Optimization, 2000, Vol. 48, pp. 409–427
Reprints available directly from the publisher
Photocopying permitted by license only

DYKSTRA'S ALGORITHM WITH BREGMAN PROJECTIONS: A CONVERGENCE PROOF

HEINZ H. BAUSCHKE[a,*] and ADRIAN S. LEWIS[b,†]

[a] *Department of Mathematics and Statistics, Okanagan University College, Kelowna, British Columbia V1V 1V7, Canada;*
[b] *Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*

(Received 18 June 1998; In final form 11 March 1999)

Dykstra's algorithm and the method of cyclic Bregman projections are often employed to solve best approximation and convex feasibility problems, which are fundamental in mathematics and the physical sciences. Censor and Reich very recently suggested a synthesis of these methods, Dykstra's algorithm with Bregman projections, to tackle a non-orthogonal best approximation problem. They obtained convergence when each constraint is a halfspace. It is shown here that this new algorithm works for general closed convex constraints; this complements Censor and Reich's result and relates to a framework by Tseng. The proof rests on Boyle and Dykstra's original work and on strong properties of Bregman distances corresponding to Legendre functions. Special cases and observations simplifying the implementation of the algorithm are also discussed.

Keywords: Best approximation; Bregman distance; Bregman projection; Convex feasibility; Cyclic projections; Dykstra's algorithm; Han's algorithm; Hildreth's method; Legendre function

Mathematics Subject Classifications 1991: Primary: 49M; Secondary: 41A29, 65J05, 90C25

*Corresponding author. This research was supported by an NSERC Postdoctoral Fellowship and by the Department of Combinatorics and Optimization, University of Waterloo. e-mail: bauschke@cecm.sfu.ca
†Research partially supported by NSERC.

1. INTRODUCTION

Throughout the paper, we assume that

E is some Euclidean space $\mathbb{R}^J$ with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$,

and that

$C_1, \ldots, C_N$ are finitely many closed convex sets in E with

$$C := \bigcap_{i=1}^{N} C_i \neq \emptyset.$$

Typically, the sets C_i are the *constraints* of a problem and C is the set of *solutions*; the problem of finding any solution is referred to as the *convex feasibility problem*. ([4, 22], and [40] are comprehensive starting points to this vast and important area.) We now describe two frequently employed algorithms for solving this problem.

1. The first algorithm solves a more ambitious *best approximation problem*: given a point $x_0 \in E$, find the (unique) point in C that is nearest to x_0. This problem is common in both mathematics and the physical sciences; it is often solved iteratively by *Dykstra's algorithm*: let $q_{-(N-1)} := \cdots := q_0 := 0$. Denote the mod N function with values in $\{1, \ldots, N\}$ by $[\cdot]$. Set $C_n := C_{[n]}$ and let P_n be the *orthogonal projection* or *nearest point mapping* onto C_n, for every $n \geq 1$. Generate sequences (x_n), (q_n) in E by

   $$x_n := P_n(x_{n-1} + q_{n-N}) \quad \text{and} \quad q_n := x_{n-1} + q_{n-N} - x_n,$$

   for every $n \geq 1$. In 1985, Boyle and Dysktra [7] proved that the sequence (x_n) converges to $P_C x_0$, *i.e.*, to the solution of the best approximation problem. (See Remark 4.2 for pointers to the literature).
2. The second algorithm employs more general *Bregman projections* onto the constraints. These not necessarily orthogonal projections are constructed as follows. We assume henceforth that (see Definition 2.1)

   f is a convex function of Legendre type on E, and that the corresponding *Bregman distance* D_f is defined by

$$\boxed{\begin{aligned} D_f : E \times \operatorname{int}(\operatorname{dom} f) \to [0, +\infty] :& \\ (x, y) \mapsto f(x) - f(y) - \langle \nabla f(y), x - y \rangle&. \end{aligned}}$$

Let S be a closed convex nonempty set in E with $S \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset$ and pick $y \in \operatorname{int}(\operatorname{dom} f)$. Then the optimization problem

$$\text{minimize} \quad D_f(x, y) \quad \text{subject to} \quad x \in S \cap \operatorname{dom} f$$

has a unique minimizer in $\operatorname{int}(\operatorname{dom} f)$, denoted $P_S^{(f)} y$, called the *Bregman projection of y onto S with respect to f*. (For a proof, see [5, Theorem 3.12.(iii)].) Under a constraint qualification such as $C \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset$, we can now describe the *method of cyclic Bregman projections*: given a starting point x_0, we generate a sequence (x_n) by (Bregman) projecting cyclically onto the constraints:

$$x_0 \xrightarrow{P_1^{(f)}} x_1 \xrightarrow{P_2^{(f)}} x_2 \xrightarrow{P_3^{(f)}} \cdots \xrightarrow{P_N^{(f)}} x_N \xrightarrow{P_1^{(f)}} x_{N+1} \xrightarrow{P_2^{(f)}} \cdots.$$

Bregman proved in 1967 [8] that the sequence (x_n) converges to some point in C, *i.e.*, to a solution of the convex feasibility problem. (Further results can be found in [1, 5, 8, 16–18, 20, 23, 43]. Bregman distances are increasingly employed in other fields; see, for instance, [10–15, 21, 28, 29, 38, 41, 42, 48].)

If we set $f := (1/2)\|\cdot\|^2$, then $\nabla f = I$ and the Bregman projection is actually the ordinary orthogonal projection and the method of cyclic Bregman projections becomes the famous *method of cyclic (orthogonal) projections*. (See also [6].) The method of cyclic projections differs from Dykstra's algorithm only by the sequence (q_n). Moreover, if each constraint set is *affine*, then the algorithms coincide. (See Deutsch's [24] exhaustive review.)

It is very tempting to combine Dykstra's algorithm with the method of cyclic Bregman projections to obtain a new algorithm that would solve the *best approximation problem with Bregman distances*. Censor and Reich [19] very recently showed that such a synthesis is indeed possible. They established convergence for the resulting *Dykstra's algorithm with Bregman projections* when the constraint sets are *halfspaces*.

