Journal of Convex Analysis
Volume 6 (1999), No. 2, 319–333

Dykstra's Algorithm as the Nonlinear Extension of Bregman's Optimization Method

Lev M. Bregman
Institute for Industrial Mathematics,
4 Yehuda Hanakhtom Street, 84311 Beer-Sheva, Israel.
e-mail: bregman@math.bgu.ac.il

Yair Censor
Department of Mathematics, University of Haifa,
Mt. Carmel, 31905 Haifa, Israel.
e-mail: yair@mathcs2.haifa.ac.il

Simeon Reich
Department of Mathematics, The Technion - Israel Institute of Technology,
Technion City, 32000 Haifa, Israel.
e-mail: sreich@techunix.technion.ac.il

Received April 2, 1998
Revised manuscript received March 28, 1999

We show that Dykstra's algorithm with Bregman projections, which finds the Bregman projection of a point onto the nonempty intersection of finitely many closed convex sets, is actually the nonlinear extension of Bregman's primal-dual, dual coordinate ascent, row-action minimization algorithm. Based on this observation we give an alternative convergence analysis and a new geometric interpretation of Dykstra's algorithm with Bregman projections which complements recent work of Censor and Reich, Bauschke and Lewis, and Tseng.

Keywords: Bregman projection, convex programming, Dykstra's algorithm

1991 Mathematics Subject Classification: 47N10, 49M30, 90C25

1. Introduction

The Dykstra algorithm is an iterative procedure which (asymptotically) finds the nearest point projection (also called the orthogonal projection) of any given point onto the intersection of a given finite family of closed convex sets. It iterates by passing sequentially over the individual sets and projecting onto each one a deflected version of the previous iterate. The algorithm was first proposed and analyzed by Dykstra [26] and rediscovered by Han [31]. Published work on Dykstra's algorithm includes: Boyle and Dykstra [5], Gaffke and Mathar [30], Iusem and De Pierro [37], Crombez [20], Combettes [17], Bauschke and Borwein [1], Deutsch and Hundal [25], Hundal and Deutsch [32], Han and Lou [33], and Escalante and Raydan [29]. See also Robertson, Wright and Dykstra [43], and Dykstra [27].

Recently Censor and Reich [14] proposed a synthesis of Dykstra's algorithm with Bregman distances and obtained a new algorithm that solves the best approximation problem with Bregman projections. However, they established convergence of the resulting *Dyk-*

ISSN 0944-6532 / $ 2.50

stra algorithm with Bregman projections only when the constraints sets are half-spaces. Shortly thereafter Bauschke and Lewis [4] provided the first proof for general closed convex constraints sets. Their analysis relies on some strong properties of Bregman distances corresponding to Legendre functions which were treated earlier by Bauschke and Borwein [3]. Bauschke and Lewis [4] also discovered the close relationship between the Dykstra algorithm with Bregman projections and the very general and powerful algorithmic framework of Tseng [46], namely the *Dual Block Coordinate Ascent (DBCA)* methods.

There is, however, a completely different point of departure which leads exactly to the same Dykstra algorithm with Bregman projections. This is Bregman's optimization algorithm for the solution of the problem

$$\begin{aligned} &\min f(x), \\ &Ax \leq b, \\ &x \in \bar{S}, \end{aligned} \tag{1.1}$$

where A is a given real $m \times n$ matrix, $b \in R^m$, the m-dimensional Euclidean space, is a given vector, and $\bar{S}$ is the closure of the set S which is the *zone* of the, so-called, Bregman function $f(x)$. Bregman's optimization algorithm was originally proposed in [6]. It is a row-action (see Censor [9]) algorithm of the primal-dual type which achieves dual coordinate ascent by performing, sequentially, Bregman projections with respect to f onto the half-spaces in (1.1) or onto certain well-defined hyperplanes parallel to them. See Bregman [6], Censor and Lent [12] or Censor and Zenios [16, Algorithm 6.3.1]. For the special choice $f(x) = \frac{1}{2}\|x\|^2$ the algorithm coincides with Hildreth's algorithm [35] and D'Esopo's [24]; see also Lent and Censor [41]. Bregman's optimization algorithm was further studied and generalized in several directions; see, e.g., Censor and Lent [12] (a storage-efficient adaptation to linear inequalities interval constraints, modeled after the work of Herman and Lent [34]), De Pierro and Iusem [23] and Iusem and Zenios [38] (introduction of underrelaxation parameters), and Iusem [36] (dual convergence and rate of primal convergence). See also Censor and Zenios [16].

However, all these studies handled only linear (equalities, inequalities, or intervals) constraints in (1.1) and an extension of Bregman's optimization algorithm to general closed convex sets remained elusive.

We claim here that Dykstra's algorithm with Bregman projections is precisely the nonlinear extension of the above-mentioned Bregman's optimization algorithm. This recognition goes beyond the fact that the two algorithms coincide in the linear constraints case, as was shown by Censor and Reich [14]. It enables us to present here a new proof and convergence analysis of Dykstra's algorithm with Bregman projections, for almost cyclic control sequences, which rests on Bregman's original work [6]. Our new analysis differs from the approaches of Bauschke and Lewis [4] and of Tseng [46] in two fundamental ways. Firstly, it offers an intuitive geometric interpretation of the iterative steps of the algorithm, helping to better "understand" its action. Secondly, it results in a convergence theorem in which the conditions on f and on the constraints sets $\{C_i\}_{i=1}^m$ are different from those in [4] and [46], extending the applicability of the algorithm.

