g(x,y) \leq \eta\}\big)_{\eta\in\mathbb{R}}$ are bounded;*
(iv) *If $(y_n)_{n\in\mathbb{N}}$ lies in S and $y_n \to y$, then $D_g(y,y_n) \to 0$;*
(v) *If $(x_n)_{n\in\mathbb{N}}$ is a bounded sequence in $\overline{S}$, $(y_n)_{n\in\mathbb{N}}$ lies in S, $y_n \to y$, and $D_g(x_n,y_n) \to 0$, then $x_n \to y$.*

The following notion is due to Censor and Reich.

DEFINITION 3.11 ([16, Definition 3.2]). *Suppose g is a Bregman function with zone $S \subset \mathbb{R}^J$ and let $T\colon S \to \mathbb{R}^J$ be an operator with domain S. A point $\bar{x} \in \mathbb{R}^J$ is called an* asymptotic fixed point *of T if there exists a sequence $(x_n)_{n\in\mathbb{N}}$ in S such that $x_n \to \bar{x}$ and $Tx_n \to \bar{x}$. The set of asymptotic fixed points is denoted by $\widehat{F}(T)$. The operator T is a* paracontraction *if $\widehat{F}(T) \neq \varnothing$ and the following two conditions hold.*

(i) $\big(\forall c \in \widehat{F}(T)\big)\big(\forall x \in S\big) \quad D_g(c,Tx) \leq D_g(c,x).$
(ii) *If $(x_n)_{n\in\mathbb{N}}$ is a bounded sequence in S and $c \in \widehat{F}(T)$ satisfies $D_g(c,x_n) - D_g(c,Tx_n) \to 0$, then $D_g(Tx_n,x_n) \to 0$.*

EXAMPLE 3.12 (Bregman retraction $\not\Rightarrow$ paracontraction). Let f and Δ be as in Example 2.8, and set $T = \overrightarrow{P}_\Delta$. Then T is a continuous Bregman retraction but not a paracontraction.

Proof. The first claim follows from Examples 2.7 and 3.7. We now show that T is not a paracontraction. First, f is a Bregman function with zone $S = \operatorname{int}\operatorname{dom} f =]0,+\infty[^2$. In addition,

$$D_f(x,y) = \xi_1\ln(\xi_1/\eta_1) - \xi_1 + \eta_1 + \xi_2\ln(\xi_2/\eta_2) - \xi_2 + \eta_2, \tag{3.6}$$

for $x = (\xi_1,\xi_2) \in \operatorname{dom} f = [0,+\infty[^2$ and $y = (\eta_1,\eta_2) \in \operatorname{int}\operatorname{dom} f =]0,+\infty[^2$. The set of asymptotic fixed points of T is seen to be

$$\widehat{F}(T) = \Delta \cap \operatorname{dom} f = \big\{(\xi_1,\xi_2) \in \mathbb{R}^2 : \xi_1 = \xi_2 \geq 0\big\} \neq \varnothing. \tag{3.7}$$

Fix $c = (0,0) \in \widehat{F}(T)$ and pick an arbitrary $x = (\xi_1,\xi_2) \in \operatorname{int}\operatorname{dom} f \smallsetminus \Delta$. By Example 2.8, $Tx = \overrightarrow{P}_\Delta x = \frac{1}{2}(\xi_1+\xi_2,\xi_1+\xi_2)$. Hence,

$$D_f(c,x) - D_f(c,Tx) = D_f(0,x) - D_f(0,Tx) = \big(\xi_1+\xi_2\big) - \big(\tfrac{1}{2}(\xi_1+\xi_2) + \tfrac{1}{2}(\xi_1+\xi_2)\big) = 0. \tag{3.8}$$

However, since $x \notin \Delta$, we have $Tx = \overrightarrow{P}_\Delta x \neq x$ and so, by Proposition 2.2.(ii), $D_f(Tx,x) > 0$. Therefore Definition 3.11.(ii) fails and it follows that T is not a paracontraction. ∎

EXAMPLE 3.13 (paracontraction $\not\Rightarrow$ Bregman retraction). Let $X = \mathbb{R}$ and $f = \frac{1}{2}|\cdot|^2$. Then f is a Bregman function with zone $S = X$ and $T\colon X \to X : x \mapsto \frac{1}{2}x$ is a paracontraction with $\widehat{F}(T) = \{0\}$. Now suppose that T is a Bregman retraction. Then, by Proposition 3.2.(ii), the underlying set must be $C = \{0\}$. However, by Definition 3.1, this is absurd since the range of T is not a subset of C. Therefore T is not a Bregman retraction.