The objective of this paper is to provide a convergence proof that makes Dykstra's algorithm with Bregman projections applicable in a more general setting.

We obtain a convergence result for general constraint sets by a nontrivial extension of Boyle and Dykstra's original proof [7] and by repeated use of the tools developed in [5].

We also note that Tseng's powerful framework [49] yields a convergence result in this setting as well. In summary, our main theorem (Theorem 3.2) is complementary to Censor and Reich's work and to the result we deduced from Tseng's framework.

The paper is organized as follows.

In Section 2, properties of Legendre functions and co-finite convex functions are recalled. We introduce the notion of a "very strictly convex function", which is useful for our analysis, and provide examples.

Our main result (Theorem 3.2) is proved in Section 3 and then compared to results by Censor and Reich and by Tseng's framework.

The final Section 4 discusses applications. We remarked above that Dykstra's algorithm coincides with the method of cyclic Bregman projections when $f = (1/2)\|\cdot\|^2$ and the constraints are affine. Theorem 4.3 shows that this correspondence holds for more general functions f. This is important as it avoids the computation of an auxiliary sequence. We conclude with a generalization of another storage-saving remark by Glunt *et al.*

The notation and language we use is fairly standard and follows Rockafellar's fundamental [46]. Given a convex function g on E, the *domain* of g (*gradient* of g, *Hessian* of g, *Fenchel conjugate* of g respectively) is denoted by dom g (∇g, $\nabla^2 g$, g^* respectively). The *interior* (*relative interior*, *boundary*, *closure*, *convex hull*, respectively) of a set S in E is abbreviated by int S (ri S, bd S, cl S, conv S, respectively). Finally, I stands for the identity mapping.

2. TOOLS

2.1. Legendre Functions

DEFINITION 2.1 Suppose f is a closed convex proper function on E with $\operatorname{int}(\operatorname{dom} f) \neq \emptyset$. Then f is *Legendre* or *a convex function of*

Legendre type, if it satisfies every one of the following conditions:

(i) f is differentiable on $\operatorname{int}(\operatorname{dom} f)$.
(ii) $\lim_{t \to 0^+} \langle \nabla f(x + t(y - x)), y - x \rangle = -\infty$, $\forall x \in \operatorname{bd}(\operatorname{dom} f)$, $\forall y \in \operatorname{int}(\operatorname{dom} f)$.
(iii) f is strictly convex on $\operatorname{int}(\operatorname{dom} f)$.

Fact 2.2 (Rockafellar's [46, Theorem 26.5]) A convex function f is Legendre if and only if its conjugate f^* is. In this case, the gradient mapping

$$\nabla f : \operatorname{int}(\operatorname{dom} f) \to \operatorname{int}(\operatorname{dom} f^*) : x \mapsto \nabla f(x)$$

is a topological isomorphism with inverse mapping $(\nabla f)^{-1} = \nabla f^*$.

Fact 2.3 ([5, Proposition 3.16]) Suppose f is Legendre on E and S is a closed convex set in E with $S \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset$. Suppose further $y \in \operatorname{int}(\operatorname{dom} f)$. Then the Bregman projection $P_S^{(f)} y$ of y onto S with respect to f is characterized by

$$P_S^{(f)} y \in S \cap \operatorname{int}(\operatorname{dom} f) \quad \text{and} \quad \langle \nabla f(y) - \nabla f(P_S^{(f)} y),\ S - P_S^{(f)} y \rangle \leq 0.$$

In addition, $D_f(P_S^{(f)} y, y) \leq D_f(s, y) - D_f(s, P_S^{(f)} y)$, $\forall s \in S \cap \operatorname{dom} f$.

Fact 2.4 (three-point-identity; Chen and Teboulle's [21, Lemma 3.1]) Suppose f is a Legendre on E. If x, $y \in \operatorname{int} \operatorname{dom} f$ and $b \in \operatorname{dom} f$, then

$$D_f(b, x) + D_f(x, y) - D_f(b, y) = \langle \nabla f(x) - \nabla f(y), x - b \rangle.$$

Remark 2.5 Note that Fact 2.4 is a special case of the following "four-point-identity":

$$\langle \nabla f(x) - \nabla f(y), a - b \rangle = D_f(b, x) + D_f(a, y) - D_f(a, x) - D_f(b, y),$$

for $x, y \in \operatorname{int}(\operatorname{dom} f)$ and $a, b \in \operatorname{dom} f$. This trivial though useful identity allows generalizations of concepts such as firm nonexpansiveness to a non-orthogonal setting; see [18, Section 5] for steps in this direction.

2.2. Co-finite Functions

DEFINITION 2.6 ([46, Section 13]) Suppose f is a closed convex proper function on E. Then f is *co-finite*, if $\lim_{r\to+\infty} f(rx)/r = +\infty$, $\forall x \in E\backslash\{0\}$.

Fact 2.7 ([46, Corollary 13.3.1]) Suppose f is a closed convex proper function on E. Then f is co-finite if and only if $\operatorname{dom} f^* = E$.

2.3. Very Strictly Convex Functions

DEFINITION 2.8 Suppose f is a closed convex proper function on E. Suppose further f is twice continuously differentiable on $\operatorname{int}(\operatorname{dom} f) \neq \emptyset$. We say that f is *very strictly convex*, if $\nabla^2 f(x)$ is positive definite, $\forall x \in \operatorname{int}(\operatorname{dom} f)$.

Remark 2.9 The class of very strictly convex functions lies strictly between the class of strictly convex functions and the class of *strongly convex* functions (see [36, Definition IV.1.1.1] for the definition of the last class). It is clear that every very strictly convex function is strictly convex; however, the converse is false: consider $x \mapsto x^4$ on $\mathbb{R}$. On the other hand, every twice continuously differentiable strongly convex function is very strictly convex (use [36, Theorem IV.4.3.1.(iii)]); the reverse implication again does not hold in general: consider exp on $\mathbb{R}$.

