CONSTRUCTION OF BEST BREGMAN APPROXIMATIONS IN REFLEXIVE BANACH SPACES

Heinz H. Bauschke* and Patrick L. Combettes†

June 28, 2002

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

(i) $\partial f$ 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 $\text{dom} \; \partial f$ (essential strict convexity).

The Bregman distance associated with $f$ is

$$D: \mathcal{X} \times \mathcal{X} \to [0, +\infty]$$

$$(x, y) \mapsto \begin{cases} f(x) - f(y) - \langle x - y, \nabla f(y) \rangle, & \text{if } y \in \text{int} \; \text{dom} \; f; \\ +\infty, & \text{otherwise.} \end{cases} \quad (1.1)$$

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

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

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

$$\text{find } \overline{x} \in S \; \text{such that} \; (\forall x \in S) \; D(\overline{x}, x_0) \leq D(x, x_0). \quad (1.3)$$
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 $P_Sx_0$. Problem (1.3) is central in many areas of mathematics and applied sciences. For instance, if $\mathcal{X}$ is Hilbertian and $f = ||x||^2/2$, (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 [11].

Existing methods for solving (1.3) [4, 6, 7, 10] are based on Dykstra’s algorithm [5, 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, 6, 10] 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 the 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 $\mathcal{B}$-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 $\mathcal{B}$-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 $\rightarrow$ 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 $\text{dom} A$ (with closure $\overline{\text{dom}} A$), $\text{gr} A$, $\text{ran} A$, and $\text{Fix} A$ (with closure $\overline{\text{Fix}} A$), respectively. The domain of a function $g: \mathcal{X} \rightarrow ]-\infty, +\infty]$ is $\text{dom} g = \{x \in \mathcal{X} \mid g(x) < +\infty\}$, its lower level set at height $\eta \in \mathbb{R}$ is $\text{lev}_{\leq \eta} g = \{x \in \mathcal{X} \mid g(x) \leq \eta\}$, and its subdifferential at $x \in \mathcal{X}$ is $\partial g(x) = \{x^* \in \mathcal{X}^* \mid (\forall y \in \mathcal{X}) \ (y-x,x^*) + g(x) \leq g(y)\}$.

2 Preliminaries

2.1 D-convergence

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

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

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

(i) $x_n \rightharpoonup x \Rightarrow x_n \xrightarrow{D} x$.

(ii) $x_n \xrightarrow{D} x \iff (x_n \rightharpoonup x$ and $f(x_n) \rightarrow f(x))$.

(iii) If $\dim \mathcal{X} < +\infty$, then $x_n \rightharpoonup x \iff x_n \xrightarrow{D} x \iff x_n \rightharpoonup x$. 

2
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 $\mathcal{X}$, $x_n \to x$ will follow from the identity $\mathcal{W}(x_n)_{n \in \mathbb{N}} = \{x\}$. To show this, take $y \in \mathcal{W}(x_n)_{n \in \mathbb{N}}$, say, $x_{k_n} \to y$. Since $f$ is weak l.s.c., so is $D(\cdot, x)$ and therefore $0 \leq D(y, x) \leq \lim D(x_{k_n}, x) = 0$. Taking into account the fact that $f$ is essentially strictly convex, we get $y = x$ [1, Lemma 7.3(vi)]. Therefore $x_n \to x$ 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 $x_n \to x$ implies $x_n \overset{D}{\to} x \Rightarrow x_n \to x$ if dim $\mathcal{X} < +\infty$. \[\]

Remark 2.3 It follows from [18, 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. [8],

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

Remark 2.4 Let $f = \| \cdot \|^2 / 2$. If $\mathcal{X}$ is Hilbertian, Proposition 2.2(ii) is the well-known Kadec-Klee property [13]. In general, Legenderness of $f$ is equivalent to $\mathcal{X}$ being strictly convex (roundup) and smooth [1, Lemma 6.2(iii)]. It follows from Proposition 2.2(ii) that D- and strong convergence coincide if and only $\mathcal{X}$ has the Kadec-Klee property. Thus, for this particular $f$, the existence of a sequence that D-converges but not in norm is equivalent to exhibiting a reflexive, strictly convex smooth Banach space which does not have the Kadec-Klee property.