3.3. New Bregman retractions via averages and products.

PROPOSITION 3.14 (averaged Bregman retractions). *Suppose f allows forward Bregman projections and C is a closed convex set in X such that $C \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$. Suppose further R_1 and R_2 are two continuous Bregman retractions of C with moduli μ_1 and μ_2. Fix $\lambda_1 > 0$ and $\lambda_2 > 0$ such that $\lambda_1 + \lambda_2 = 1$, and set $R = \lambda_1 R_1 + \lambda_2 R_2$. Then R is a Bregman retraction of C with modulus $\mu = \lambda_1\mu_1 + \lambda_2\mu_2$.*

Proof. It is clear that the range of R is contained in $C \cap \operatorname{int}\operatorname{dom} f$ and that $\operatorname{dom} R = \operatorname{int}\operatorname{dom} f$. Fix $c \in C \cap \operatorname{int}\operatorname{dom} f$ and $x \in \operatorname{int}\operatorname{dom} f$. Since both R_1 and R_2 are Bregman retractions of C and since $D_f(c,\cdot)$ is convex on $\operatorname{int}\operatorname{dom} f$, we have

$$\begin{aligned}
D_f(c,x) &= \lambda_1 D_f(c,x) + \lambda_2 D_f(c,x)\\
&\geq \lambda_1\big(D_f(c,R_1x) + \mu_1(c,x)\big) + \lambda_2\big(D_f(c,R_2x) + \mu_2(c,x)\big)\\
&= \big(\lambda_1 D_f(c,R_1x) + \lambda_2 D_f(c,R_2x)\big) + \mu(c,x)\\
&\geq D_f(c,Rx) + \mu(c,x).
\end{aligned}$$

Hence condition (i) of Definition 3.1 holds. Next, assume $(x_n)_{n\in\mathbb{N}}$ is a sequence in $\operatorname{int}\operatorname{dom} f$ converging to x such that $Rx_n \to y \in \operatorname{int}\operatorname{dom} f$ and $\mu(c,x_n) \to 0$. Then $\mu_1(c,x_n) \to 0$ and $\mu_2(c,x_n) \to 0$. On the other hand, since R_1 and R_2 are continuous on $\operatorname{int}\operatorname{dom} f$, $(R_1x_n, R_2x_n) \to (R_1x, R_2x)$ and hence $Rx_n \to Rx$. Thus $y = Rx$. Using condition (ii) of Definition 3.1 on each R_i, we also have $x = R_1x = R_2x$ and thus $x = Rx$. Altogether, $x = y$ and condition (ii) of Definition 3.1 is verified as well. ∎

EXAMPLE 3.15 (averaged backward-forward Bregman projections). Suppose f allows forward Bregman projections and C is a closed convex set in X such that $C \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$. Denote the Bregman retraction and its modulus from Example 3.6 (respectively Example 3.7) by R_1 and μ_1 (respectively R_2 and μ_2). Fix $\lambda_1 > 0$ and $\lambda_2 > 0$ such that $\lambda_1 + \lambda_2 = 1$, and set $R = \lambda_1R_1 + \lambda_2R_2$. Then R is a Bregman retraction of C with modulus $\mu = \lambda_1\mu_1 + \lambda_2\mu_2$.

We conclude this section with a product space construction first introduced by Pierra in [28] (see also [29]). The extension to a Bregman distance setting is due to Censor and Elfving [13]. The product space technique will be extremely useful for analyzing the parallel projection methods presented in Section 5.

