PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 131, Number 12, Pages 3757–3766
S 0002-9939(03)07050-3
Article electronically published on April 24, 2003

CONSTRUCTION OF BEST BREGMAN APPROXIMATIONS IN REFLEXIVE BANACH SPACES

HEINZ H. BAUSCHKE AND PATRICK L. COMBETTES

(Communicated by Jonathan M. Borwein)

ABSTRACT. An iterative method is proposed to construct the Bregman projection of a point onto a countable intersection of closed convex sets in a reflexive Banach space.

1. PROBLEM STATEMENT

Let $(\mathcal{X}, \|\cdot\|)$ be a reflexive real Banach space with dual $(\mathcal{X}^*, \|\cdot\|_*)$ and let $f\colon \mathcal{X} \to]-\infty, +\infty]$ be a lower semicontinuous (l.s.c.) convex function which is Gâteaux differentiable on $\operatorname{int}\operatorname{dom} f \neq \varnothing$ and Legendre [1, Def. 5.2], i.e., it satisfies the following two properties:

(i) ∂f is both locally bounded and single-valued on its domain (essential smoothness);
(ii) $(\partial f)^{-1}$ is locally bounded on its domain and f is strictly convex on every convex subset of $\operatorname{dom}\partial f$ (essential strict convexity).

The Bregman distance associated with f is

$$D\colon \mathcal{X}\times\mathcal{X} \to [0,+\infty],$$
$$(x,y) \mapsto \begin{cases} f(x) - f(y) - \langle x-y, \nabla f(y)\rangle, & \text{if } y \in \operatorname{int}\operatorname{dom} f, \\ +\infty, & \text{otherwise.} \end{cases} \tag{1.1}$$

Let x_0 be a point in $\mathcal{X}$ and $(S_i)_{i\in I}$ a countable family of closed and convex subsets of $\mathcal{X}$ such that

$$x_0 \in \operatorname{int}\operatorname{dom} f, \quad (\operatorname{int}\operatorname{dom} f) \cap \bigcap_{i\in I} S_i \neq \varnothing, \text{ and } S = \overline{\operatorname{dom}}\, f \cap \bigcap_{i\in I} S_i. \tag{1.2}$$

The goal of this paper is to present a method for finding the best Bregman approximation (best D-approximation for short) to x_0 from S, i.e., for solving the problem

$$\text{find } \overline{x} \in S \text{ such that } (\forall x \in S)\ D(\overline{x}, x_0) \leq D(x, x_0). \tag{1.3}$$

It follows from [1, Cor. 7.9] that (1.3) possesses a unique solution, which is called the D-projection of x_0 onto S and is denoted by $P_S x_0$. Problem (1.3) is central in

Received by the editors June 28, 2002.
2000 *Mathematics Subject Classification.* Primary 41A65, 90C25; Secondary 41A29, 41A50.
Key words and phrases. Best approximation, Bregman distance, decomposition, Haugazeau.
The first author was supported by the Natural Sciences and Engineering Research Council of Canada.

many areas of mathematics and applied sciences. For instance, if $\mathcal{X}$ is Hilbertian and $f = \|\cdot\|^2/2$, (1.3) is a metric best approximation problem [14]; if $\mathcal{X} = \mathbb{R}^N$ and f is the negative entropy, then (1.3) is a best Kullback-Leibler approximation (in particular, maximum entropy) problem [12].

Existing methods for solving (1.3) [4, 7, 8, 11] are based on Dykstra's algorithm [6, 14] for the metric best approximation problems in Hilbert spaces. They are limited to scenarios in which $\mathcal{X}$ is a Euclidean space and I is finite, and require in addition the storage of an auxiliary vector for each set S_i at every iteration. Furthermore, the methods of [4, 7, 11] demand the ability to perform D-projections onto each S_i.

The method presented in this paper extends a fixed point algorithmic framework proposed in [3] from metric best approximation in Hilbert spaces to Bregman best approximation in reflexive Banach spaces. An iteration of our method requires only a step towards one of the sets in $(S_i)_{i\in I}$, followed by the D-projection of x_0 onto the intersection of two half-spaces. The step towards the set is implemented by applying a $\mathfrak{B}$-class operator to the current iterate. The main result shows the strong convergence of this algorithm to the solution of (1.3). The method owes its flexibility and generality to the large pool of operators available in the $\mathfrak{B}$-class.

The remainder of the paper is organized as follows. Section 2 contains preliminary results. A conceptual fixed point method for finding the D-projection onto a single closed convex set is developed in Section 3; it serves as the core of the main algorithm for solving (1.3). In Section 4, the main algorithm is developed and analyzed. Applications to various best D-approximation problems are described in Section 5.

Throughout, the symbols $\rightharpoonup$ and $\to$ stand for weak and strong convergence, respectively, and $\mathfrak{W}(x_n)_{n\in\mathbb{N}}$ denotes the set of weak cluster points of a sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{X}$. The domain, graph, range, and fixed point set of a set-valued operator A are denoted by $\operatorname{dom} A$ (with closure $\overline{\operatorname{dom} A}$), $\operatorname{gr} A$, $\operatorname{ran} A$, and $\operatorname{Fix} A$ (with closure $\overline{\operatorname{Fix}}\, A$), respectively. The domain of a function $g\colon \mathcal{X} \to \left]-\infty,+\infty\right]$ is $\operatorname{dom} g = \{x \in \mathcal{X} \mid g(x) < +\infty\}$, its lower level set at height $\eta \in \mathbb{R}$ is $\operatorname{lev}_{\leq \eta} g = \{x \in \mathcal{X} \mid g(x) \leq \eta\}$, and its subdifferential at $x \in \mathcal{X}$ is $\partial g(x) = \big\{x^* \in \mathcal{X}^* \mid (\forall y \in \mathcal{X})\ \ \langle y - x, x^*\rangle + g(x) \leq g(y)\big\}$.