PROPOSITION 2.10 *Suppose f is a very strictly convex function on E. Then for every compact convex subset K of $\operatorname{int}(\operatorname{dom} f)$, there exist reals $0 < \theta$ and $\Theta < +\infty$ such that for every $x, y \in K$*

(i) $D_f(x, y) \geq \theta \|x - y\|^2$ *and*
(ii) $\|\nabla f(x) - \nabla f(y)\| \leq \Theta \|x - y\|$.

Proof Abbreviate the symmetric matrices in $\mathbb{R}^{J\times J}$ by S_J. For a given symmetric matrix $A \in S_J$, denote the largest (resp. smallest) eigenvalue of A by $\Lambda(A)$ (resp. $\lambda(A)$). By [36, Section IV.1.2.(e), page 155], the induced mapping Λ (resp. λ) in convex (resp. concave) on S_J. Hence Λ and λ are continuous on S_J. Thus, by assumption, the mappings $\Lambda \circ \nabla^2 f$ and $\lambda \circ \nabla^2 f$ are continuous on $\operatorname{int}(\operatorname{dom} f)$. Now fix an arbitrary compact convex subset K of $\operatorname{int}(\operatorname{dom} f)$. Then, by continuity, $0 < \lambda_K := \min_{x\in K}(\lambda \circ \nabla^2 f)(x) \leq \Lambda_K := \max_{x\in K}(\Lambda \circ \nabla^2 f)(x) < +\infty$. Hence $\lambda_K\|h\|^2 \leq \langle h, \nabla^2 f(z)h\rangle \leq \Lambda_K\|h\|^2$, $\forall z \in K$, $\forall h \in X$. Pick any two $x, y \in K$.

Then Taylor's Theorem (see [45, Theorem 1.2.4] or [47, Exercise 5.(b) on page 378]) yields some z on the line segment between x and y such that

$$D_f(x,y) = f(x) - f(y) - \langle \nabla f(y), x - y \rangle = \frac{1}{2}\langle x - y, \nabla^2 f(z)(x - y) \rangle.$$

Hence (i) follows with $\theta := (1/2)\lambda_K$. To prove (ii), note first that the operator norm $\|\nabla^2 f(z)\| \le \Lambda_K$, $\forall z \in K$ (see [44, Theorem XVIII.2.2]). By a consequence of the Mean Value Theorem (see [44, Corollary XIII.4.3]), $\|\nabla f(x) - f(y)\| \le \sup_{z \in K} \|\nabla^2 f(z)\| \|x - y\|$. Therefore, $\Theta := \Lambda_K$ does the job. ■

2.4. "All of the Above"

Let $\mathcal{C}(E)$ denote the class of closed convex proper functions on E that are Legendre, co-finite, and very strictly convex. The class $\mathcal{C}(E)$ is important to us, since our main result (Theorem 3.2) requires that the function determining the Bregman distance be a member of $\mathcal{C}(E)$. Suppose now f is a *separable* function on E, *i.e.*, it can be written as $f(x) = \sum_j f_j(x_j)$, where each f_j is a function on $\mathbb{R}$ and $x = (x_1, \ldots, x_J)$. Separable functions are almost always used in practice. Since

f belongs to $\mathcal{C}(E)$ if and only if each f_j is in $\mathcal{C}(\mathbb{R})$,

we are particularly interested in $\mathcal{C}(\mathbb{R})$.

Examples 2.11 Every function listed below is closed convex proper on $\mathbb{R}$, Legendre, co-finite, and very strictly convex:

(i) ("norm^2") $f(x) := (1/2)x^2$.
(ii) ("Boltzmann/Shannon") $f(x) := x \ln x - x$ on $[0, +\infty[$.
(iii) ("Hellinger") $f(x) := -\sqrt{1 - x^2}$ on $[-1, +1]$.
(iv) ("Fermi/Dirac") $f(x) := x \ln x + (1 - x) \ln(1 - x)$ on $[0, 1]$.
(v) ("De Pierro and Iusem") $f(x) := (1/2)x^2 + 2x + (1/2)$, if $x \le -1$; $-1 - \ln(-x)$, if $-1 \le x < 0$; $+\infty$, otherwise.

Proof Follows from the discussion of these functions in [5, Section 6.1], Fact 2.7, and some calculus. ■

Remarks 2.12 (i) Analogous to [5, Proposition 5.1], one can develop criteria under which sums or infimal convolutions of members of $\mathcal{C}(E)$ belong to $\mathcal{C}(E)$ again. (ii) As an aside, we note that the function by De Pierro and Iusem in Examples 2.11 is *not* a "Bregman function (with the zone being the negative reals)" in the sense of Censor and Lent [17].

3. MAIN RESULT

PROPOSITION 3.1 *Suppose (ρ_k) is a sequence of nonnegative reals with $\sum_k \rho_k^2 < +\infty$. Let $R_n := \sum_{k=1}^n \rho_k$ be the nth partial sum and m be an arbitrary nonnegative integer. Then $\lim_n R_n(\rho_{n-m} + \cdots + \rho_{n-1} + \rho_n) = 0$.*

Proof By Cauchy/Schwarz, $R_n(\rho_{n-m}+\cdots+\rho_n) \leq \sqrt{n}\sqrt{\sum_{k=1}^n \rho_k^2}(\rho_{n-m}+\cdots+\rho_n)$; hence it suffices to show that $\lim_n \sqrt{n}(\rho_{n-m}+\cdots+\rho_n)=0$. Suppose not. Then there exists $\varepsilon > 0$ such that eventually $\rho_{n-m}+\cdots+\rho_n \geq \varepsilon/\sqrt{n}$, which implies, again by Cauchy/Schwarz,

$$\frac{\varepsilon^2}{n} \leq (\rho_{n-m} + \cdots + \rho_n)^2 \leq (m+1)(\rho_{n-m}^2 + \cdots + \rho_n^2).$$

Summing over sufficiently large n yields the desired contradiction. ■

We are now ready for the main result.

THEOREM 3.2 *Suppose f is a closed convex proper function on E and $C_1, \ldots, C_N$ are finitely many closed convex sets in E with $C := \cap_i C_i \neq \emptyset$.*

Assumption on the Bregman distance: f is very strictly convex, cofinite, and Legendre.
Constraint qualification: $C \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset$.