EXAMPLE 3.16 (product space setup). For convenience, let $I = \{1,\ldots,N\}$ in (1.3). Denote the standard Euclidean product space X^N by $\mathbf{X}$ and write $\mathbf{x} = (x_i)_{i\in I}$, for $\mathbf{x} \in \mathbf{X}$. Let

$$\boldsymbol{\Delta} = \big\{(x,\ldots,x) \in \mathbf{X} : x \in X\big\} \quad \text{and} \quad \mathbf{C} = C_1 \times \cdots \times C_N. \tag{3.9}$$

Fix $(\lambda_i)_{i\in I}$ in $]0,1]$ such that $\sum_{i\in I}\lambda_i = 1$, and set

$$\mathbf{f}\colon \mathbf{X} \to\ ]-\infty,+\infty] : \mathbf{x} \mapsto \textstyle\sum_{i\in I}\lambda_i f(x_i). \tag{3.10}$$

Then $\mathbf{f}$ is Legendre, $\operatorname{dom}\mathbf{f}^*$ is open, and $\boldsymbol{\Delta} \cap \mathbf{C} \cap \operatorname{int}\operatorname{dom}\mathbf{f} \neq \varnothing$. In addition, if $\mathbf{x} \in \operatorname{dom}\mathbf{f}$ and $\mathbf{y} \in \operatorname{int}\operatorname{dom}\mathbf{f}$, then $D_{\mathbf{f}}(\mathbf{x},\mathbf{y}) = \sum_{i\in I}\lambda_i D_f(x_i,y_i)$. Moreover:

(i) The operators $\overleftarrow{P}_{\mathbf{\Delta}}$ and $\overleftarrow{P}_{\mathbf{C}}$ are continuous Bregman retractions of $\mathbf{\Delta}$ and $\mathbf{C}$, respectively, and

$$\begin{aligned} &\overleftarrow{P}_{\mathbf{\Delta}}\mathbf{y} = (z,\ldots,z), \quad \text{where} \quad z = \nabla f^*\big(\textstyle\sum_{i\in I}\lambda_i\nabla f(y_i)\big),\\ &\overleftarrow{P}_{\mathbf{C}}\mathbf{y} = \big(\overleftarrow{P}_{C_i}y_i\big)_{i\in I}. \end{aligned} \tag{3.11}$$

(ii) Suppose f allows forward Bregman projections. Then so does $\mathbf{f}$. The operators $\overrightarrow{P}_{\mathbf{\Delta}}$ and $\overrightarrow{P}_{\mathbf{C}}$ are continuous Bregman retractions of $\mathbf{\Delta}$ and $\mathbf{C}$, respectively, and

$$\begin{aligned} &\overrightarrow{P}_{\mathbf{\Delta}}\mathbf{y} = (z,\ldots,z), \quad \text{where} \quad z = \textstyle\sum_{i\in I}\lambda_i y_i,\\ &\overrightarrow{P}_{\mathbf{C}}\mathbf{y} = \big(\overrightarrow{P}_{C_i}y_i\big)_{i\in I}. \end{aligned} \tag{3.12}$$

Proof. The fact that the operators $\overleftarrow{P}_{\mathbf{\Delta}}$, $\overleftarrow{P}_{\mathbf{C}}$ (and $\overrightarrow{P}_{\mathbf{\Delta}}$, $\overrightarrow{P}_{\mathbf{C}}$ provided they exist) are continuous Bregman retractions follows from Example 3.6 (and Example 3.7, respectively). (i): See [5, Corollary 7.2] or [13, Lemmata 4.1 and 4.2]. (ii): Using Definition 2.4, it is straightforward to check that $\mathbf{f}$ allows forward Bregman projections. Next, let $z = \sum_{i\in I}\lambda_i y_i$ and $\mathbf{z} = (z,\ldots,z) \in \mathbf{X}$. Then $\mathbf{z} \in \mathbf{\Delta}$. Observe that $\nabla^2\mathbf{f}(\mathbf{z})\mathbf{y} = \big(\lambda_i\nabla^2 f(z)y_i\big)_{i\in I}$ and $\nabla^2\mathbf{f}(\mathbf{z})\mathbf{z} = \big(\lambda_i\nabla^2 f(z)z\big)_{i\in I}$. Hence $\nabla^2\mathbf{f}(\mathbf{z})(\mathbf{y}-\mathbf{z}) \in \mathbf{\Delta}^\perp = \{\mathbf{x} \in \mathbf{X} : \sum_{i\in I}x_i = 0\}$, because $\sum_{i\in I}\lambda_i\nabla^2 f(z)y_i = \nabla^2 f(z)(\sum_{i\in I}\lambda_i y_i) = \nabla^2 f(z)z = \sum_{i\in I}\lambda_i\nabla^2 f(z)z$. Thus, it follows from (2.5) that $\mathbf{z} = \overrightarrow{P}_{\mathbf{\Delta}}\mathbf{y}$. In view of the separability of $D_{\mathbf{f}}$ and $\mathbf{C}$, the formula for $\overrightarrow{P}_{\mathbf{C}}$ is clear. □