2. Preliminaries

2.1. **D-convergence.**

Definition 2.1. A sequence $(x_n)_{n\in\mathbb{N}}$ in $\operatorname{dom} f$ *D-converges* to $y \in \operatorname{int}\operatorname{dom} f$ if $D(x_n, y) \to 0$, in symbols $x_n \overset{D}{\to} y$.

The following proposition clarifies the relationships between weak, strong, and D-convergence.

Proposition 2.2. *Let x be a point in* $\operatorname{int}\operatorname{dom} f$ *and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in* $\operatorname{dom} f$. *Then:*

(i) $x_n \to x \Rightarrow x_n \overset{D}{\to} x$.
(ii) $x_n \overset{D}{\to} x \Leftrightarrow \big(x_n \rightharpoonup x$ *and* $f(x_n) \to f(x)\big)$.
(iii) *If* $\dim \mathcal{X} < +\infty$, *then* $x_n \to x \Leftrightarrow x_n \overset{D}{\to} x \Leftrightarrow x_n \rightharpoonup x$.

Proof. (i): $x_n \to x$ implies $\langle x_n - x, \nabla f(x)\rangle \to 0$ and, by continuity of f at x, $f(x_n) - f(x) \to 0$. Altogether, $x_n \overset{D}{\to} x$. (ii): Suppose $x_n \overset{D}{\to} x$. It follows from the essential strict convexity of f that $D(\cdot, x)$ is coercive [1, Lemma 7.3(v)]. Hence $(x_n)_{n\in\mathbb{N}}$ is bounded and, in view of the reflexivity of $\mathcal{X}$, $x_n \rightharpoonup x$ will follow from the identity $\mathfrak{W}(x_n)_{n\in\mathbb{N}} = \{x\}$. To show this, take $y \in \mathfrak{W}(x_n)_{n\in\mathbb{N}}$, say, $x_{k_n} \rightharpoonup y$. Since f is weak l.s.c., so is $D(\cdot, x)$ and therefore $0 \le D(y, x) \le \underline{\lim}\, D(x_{k_n}, x) = 0$. Taking into account the fact that f is essentially strictly convex, we get $y = x$ [1, Lemma 7.3(vi)]. Therefore $x_n \rightharpoonup x$ and, in turn, $f(x_n) - f(x) = D(x_n, x) + \langle x_n - x, \nabla f(x)\rangle \to 0$. The reverse implication is clear. (iii) follows from (i) and (ii) since $x_n \rightharpoonup x \Rightarrow x_n \to x$ if $\dim \mathcal{X} < +\infty$. $\square$

Remark 2.3. It follows from [20, Prop. 2.2] that the implication $x_n \overset{D}{\to} x \Rightarrow x_n \to x$ holds when f is totally convex at x, i.e., [9],

$$(\forall t \in\ ]0, +\infty[)\quad \inf\big\{D(u, x) \mid u \in \operatorname{dom} f \ \text{ and } \ \|u - x\| = t\big\} > 0.$$

Remark 2.4. Let $f = \|\cdot\|^2/2$. If $\mathcal{X}$ is Hilbertian, Proposition 2.2(ii) is the well-known Kadec-Klee property [15]. In general, f is Legendre if and only if $\mathcal{X}$ is strictly convex (rotund) and Gâteaux smooth [1, Lemma 6.2(iii)]. It follows from Proposition 2.2(ii) that D- and strong convergence coincide in this case if and only $\mathcal{X}$ has the Kadec-Klee property.

The following example shows that $\mathcal{X}$ can be endowed with an equivalent norm $|||\cdot|||$ so that $f = |||\cdot|||^2/2$ is Legendre while D- and strong convergence do not coincide. Moreover, this function f is apparently the first example of a Legendre function that has full domain but fails to be everywhere totally convex (see [20] for further information).

Example 2.5 (Vanderwerff [23])**.** There exists an equivalent norm $|||\cdot|||$ on $\mathcal{X}$ such that $(\mathcal{X}, |||\cdot|||)$ is strictly convex, Gâteaux smooth, but fails to have the Kadec-Klee property.

Proof. It follows from [5, Corollary 1] that there exists a norm $\|\cdot\|_{1,*}$ on $\mathcal{X}^*$ which is equivalent to $\|\cdot\|_*$, Gâteaux differentiable on $\mathcal{X}^* \smallsetminus \{0\}$, and not Fréchet differentiable at some $x_0^* \in \mathcal{X}^* \smallsetminus \{0\}$. Furthermore, there exists a norm $\|\cdot\|_{2,*}$ on $\mathcal{X}^*$ which is equivalent to $\|\cdot\|_*$ and such that $(\mathcal{X}^*, \|\cdot\|_{2,*})$ is both strictly convex and Gâteaux smooth (see [15, Theorem VII.2.7] for a much more general result). Now set $|||\cdot|||_* = \big(\|\cdot\|_{1,*} + \|\cdot\|_{2,*}\big)/\big(\|x_0^*\|_{1,*} + \|x_0^*\|_{2,*}\big)$. Then the norm $|||\cdot|||_*$ is equivalent to $\|\cdot\|_*$ on $\mathcal{X}^*$. Also, $(\mathcal{X}^*, |||\cdot|||_*)$ is strictly convex and Gâteaux smooth. Now let $|||\cdot|||$ be the dual norm of $|||\cdot|||_*$ on $\mathcal{X}^{**} = \mathcal{X}$. Then $|||\cdot|||$ is equivalent to $\|\cdot\|$ on $\mathcal{X}$ and $(\mathcal{X}, |||\cdot|||)$ is both strictly convex and Gâteaux smooth (see, e.g., [15, Proposition II.1.6]). However, since $\|\cdot\|_{1,*}$ is not Fréchet differentiable at x_0^*, neither is $|||\cdot|||_*$. Consequently, since $|||x_0^*|||_* = 1$, [15, Theorem I.1.4(ii)] implies the existence of sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ in $\mathcal{X}$ and of a number $\varepsilon > 0$ such that

$$\begin{cases} \langle x_n, x_0^*\rangle \to 1, \\ \langle y_n, x_0^*\rangle \to 1, \end{cases} \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad \begin{cases} |||x_n||| = |||y_n||| = 1, \\ |||x_n - y_n||| \ge \varepsilon. \end{cases} \tag{2.1}$$