Let $q_{-(N-1)} := \cdots := q_{-1} := q_0 := 0$. Denoting the mod N function with values in $\{1, \ldots, N\}$ by $[\cdot]$, set $C_n := C_{[n]}$ and $P_n := P_{C_n}^{(f)}$, for every $n \geq 1$. Finally, let $x_0 \in \operatorname{int}(\operatorname{dom} f)$ and define sequences (x_n), (q_n) by

$$x_n := (P_n \circ \nabla f^*)(\nabla f(x_{n-1}) + q_{n-N}) \quad \text{and}$$
$$q_n := \nabla f(x_{n-1}) + q_{n-N} - \nabla f(x_n),$$

for every $n \geq 1$. Then the sequence (x_n) converges to $P_C^{(f)} x_0$.

Proof Because f is co-finite, we have $\operatorname{dom} f^* = E$ (Fact 2.7) and so the sequences are well-defined. The following hold true for every $n \geq 1$:

$$x_n \in C_n \cap \operatorname{int}(\operatorname{dom} f), \ \langle x_n - C_n, q_n \rangle \geq 0, \ \text{so } \langle x_n, q_n \rangle = \iota^*_{C_n}(q_n); \tag{1}$$

$$\nabla f(x_{n-1}) - \nabla f(x_n) = q_n - q_{n-N}; \tag{2}$$

$$\nabla f(x_0) - \nabla f(x_n) = \sum_{k=n-N+1}^{n} q_k. \tag{3}$$

Facts 2.2 and 2.3 imply (1). The definition of the sequences (q_n) yields (2). Equation (3) follows from (2) by induction.

The next identity is crucial; it is true for $0 \leq m \leq n$ and every $c \in \operatorname{dom} f$:

$$\begin{aligned} D_f(c, x_m) = D_f(c, x_n) &+ \sum_{k=m+1}^{n} \left(D_f(x_k, x_{k-1}) + \langle q_{k-N}, x_{k-N} - x_k \rangle \right) \\ &+ \sum_{k=n-N+1}^{n} \langle x_k - c, q_k \rangle - \sum_{k=m-(N-1)}^{m} \langle x_k - c, q_k \rangle. \end{aligned} \tag{4}$$

We prove (4) by induction on n, where m is arbitrary but fixed: clearly, (4) holds when $n = m$. Now let $n \geq m$. We observe that

$$\begin{aligned} D_f(c, x_n) &= D_f(c, x_{n+1}) + D_f(x_{n+1}, x_n) \\ &\quad - \langle \nabla f(x_{n+1}) - \nabla f(x_n), x_{n+1} - c \rangle \\ &= D_f(c, x_{n+1}) + D_f(x_{n+1}, x_n) \\ &\quad + \langle x_{n+1} - c, q_{n+1} - q_{n+1-N} \rangle \\ &= D_f(c, x_{n+1}) + D_f(x_{n+1}, x_n) \\ &\quad + \langle q_{n+1-N}, x_{n+1-N} - x_{n+1} \rangle \\ &\quad + \langle x_{n+1} - c, q_{n+1} \rangle - \langle x_{n-N+1} - c, q_{n-N+1} \rangle; \end{aligned}$$

here, the first equality follows from the three-point-identity (Fact 2.4) and the second one from (2). The last displayed equality yields the induction step and so identity (4) holds. Let us start exploring (4) by choosing $m = 0$ and $c \in C \cap \operatorname{int}(\operatorname{dom} f)$. Then for every $n \geq 0$:

$$D_f(c, x_0) = D_f(c, x_n) + \sum_{k=1}^{n} (D_f(x_k, x_{k-1}) + \langle q_{k-N}, x_{k-N} - x_k \rangle) + \sum_{k=n-N+1}^{n} \langle x_k - c, q_k \rangle.$$

By (2), *all* terms on the right-hand side are nonnegative. Hence

$$(D_f(c, x_n)) \text{ is bounded} \quad \text{and} \quad \sum_{k=1}^{\infty} D_f(x_k, x_{k-1}) < +\infty.$$

Now $\operatorname{dom} f^*$ is open and $(D_f(c, x_n))$ is bounded, hence (x_n) is bounded (by [5, Corollary 3.11]). The boundedness of $(D_f(c, x_n))$ and the fact that $c \in \operatorname{int}(\operatorname{dom} f)$ together imply that cluster points of (x_n) must lie in $\operatorname{int}(\operatorname{dom} f)$ (by [5, Theorem 3.8.(i)]). But now $D_f(x_k, x_{k-1}) \to 0$ and [5, Theorem 3.9.(iii)] imply that $x_k - x_{k-1} \to 0$. Let us record what we just learnt:

$$(x_k) \text{ is a bounded sequence in } \operatorname{int}(\operatorname{dom} f), \text{ all of its cluster points lie in } \operatorname{int}(\operatorname{dom} f), \quad \text{and} \quad x_k - x_{k-1} \to 0.$$

Now (3) implies the identity

$$\langle c - x_n, \nabla f(x_0) - \nabla f(x_n) \rangle = \sum_{k=n-N+1}^{n} \langle c - x_k, q_k \rangle + \sum_{k=n-N+1}^{n} \langle x_k - x_n, q_k \rangle. \tag{5}$$

Denote the first (resp. second) sum on the right-hand side of (5) by $S_1(n)$ (resp. $S_2(n)$). By (1), $S_1(n)$ is always nonpositive. Concerning $S_2(n)$, we will prove that

$$\lim_n \sum_{k=n-(N-1)}^{n} |\langle x_k - x_n, q_k \rangle| \stackrel{?}{=} 0. \tag{6}$$