REMARK 3.17. The case when $\mathbf{f}$ is replaced by $\sum_{i\in I}\lambda_i g_i(x_i)$, where $(g_i)_{i\in I}$ is a family of possibly different Bregman functions, was considered in [12] and [13]. This setup is too general to permit closed forms for $\overleftarrow{P}_{\mathbf{\Delta}}$ or $\overrightarrow{P}_{\mathbf{\Delta}}$. Furthermore, since Bregman functions are not necessarily Legendre, the existence of Bregman projections is not guaranteed and must therefore be imposed.

4. Main result. Going back to (1.4), we henceforth set

$$C = \textstyle\bigcap_{i\in I} C_i \tag{4.1}$$

and assume that (the existence of the Bregman retractions is guaranteed by (1.3) and Example 3.6)

$$(\forall i \in I) \quad R_i \text{ is a Bregman retraction of } C_i \text{ with modulus } \mu_i. \tag{4.2}$$

We now formulate our main result.

THEOREM 4.1 (method of Bregman retractions). *Given an arbitrary starting point $y_0 \in \operatorname{int}\operatorname{dom} f$, generate a sequence by*

$$(\forall n \in \mathbb{N}) \quad y_{n+1} = R_{\mathrm{i}(n+1)}y_n, \tag{4.3}$$

where $\mathrm{i}\colon \mathbb{N} \to I$ takes on each value in I infinitely often. Then the sequence $(y_n)_{n\in\mathbb{N}}$ converges to a point in $C \cap \operatorname{int}\operatorname{dom} f$.

Proof. We proceed in several steps.

Step 1: We have

$$(\forall n \in \mathbb{N})(\forall c \in C_{\mathrm{i}(n+1)} \cap \operatorname{int}\operatorname{dom} f) \quad D_f(c,y_n) \geq D_f(c,y_{n+1}) + \mu_{\mathrm{i}(n+1)}(c,y_n).$$

Indeed, $y_{n+1} = R_{\mathrm{i}(n+1)} y_n$ and $R_{\mathrm{i}(n+1)}$ is a Bregman retraction of $C_{\mathrm{i}(n+1)}$.

Step 2: $(y_n)_{n\in\mathbb{N}}$ is Bregman monotone with respect to $C \cap \operatorname{int}\operatorname{dom} f$.
This is clear from *Step 1*.

Step 3: $(y_n)_{n\in\mathbb{N}}$ is bounded and all its cluster points belong to $\operatorname{int}\operatorname{dom} f$.
Fix $c \in C \cap \operatorname{int}\operatorname{dom} f$. In view of *Step 2*, the sequence $\big(D_f(c, y_n)\big)_{n\in\mathbb{N}}$ is decreasing and hence bounded. Now apply Proposition 2.2.(iv).

Next, let us consider an arbitrary cluster point of $(y_n)_{n\in\mathbb{N}}$, say $y_{k_n} \to y$.

Step 4: $y \in \operatorname{int}\operatorname{dom} f$ and $D_f(y, y_{k_n}) \to 0$.
This follows from *Step 3* and Proposition 2.2.(i).

Because I is finite, after passing to a further subsequence and relabelling if necessary, we assume that $\mathrm{i}(k_n) \equiv i_{\mathrm{in}}$. Since $C_{i_{\mathrm{in}}}$ is closed, we have $y \in C_{i_{\mathrm{in}}}$.

We now define $I_{\mathrm{in}} = \{i \in I : y \in C_i\}$ and $I_{\mathrm{out}} = \{i \in I : y \notin C_i\}$.