The present paper is laid out as follows. In Section 2 we present the algorithm and present several preliminary lemmas. The convergence theorem is proven in Section 3. We conclude by showing, in Section 4, the precise relation of the algorithm with the problem of finding the projection onto the intersection of closed convex sets, giving our new geometric interpretation, and describing the differences between the conditions in our convergence theorem and those in Bauschke and Lewis [4] and Tseng [46].

Projection methods have been studied intensively. They are applicable and useful in best approximation theory, optimization, statistics, partial differential equations, image reconstruction from projections, signal processing and other fields. See, e.g., the review and tutorial of Bauschke and Borwein [2] or Combettes [18, 19] and the references therein.

The use of Bregman projections, which seek to minimize some generalized distance other than the Euclidean, has in the past two decades propagated into a variety of problem areas, generating theoretical as well as practical consequences. These include: proximal point algorithms, e.g., Eckstein [28], Censor and Zenios [15], Kiwiel [40], and Teboulle [45]; variational inequalities, see, e.g., Censor, Iusem and Zenios [11] or Burachik and Scheimberg [8]; multiprojections, see Censor and Elfving [10]; paracontractions and other operators, see Censor and Reich [13]. Csiszár presented in [21] an axiomatic theory which leads to Bregman distances; in that connection see also Bregman and Naumova [7]. Bauschke and Borwein [3] introduced the Legendre/Bregman functions. Some of these developments are included also in Censor and Zenios [16], but the above list is by no means exhaustive.

Notions and notations from convex analysis are as in Rockafellar [44]. For any set C in R^n, $\bar{C}$, $\mathrm{int}(C)$, $\mathrm{ri}(C)$ and $\mathrm{bd}(C)$ denote the closure, interior, relative interior and boundary of C, respectively. For any closed proper convex function f on R^n, $\mathrm{dom}\, f = \{x \in R^n \mid f(x) < \infty\}$ and $\nabla f(x)$ denotes the gradient of f at x. A proper closed convex function f on R^n with $\mathrm{int}(\mathrm{dom}\, f) \neq \emptyset$ is said to be *Legendre* if it is strictly convex and differentiable on $\mathrm{int}(\mathrm{dom}\, f)$, and

$$\lim_{t \to +0} \langle \nabla f(x + t(y - x)), y - x \rangle = -\infty$$

for every $x \in \mathrm{bd}(\mathrm{dom}\, f)$ and $y \in \mathrm{int}(\mathrm{dom}\, f)$. A proper closed convex function f on R^n is called *co-finite* if

$$\lim_{t \to +\infty} f(tx)/t = +\infty$$

for every $x \in R^n$, $x \neq 0$.

2. The algorithm and preliminary results

Let f be a proper closed strictly convex function on $\mathrm{dom}\, f \subseteq R^n$ where $\mathrm{dom}\, f$ is a closed set and $\mathrm{int}(\mathrm{dom}\, f) \neq \emptyset$. The function f is further assumed to be continuous on $\mathrm{dom}\, f$ and differentiable on $\mathrm{int}(\mathrm{dom}\, f)$. Set $D_f(x,y) \triangleq f(x) - f(y) - \langle \nabla f(y), x - y \rangle$, defined for all $x \in \mathrm{dom}\, f$, $y \in \mathrm{int}(\mathrm{dom}\, f)$, where $\langle \cdot, \cdot \rangle$ denotes the inner product in R^n. The function D_f has a metric property: $D_f(x,y) \geq 0$, and $D_f(x,y) = 0$ iff $x = y$. See, e.g., [16, Lemma 2.1.1].

Assume also that D_f satisfies the following conditions:

D1 For every $\alpha \in R$ and for every $z \in \mathrm{dom}\, f$ the partial level set of D_f

$$L(z, \alpha) = \{x \in \mathrm{int}(\mathrm{dom}\, f) \mid D_f(z,x) \leq \alpha\}$$

is bounded.

D2 If $z^k \in \operatorname{dom} f$, $x^k \in \operatorname{int}(\operatorname{dom} f)$, $\lim_{k\to\infty} D_f(z^k, x^k) = 0$, $\lim_{k\to\infty} x^k = x^*$ and $\{z^k\}$ is bounded, then $\lim_{k\to\infty} z^k = x^*$.

D3 If $z^k \in \operatorname{int}(\operatorname{dom} f)$ and $z^k \to z$, then $\lim_{k\to\infty} D_f(z, z^k) = 0$.

Functions f that fulfill all the above conditions were termed in [12] Bregman functions with zone $S = \operatorname{int}(\operatorname{dom} f)$ and their class was denoted by $f \in \mathcal{B}(S)$. (In [12] and several subsequent publications this definition included an additional condition, concerning the boundedness of the second partial level set of D_f, first observed by Kiwiel [39] to be superfluous). See also [16, Chapter 2] or Bauschke and Borwein [3] for comprehensive analyses of the properties of D_f.

Let $C_i \subseteq R^n$, $i = 1, 2, \ldots, m$, be closed convex subsets of R^n, and let $C = \bigcap_{i=1}^m C_i$ be nonempty. We assume that $C \cap \operatorname{dom} f$ is not empty, and that $C_i \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset$, for all i (but $C \cap \operatorname{int}(\operatorname{dom} f)$ may be empty).