Let $K := \operatorname{cl}(\operatorname{conv}\{x_n : n \geq 0\})$. Then, by [46, Theorem 17.2], $K = \operatorname{conv}(\operatorname{cl}\{x_n : n \geq 0\}) = \operatorname{conv}(\{x_n : n \geq 0\} \cup \text{cluster points of } (x_n)) \subseteq \operatorname{conv}(\operatorname{int}(\operatorname{dom} f)) = \operatorname{int}(\operatorname{dom} f)$. Hence, by assumption and Proposition 2.10, there exist reals $\theta > 0$ and $\Theta < +\infty$ such that $D_f(x, y) \geq \theta \| x - y \|^2$ and

$\| \nabla f(x) - \nabla f(y) \| \leq \Theta \| x - y \|, \forall x, y \in K$. In particular, for every $k \geq 1$,

$$D_f(x_k, x_{k-1}) \geq \theta \|x_k - x_{k-1}\|^2 \quad \text{and}$$
$$\|\nabla f(x_{k-1}) - \nabla f(x_k)\| \leq \Theta \|x_{k-1} - x_k\|.$$

In view of $\sum_{k=1}^{\infty} D_f(x_k, x_{k-1}) < +\infty$, we deduce

$$\sum_{k=1}^{n} \|x_{k-1} - x_k\|^2 < +\infty.$$

Now note the following telescoping identity:

$$\begin{aligned} q_k &= q_k - 0 = q_k - q_{[k]-N} \\ &= (q_k - q_{k-N}) + (q_{k-N} - q_{k-2N}) + \cdots + (q_{[k]} - q_{[k]-N}). \end{aligned}$$

This yields

$$\|q_k\| \leq \|q_k - q_{k-N}\| + \|q_{k-N} - q_{k-2N}\| + \cdots + \|q_{[k]} - q_{[k]-N}\|$$

and hence, using (2) and the defining property of Θ,

$$\begin{aligned} \sum_{k=n-(N-1)}^{n-1} \|q_k\| &\leq \sum_{k=1}^{n-1} \|q_k - q_{k-N}\| \\ &= \sum_{k=1}^{n-1} \|\nabla f(x_{k-1}) - \nabla f(x_k)\| \leq \Theta \sum_{k=1}^{n-1} \|x_{k-1} - x_k\|. \end{aligned} \tag{7}$$

With (7), we obtain

$$\begin{aligned} \sum_{k=n-(N-1)}^{n} |\langle x_k - x_n, q_k \rangle| &\leq \sum_{k=n-(N-1)}^{n-1} \|x_k - x_n\| \|q_k\| \\ &\leq \sum_{k=n-(N-1)}^{n-1} \|q_k\| \sum_{l=n-N+2}^{n} \|x_{l-1} - x_l\| \\ &\leq \Theta \sum_{k=1}^{n-1} \|x_{k-1} - x_k\| \sum_{l=n-N+2}^{n} \|x_{l-1} - x_l\|. \end{aligned}$$

By Proposition 3.1, the limit inferior of the sequence generated by the last expression is equal to 0. Therefore, (6) holds.

Hence we obtain a subsequence (k_n) of (n) such that (recall (5) and the nonpositivity of $S_1(n)$) for every $c \in C$:

$$\sum_{k=k_n-(N-1)}^{k_n} |\langle x_k - x_{k_n}, q_k\rangle| \to 0 \quad \text{and} \qquad (8)$$
$$0 \geq \overline{\lim_n} \langle c - x_{k_n}, \nabla f(x_0) - \nabla f(x_{k_n})\rangle.$$

After passing to another subsequence if necessary, we assume without loss of generality that

$x_{k_n} \to c^*$, for some $c^* \in \operatorname{int}(\operatorname{dom} f)$, and $[k_n] \equiv j$, for some index j.

It follows that $c^* \in C_j$ and thus $c^* \in C$, since $x_n - x_{n+1} \to 0$. Now $D_f(\cdot, \cdot)$ is separately continuous on int (dom f) × int (dom f), because ∇f and f are continuous on int (dom f). Hence $D_f(c, x_{k_n}) \to D_f(c, c^*)$ and $D_f(x_{k_n}, x_0) \to D_f(c^*, x_0)$. Thus for an arbitrary but fixed $c \in C$, we conclude:

$$\begin{aligned} &\langle c - c^*, \nabla f(x_0) - \nabla f(c^*)\rangle \\ &\quad = D_f(c, c^*) + D_f(c^*, x_0) - D_f(c, x_0) \\ &\quad = \lim_n (D_f(c, x_{k_n}) + D_f(x_{k_n}, x_0) - D_f(c, x_0)) \\ &\quad = \lim_n \langle c - x_{k_n}, \nabla f(x_0) - \nabla f(x_{k_n})\rangle \\ &\quad \leq 0, \end{aligned} \qquad (9)$$

here, the first and the third equality come from the three-point-identity (Fact 2.4) and the inequality comes from (8). By Fact 2.3,

$$c^* = P_C^{(f)}(x_0).$$

Now (9) and (5) yield (with c^* and k_n)

$$\begin{aligned} 0 &\leftarrow \langle c^* - x_{k_n}, \nabla f(x_0) - \nabla f(x_{k_n})\rangle \\ &= \sum_{k=k_n-N+1}^{k_n} \langle c - x_k, q_k\rangle + \sum_{k=k_n-N+1}^{k_n-1} \langle x_k - x_{k_n}, q_k\rangle. \end{aligned}$$

The second sum tends to 0 by (8). Hence so must the first sum, which we know to possess exclusively nonpositive terms:

$$0 \leftarrow \sum_{k=k_n-N+1}^{k_n} \langle c^* - x_k, q_k\rangle \leq 0. \qquad (10)$$