2.2 $\mathcal{B}$-class

For every $x$ and $u$ in int dom $f$, set

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

Definition 2.5 [2] An operator $T : \mathcal{X} \to 2^{\mathcal{X}}$ belongs to $\mathcal{B}$ if ran $T \subset \text{dom } f$ and ($\forall (x, u) \in \text{gr } T$) $\text{Fix } T \subset H(x, u)$.

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 2.6 [2, Section 3] In each of the following cases, the operator $T : \mathcal{X} \to 2^{\mathcal{X}}$ belongs to $\mathcal{B}$.

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

$$P_C x \in \text{int dom } f \cap C \quad \text{and} \quad C \subset H(x, P_C x).$$ (2.1)

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

(ii) $g : \mathcal{X} \to [\infty, +\infty]$ is an l.s.c. convex function such that $\text{lev}_{\leq 0} g \cap \text{int dom } f \neq \emptyset$ and dom $f \subset \text{dom } g$.

For every $x \in \text{int dom } f$ and $x^* \in \partial g(x)$, set $G(x, x^*) = \{y \in \mathcal{X} \mid \langle x - y, x^* \rangle \geq g(x)\}$. $T$ is the subgradient D-projector onto $\text{lev}_{\leq 0} g$, i.e., for every $x \in \text{int dom } f$,

$$T x = \{P_{G(x, x^*)} x \mid x^* \in \partial g(x)\}.$$\[\]

In this case, Fix $T = \text{lev}_{\leq 0} g \cap \text{int dom } f$. 3
(iii) \( A: X \to 2^X \) is a maximal monotone operator such that \( 0 \in \text{ran } A \) and \( \text{dom } A \subset \text{int dom } f \). \( T \) is the D-resolvent of \( A \) of index \( \gamma \in [0, +\infty) \), i.e., \( T = (\nabla f + \gamma A)^{-1} \circ \nabla f \). In this case, \( T \) is single-valued and \( \text{Fix } T = A^{-1}0 \cap \text{int dom } f \).

3 D-projection à la Haugazeau

We develop a conceptual fixed point method for finding the D-projection of \( x_0 \in \text{int dom } f \) onto a closed convex set \( C \subset X \) in the spirit of a method initially proposed by Haugazeau for metric projections in Hilbert spaces [16] and further studied in this context in [3, 12, 15, 17].

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

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

**Condition 3.2** \( C \cap \text{int dom } f \neq \emptyset \), \( C \subset \bigcap_{n \in \mathbb{N}} \text{Fix } T_n \), and \( \bigcap_{n \in \mathbb{N}} \text{Fix } T_n \neq \emptyset \).

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

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

Since by Condition 3.2 \( C \cap \text{int dom } f \neq \emptyset \), 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 \( \mathcal{B} \), we have

\[ C \subset \bigcap_{n \in \mathbb{N}} \text{Fix } T_n \subset \bigcap_{n \in \mathbb{N}} H(x_n, u_n). \]  

(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 \( C \subset H(x_0, x_0) = X \). Furthermore, for every \( n \in \mathbb{N} \), it results from (3.1) and (2.1) that

\[ C \subset H(x_0, x_n) \Rightarrow C \subset E_n \Rightarrow C \subset H(x_0, P_{E_n} x_0) = H(x_0, x_{n+1}), \]  

(3.2)

which completes the proof. \( \Box \)

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:

1. \((\forall n \in \mathbb{N}) \) \( D(x_n, x_0) \leq D(x_{n+1}, x_0) \leq D(P_C x_0, x_0) \).
2. \((x_n)_{n \in \mathbb{N}} \) is bounded.
3. \((D(x_n, x_0))_{n \in \mathbb{N}} \) converges and \( \lim_{n \to \infty} D(x_n, x_0) \leq D(P_C x_0, x_0) \).
4. \((\forall n \in \mathbb{N}) \) \( x_n \in C \iff x_n = P_C x_0 \).
5. \( x_n \overset{D}{\rightarrow} P_C x_0 \iff \forall (x_n)_{n \in \mathbb{N}} \subset C \).
(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 (2.1), 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 \( \mathcal{M}(x_n)_{n \in \mathbb{N}} \subset C \) and fix \( x \in \mathcal{M}(x_n)_{n \in \mathbb{N}}, \) say \( x_k \to x \) (the existence of \( x \) follows from (ii) and the reflexivity of \( X \)). It results from the weak lower semicontinuity of \( f \) and (iii) that

\[
D(x, x_0) \leq \lim D(x_k, x_0) = \lim D(x_n, x_0) \leq D(P_C x_0, x_0).
\]

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

\[
D(P_C x_0, x_0) \leq \lim D(x_n, x_0) \leq D(P_C x_0, x_0),
\]

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

\[
D(u, y) = D(u, z) + D(z, y) + \langle u - z, \nabla f(z) - \nabla f(y) \rangle.
\]

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

\[
D(x_{n+1}, x_0) - D(x_n, x_0) = D(x_{n+1}, x_n) + \langle x_{n+1} - x_n, \nabla f(x_n) - \nabla f(x_0) \rangle \\
\geq D(x_{n+1}, x_n).
\]

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

\[
D(x_{n+1}, x_n) = D(x_{n+1}, u_n) + D(u_n, x_n) - \langle x_{n+1} - u_n, \nabla f(x_n) - \nabla f(u_n) \rangle \\
\geq D(u_n, x_n).
\]

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

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

If \( X \) 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 \( x_n \xrightarrow{P} P_C x_0. \)

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 $\mathcal{B}$-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 $i(n) \in I$ and $T_n \in \mathcal{B}$ such that $S_{i(n)} \cap \text{int dom } f \subset \text{Fix } T_n$. 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 $S_{i(n)}$. Indeed, since $T_n \in \mathcal{B}$ and $S_{i(n)} \cap \text{int dom } f \subset \text{Fix } T_n$, [2, Prop. 3.3] yields

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

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 halfspaces $H(x_0, x_n)$ and $H(x_n, u_n)$, which is a standard convex optimization problem.

Condition 4.3

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

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

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

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

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

$$\begin{cases}
\forall n \in \mathbb{N} \ i = i(p_n) \\
x_{p_n} \to x \\
x_{n+1} - x_n \to 0 \\
u_n - x_n \to 0
\end{cases}\implies x \in S_i.$$  

(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 [9, Section 4] (for examples, see [20]). 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 $X$ such that $C_1$ is closed and $C_1 \cap \text{int } C_2 \neq \emptyset$. Then $C_1 \cap \text{int } C_2 = C_1 \cap C_2$.

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 $x_n \to P_Sx_0$. 

6
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

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

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

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

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

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

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

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

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

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

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

Furthermore,

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

Consequently, if $x_{n+1} - x_n \to 0$, then $x_{p_n} - x_{k_n} \to 0$ and, in turn, $x_{p_n} \to x$. 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 2.6(i), $P_i$ is a single-valued operator in $\mathcal{B}$ with $\text{Fix } P_i = S_i \cap \text{int dom } f$ and we can implement Algorithm 4.1 as

**Algorithm 5.1** For every $n \in \mathbb{N}$, take $i(n) \in I$ and set $x_{n+1} = Q(x_0, x_n, P_{i(n)} x_n).$
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 \(i(p_n) \equiv i\), \(x_{p_n} \to x\), and \(P_S x_{p_n} - x_{p_n} \to 0\). Then \(S_i \ni P_S x_{p_n} \to x\) and, since \(S_i\) is weakly closed, \(x \in S_i\). 