By reflexivity of $\mathcal{X}$, we further assume that $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ converge weakly, say to x and y, respectively. By weak lower semicontinuity of $|||\cdot|||$, (2.1) yields

$1 = \langle x, x_0^* \rangle \leq |||x||| \cdot |||x_0^*|||_* = |||x||| \leq \underline{\lim} \, |||x_n||| = 1$, whence $|||x|||^2 = \langle x, x_0^* \rangle = |||x_0^*|||_*^2$. Because $(\mathcal{X}^*, ||| \cdot |||_*)$ is Gâteaux smooth, it follows from [25, Theorem 47.19(1)] that $J^* x_0^* = \{x\}$, where J^* denotes the normalized duality map of $(\mathcal{X}^*, ||| \cdot |||_*)$. Likewise, $J^* x_0^* = \{y\}$, whence $y = x$. In summary,

$$\begin{cases} x_n \rightharpoonup x, \\ y_n \rightharpoonup x, \end{cases} \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad \begin{cases} |||x_n||| = |||y_n||| = |||x|||, \\ |||x_n - y_n||| \geq \varepsilon. \end{cases} \tag{2.2}$$

Thus, if $(\mathcal{X}, |||| \cdot |||)$ had the Kadec-Klee property, we would have $x_n \to x$ and $y_n \to x$, whence $x_n - y_n \to 0$ in contradiction to the inequality $\inf_{n \in \mathbb{N}} |||x_n - y_n||| \geq \varepsilon$. $\square$

2.2. **$\mathfrak{B}$-class.** For every x and u in $\operatorname{int} \operatorname{dom} f$, set

$$H(x, u) = \big\{ y \in \mathcal{X} \mid \langle y - u, \nabla f(x) - \nabla f(u) \rangle \leq 0 \big\}.$$

Definition 2.6 ([2])**.** An operator $T \colon \mathcal{X} \to 2^{\mathcal{X}}$ belongs to $\mathfrak{B}$ if $\operatorname{ran} T \subset \operatorname{dom} T = \operatorname{int} \operatorname{dom} f$ and $(\forall (x, u) \in \operatorname{gr} T)$ $\operatorname{Fix} T \subset H(x, u)$.

The operators in this class have properties that are crucial to the convergence analysis of our method. The common types of operators encountered in numerical methods based on Bregman distances are also found in this class. The following proposition supplies some examples; it also introduces key notation and definitions.

Proposition 2.7 ([2, Section 3])**.** *In each of the following cases, the operator $T \colon \mathcal{X} \to 2^{\mathcal{X}}$ belongs to $\mathfrak{B}$:*

(i) *C is a closed convex subset of $\mathcal{X}$ such that $C \cap \operatorname{int} \operatorname{dom} f \neq \varnothing$. T is the* D-projector *onto C, i.e., $T = P_C$, where for every $x \in \operatorname{int} \operatorname{dom} f$, $P_C x = \operatorname{argmin} D(C, x)$ or, equivalently,*

$$P_C x \in C \cap \operatorname{int} \operatorname{dom} f \quad \text{and} \quad C \subset H(x, P_C x). \tag{2.3}$$

In this case, T is single-valued and $\operatorname{Fix} T = C \cap \operatorname{int} \operatorname{dom} f$.

(ii) *$g \colon \mathcal{X} \to \left]-\infty, +\infty\right]$ is an l.s.c. convex function such that $\operatorname{lev}_{\leq 0} g \cap \operatorname{int} \operatorname{dom} f \neq \varnothing$ and $\operatorname{dom} f \subset \operatorname{dom} g$. For every $x \in \operatorname{int} \operatorname{dom} f$ and $x^* \in \partial g(x)$, set $G(x, x^*) = \big\{ y \in \mathcal{X} \mid \langle x - y, x^* \rangle \geq g(x) \big\}$. T is the* subgradient D-projector *onto $\operatorname{lev}_{\leq 0} g$, i.e., for every $x \in \operatorname{int} \operatorname{dom} f$,*

$$Tx = \big\{ P_{G(x, x^*)} x \mid x^* \in \partial g(x) \big\}.$$

In this case, $\operatorname{Fix} T = \operatorname{lev}_{\leq 0} g \cap \operatorname{int} \operatorname{dom} f$.

(iii) *$A \colon \mathcal{X} \to 2^{\mathcal{X}^*}$ is a maximal monotone operator such that $0 \in \operatorname{ran} A$ and $\operatorname{dom} A \subset \operatorname{int} \operatorname{dom} f$. T is the* D-resolvent *of A of index $\gamma \in \left]0, +\infty\right[$, i.e., $T = (\nabla f + \gamma A)^{-1} \circ \nabla f$. In this case, T is single-valued and* $\operatorname{Fix} T = A^{-1} 0 \cap \operatorname{int} \operatorname{dom} f$.