It remains to show that the entire sequence (x_n) converges to c^*. Assume the contrary. Then there exist $\varepsilon > 0$ and a subsequence (l_n) of (n) such that $\|c^* - x_{l_n}\| \geq \varepsilon$, $\forall n$. After passing to a subsequence of (l_n)

if necessary, we assume WLOG that $l_n > k_n$, $\forall n$. Since K contains c^* and (x_{l_n}), we conclude $D_f(c^*, x_{l_n}) \geq \varepsilon^2\theta$, $\forall n$. Equation (4) (with c^* and $k_n < l_n$), (10) and [5, Proposition 3.2.(ii)] yield

$$0 \leftarrow D_f(c^*, x_{k_n}) \geq D_f(c^*, x_{l_n}) - \sum_{k=k_n-(N-1)}^{k_n} \langle x_k - c^*, q_k \rangle$$

$$\geq \sum_{k=k_n-(N-1)}^{k_n} \langle c^* - x_k, q_k \rangle \to 0.$$

Thus $D_f(c^*, x_{l_n}) - \sum_{k=k_n-(N-1)}^{k_n} \langle x_k - c^*, q_k \rangle \to 0$ Hence, by (10), $D_f(c^*, x_{l_n}) \to 0$, which is a contradiction. ∎

Remark 3.3

- The proof of Theorem 3.2 follows the [2, Proof of Theorem 11.2.1], which in turn rests on Boyle and Dykstra's original work [7].
- The analyses of Dykstra's algorithm for two (possibly nonintersecting) sets by Bauschke and Borwein and by Iusem and De Pierro relies on identities which can be interpreted as special cases of the key identity (4) above; see [3, Lemma 4.5] and [39, Proposition 4].
- Every function listed in Examples 2.11 satisfies the assumption on the distance.
- Examples of Bregman projections appear in [5].

Remark 3.4 A second look at the proof of Theorem 3.2 reveals the following.

- The assumption that f be co-finite can be replaced by "domf^* is open and the sequences are well-defined". This would cover the "Burg entropy", $-\sum_j \ln x_j$, for instance; however, we do not know whether or not the sequences are well-defined.
- The assumption that f be very strictly convex can be replaced by the conclusion of Proposition 2.10.

Remark 3.5 We now compare Theorem 3.2 to the two related complementary though different results mentioned in Section 1.

Tseng [49] proposed a very general and powerful algorithmic framework that covers Han's algorithm (Remark 4.2). Analogous to his derivation of Han's result in [49, Section 4], we can apply his framework to our algorithmic setting. We find that Tseng's framework:

- allows *essentially cyclic control* (which is more general than our cyclic control though more cumbersome to describe);
- has less stringent assumptions on f; but
- requires the constraint qualification $\text{int}(\text{dom} f) \cap \bigcap_{i=1}^{N} \text{ri}(C_i) \neq \emptyset$ (the relative interior can be dropped for *polyhedral* sets). On the one hand, his framework is clearly more flexible in some circumstances. On the other hand, our framework has a genuinely less restrictive constraint qualification. (For instance, consider $f = 1/2 \|\cdot\|^2$ on $\mathbb{R}^2$ and two disks that have exactly one point in common.)

Censor and Reich essentially suggested the framework we investigated here in [19]. Their convergence result

- works when each set is a *halfspace*; and
- the function f is a Bregman function (which excludes some Legendre functions; see Remarks 2.12. (ii) and also [5, Section 4]); but
- their constraint qualification ("strong zone consistency") is less restrictive than ours; and
- neither very strict convexity nor co-finiteness is needed. (In practice, it appears to be quite hard to verify some of their assumptions; see [17, Subsections 6.1 and 6.4].)

4. APPLICATIONS

Example 4.1 ([7, Theorem 4.1]) Suppose we let $f := 1/2 \|\cdot\|^2$ in Theorem 3.2. Then we obtain *Dykstra's algorithm in Euclidean space.*

Remark 4.2

- Boyle and Dykstra's convergence proof actually works in general Hilbert space. For applications and generalizations, see [3], [2, Theorem 11.2.1], [26, 25, 30, 37, 31–33, 39].
- *Han's algorithm* [34] is Example 4.1 with an additional constraint qualification imposed.
- If we assume that the constraint sets in Example 4.1 are *halfspaces*, then we obtain *Hildreth's method* [35] for quadratic programming; see [39, Section 3].

Dykstra pointed out in [27] that the *method of cyclic orthogonal projections*, which allows us to ignore the sequence (q_n), arises when the constraints in Example 4.1 are all *affine subspaces*. Fortunately, as the next result shows, this is true not only for $f = 1/2\,\|\cdot\|^2$ but also in our setting.

THEOREM 4.3 *Suppose f is Legendre, co-finite, and very strictly convex on E. Suppose further $C_1, \ldots, C_N$ are finitely many affine subspaces of E with $C := \bigcap_i C_i$ and $C \cap \operatorname{int}\operatorname{dom} f \neq \emptyset$. Denoting the mod N function with values in $\{1, \ldots, N\}$ by $[\cdot]$, set $C_n := C_{[n]}$ and $P_n := P^{(f)}_{C_n}$, for every $n \geq 1$. Let $y_0 \in \operatorname{int}\operatorname{dom} f$ and define the sequence of cyclic Bregman projections by*

$$y_n := P_n y_{n-1},$$

for every $n \geq 1$. Then (y_n) converges to $P^{(f)}_C y_0$.

Proof Denote the linear subspace parallel to C_n by L_n, for every $n \geq 1$. Consider the sequence (x_n) generated by the Dykstra algorithm with starting point $x_0 := y_0$. In view of Theorem 3.2, it suffices to show that $x_n = y_n$, for every $n \geq 1$. We do this by induction on n. Because $q_{-(N-1)} = \cdots = q_0 = 0$, we have $x_n = y_n$, for every $n \in \{0, 1, \ldots, N\}$. To do the induction step, we assume that $x_{n-1} = y_{n-1}$, for some $n > N$. By Eq. (1) in the proof of Theorem 3.2, we have $0 \leq \langle x_{n-N} - C_n, q_{n-N}\rangle = \langle L_n, q_{n-N}\rangle$. Hence $q_{n-N} \in L_n^{\perp}$. This implies (because $y_n \in C_n$ so that $C_n - y_n = L_n$ as well)

$$\langle C_n - y_n, q_{n-N}\rangle = 0.$$

We claim that $y_n = P_n(\nabla f^*(\nabla f(x_{n-1}) + q_{n-N}))$. (This would complete the proof, as the latter point is x_n by definition.) Clearly, $y_n \in \operatorname{int}(\operatorname{dom} f) \cap C_n$. Fix an arbitrary $c_n \in C_n$. Then, using Fact 2.3, $x_{n-1} = y_{n-1}$, and the last displayed equation,

$$\begin{aligned}
&\langle c_n - y_n, \nabla f(\nabla f^*(\nabla f(x_{n-1}) + q_{n-N})) - \nabla f(y_n)\rangle \\
&\quad = \langle c_n - y_n, \nabla f(x_{n-1}) + q_{n-N} - \nabla f(y_n)\rangle \\
&\quad = \langle c_n - y_n, \nabla f(y_{n-1}) - \nabla f(y_n)\rangle + \langle c_n - y_n, q_{n-N}\rangle \\
&\quad = \langle c_n - y_n, \nabla f(y_{n-1}) - \nabla f(y_n)\rangle \\
&\quad \leq 0.
\end{aligned}$$