Step 5: $I_{\mathrm{out}} = \varnothing$.
Suppose to the contrary that $I_{\mathrm{out}} \neq \varnothing$. After passing to a further subsequence and relabelling if necessary, we assume that $\{\mathrm{i}(k_n), \mathrm{i}(k_n+1), \ldots, \mathrm{i}(k_{n+1}-1)\} = I$ — this is possible, by our assumptions on the index selector i. For each $n \in \mathbb{N}$, let

$$m_n = \min\big\{k_n \le k \le k_{n+1} - 1 : \mathrm{i}(k) \in I_{\mathrm{out}}\big\} - 1. \tag{4.4}$$

The current assumptions imply that each m_n is a well-defined integer in $[k_n, k_{n+1}-2]$ satisfying $y \in \bigcap_{k_n \le k \le m_n} C_{\mathrm{i}(k)}$. Repeated use of *Step 1* thus yields

$$(\forall n \in \mathbb{N}) \quad D_f(y, y_{m_n}) \le D_f(y, y_{k_n}). \tag{4.5}$$

Using *Step 4* and Proposition 2.2.(v), we deduce $y_{m_n} \to y$. After passing to a further subsequence and relabelling if necessary, we assume that $\mathrm{i}(m_n+1) \equiv i_{\mathrm{out}}$ and that $y_{m_n+1} = R_{i_{\mathrm{out}}} y_{m_n} \to z \in C_{i_{\mathrm{out}}} \cap \operatorname{int}\operatorname{dom} f$ (using *Step 3* again). Now fix $c \in C \cap \operatorname{int}\operatorname{dom} f$. *Step 1* implies that $\big(\mu_{\mathrm{i}(n+1)}(c, y_n)\big)_{n\in\mathbb{N}}$ is summable; in particular,

$$\mu_{i_{\mathrm{out}}}(c, y_{m_n}) = \mu_{\mathrm{i}(m_n+1)}(c, y_{m_n}) \to 0. \tag{4.6}$$

Since $R_{i_{\mathrm{out}}}$ is a Bregman retraction, we obtain $y = z \in C_{i_{\mathrm{out}}}$. But this in turn implies $i_{\mathrm{out}} \in I_{\mathrm{in}}$, which is the desired contradiction.

Last step: We have shown that $(y_n)_{n\in\mathbb{N}}$ is Bregman monotone with respect to $C \cap \operatorname{int}\operatorname{dom} f$ (*Step 2*), and that all its cluster points lie in $C \cap \operatorname{int}\operatorname{dom} f$ (*Step 4* and *Step 5*). Therefore, by Proposition 2.9, the entire sequence $(y_n)_{n\in\mathbb{N}}$ converges to some point in $C \cap \operatorname{int}\operatorname{dom} f$. ☐

Remark 4.2. The proof of Theorem 4.1 is guided by the proof of [7, Theorem 4.1] and similar convergence results on iterating operators under such general control; see [5], [16], and [26]. The present proof clearly shows when properties of the Bregman distance are used, as opposed to those of the modulus. This distinction is blurred in other proofs, because the implicit surrogates for the modulus depend on D_f: see the roles of $D_f\big(\overleftarrow{P}_{C_{r(n+1)}} y_n, y_n\big)$, $D_{D_f}\big((c,c), (y_n, \overrightarrow{P}_{C_{r(n+1)}} y_n)\big)$, $D_f\big(T_s z(t), z(t)\big)$, and $D_h^k(x^{k+1}, x^k)$ in the proofs of [5, Theorem 8.1], [7, Theorem 4.1], and [16, Theorem 3.1], [26, Theorem 4.1], respectively.

Remark 4.3 (Bregman retractions must correspond to the same Bregman distance). It is natural to ask whether it is possible to use iterates of Bregman retractions coming from possibly *different* underlying Bregman distances to solve convex feasibility problems. Unfortunately, this approach is not successful in general. To see this, let $X = \mathbb{R}^2$ and set $R_C = \overleftarrow{P}_C = \overrightarrow{P}_C$, where f is the *negative entropy* and C is as in Example 3.8 (with $J = 2$). Further, let L be the straight line through the points $(0, \frac{53}{56})$ and $(\frac{1}{4}, \frac{7}{8})$ and let R_L be the orthogonal projection, i.e., the backward or forward Bregman projection with respect to the *energy.* Then, although $L \cap \operatorname{int} C \neq \varnothing$, iterating the map $T = R_L \circ R_C$ may not lead to a point in $C \cap L$: indeed, $(\frac{1}{4}, \frac{7}{8})$ is a fixed point of T *outside* C.

5. Applications. Continuing to work under Assumptions (4.1) and (4.2), we now discuss various sequential and parallel algorithms derived from Theorem 4.1.