3. D-projection à la Haugazeau

We develop a conceptual fixed point method for finding the D-projection of $x_0 \in \operatorname{int} \operatorname{dom} f$ onto a closed convex set $C \subset \mathcal{X}$ in the spirit of a method initially proposed by Haugazeau for metric projections in Hilbert spaces [18] and further studied in this context in [3, 13, 16, 19, 22].

Given a triple (x, y, z) in $(\operatorname{int} \operatorname{dom} f)^3$ such that $H(x, y) \cap H(y, z) \cap \operatorname{int} \operatorname{dom} f \neq \varnothing$, the D-projection of x_0 onto $H(x, y) \cap H(y, z)$ is a well-defined point in $\operatorname{int} \operatorname{dom} f$ by [1, Cor. 7.9]. We denote this point by $Q(x, y, z)$.

Algorithm 3.1. *At every iteration $n \in \mathbb{N}$, select $T_n \in \mathfrak{B}$, $u_n \in T_n x_n$, and set $x_{n+1} = Q(x_0, x_n, u_n)$.*

Condition 3.2. $C \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$, $C \subset \bigcap_{n\in\mathbb{N}} \overline{\operatorname{Fix}}\, T_n$, and $\bigcap_{n\in\mathbb{N}} \operatorname{Fix} T_n \neq \varnothing$.

Proposition 3.3 (viability)**.** *Under Condition 3.2, Algorithm 3.1 generates an infinite sequence $(x_n)_{n\in\mathbb{N}}$ in $\operatorname{int}\operatorname{dom} f$.*

Proof. By assumption, $x_0 \in \operatorname{int}\operatorname{dom} f$. Now suppose that, at some iteration $n \in \mathbb{N}$, $x_n \in \operatorname{int}\operatorname{dom} f$. Since $T_n \in \mathfrak{B}$, $u_n \in \operatorname{int}\operatorname{dom} f$ and $E_n = H(x_0, x_n) \cap H(x_n, u_n)$ is well-defined. In view of [1, Cor. 7.9], it suffices to show that $E_n \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$ to guarantee that $x_{n+1} = P_{E_n} x_0$ is a well-defined point in $\operatorname{int}\operatorname{dom} f$.

Since by Condition 3.2 $C \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$, we shall actually show that $C \subset \bigcap_{n\in\mathbb{N}} E_n$. Because Condition 3.2 holds and $(T_n)_{n\in\mathbb{N}}$ lies in $\mathfrak{B}$, we have

$$C \subset \bigcap_{n\in\mathbb{N}} \overline{\operatorname{Fix}}\, T_n \subset \bigcap_{n\in\mathbb{N}} H(x_n, u_n). \tag{3.1}$$

Consequently, it remains to show that, for every $n \in \mathbb{N}$, $C \subset H(x_0, x_n)$. Let us proceed by induction. For $n = 0$, it is clear that $C \subset H(x_0, x_0) = \mathcal{X}$. Furthermore, for every $n \in \mathbb{N}$, it results from (3.1) and (2.3) that

$$C \subset H(x_0, x_n) \Rightarrow C \subset E_n \Rightarrow C \subset H(x_0, P_{E_n} x_0) = H(x_0, x_{n+1}), \tag{3.2}$$

which completes the proof. ☐

Some basic properties of Algorithm 3.1 can now be collected.

Proposition 3.4. *Let $(x_n)_{n\in\mathbb{N}}$ be an arbitrary orbit of Algorithm 3.1 generated under Condition 3.2. Then:*

(i) $(\forall n \in \mathbb{N})$ $D(x_n, x_0) \leq D(x_{n+1}, x_0) \leq D(P_C x_0, x_0)$.
(ii) $(x_n)_{n\in\mathbb{N}}$ *is bounded.*
(iii) $\big(D(x_n, x_0)\big)_{n\in\mathbb{N}}$ *converges and* $\lim D(x_n, x_0) \leq D(P_C x_0, x_0)$.
(iv) $(\forall n \in \mathbb{N})$ $x_n \in C \quad\Leftrightarrow\quad x_n = P_C x_0$.
(v) $x_n \overset{D}{\to} P_C x_0 \Leftrightarrow \mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset C$.
(vi) $\sum_{n\in\mathbb{N}} D(x_{n+1}, x_n) < +\infty$.
(vii) $\sum_{n\in\mathbb{N}} D(u_n, x_n) < +\infty$.

Proof. (i): By (2.3), for every $n \in \mathbb{N}$, x_n is the D-projection of x_0 onto $H(x_0, x_n)$. Hence, the first inequality follows from the inclusion $x_{n+1} \in H(x_0, x_n)$ and the second from the inclusions $P_C x_0 \in C \subset H(x_0, x_{n+1})$ (see (3.2)). (ii): Since $D(\cdot, x_0)$ is coercive by [1, Lemma 7.3(v)], it results from (i) that $(x_n)_{n\in\mathbb{N}}$ is bounded. (iii) and (iv) follow from (i). (v): The forward implication follows from Proposition 2.2(ii). For the reverse implication, assume $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset C$ and fix $x \in \mathfrak{W}(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \rightharpoonup x$ (the existence of x follows from (ii) and the reflexivity of $\mathcal{X}$). It results from the weak lower semicontinuity of f and (iii) that

$$D(x, x_0) \leq \underline{\lim}\, D(x_{k_n}, x_0) = \lim D(x_n, x_0) \leq D(P_C x_0, x_0). \tag{3.3}$$