Consider the following problem, which generalizes (1.1):

$$\begin{gathered} \min f(x), \\ x \in C \cap \operatorname{dom} f. \end{gathered} \tag{2.1}$$

The best approximation problem, in the sense of Bregman's generalized distance D_f, from a point x^0 to the set intersection C is:

$$\begin{gathered} \min D_f(x, x^0), \\ x \in C \cap \operatorname{dom} f. \end{gathered} \tag{2.2}$$

However, we choose to make the additional assumption that f has a global minimum at $x^0 \in \operatorname{int}(\operatorname{dom} f)$. Thus we assume that

D4 There exists an $x^0 \in \operatorname{int}(\operatorname{dom} f)$ for which $\nabla f(x^0) = 0$.

Sufficient conditions for this usually involve certain coercivity assumptions on f, see, e.g., Peressini, Sullivan and Uhl [42, Section 1.4]. When **D4** is imposed, the problem (2.2) reduces to (2.1), which looks somewhat more generic, and the subsequent presentation and treatment become simpler. However, there is no loss of generality involved because it is possible to treat (2.2) directly without assumption **D4** along the same lines of reasoning, but at the expense of complicating the presentation.

To guarantee that Algorithm 2.4 below is well-defined we need also the following condition:

D5 The function f is Legendre.

Assumption **D5** implies the, so called, *zone consistency*, see, e.g., [16, Definition 2.2.1] and [3, Section 3].

Consider some examples of functions satisfying **D1** – **D5**. In each one of these examples we suppose the function f on R^n is of the form $f(x) = \sum_{j=1}^n g(x_j)$, where g is a function on R^1. It is not difficult to verify that the functions f satisfy **D1** – **D5**, for the following g:

Example 2.1. $g(x) = \frac{1}{p}|x|^p$, for $1 < p < \infty$; $\operatorname{dom} g = R^1$.

Example 2.2. $g(x) = -\frac{1}{p}x^p + ax$, for $0 < p < 1$, $a > 0$; $\operatorname{dom} g = [0, +\infty)$.

Example 2.3 (negative Shannon entropy). $g(x) = x \ln x - x$, (with the definition $0 \log 0 = 0$); $\operatorname{dom} g = [0, +\infty)$.

Some more examples of functions satisfying **D1** – **D3** and **D5** can be found in [3, Section 6] which appeared recently in this journal.

The nonlinear extension of Bregman's optimization algorithm is as follows.

Algorithm 2.4.

1. **Data at the beginning of the k-th iterative step:**
   1.1. Current approximation $x^k \in \operatorname{int}(\operatorname{dom} f)$;
   1.2. m vectors $a_i^k \in R^n$ and m real numbers α_i^k, $i = 1, 2, \dots, m$, such that each pair (a_i^k, α_i^k) defines a supporting hyperplane H_i^k for C_i, i.e.,

$$H_i^k = \{x \in R^n | \ \langle a_i^k, x \rangle = \alpha_i^k\},$$

   such that $\langle a_i^k, x \rangle \le \alpha_i^k$, for all $x \in C_i$, and $\langle a_i^k, \bar{x} \rangle = \alpha_i^k$, for some $\bar{x} \in C_i$.

2. **Initialization:**
   2.1. x^0 is the (global) minimum point of $f(x)$ on R^n, i.e.,

$$\nabla f(x^0) = 0. \tag{2.3}$$

   2.2. Set $a_i^0 = 0$ and $\alpha_i^0 = 0$, for all $i = 1, 2, \dots, m$.

3. **Iterative step:**
   3.1. Choose an operating control index $i_k \in \{1, 2, \dots, m\}$;
   3.2. Construct the function $F_k(x) \triangleq f(x) - \langle \nabla f(x^k) + a_{i_k}^k, x \rangle$ and solve the subproblem:

$$\begin{gathered} \min F_k(x), \\ x \in C_{i_k} \cap \operatorname{dom} f. \end{gathered} \tag{2.4}$$

   We show below (Lemma 2.7) that this problem has a solution $x^{k+1} \in \operatorname{int}(\operatorname{dom} f)$. For $i \neq i_k$ put $a_i^{k+1} = a_i^k$ and $\alpha_i^{k+1} = \alpha_i^k$; for $i = i_k$ let

$$a_{i_k}^{k+1} = a_{i_k}^k + \nabla f(x^k) - \nabla f(x^{k+1}) \tag{2.5}$$

$$\alpha_{i_k}^{k+1} = \langle a_{i_k}^{k+1}, x^{k+1} \rangle. \tag{2.6}$$

4. **Control sequence:**
   Suppose the indices i_k are chosen in an almost cyclic order, i.e., there exists an integer $T \ge m$ such that $\{1, 2, \dots, m\} \subseteq \{i_{r+1}, \dots, i_{r+T}\}$, for all r.

Remark 2.5. In the cyclic case, i.e., if $i_k - 1 \equiv k \bmod m$, this algorithm produces the same sequence $\{x^k\}$ as the Dykstra algorithm with Bregman projections of Censor and Reich [14] and Bauschke and Lewis [4].

Remark 2.6. If the function f is co-finite, then the mapping $y = \nabla f(x)$ is a one-to-one mapping of $\operatorname{int}(\operatorname{dom} f)$ onto R^n; see, e.g., Rockafellar [44, Theorem 26.5]. Thus there exists a point $z^k \in \operatorname{int}(\operatorname{dom} f)$ such that $\nabla f(z^k) = \nabla f(x^k) + a_{i_k}$, and x^{k+1}, defined in Step 3.2 of Algorithm 2.4, solves the problem $\min_{x \in C_{i_k} \cap \operatorname{dom} f} D_f(x, z^k)$.