5.1. Sequential algorithms.

APPLICATION 5.1 (sequential Bregman projections). For each $i \in I$, let $R_i = \overleftarrow{P}_{C_i}$. Then Theorem 4.1 coincides with [5, Theorem 8.1.(ii)]; see also [16, Theorem 3.2]. For cyclic Bregman projections, see Bregman's classical [8].

APPLICATION 5.2 (new method of mixed backward-forward Bregman projections). Suppose f allows forward Bregman projections. For each $i \in I$, let either $R_i = \overleftarrow{P}_{C_i}$ or $R_i = \overrightarrow{P}_{C_i}$. Then Theorem 4.1 yields a convergence result on iterating a mixture of backward and forward Bregman projections. *Note:* If desired, it is possible to use both $\overleftarrow{P}_{C_i}$ and $\overrightarrow{P}_{C_i}$ for a given set C_i infinitely often, by counting this set twice.

The following three algorithms are special instances of Application 5.2.

APPLICATION 5.3 (sequential forward Bregman projections). Suppose f allows forward Bregman projections and let $R_i = \overrightarrow{P}_{C_i}$, for every $i \in I$. Then Theorem 4.1 reduces to [7, Theorem 4.1].

APPLICATION 5.4 (sequential orthogonal projections). Suppose $f = \frac{1}{2}\|\cdot\|^2$, and let each R_i be the orthogonal projection P_{C_i}. Then Theorem 4.1 turns into a convergence result on (chaotic or random) iterations of orthogonal projections; see also [2], [19], and references therein.

APPLICATION 5.5 (alternating backward-forward Bregman projections). Suppose f allows forward Bregman projections and let $I = \{1, 2\}$, $R_1 = \overrightarrow{P}_{C_1}$, and $R_2 = \overleftarrow{P}_{C_2}$. Then the method of Bregman retractions (4.3) corresponds to an alternating backward-forward Bregman projection method, which can be viewed as Csiszár and Tusnády's *alternating minimization procedure* [22] applied to D_f (this covers the *Expectation-Maximization method* for a specific Poisson model; see [22] and [24]).

5.2. Parallel algorithms. Various parallel algorithms arise by specializing Application 5.2 to the product space setting of Example 3.16. Using Example 3.16 and its notation, we deduce that the sequence $(\mathbf{T}^n \mathbf{x}_0)_{n \in \mathbb{N}}$, where $\mathbf{x}_0 \in \mathbf{\Delta}$ and $\mathbf{T} = \overleftarrow{P}_{\mathbf{\Delta}} \circ \overleftarrow{P}_{\mathbf{C}}$, converges to some point in $\mathbf{\Delta} \cap \mathbf{C} \cap \operatorname{int} \operatorname{dom} \mathbf{f}$. The same holds true when $\mathbf{T} \in \big\{\overrightarrow{P}_{\mathbf{\Delta}} \circ \overleftarrow{P}_{\mathbf{C}}, \overleftarrow{P}_{\mathbf{\Delta}} \circ \overrightarrow{P}_{\mathbf{C}}, \overrightarrow{P}_{\mathbf{\Delta}} \circ \overrightarrow{P}_{\mathbf{C}}\big\}$ provided that f allows forward Bregman projections.

Translating back to the original space X, we obtain the following four parallel algorithms.

APPLICATION 5.6 (parallel projections à la Censor and Elfving). Given $x_0 \in \operatorname{int} \operatorname{dom} f$, the sequence generated by

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \nabla f^* \big(\textstyle\sum_{i \in I} \lambda_i \nabla f(\overleftarrow{P}_{C_i} x_n) \big) \tag{5.1}$$

converges to a point in $C \cap \operatorname{int} \operatorname{dom} f$. This method, which amounts to iterating $\overleftarrow{P}_{\mathbf{\Delta}} \circ \overleftarrow{P}_{\mathbf{C}}$ in $\mathbf{X}$, was first suggested implicitly in [13]; see also [5] and Remark 3.17.

APPLICATION 5.7 (parallel projections à la Byrne and Censor I). Suppose f allows forward Bregman projections. Given $x_0 \in \operatorname{int} \operatorname{dom} f$, the sequence generated by

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \textstyle\sum_{i\in I} \lambda_i \overleftarrow{P}_{C_i} x_n \tag{5.2}$$

converges to a point in $C \cap \operatorname{int}\operatorname{dom} f$. This method, which amounts to iterating $\overrightarrow{P}_{\boldsymbol{\Delta}} \circ \overleftarrow{P}_{\mathbf{C}}$ in $\mathbf{X}$, can be found implicitly in [11, Section 4.1] (see also Remark 3.17).