Consequently, since $x \in C$, $x = P_C x_0$ and, in turn, $\mathfrak{W}(x_n)_{n\in\mathbb{N}} = \{P_C x_0\}$. Next, since $(x_n)_{n\in\mathbb{N}}$ is bounded, we obtain $x_n \rightharpoonup P_C x_0$. Since (3.3) yields

$$D(P_C x_0, x_0) \leq \lim D(x_n, x_0) \leq D(P_C x_0, x_0), \tag{3.4}$$

we have $D(x_n, x_0) \to D(P_C x_0, x_0)$ and, as a result, $x_n \overset{D}{\to} P_C x_0$. (vi): It follows easily from (1.1) that, for every $u \in \mathcal{X}$ and every $(y, z) \in (\operatorname{int} \operatorname{dom} f)^2$,

$$D(u, y) = D(u, z) + D(z, y) + \langle u - z, \nabla f(z) - \nabla f(y) \rangle . \tag{3.5}$$

For every $n \in \mathbb{N}$, this identity and the inclusion $x_{n+1} \in H(x_0, x_n)$ imply

$$\begin{aligned} D(x_{n+1}, x_0) - D(x_n, x_0) &= D(x_{n+1}, x_n) + \langle x_{n+1} - x_n, \nabla f(x_n) - \nabla f(x_0) \rangle \\ &\geq D(x_{n+1}, x_n). \end{aligned}$$

Hence, $\sum_{n\in\mathbb{N}} D(x_{n+1}, x_n) \leq D(P_C x_0, x_0)$ by (i). (vii): For every $n \in \mathbb{N}$, (3.5) and the inclusion $x_{n+1} \in H(x_n, u_n)$ yield

$$\begin{aligned} D(x_{n+1}, x_n) &= D(x_{n+1}, u_n) + D(u_n, x_n) - \langle x_{n+1} - u_n, \nabla f(x_n) - \nabla f(u_n) \rangle \\ &\geq D(u_n, x_n). \end{aligned}$$

In view of (vi), we conclude $\sum_{n\in\mathbb{N}} D(u_n, x_n) < +\infty$. □

It will be convenient to repackage the main convergence properties of Algorithm 3.1 as follows.

Condition 3.5. For every orbit $(x_n)_{n\in\mathbb{N}}$ of Algorithm 3.1 one has

$$\begin{cases} \sum_{n\in\mathbb{N}} D(x_{n+1}, x_n) < +\infty, \\ \sum_{n\in\mathbb{N}} D(u_n, x_n) < +\infty \end{cases} \quad\Rightarrow\quad \mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset C.$$

If $\mathcal{X}$ is Hilbertian and $f = \|\cdot\|^2/2$, the theorem below is [3, Thm. 4.2(ii)(a)].

Theorem 3.6. *Let $(x_n)_{n\in\mathbb{N}}$ be an arbitrary orbit of Algorithm 3.1 generated under Conditions 3.2 and 3.5. Then $x_n \overset{D}{\to} P_C x_0$.*

Proof. Apply Proposition 3.4(v)–(vii). □

4. Main result

In order to solve (1.3), we implement Algorithm 3.1 in the following sequential format (since the $\mathfrak{B}$-class is closed under certain averaging operations [2], it is also possible to devise parallel block-iterative implementations along the lines of those described in [2] for feasibility problems at the expense of more technical arguments).

Algorithm 4.1. *At every iteration $n \in \mathbb{N}$, take $\mathrm{i}(n) \in I$ and $T_n \in \mathfrak{B}$ such that $S_{\mathrm{i}(n)} \cap \operatorname{int} \operatorname{dom} f \subset \operatorname{Fix} T_n$. Then select $u_n \in T_n x_n$ and set $x_{n+1} = Q(x_0, x_n, u_n)$.*

Remark 4.2. At iteration n, the selection of u_n amounts to taking a step towards $S_{\mathrm{i}(n)}$. Indeed, since $T_n \in \mathfrak{B}$ and $S_{\mathrm{i}(n)} \cap \operatorname{int} \operatorname{dom} f \subset \operatorname{Fix} T_n$, [2, Prop. 3.3] yields

$$(\forall y \in S_{\mathrm{i}(n)}) \ \ D(y, u_n) \leq D(y, x_n). \tag{4.1}$$

The update $x_{n+1} = Q(x_0, x_n, u_n)$ is then obtained as the minimizer of $f - \nabla f(x_0)$ over the intersection of the two half-spaces $H(x_0, x_n)$ and $H(x_n, u_n)$, which is a standard convex optimization problem.

Condition 4.3.

(i) The index control mapping $\mathrm{i}\colon \mathbb{N} \to I$ satisfies

$$(\forall i \in I)(\exists\, M_i > 0)(\forall n \in \mathbb{N}) \ \ i \in \{\mathrm{i}(n), \ldots, \mathrm{i}(n + M_i - 1)\}.$$

(ii) For every sequence $(y_n)_{n\in\mathbb{N}}$ in $\operatorname{int}\operatorname{dom} f$ and every bounded sequence $(z_n)_{n\in\mathbb{N}}$ in $\operatorname{int}\operatorname{dom} f$, one has

$$D(y_n, z_n) \to 0 \quad \Rightarrow \quad y_n - z_n \to 0.$$

Condition 4.4. For every orbit $(x_n)_{n\in\mathbb{N}}$ of Algorithm 4.1, every strictly increasing sequence $(p_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$, and every index $i \in I$, one has

$$\begin{cases} (\forall n \in \mathbb{N}) \;\; i = \mathrm{i}(p_n), \\ x_{p_n} \rightharpoonup x, \\ x_{n+1} - x_n \to 0, \\ u_n - x_n \to 0 \end{cases} \quad \Rightarrow \quad x \in S_i. \tag{4.2}$$

Remark 4.5. Condition 4.3(i) states that each set S_i must be activated at least once within any M_i consecutive iterations. Condition 4.3(ii) holds when f is uniformly convex on bounded sets [10, Section 4] (for examples, see [24]). Finally, concrete examples in which Condition 4.4 holds will be described in Section 5.