The lemma below shows that Algorithm 2.4 is well-defined.

Lemma 2.7. *Let the conditions* **D1** – **D5** *hold. Then*

(i) *The problem* (2.4) *has a unique solution* $x^{k+1} \in \operatorname{int}(\operatorname{dom} f)$*;*

(ii) *The point* x^k *and the pairs* (a_i^k, α_i^k), $i = 1, 2, \dots, m$, *satisfy the conditions of Step* 1 *of* Algorithm 2.4, *for every* k.

Proof. By **D4**, $x^0 \in \operatorname{int}(\operatorname{dom} f)$; hence the statement (ii) is true for $k = 0$. Suppose (ii) is true for some $k \geq 0$. We show that the problem (2.4) has a unique solution x^{k+1} which satisfies the condition of Step 1.1, and a_i^{k+1} and α_i^{k+1}, defined in Step 3.2, satisfy the conditions of Step 1.2 of Algorithm 2.4.

Choose $\bar{x} \in C_{i_k} \cap \operatorname{int}(\operatorname{dom} f)$ (this intersection is assumed to be not empty), and let $\gamma \triangleq F_k(\bar{x})$. Denote $Q = \{x \in R^n \mid F_k(x) \leq \gamma\} \cap C_{i_k}$. It is clear that Q is nonempty and closed. Further,

$$D_f(x, x^k) = f(x) - f(x^k) - \langle \nabla f(x^k), x - x^k \rangle = F_k(x) + \langle a_{i_k}^k, x \rangle + \beta_k,$$

where $\beta_k = \langle \nabla f(x^k), x^k \rangle - f(x^k)$.

For every $x \in Q$, we have $F_k(x) \leq \gamma$, and $\langle a_{i_k}^k, x \rangle \leq \alpha_{i_k}^k$ because, by the induction hypothesis, the pair $(a_{i_k}^k, \alpha_{i_k}^k)$ defines a supporting hyperplane for C_{i_k}. Therefore $D_f(x, x^k) \leq \gamma + \alpha_{i_k}^k + \beta_k$, and Q is contained in the level set $\bar{L}(x^k, \gamma + \alpha_{i_k}^k + \beta_k) = \{x \in \operatorname{dom} f \mid D_f(x, x^k) \leq \gamma + \alpha_{i_k}^k + \beta_k\}$. Since $\bar{L}(x^k, r)$ is bounded for every r (see, e.g., [3, Theorem 3.7]), the set Q is also bounded. Hence the function $F_k(x)$ attains its minimum on Q at some point x^{k+1} which is also its minimal point on C_{i_k}. The minimal point x^{k+1} is unique because the function $F_k(x)$ is strictly convex; $x^{k+1} \in \operatorname{int}(\operatorname{dom} f)$ because $C_{i_k} \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset$, and the function $F_k(x)$ is Legendre.

Next we verify that the pair $(a_{i_k}^{k+1}, \alpha_{i_k}^{k+1})$ defines a supporting hyperplane $H_{i_k}^{k+1}$ for C_{i_k} at the point x^{k+1}.

Indeed, since x^{k+1} is the minimum point of $F_k(x)$ on C_{i_k}, we have:

$$\langle \nabla f(x^{k+1}) - \nabla f(x^k) - a_{i_k}^k, x - x^{k+1} \rangle \geq 0, \text{ for all } x \in C_{i_k}, \tag{2.7}$$

or

$$\langle a_{i_k}^{k+1}, x \rangle \leq \alpha_{i_k}^{k+1} \text{ for all } x \in C_{i_k}. \tag{2.8}$$

This completes the proof. □

Lemma 2.8. *For all* $k \geq 0$,

$$\nabla f(x^k) + \sum_{i=1}^{m} a_i^k = 0. \tag{2.9}$$

Proof. This is true for $k = 0$ by initialization and remains true, by induction, in accordance with (2.5). □

3. Convergence

Theorem 3.1. *Let all the assumptions made in* Section 2 *hold. Let the set* $I = \{1, 2, \ldots, m\}$ *be composed of two subsets:* $I_1 \subseteq I$ *such that* $(\operatorname{dom} f) \cap C \cap (\bigcap_{i \in I_1} \operatorname{ri} C_i)$ *is not empty, and* $I_2 = I \setminus I_1$ *such that* C_i *are polyhedral for* $i \in I_2$*. Then any sequence* $\{x^k\}$ *generated by* Algorithm 2.4 *converges to the solution of* (2.1).

Proof. Our proof consists of the following six steps.