APPLICATION 5.8 (parallel projections à la Byrne and Censor II). Suppose f allows forward Bregman projections. Given $x_0 \in \operatorname{int}\operatorname{dom} f$, the sequence generated by

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \nabla f^*\big(\textstyle\sum_{i\in I} \lambda_i \nabla f(\overrightarrow{P}_{C_i} x_n)\big) \tag{5.3}$$

converges to a point in $C \cap \operatorname{int}\operatorname{dom} f$. This method, which amounts to iterating $\overleftarrow{P}_{\boldsymbol{\Delta}} \circ \overrightarrow{P}_{\mathbf{C}}$ in $\mathbf{X}$, can be found implicitly in [11, Section 4.2] (see also Remark 3.17).

APPLICATION 5.9 (new parallel method). Suppose f allows forward Bregman projections. Given $x_0 \in \operatorname{int}\operatorname{dom} f$, the sequence generated by

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \textstyle\sum_{i\in I} \lambda_i \overrightarrow{P}_{C_i} x_n \tag{5.4}$$

converges to a point in $C \cap \operatorname{int}\operatorname{dom} f$. This corresponds to iterating $\overrightarrow{P}_{\boldsymbol{\Delta}} \circ \overrightarrow{P}_{\mathbf{C}}$ in $\mathbf{X}$.

The negative entropy and the energy lead to concrete examples:

APPLICATION 5.10 (averaged entropic projections à la Butnariu, Censor, and Reich). Let f be the negative entropy. Given $x_0 \in \operatorname{int}\operatorname{dom} f$, the sequence generated by

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \textstyle\sum_{i\in I} \lambda_i \overleftarrow{P}_{C_i} x_n \tag{5.5}$$

converges to a point in $C \cap \operatorname{int}\operatorname{dom} f$. Convergence is guaranteed by [10, Theorem 3.3], which holds true in more general settings, or by Application 5.7.

We conclude with a classical method which can be obtained from Application 5.6, 5.7, 5.8, or 5.9 by setting $f = \frac{1}{2}\|\cdot\|^2$.

APPLICATION 5.11 (parallel orthogonal projections à la Auslender). For each $i \in I$, let P_{C_i} be the orthogonal projection onto C_i. Given $x_0 \in X$, the sequence generated by

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \textstyle\sum_{i\in I} \lambda_i P_{C_i} x_n \tag{5.6}$$

converges to some point in C [1] (see also [3], [18], and [23] for the case when $C = \varnothing$).

Acknowledgment. We wish to thank Charlie Byrne and Yair Censor for sending us [11] and an anonymous referee for careful reading and helpful comments.

REFERENCES

[1] A. AUSLENDER, *Optimisation – Méthodes Numériques*, Paris, Masson, 1976.

[2] H. H. BAUSCHKE, *A norm convergence result on random products of relaxed projections in Hilbert space*, Trans. Amer. Math. Soc., 347 (1995), pp. 1365–1374.

[3] H. H. BAUSCHKE AND J. M. BORWEIN, *On the convergence of von Neumann's alternating projection algorithm for two sets*, Set-Valued Anal., 1 (1993), pp. 185–212.

[4] ———, *On projection algorithms for solving convex feasibility problems*, SIAM Rev., 38 (1996), pp. 367–426.

[5] ———, *Legendre functions and the method of random Bregman projections*, J. Convex Anal., 4 (1997), pp. 27–67.