Lemma 4.6 ([2, Lemma 3.2])**.** *Let C_1 and C_2 be two convex subsets of $\mathcal{X}$ such that C_1 is closed and $C_1 \cap \operatorname{int} C_2 \neq \varnothing$. Then $\overline{C_1 \cap \operatorname{int} C_2} = C_1 \cap \overline{C_2}$.*

Our main result states that every orbit of Algorithm 4.1 converges strongly to the solution of (1.3).

Theorem 4.7. *Let $(x_n)_{n\in\mathbb{N}}$ be an arbitrary orbit of Algorithm 4.1 generated under Conditions 4.3 and 4.4. Then $x_n \to P_S x_0$.*

Proof. Since Algorithm 4.1 is a special case of Algorithm 3.1, we shall apply Theorem 3.6 to $C = S$. Let us first verify Condition 3.2. Assumption (1.2) gives

(4.3)
$$(\forall n \in \mathbb{N}) \;\; \varnothing \neq (\operatorname{int}\operatorname{dom} f) \cap \bigcap_{i\in I} S_i = (\operatorname{int}\operatorname{dom} f) \cap S \subset (\operatorname{int}\operatorname{dom} f) \cap S_{\mathrm{i}(n)} \subset \operatorname{Fix} T_n.$$

Hence $\bigcap_{n\in\mathbb{N}} \operatorname{Fix} T_n \neq \varnothing$. Next, we derive from (1.2), Lemma 4.6, and (4.3) that

$$(\forall n \in \mathbb{N}) \;\; S = \overline{\operatorname{dom}}\, f \cap \bigcap_{i\in I} S_i \subset \overline{\operatorname{Fix}}\, T_n. \tag{4.4}$$

Consequently, $S \subset \bigcap_{n\in\mathbb{N}} \overline{\operatorname{Fix}}\, T_n$. Altogether, Condition 3.2 holds.

Next, we verify Condition 3.5. To this end, fix $i \in I$ and $x \in \mathfrak{W}(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \rightharpoonup x$. Because $x \in \overline{\operatorname{dom}}\, f$, it is sufficient to show

$$\begin{cases} \sum_{n\in\mathbb{N}} D(x_{n+1}, x_n) < +\infty, \\ \sum_{n\in\mathbb{N}} D(u_n, x_n) < +\infty \end{cases} \quad \Rightarrow \quad x \in S_i. \tag{4.5}$$

In view of Proposition 3.4(ii) and Condition 4.3(ii), it is actually enough to show

$$\begin{cases} x_{n+1} - x_n \to 0, \\ u_n - x_n \to 0 \end{cases} \quad \Rightarrow \quad x \in S_i. \tag{4.6}$$

Let M_i be as in Condition 4.3(i). After passing to a subsequence of $(x_{k_n})_{n\in\mathbb{N}}$ if necessary, we assume that, for every $n \in \mathbb{N}$, $k_{n+1} \geq k_n + M_i$. Accordingly, there exists a subsequence $(x_{p_n})_{n\in\mathbb{N}}$ of $(x_n)_{n\in\mathbb{N}}$ such that

$$(\forall n \in \mathbb{N}) \;\; k_n \leq p_n \leq k_n + M_i - 1 < k_{n+1} \leq p_{n+1} \quad \text{and} \quad i = \mathrm{i}(p_n). \tag{4.7}$$

Furthermore,

$$(\forall n \in \mathbb{N}) \ \ \|x_{p_n} - x_{k_n}\| \leq \sum_{l=k_n}^{k_n+M_i-2} \|x_{l+1} - x_l\| \leq (M_i - 1) \max_{k_n \leq l \leq k_n+M_i-2} \|x_{l+1} - x_l\|.$$

Consequently, if $x_{n+1} - x_n \to 0$, then $x_{p_n} - x_{k_n} \to 0$ and, in turn, $x_{p_n} \rightharpoonup x$. If, in addition, $u_n - x_n \to 0$, then Condition 4.4 yields $x \in S_i$. Thus, (4.6) holds true.

We can now apply Theorem 3.6 to get $D(x_n, P_S x_0) \to 0$. In turn, Condition 4.3(ii) yields $x_n \to P_S x_0$. $\square$

5. Applications

The versatility of Algorithm 4.1 is illustrated through its application to three specific versions of the best D-approximation problem (1.2)–(1.3).

5.1. **Best D-approximation via D-projections.** For every $i \in I$, let P_i be the D-projector onto the set S_i. By Proposition 2.7(i), P_i is a single-valued operator in $\mathfrak{B}$ with $\operatorname{Fix} P_i = S_i \cap \operatorname{int}\operatorname{dom} f$ and we can implement Algorithm 4.1 as

Algorithm 5.1. *For every $n \in \mathbb{N}$, take $\mathrm{i}(n) \in I$ and set $x_{n+1} = Q\big(x_0, x_n, P_{\mathrm{i}(n)} x_n\big)$.*

Corollary 5.2. *Let $(x_n)_{n\in\mathbb{N}}$ be an arbitrary orbit of Algorithm 5.1 generated under Condition 4.3. Then $x_n \to P_S x_0$.*

Proof. In view of Theorem 4.7, it is enough to check that Condition 4.4 is satisfied. To this end, take $i \in I$ and a suborbit $(x_{p_n})_{n\in\mathbb{N}}$ such that $\mathrm{i}(p_n) \equiv i$, $x_{p_n} \rightharpoonup x$, and $P_i x_{p_n} - x_{p_n} \to 0$. Then $S_i \ni P_i x_{p_n} \rightharpoonup x$ and, since S_i is weakly closed, $x \in S_i$. $\square$

If $\mathcal{X}$ is Hilbertian, $f = \|\cdot\|^2/2$, $I = \{1, \ldots, m\}$ is finite, and $\mathrm{i}\colon n \mapsto n (\operatorname{mod} m) + 1$, then Algorithm 5.1 is Haugazeau's original best approximation method and Corollary 5.2 relapses to [18, Thm. 3-2].