Again by Fact 2.3, the claim follows. ■

Remark 4.4 Theorem 4.3 is related to [5, Theorem 8.4]; see also [5, Remark 8.5].

Remark 4.5 Suppose f is Legendre, co-finite, and very strictly convex on E. Suppose further we only have *two* constraints, say A and B, where B is an affine subspace. (A standard product space construction always allows a reduction to this case; see [3, Section 6] and [5, Section 7.1].) Assume $\operatorname{int}(\operatorname{dom} f) \cap A \cap B \neq \emptyset$ and let $x_0 \in \operatorname{int}(\operatorname{dom} f)$. A second glance at the proof of Theorem 4.3 reveals that then Dykstra's algorithm with Bregman projections with respect to f simplifies to: $b_0 := x_0$, $p_0 := 0$,

$$\begin{aligned} a_n &:= (P_A^{(f)} \circ \nabla f^*)(\nabla f(b_{n-1} + p_{n-1})), \\ p_n &:= \nabla f(b_{n-1}) + p_{n-1} - \nabla f(a_n), \\ b_n &:= P_B^{(f)} a_n, \end{aligned}$$

for every $n \geq 1$. Glunt *et al.* [32] pointed out (see also [33, page 292]) that when $f := 1/2\,\|\cdot\|^2$, it is not necessary to store the sequence (p_n). This nice observation works also in our setting: let $w_0 := x_0$ and update according to

$$w_n := \nabla f^*(\nabla f(w_{n-1}) + \nabla f(P_B^{(f)} P_A^{(f)} w_{n-1}) - \nabla f(P_A^{(f)} w_{n-1})),$$

for every $n \geq 1$. An easy induction yields $\nabla f(w_n) = \nabla f(b_n) + p_n$ identically, and hence, by Theorem 3.2, the sequence $(P_A^{(f)} w_n)$ converges to $P_{A\cap B}^{(f)} x_0$.

Acknowledgments

We wish to thank Yair Censor and Simeon Reich for sending us [19].

Note After preparation of this manuscript, the authors received an interesting related manuscript [9] by Lev Bregman, Yair Censor, and Simeon Reich.

References

[1] Alber, Y. and Butnariu, D. (1997). Convergence of Bregman-projection methods for solving consistent convex feasibility problems in reflexive Banach spaces. *Journal of Optimization Theory and Applications*, **92**(1), 33–61.

[2] Bauschke, H. H., *Projection Algorithms and Monotone Operators. Ph. D. Thesis*, Simon Fraser University, Department of Mathematics, Burnaby, British Columbia V5A 1S6, Canada, August 1996. Available at http://www.cecm.sfu.ca/preprints/1996pp.html.

[3] Bauschke, H. H. and Borwein, J. M., Dykstra's alternating projection algorithm for two sets. *Journal of Approximation Theory*, **79**(3), 418–443, December 1994.
[4] Bauschke, H. H. and Borwein, J. M., On projection algorithms for solving convex feasibility problems. *SIAM Review*, **38**(3), 367–426, September 1996.
[5] Bauschke, H. H. and Borwein, J. M. (1997). Legendre functions and the method of random Bregman projections. *Journal of Convex Analysis*, **4**(1), 27–67.
[6] Bauschke, H. H., Borwein, J. M. and Lewis, A. S., The method of cyclic projections for closed convex sets in Hilbert space. In: Censor, Y. and Reich, S. Eds., *Optimization and Nonlinear Analysis*, volume 204 of *Contemporary Mathematics*, pp. 1–38. American Mathematical Society, 1997. *Proceedings on the Special Session on Optimization and Nonlinear Analysis*, Jerusalem, May 1995.
[7] Boyle, J. P. and Dykstra, R. L. (1985). A method for finding projections onto the intersection of convex sets in Hilbert spaces. In: Dykstra, R. L., Robertson, T. and Wright, F. T. Eds., *Advances in Order Restricted Statistical Inference*, pp. 28–47. Springer-Verlag, *Lecture Notes in Statistics*; Vol. 37. Proceedings, Iowa City.
[8] Bregman, L. M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *U.S.S.R. Computational Mathematics and Mathematical Physics*, **7**(3), 200–217.
[9] Bregman, L. M., Censor, Y. and Reich, S. (1998). Dykstra's algorithm as the nonlinear extension of Bregman's optimization method. To appear in *Journal of Convex Analysis*.
[10] Burachik, R. S. (1995). Generalized Proximal Point Methods of the Variational Inequality Problem. *Ph. D. Thesis*, Rio de Janeiro.
[11] Butnariu, D., Censor, Y. and Reich, S. (1997). Iterative averaging of entropic projections for solving stochastic convex feasibility problems. *Computational Optimization and Applications*, **8**, 21–39.
[12] Butnariu, D. and Iusem, A. N., Local moduli of convexity and their application to finding almost common points of measurable families of operators. In: Censor, Y. and Reich, S. Eds., *Recent Developments in Optimization Theory and Nonlinear Analysis*, volume 204 of *Contemporary Mathematics*, pp. 61–91. American Mathematical Society, 1996. *Proceedings on the Special Session on Optimization and Nonlinear Analysis*, Jerusalem, May 1995.
[13] Butnariu, D. and Iusem, A. N. (1997). On a proximal point method for convex optimization in Banach spaces. *Numerical Functional Analysis and Optimization*, **18**(7–8), 723–744.
[14] Butnariu, D., Iusem, A. N. and Burachik, R. S. (1996). Iterative methods of solving stochastic convex feasibility problems and applications, To appear in *Computational Optimization and Applications*.
[15] Byrne, C. and Censor, Y. (1997). Proximity function minimization and the convex feasibility problem for jointly convex Bregman distances. Preprint.
[16] Censor, Y. and Elfving, T. (1994). A multiprojection algorithm using Bregman projections in a product space. *Numerical Algorithms*, **8**, 221–239.
[17] Censor, Y. and Lent, A., An iterative row-action method for interval convex programming. *Journal of Optimization Theory and Applications*, **34**(3), 321–353, July 1981.
[18] Censor, Y. and Reich, S. (1996). Iterations of paracontractions and firmly nonexpansive operators with applications to feasibility and optimization. *Optimization*, **37**(4), 323–339.
[19] Censor, Y. and Reich, S. (1998). The Dykstra algorithm with Bregman projections. *Communications in Applied Analysis*, **2**(3), 407–419.
[20] Censor, Y. and Zenios, S. A. (1997). Parallel optimization. *Numerical Mathematics and Scientific Computation*. Oxford University Press, New York.
[21] Chen, G. and Teboulle, M., Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM Journal on Optimization*, **3**(3), 538–543, August 1993.