[6] ———, *Joint and separate convexity of the Bregman distance*, in Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications, D. Butnariu, Y. Censor, and S. Reich, eds., pp. 23–36, Elsevier, Amsterdam, The Netherlands, 2001.
[7] H. H. Bauschke and D. Noll, *The method of forward projections*, J. Nonlinear Convex Anal., 3 (2002), pp. 191–205.
[8] L. M. Bregman, *The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming*, USSR Comput. Math. Math. Phys., 7 (1967), pp. 200–217.
[9] D. Butnariu, C. Byrne, and Y. Censor, *Redundant axioms in the definition of Bregman functions,* J. Convex Anal., 10 (2003), to appear.
[10] D. Butnariu, Y. Censor, and S. Reich, *Iterative averaging of entropic projections for solving stochastic convex feasibility problems*, Comput. Optim. Appl., 8 (1997), pp. 21–39.
[11] C. Byrne and Y. Censor, *Proximity function minimization using multiple Bregman projections, with applications to entropy optimization and Kullback-Leibler distance minimization*, unpublished research report (June 7, 1999), 37 pages. An abridged version of this report was published as [12]. Available at `http://math.haifa.ac.il/yair/bc29b070699.ps`.
[12] ———, *Proximity function minimization using multiple Bregman projections, with applications to split feasibility and Kullback-Leibler distance minimization*, Ann. Oper. Res., 105 (2001), pp. 77–98.
[13] Y. Censor and T. Elfving, *A multiprojection algorithm using Bregman projections in a product space*, Numer. Algorithms, 8 (1994), pp. 221–239.
[14] Y. Censor and G. T. Herman, *Block-iterative algorithms with underrelaxed Bregman projections*, SIAM J. Optim., 13 (2003), pp. 283–297.
[15] Y. Censor and A. Lent, *An iterative row-action method for interval convex programming*, J. Optim. Theory Appl., 34 (1981), pp. 321–353.
[16] Y. Censor and S. Reich, *Iterations of paracontractions and firmly nonexpansive operators with applications to feasibility and optimization*, Optimization, 37 (1996), pp. 323–339.
[17] Y. Censor and S. A. Zenios, *Parallel Optimization: Theory, Algorithms, and Applications,* Oxford University Press, New York, 1997.
[18] P. L. Combettes, *Inconsistent signal feasibility problems: Least-squares solutions in a product space*, IEEE Trans. Signal Process., 42 (1994), pp. 2955–2966.
[19] ———, *Construction d'un point fixe commun à une famille de contractions fermes*, C. R. Acad. Sci. Paris Sér. I Math., 320 (1995), pp. 1385–1390.
[20] ———, *Hilbertian convex feasibility problem: Convergence of projection methods*, Appl. Math. Optim., 35 (1997), pp. 311–330.
[21] ———, *Fejér monotonicity in convex optimization*, in Encyclopedia of Optimization, C. A. Floudas and P. M. Pardalos, eds., Vol. 2, pp. 106–114, Kluwer, Boston, MA, 2001.
[22] I. Csiszár and G. Tusnády, *Information geometry and alternating minimization procedures*, Statist. Decisions, (Supplement 1) (1984), pp. 205–237.
[23] A. R. De Pierro and A. N. Iusem, *A parallel projection method for finding a common point of a family of convex sets*, Pesqui. Oper., 5 (1985), pp. 1–20.
[24] A. N. Iusem, *A short convergence proof of the EM algorithm for a specific Poisson model*, Rev. Brasil. Prob. Estatística, 6 (1992), pp. 57–67.
[25] K. C. Kiwiel, *Free-steering relaxation methods for problems with strictly convex costs and linear constraints*, Math. Oper. Res., 22 (1997), pp. 326–349.
[26] ———*Generalized Bregman projections in convex feasibility problems*, J. Optim. Theory Appl., 96 (1998), pp. 139–157.
[27] K. C. Kiwiel and B. Łopuch, *Surrogate projection methods for finding fixed points of firmly nonexpansive mappings*, SIAM J. Optim., 7 (1997), pp. 1084–1102.
[28] G. Pierra, *Eclatement de contraintes en parallèle pour la minimisation d'une forme quadratique*, in Lecture Notes in Computer Science, Vol. 41, Springer-Verlag, New York, 1976, pp. 200–218.
[29] ———, *Decomposition through formalization in a product space*, Math. Programming, 28 (1984), pp. 96–115.
[30] B. T. Polyak, *Random algorithms for solving convex inequalities*, in Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications, D. Butnariu, Y. Censor, and S. Reich, eds., pp. 409–422, Elsevier, Amsterdam, The Netherlands, 2001.
[31] R. T. Rockafellar, *Convex Analysis,* Princeton University Press, Princeton, NJ, 1970.
[32] M. V. Solodov and B. F. Svaiter, *An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions,* Math. Oper. Res., 25 (2000), pp. 214–230.