Step 1. Introduce the numbers φ_k, defined by

$$\varphi_k \triangleq f(x^k) + \sum_{i=1}^{m} (\langle a_i^k, x^k \rangle - \alpha_i^k). \tag{3.1}$$

We show that the sequence $\{\varphi_k\}$ is increasing. Let $d_k \triangleq \varphi_{k+1} - \varphi_k$. We have

$$d_k = f(x^{k+1}) - f(x^k) + \sum_{i=1}^{m} (\langle a_i^{k+1}, x^{k+1} \rangle - \alpha_i^{k+1}) - \sum_{i=1}^{m} (\langle a_i^k, x^k \rangle - \alpha_i^k).$$

Since $a_i^{k+1} = a_i^k$ and $\alpha_i^{k+1} = \alpha_i^k$, for $i \neq i_k$, we can write

$$\begin{aligned} d_k \;&=\; f(x^{k+1}) - f(x^k) + \textstyle\sum_{i \neq i_k} \langle a_i^k, x^{k+1} - x^k \rangle \\ &\quad + (\langle a_{i_k}^{k+1}, x^{k+1} \rangle - \alpha_{i_k}^{k+1}) - (\langle a_{i_k}^k, x^k \rangle - \alpha_{i_k}^k). \end{aligned}$$

From the definition (2.6) of $\alpha_{i_k}^{k+1}$ we have

$$d_k = f(x^{k+1}) - f(x^k) + \sum_{i \neq i_k} \langle a_i^k, x^{k+1} - x^k \rangle - (\langle a_{i_k}^k, x^k \rangle - \alpha_{i_k}^k).$$

Taking into account (2.9), we get

$$d_k = f(x^{k+1}) - f(x^k) - \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \alpha_{i_k}^k - \langle a_{i_k}^k, x^{k+1} \rangle$$

or

$$d_k = D_f(x^{k+1}, x^k) + \alpha_{i_k}^k - \langle a_{i_k}^k, x^{k+1} \rangle. \tag{3.2}$$

Since $x^{k+1} \in C_{i_k}$, and the pair $(a_{i_k}^k, \alpha_{i_k}^k)$ defines a supporting hyperplane for C_{i_k}, we have

$$d_k \geq D_f(x^{k+1}, x^k) \geq 0, \tag{3.3}$$

and so the sequence $\{\varphi_k\}$ is increasing.

Step 2. Let $y \in C \cap \operatorname{dom} f$. Then

$$\begin{aligned} D_f(y, x^k) &= f(y) - f(x^k) - \langle \nabla f(x^k), y - x^k \rangle \\ &= f(y) - f(x^k) + \langle \textstyle\sum_{i=1}^m a_i^k, y - x^k \rangle \\ &= f(y) - f(x^k) - \textstyle\sum_{i=1}^m (\langle a_i^k, x^k \rangle - \alpha_i^k) \\ &\quad + \textstyle\sum_{i=1}^m \langle a_i^k, y - x^k \rangle + \sum_{i=1}^m (\langle a_i^k, x^k \rangle - \alpha_i^k) \\ &= f(y) - \varphi_k + \textstyle\sum_{i=1}^m (\langle a_i^k, y \rangle - \alpha_i^k). \end{aligned} \tag{3.4}$$

Since $y \in C$, we have $\langle a_i^k, y \rangle - \alpha_i^k \leq 0$, for all i. Hence, by Step 1,

$$D_f(y, x^k) \leq f(y) - \varphi_k \leq f(y) - \varphi_0, \tag{3.5}$$

and the sequence $\{x^k\}$ is bounded, by assumption **D1**. Also,

$$\varphi_k \leq f(y), \text{ for all } y \in C \cap \operatorname{dom} f, \tag{3.6}$$

i.e., $\{\varphi_k\}$ is bounded, and $\lim_{k\to\infty} \varphi_k$ exists. This, together with (3.3), implies that $D_f(x^{k+1}, x^k) \to 0$ as $k \to \infty$.

One more inequality which we can deduce from (3.4) is

$$\sum_{i=1}^m (\alpha_i^k - \langle a_i^k, y \rangle) \leq f(y) - \varphi_0, \text{ for all } y \in C \cap \operatorname{dom} f, \tag{3.7}$$

or, because the numbers $\alpha_i^k - \langle a_i^k, y \rangle$ are nonnegative, for all i,

$$\alpha_i^k - \langle a_i^k, y \rangle \leq f(y) - \varphi_0, \text{ for all } y \in C \cap \operatorname{dom} f \text{ and for all } i. \tag{3.8}$$

Step 3. Since the sequence $\{x^k\}$ is bounded, it has cluster points. We show that all such cluster points belong to C.

Let $\{x^{k_t}\}$ be a convergent subsequence of $\{x^k\}$ and $x^* = \lim_{t\to\infty} x^{k_t}$. Since $D_f(x^{k+1}, x^k) \to 0$, we have, by assumption **D2**, that the subsequence $\{x^{k_t+1}\}$ converges to the same point x^*. Repeating this, we get that the subsequences $\{x^{k_t+2}\}, \ldots, \{x^{k_t+T-1}\}$ all converge to x^*. Therefore, the union of these sequences also converges to x^*, and according to the almost cyclic choice of the operating control index, for each $i \in \{1, 2, \ldots, m\}$, there exists an infinite sequence of integers $\{j(i,t)\}_{t\geq 0}$, such that $0 \leq j(i,t) \leq T-1$, and $x^{k_t+j(i,t)} \in C_i$, for all $t \geq 0$. Hence, $\lim_{t\to\infty} x^{k_t+j(i,t)} = x^*$, and $x^* \in C_i$. So $x^* \in C$.

Step 4. Take $i \in I_1$. Let L_i be the subspace which is parallel to the affine hull *aff* C_i. Then there exists a unique representation $a_i^k = h_i^k + g_i^k$, where $h_i^k \in L_i$ and g_i^k is orthogonal to L_i. We show first that $\{h_i^k\}_{k\geq 0}$ are bounded sequences, for $i \in I_1$.