5.2. **Best D-approximation from convex inequalities.** For every $i \in I$, let $S_i = \operatorname{lev}_{\leq 0} g_i$, where $g_i\colon \mathcal{X} \to \,]-\infty, +\infty]$ is an l.s.c. convex function such that ∂g_i maps bounded sets to bounded sets and $\operatorname{dom} f \subset \operatorname{dom} g_i$, and let R_i be the subgradient D-projector onto S_i. By Proposition 2.7(ii), $R_i \in \mathfrak{B}$ and $\operatorname{Fix} R_i = \operatorname{lev}_{\leq 0} g_i \cap \operatorname{int}\operatorname{dom} f$ ($\neq \varnothing$ by (1.2)), and we can implement Algorithm 4.1 as

Algorithm 5.3. *For every $n \in \mathbb{N}$, take $\mathrm{i}(n) \in I$, $u_n \in R_i x_n$, and set $x_{n+1} = Q\big(x_0, x_n, u_n\big)$.*

Corollary 5.4. *Let $(x_n)_{n\in\mathbb{N}}$ be an arbitrary orbit of Algorithm 5.3 generated under Condition 4.3. Then $x_n \to P_S x_0$.*

Proof. Again, to apply Theorem 4.7, it suffices to check Condition 4.4. Take $i \in I$ and a suborbit $(x_{p_n})_{n\in\mathbb{N}}$ such that $\mathrm{i}(p_n) \equiv i$, $x_{p_n} \rightharpoonup x$, and $u_{p_n} - x_{p_n} \to 0$. For every $n \in \mathbb{N}$, u_{p_n} is the D-projection of x_n onto $G_i(x_{p_n}, x_n^*) = \big\{y \in \mathcal{X} \mid \langle x_{p_n} - y, x_n^* \rangle \geq g_i(x_{p_n})\big\}$ for some $x_n^* \in \partial g_i(x_{p_n})$. Since $u_{i,p_n} \in G_i(x_{p_n}, x_n^*)$, we have

$$\|u_{p_n} - x_{p_n}\| \geq d_{G_i(x_{p_n}, x_n^*)}(x_{p_n}) = \begin{cases} g_i^+(x_{p_n})/\|x_n^*\|_*, & \text{if } x_n^* \neq 0, \\ 0, & \text{otherwise}, \end{cases} \tag{5.1}$$

where $d_{G_i(x_{p_n}, x_n^*)}$ is the metric distance function to $G_i(x_{p_n}, x_n^*)$, $g_i^+ = \max\{0, g_i\}$, and the equality follows from [21, Lemma I.1.2]. Since $(x_{p_n})_{n\in\mathbb{N}}$ is bounded by Proposition 3.4(ii), $(x_n^*)_{n\in\mathbb{N}}$ is bounded by assumption. Therefore, $u_{p_n} - x_{p_n} \to 0$

implies $g_i^+(x_{p_n}) \to 0$. However, since g_i^+ is convex and l.s.c., it is weak l.s.c. and thus $g_i^+(x) \le \underline{\lim}\, g_i^+(x_{p_n}) = 0$. We conclude $g_i(x) \le 0$, i.e., $x \in S_i$. $\square$

5.3. **Best D-approximation from zeros of monotone operators.** Suppose $\operatorname{dom} f = \mathcal{X}$ and, for every $i \in I$, let $S_i = A_i^{-1}0$ be the set of zeros of a maximal monotone operator $A_i\colon \mathcal{X} \to 2^{\mathcal{X}^*}$. For every $\gamma \in\,]0,+\infty[$, Proposition 2.7(iii) asserts that the D-resolvent $(\nabla f + \gamma A_i)^{-1} \circ \nabla f$ is a single-valued operator in $\mathfrak{B}$ with fixed point set $A_i^{-1}0$. Accordingly, we can implement Algorithm 4.1 as

Algorithm 5.5. *For every $n \in \mathbb{N}$, take $\mathrm{i}(n) \in I$, $\gamma_n \in\,]0,+\infty[$, and set $x_{n+1} = Q\big(x_0, x_n, (\nabla f + \gamma_n A_{\mathrm{i}(n)})^{-1} \circ \nabla f(x_n)\big)$.*

Remark 5.6. When I is a singleton, Algorithm 5.5 corresponds to the exact version of the algorithm announced in [17].

Corollary 5.7. *Let $(x_n)_{n\in\mathbb{N}}$ be an arbitrary orbit of Algorithm 5.5 generated under Condition 4.3. Suppose ∇f is uniformly continuous on bounded subsets of $\mathcal{X}$ and, for every $i \in I$ and every strictly increasing sequence $(p_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ such that $\mathrm{i}(p_n) \equiv i$, one has $\inf_{n\in\mathbb{N}} \gamma_{p_n} > 0$. Then $x_n \to P_S x_0$.*