If \(\mathcal{X}\) is Hilbertian, \(f = \| \cdot \|^2 / 2\), \(I = \{1, \ldots, m\}\) is finite, and \(i : n \mapsto n \mod{m} + 1\), then Algorithm 5.1 is Haugazeau’s original best approximation method and Corollary 5.2 relaxes to [16, Thm. 3-2].

5.2 Best D-approximation from convex inequalities

For every \(i \in I\), let \(S_i = \text{lev}_{\leq 0} g_i\), where \(g_i : \mathcal{X} \to [-\infty, +\infty]\) is an l.s.c. convex function such that \(\partial g_i\) maps bounded sets to bounded sets and dom \(f \subset \text{dom} \, g_i\), and let \(R_i\) be the subgradient D-projector onto \(S_i\). By Proposition 2.6(ii), \(R_i \in \mathfrak{B}\) and Fix \(R_i = \text{lev}_{\leq 0} g_i \cap \text{int} \, \text{dom} \, f \neq \emptyset\) by (1.2)), and we can implement Algorithm 4.1 as

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

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 \(i(p_n) \equiv i\), \(x_{p_n} \to x\), 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 \(G_i(x_{p_n}, x_n^*) = \{y \in \mathcal{X} \mid (x_{p_n} - y, x_n^*) \geq g_i(x_{p_n})\}\) 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

\[
\|u_{p_n} - x_{p_n}\| \geq d_{G_i(x_{p_n}, x_n^*)}(x_{p_n}) = \begin{cases} \frac{g_i^+(x_{p_n})}{\|x_n^*\|}, & \text{if } x_n^* \neq 0; \\ 0, & \text{otherwise,} \end{cases}
\]

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 [19, 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 \(g_i^+(x) \leq \lim_{n \to \infty} g_i^+(x_{p_n}) = 0\). We conclude \(g_i(x) \leq 0\), i.e., \(x \in S_i\). 

5.3 Best D-approximation from zeros of monotone operators

Suppose \(\text{dom} \, f = \mathcal{X}\) and, for every \(i \in I\), let \(S_i = A_i^{-1} 0\) be the set of zeros of a maximal monotone operator \(A_i : \mathcal{X} \to 2^{\mathcal{X}}\). For every \(\gamma \in [0, +\infty]\), Proposition 2.6(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 \(i(n) \in I\), \(\gamma_n \in [0, +\infty]\), and set \(x_{n+1} = Q(x_0, x_n, (\nabla f + \gamma_n A_i(n))^{-1} \circ \nabla f(x_n))\).

Corollary 5.6 Let \((x_n)_{n \in \mathbb{N}}\) be an arbitrary orbit of Algorithm 5.5 generated under Condition 4.3. Suppose \(\nabla f\) 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 \(i(p_n) \equiv i\), 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 subgradient $(x_{p_n})_{n \in \mathbb{N}}$ such that $i(p_n) \equiv i$, $x_{p_n} \rightharpoonup x$, and $u_{p_n} \rightarrow 0$, where $u_{p_n} = (\nabla f + \gamma_{p_n} A_i)^{-1} \circ \nabla f(x_{p_n})$. Then $u_{p_n} \rightarrow x$. Now set, for every $n \in \mathbb{N}$, $u_n = (\nabla f(x_{p_n}) - \nabla f(u_{p_n})) / \gamma_{p_n}$. Then $(u_n, \gamma_n)_{n \in \mathbb{N}}$ lies in $\text{gr} A_i$. On the other hand, $u_n \rightharpoonup 0$ since $u_{p_n} \rightarrow 0$, $\nabla f$ is uniformly continuous on int dom $f$, and $\inf_{n \in \mathbb{N}} \gamma_{p_n} > 0$. Since $A_i$ is maximal monotone, $\text{gr} A_i$ is weakly-strongly closed and must therefore contain $(x, 0)$, i.e., $x \in S_i$. \qed

Remark 5.7 The function $f$ in Corollary 5.6 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 5.6: it is Legendre [1, Lemma 6.2(iii)] and $\nabla f$ is uniformly continuous on bounded sets [21, Prop. 47.19(2)(ii)].

References