Denote by x_i^k the vector x^l, where $l \leq k$ is the maximal index for which $i_{l-1} = i$. It is clear that $x_i^k \in C_i$ and that $\alpha_i^k = \langle a_i^k, x_i^k \rangle$.

Let $y \in C \cap \operatorname{dom} f$. Then

$$\left\langle a_i^k, y\right\rangle - \alpha_i^k = \left\langle a_i^k, y - x_i^k\right\rangle = \left\langle h_i^k, y - x_i^k\right\rangle. \tag{3.9}$$

The latter equality is valid because both y and x_i^k belong to C_i, $y - x_i^k \in L_i$ and $\left\langle g_i^k, y - x_i^k\right\rangle = 0$.

Take a vector $\bar{y} \in (\operatorname{dom} f) \cap C \cap (\bigcap_{i \in I_1} \operatorname{ri} C_i)$. Then there exists a positive ε such that for every $r \in L_i$, $i \in I_1$, with $\|r\| \leq \varepsilon$, we have $\bar{y} + r \in C_i$. Let $r \triangleq \varepsilon h_i^k / \|h_i^k\|$. Then $\tilde{y} \triangleq \bar{y} + r \in C_i$, and since $a_i^k = h_i^k + g_i^k$,

$$0 \geq \left\langle a_i^k, \tilde{y}\right\rangle - \alpha_i^k = \left\langle a_i^k, \bar{y}\right\rangle - \alpha_i^k + \varepsilon \left\langle a_i^k, h_i^k / \|h_i^k\|\right\rangle = \left\langle a_i^k, \bar{y}\right\rangle - \alpha_i^k + \varepsilon \|h_i^k\|.$$

Since $\bar{y} \in C$, we get, by (3.8),

$$\varepsilon \|h_i^k\| \leq \alpha_i^k - \left\langle a_i^k, \bar{y}\right\rangle \leq f(\bar{y}) - \varphi_0,$$

implying that the sequences $\{h_i^k\}$ are bounded, i.e., there exists a real number τ such that $\|h_i^k\| \leq \tau$, for all $k \geq 0$ and $i \in I_1$. Hence, by (3.9),

$$|\left\langle a_i^k, y\right\rangle - \alpha_i^k| = |\left\langle h_i^k, y - x_i^k\right\rangle| \leq \tau \|y - x_i^k\|. \tag{3.10}$$

Step 5. Take $i \in I_2$. Then C_i is polyhedral, i.e., an intersection of a finite number of half-spaces,

$$C_i = \{x \in R^n \mid \langle b_i^p, x\rangle \leq \beta_i^p,\ p = 1, 2, \ldots, P_i\}.$$

Consider the linear programming problem:

$$\begin{gathered} \max \left\langle a_i^k, x\right\rangle, \\ x \in C_i. \end{gathered} \tag{3.11}$$

This problem has x_i^k as a solution, and $\max \left\langle a_i^k, x\right\rangle = \alpha_i^k$. By the duality theorem (see, e.g., Dantzig [22, Chapter 6]), there exist nonnegative numbers λ_{pi}^k, $p = 1, 2, \ldots, P_i$, such that

$$a_i^k = \sum_{p=1}^{P_i} \lambda_{pi}^k b_i^p, \tag{3.12}$$

$$\alpha_i^k = \sum_{p=1}^{P_i} \lambda_{pi}^k \beta_i^p, \tag{3.13}$$

$$\lambda_{pi}^k(\left\langle b_i^p, x_i^k\right\rangle - \beta_i^p) = 0. \tag{3.14}$$

Divide the set of indices $\{1, 2, \ldots, P_i\}$ into two subsets:

$$P_i^1 \triangleq \{p \in \{1, 2, \ldots, P_i\} \mid \langle b_i^p, y \rangle < \beta_i^p, \text{ for some } y \in C \cap \operatorname{dom} f\},$$

and

$$P_i^2 \triangleq \{p \in \{1, 2, \ldots, P_i\} \mid \langle b_i^p, y \rangle = \beta_i^p, \text{ for all } y \in C \cap \operatorname{dom} f\}.$$

Let $y \in C \cap \operatorname{dom} f$. Using (3.12) and (3.13), we obtain

$$\left\langle a_i^k, y \right\rangle - \alpha_i^k = \sum_{p=1}^{P_i} \lambda_{pi}^k (\langle b_i^p, y \rangle - \beta_i^p) = \sum_{p \in P_i^1} \lambda_{pi}^k (\langle b_i^p, y \rangle - \beta_i^p). \tag{3.15}$$

There exist $\varepsilon_i > 0$ and $\bar{y}^i \in C \cap \operatorname{dom} f$ such that $\beta_i^p - \langle b_i^p, \bar{y}^i \rangle > \varepsilon_i$, for all $p \in P_i^1$. Defining $\varepsilon = \min_i \varepsilon_i$ and $\bar{y} \triangleq \sum_{i \in I_2} \frac{1}{|I_2|} \bar{y}^i$ we conclude that $\varepsilon > 0$ and $\bar{y} \in C \cap \operatorname{dom} f$ are such that $\beta_i^p - \langle b_i^p, \bar{y} \rangle > \varepsilon$, for all $p \in P_i^1$ and all $i \in I_2$.