Proof. As before, it is enough to check Condition 4.4. Take $i \in I$ and a suborbit $(x_{p_n})_{n\in\mathbb{N}}$ such that $\mathrm{i}(p_n) \equiv i$, $x_{p_n} \rightharpoonup x$, and $u_{p_n} - x_{p_n} \to 0$, where $u_{p_n} = (\nabla f + \gamma_{p_n} A_i)^{-1} \circ \nabla f(x_{p_n})$. Then $u_{p_n} \rightharpoonup x$. Now set, for every $n \in \mathbb{N}$,

$$u_n^* = (\nabla f(x_{p_n}) - \nabla f(u_{p_n}))/\gamma_{p_n}.$$

Then $\big((u_{p_n}, u_n^*)\big)_{n\in\mathbb{N}}$ lies in $\operatorname{gr} A_i$. On the other hand, $u_n^* \to 0$ since $u_{p_n} - x_{p_n} \to 0$, ∇f is uniformly continuous on $\operatorname{int}\operatorname{dom} f$, and $\inf_{n\in\mathbb{N}} \gamma_{p_n} > 0$. Since A_i is maximal monotone, $\operatorname{gr} A_i$ is weakly-strongly closed and must therefore contain $(x, 0)$, i.e., $x \in S_i$. $\square$

Remark 5.8. The function f in Corollary 5.7 is assumed to have full domain because, by [1, Thm. 5.6(v)], every essentially smooth function whose gradient is uniformly continuous on bounded subsets of the interior of its domain has full domain. If $\mathcal{X}$ is strictly convex and uniformly smooth, then $f = \|\cdot\|^2/2$ satisfies the requirements of Corollary 5.7: it is Legendre [1, Lemma 6.2(iii)] and ∇f is uniformly continuous on bounded sets [25, Prop. 47.19(2)(ii)].

Acknowledgement

We thank Jon Vanderwerff for providing Example 2.5, which completely settles a question implicit in the original version of this manuscript.

References

[1] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, *Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces*, Comm. Contemp. Math., 3 (2001), 615–647. MR **2002k:**49040

[2] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, *Bregman monotone optimization algorithms*, SIAM J. Control Optim., to appear.

[3] H. H. Bauschke and P. L. Combettes, *A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces*, Math. Oper. Res., 26 (2001), 248–264.

[4] H. H. Bauschke and A. S. Lewis, *Dykstra's algorithm with Bregman projections: A convergence proof*, Optimization, 48 (2000), 409–427. MR **2002e:**90062

[5] J. M. Borwein and M. Fabian, *On convex functions having points of Gâteaux differentiability which are not points of Fréchet differentiability*, Canad. J. Math., 45 (1993), 1121–1134. MR **94i:**46018

[6] J. P. Boyle and R. L. Dykstra, *A method for finding projections onto the intersection of convex sets in Hilbert spaces*, in Lecture Notes in Statistics, Vol. 37, Springer-Verlag, New York, 1986, 28–47. MR **87m:**49024

[7] L. M. Bregman, Y. Censor, and S. Reich, *Dykstra's algorithm as the nonlinear extension of Bregman's optimization method,* J. Convex Anal., 6 (1999), 319–333. MR **2001e:**90054

[8] L. M. Bregman, Y. Censor, S. Reich, and Y. Zepkowitz-Malachi, *Finding the projection of a point onto the intersection of convex sets via projections onto halfspaces,* preprint, 2002.

[9] D. Butnariu and A. N. Iusem, *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization,* Kluwer, Boston, MA, 2000. MR **2002k:**90002

[10] D. Butnariu, A. Iusem, and C. Zălinescu, *On uniform convexity, total convexity and the convergence of the proximal point and outer Bregman projection algorithms in Banach spaces*, J. Convex Anal., to appear.

[11] Y. Censor and S. Reich, *The Dykstra algorithm with Bregman projections*, Commun. Appl. Anal., 2 (1998), 407–419. MR **99e:**90111

[12] Y. Censor and S. A. Zenios, *Parallel Optimization – Theory, Algorithms and Applications*, Oxford University Press, New York, 1997. MR **2000a:**90002

[13] P. L. Combettes, *Strong convergence of block-iterative outer approximation methods for convex optimization*, SIAM J. Control Optim., 38 (2000), 538–565. MR **2001b:**90064

[14] F. Deutsch, *Best Approximation in Inner Product Spaces,* Springer-Verlag, New York, 2001. MR **2002c:**41001

[15] R. Deville, G. Godefroy, and V. Zizler, *Smoothness and Renormings in Banach Spaces*, Wiley, New York, 1993. MR **94d:**46012

[16] T. Dumont, *Décomposition par Projection de Certains Problèmes aux Limites Elliptiques non Linéaires*, Thèse, Université Claude Bernard, Lyon, France, 1978.

[17] R. Gárciga-Otero, *A strongly convergent hybrid proximal point method in Banach spaces*, conference talk (presented at the IV Brazilian Workshop on Continuous Optimization, Rio de Janeiro, July 15, 2002) reporting on a forthcoming paper with B. F. Svaiter.

[18] Y. Haugazeau, *Sur les Inéquations Variationnelles et la Minimisation de Fonctionnelles Convexes*, Thèse, Université de Paris, Paris, France, 1968.

[19] 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, 200–218.

[20] E. Resmerita, *On total convexity, Bregman projections and stability in Banach spaces*, preprint, 2002.

[21] I. Singer, *Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces,* Springer-Verlag, New York, 1970. MR **42:**4937

[22] M. V. Solodov and B. F. Svaiter, *Forcing strong convergence of proximal point iterations in a Hilbert space,* Math. Programming A, 87 (2000), 189–202. MR **2000j:**90077

[23] J. D. Vanderwerff, personal communication, 2002.

[24] C. Zălinescu, *On uniformly convex functions*, J. Math. Anal. Appl., 95 (1983), 344–374. MR **85a:**26018

[25] E. Zeidler, *Nonlinear Functional Analysis and Its Applications III – Variational Methods and Optimization,* Springer-Verlag, New York, 1985. MR **90b:**49005

Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada N1G 2W1

E-mail address: `hbauschk@uoguelph.ca`

Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie – Paris 6, 75005 Paris, France

E-mail address: `plc@math.jussieu.fr`