[22] Combettes, P. L. (1996). *The Convex Feasibility Problem in Image Recovery*, volume 95 of *Advances in Imaging and Electron Physics*, pp. 155–270. Academic Press, Inc.

[23] De Pierro, A. R. and Iusem, A. N., A relaxed version of Bregman's method for convex programming. *Journal of Optimization Theory and Applications*, **51**(3), 421–440, December 1986.

[24] Deutsch, F. (1992). The method of alternating orthogonal projections. In: Singh, S. P. Ed., *Approximation Theory, Spline Functions and Applications*, pp. 105–121. Kluwer Academic Publ.

[25] Deutsch, F. (1995). Dykstra's cyclic projections algorithm: the rate of convergence. In: Singh, S. P. Ed., *Approximation Theory, Wavelets and Applications*, pp. 87–94. Kluwer Academic Publishers.

[26] Deutsch, F. and Hundal, H. (1994). The rate of convergence of Dykstra's cyclic projections algorithm: the polyhedral case. *Numerical Functional Analysis and Optimization*, **15**(5–6), 537–565.

[27] Dykstra, R. L., An algorithm for restricted least squares regression. *Journal of the American Statistical Association*, **78**(384), 837–842, December 1983.

[28] Eckstein, J., Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Mathematics of Operations Research*, **18**(1), 202–226, February 1993.

[29] Eckstein, J. (1998). Approximate iterations in Bregman-function-based proximal algorithms. *Mathematical Programming*, **83**(1), 113–123.

[30] Escalante, R. and Raydan, M. (1996). Dykstra's algorithm for a constrained least-squares matrix problem. *Numerical Linear Algebra with Applications*, **3**(6), 459–471.

[31] Gaffke, N. and Mathar, R. (1989). A cyclic projection algorithm via duality. *Metrika*, **36**, 29–54.

[32] Glunt, W., Hayden, T. L., Hong, S. and Wells, J. (1990). An alternating projection algorithm for computing the nearest Euclidean distance matrix. *SIAM Journal on Matrix Analysis and Applications*, **11**(4), 589–600.

[33] Glunt, W. K. (1995). An alternating projections method for certain linear problems in a Hilbert space. *IMA Journal on Numerical Analysis*, **15**(2), 291–305.

[34] Han, S.-P. (1988). A successive projection method. *Mathematical Programming*, **40**, 1–14.

[35] Hildreth, C. (1957). A quadratic programming procedure. *Naval Research Logistics Quarterly*, **4**, 79–85, See also Erratum on page 361.

[36] Hiriart-Urruty, J.-B. and Lemaréchal, C. (1993). *Convex Analysis and Minimization Algorithms I*, volume 305 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag.

[37] Hundal, H. and Deutsch, F. (1997). Two generalizations of Dykstra's cyclic projection algorithm. *Mathematical Programming*, **77**(3), 335–355.

[38] Iusem, A. N., Proximal point methods in optimization, May 1994. Unpublished manuscript.

[39] Iusem, A. N. and De Pierro, A. R. (1991). On the convergence of Han's method for convex programming with quadratic objective. *Mathematical Programming*, **52**, 265–284.

[40] Kiwiel, K. C. (1995). Block-iterative surrogate projection methods for convex feasibility problems. *Linear Algebra and its Applications*, **215**, 225–260.

[41] Kiwiel, K. C. (1997). Free-steering relaxation methods for problems with strictly convex costs and linear constraints. *Mathematics of Operations Research*, **22**(2), 326–349.

[42] Kiwiel, K. C. (1997). Proximal minimization methods with generalized Bregman functions. *SIAM Journal on Control and Optimization*, **35**(4), 1142–1168.

[43] Kiwiel, K. C. (1998). Generalized Bregman projections in convex feasibility problems. *Journal of Optimization Theory and Applications*, **96**(1), 139–157.

[44] Lang, S. (1993). *Real and Functional Analysis*, volume 142 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition.

[45] Peressini, A. L., Sullivan, F. E. and Uhl, J. J. Jr. (1988). *The Mathematics of Nonlinear Programming*. Undergraduate texts in mathematics. Springer-Verlag, New York.

[46] Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton, NJ.

[47] Stromberg, K. R. (1981). *An Introduction to Classical Real Analysis*. Wadsworth International Group, Belmont, California.

[48] Teboulle, M. (1992). Entropic proximal mappings with applications to nonlinear programming. *Mathematics of Operations Research*, **17**(3), 670–690.

[49] Tseng, P. (1993). Dual coordinate ascent methods for non-strictly convex minimization. *Mathematical Programming*, **59**, 231–247.