Using (3.8) and (3.15) and taking into account that λ_{pi}^k and $\beta_i^p - \langle b_i^p, \bar{y} \rangle$ are nonnegative, we have, for all $p \in P_i^1$,

$$f(\bar{y}) - \varphi_0 \geq \alpha_i^k - \left\langle a_i^k, \bar{y} \right\rangle = \sum_{p \in P_i^1} \lambda_{pi}^k (\beta_i^p - \langle b_i^p, \bar{y} \rangle) \geq \lambda_{pi}^k (\beta_i^p - \langle b_i^p, \bar{y} \rangle) \geq \varepsilon \lambda_{pi}^k.$$

Hence λ_{pi}^k are bounded by $(f(\bar{y}) - \varphi_0)/\varepsilon$.

Further, (3.14) implies that $\left\langle b_i^p, x_i^k \right\rangle - \beta_i^p = 0$ if $\lambda_{pi}^k > 0$, so we obtain, from (3.15),

$$\left\langle a_i^k, y \right\rangle - \alpha_i^k = \sum_{p \in P_i^1} \lambda_{pi}^k \left\langle b_i^p, y - x_i^k \right\rangle. \tag{3.16}$$

Since λ_{pi}^k are bounded for $p \in P_i^1$, there exists a real number ρ such that

$$| \left\langle a_i^k, y \right\rangle - \alpha_i^k | \leq \rho \| y - x_i^k \|, \tag{3.17}$$

for all $y \in C \cap \operatorname{dom} f$.

Comparing this with (3.10), we see that the inequality (3.17) is valid for all $i = 1, 2, \ldots, m$. Note that we may take a common value $\tau = \rho$ for both (3.10) and (3.17).

Step 6. Let x^* be a cluster point of $\{x^k\}$, and let $\{x^{k_t}\}$ be a subsequence converging to x^*. According to Step 3 of the proof, $x^* \in C$ and $x^* \in \operatorname{dom} f$ because $\operatorname{dom} f$ is closed. As we have seen in Step 3, the subsequences $\{x^{k_t+j}\}$, where $0 \leq j < T$, converge to x^* as well. Since the set $\{i_{k_t}, i_{k_t+1}, \ldots, i_{k_t+T-1}\}$ contains the set of all indices $\{1, 2, \ldots, m\}$, the subsequence $\{x^{k_s}\} \triangleq \{x^{k_t+T-1}\}$ is such that the subsequences $\{x_i^{k_s}\}$ converge to x^*, for every $i = 1, 2, \ldots, m$.

Applying (3.17) for $y = x^*$, we get

$$| \left\langle a_i^k, x^* \right\rangle - \alpha_i^k | \leq \rho \| x^* - x_i^k \|.$$

Since the subsequences $\{x_i^{k_s}\}$ converge to x^*, we have

$$\lim_{s\to\infty} |\langle a_i^{k_s}, x^*\rangle - \alpha_i^{k_s}| = 0, \tag{3.18}$$

for every i.

Applying (3.4) for $y = x^*$, we have

$$D_f(x^*, x^k) = f(x^*) - \varphi_k + \sum_{i=1}^{m}(\langle a_i^k, x^*\rangle - \alpha_i^k).$$

According to Step 2 of the proof, $\lim_{k\to\infty}\varphi_k = \lim_{s\to\infty}\varphi_{k_s}$ exists. Hence, by assumption **D3** and (3.18),

$$0 = \lim_{s\to\infty} D_f(x^*, x^{k_s}) + \lim_{s\to\infty} \sum_{i=1}^{m}(\alpha_i^{k_s} - \langle a_i^{k_s}, x^*\rangle) = f(x^*) - \lim_{s\to\infty}\varphi_{k_s},$$

i.e.,

$$\lim_{k\to\infty}\varphi_k = f(x^*).$$

Since, by (3.6), $\lim_{k\to\infty}\varphi_k \leq \min_{x\in C\cap \operatorname{dom} f} f(x)$, we have

$$f(x^*) = \min_{x\in C\cap \operatorname{dom} f} f(x),$$

and because f is strictly convex and has a unique minimum on $C \cap dom\, f$, it follows that the whole sequence $\{x^k\}$ converges to x^*. This completes the proof. □

4. Connections, comparisons and interpretation

In this section we compare our Theorem 3.1 with the convergence theorems previously obtained by Censor and Reich [14, Theorem 3.1] and by Bauschke and Lewis [4, Theorem 3.2], as well as with the result that can be obtained by applying Tseng's general framework [46]. We first note that although, at first glance, the problem considered in [14] and in [4], namely the minimization of $D_f(\cdot, x^0)$ over $C \cap dom\, f$, differs from our problem (2.1), they are, in fact, equivalent. This is because, on the one hand, $D_f(\cdot, x^0)$ is a strictly convex function on $\operatorname{dom} f$, and on the other hand, it coincides with f itself when $\nabla f(x^0) = 0$. We also remark that although Algorithm 2.1 in [14] employs a double iteration, it actually generates, in the cyclic order case, the same sequence which is generated by Algorithm 2.4 in the present paper.

The assumptions and the proofs of these four convergence results are all different. As mentioned in the Introduction, Censor and Reich used Bregman's original method to establish convergence only when the constraints sets are half-spaces. Bauschke and Lewis used a non-trivial extension of Boyle and Dykstra's proof [5], as well as the tools developed by Bauschke and Borwein in [3]. Their control sequence is cyclic, and their constraint qualification condition is

$$\operatorname{int}(\operatorname{dom} f) \cap C \neq \emptyset. \tag{4.1}$$

A convergence result for our Algorithm 2.4 can be derived from Tseng's very general framework by extending his derivation of Han's [31] theorem in [46, Section 4]. The resulting constraint qualification condition is

$$\operatorname{int}(\operatorname{dom} f) \cap C \cap \Big(\bigcap_{i \in I_1} \operatorname{ri} C_i\Big) \neq \emptyset. \tag{4.2}$$

Finally, our condition (see Theorem 3.1) is

$$\operatorname{dom} f \cap C \cap \Big(\bigcap_{i \in I_1} \operatorname{ri} C_i\Big) \neq \emptyset. \tag{4.3}$$

Condition (4.2) is clearly more restrictive than (4.3), but this does not mean that our result contains Tseng's result. This is because Tseng's framework is more general in other respects, not addressed in our study. He also has some results on dual convergence.

All the above-mentioned results also impose different requirements on the objective function f. Both Bauschke and Lewis [4] and Tseng [46] require co-finiteness of the function f, but we do not need this condition. For example, the function f in Example 2.2 satisfies the condition **D1** $-$ **D5**, but it is not co-finite.

The proof of our Theorem 3.1 is in the spirit of Bregman's original approach [6]. It offers the following useful geometric interpretation (cf. Lemma 2.7): At each step of our algorithm, having chosen an index $i_k \in \{1, 2, \ldots, m\}$, we replace the set C_{i_k} with one particular half-space drawn from the (in general, infinite) family of all the half-spaces the intersection of which equals C_{i_k}. Note, in particular, that if, at the k-th stage, the problem (2.1) is replaced by the problem

$$\begin{gathered} \min f(x), \\ \langle a_i^k, x \rangle \leq \alpha_i^k,\ i = 1, 2, \ldots, m, \end{gathered} \tag{4.4}$$

$\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ is the Lagrange multipliers vector, and

$$L_k(x, \lambda) \triangleq f(x) + \sum_{i=1}^{m} \lambda_i(\langle a_i^k, x \rangle - \alpha_i^k)$$

is the corresponding Lagrangian function, then (2.9) shows that $\nabla_x L_k(x^k, \mathbf{1}) = 0$, where $\mathbf{1} = (1, 1, \ldots, 1)$. Moreover, $\varphi_k = L_k(x^k, \mathbf{1})$ by (3.1).

It is clear that [14, Theorem 3.1] is a special case of Theorem 3.1 in the present paper. However, the two other results mentioned above neither contain nor are they contained in Theorem 3.1. For example, for non-polyhedral sets, (4.1) is, in some sense, the weakest constraint qualification, but the conditions imposed on f in [4] are rather strong. Furthermore, our condition (4.3) covers, for instance, the case when f is the negative Shannon entropy function (Example 2.3) and C does not contain any positive vector (in this connection see the discussion in [14, Section 4]), while (4.1) and (4.2) do not. On the other hand, (4.1) covers the case when $f = \frac{1}{2}\|x\|^2$ and $\bigcap_{i \in I_1} \operatorname{ri} C_i = \emptyset$, but (4.2) and (4.3) do not.

Acknowledgements. We thank two anonymous referees for many constructive comments which helped to improve the first version of this manuscript. We also thank Krzysztof Kiwiel for reading the paper and sending us useful comments.

The work of Y. Censor and S. Reich was partially supported by the Israel Science Foundation founded by the Israel Academy of Sciences and Humanities. The work of Y. Censor was also partially supported by NIH grant HL-28438 at the Medical Image Processing Group (MIPG), Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA. The work of S. Reich was also partially supported by the Fund for the Promotion of Research at the Technion and by the Technion VPR Fund.

References

[1] H. H. Bauschke, J. M. Borwein: Dykstra's alternating projection algorithm for two sets, Journal of Approximation Theory 79 (1994) 418–443.

[2] H. H. Bauschke, J. M. Borwein: On projection algorithms for solving convex feasibility problems, SIAM Review 38 (1996) 367–426.

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

[4] H. H. Bauschke, A. S. Lewis: Dykstra's algorithm with Bregman projections: a convergence proof, Technical Report, Dept. of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, N2L3G1, Canada, June 1998. Optimization, accepted for publication.

[5] J. P. Boyle, R. L. Dykstra: A method for finding projections onto the intersection of convex sets in Hilbert spaces, In: Advances in Order Restricted Statistical Inference, Lecture Notes in Statistics, Springer-Verlag, Berlin, Germany, Volume 37 (1986) 28–47.

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

[7] L. M. Bregman, N. I. Naumova: Arbitration solutions with ideal point generated by systems of functions, Soviet Mathematics Doklady 30 (1984) 583–587.

[8] R. S. Burachik, S. Scheimberg: A proximal point method for the variational inequality problem in Banach spaces, Technical Report ES-470, COPPE-UFRJ, Rio de Janeiro, Brazil, June 1998.

[9] Y. Censor: Row-action methods for huge and sparse systems and their applications, SIAM Review 23 (1981) 444–466.

[10] Y. Censor, T. Elfving: A multiprojection algorithm using Bregman projections in a product space, Numerical Algorithms 8 (1994) 221–239.

[11] Y. Censor, A. N. Iusem, S. A. Zenios: An interior point method with Bregman functions for the variational inequality problem with paramonotone operators, Mathematical Programming 81 (1998) 373–400.

[12] Y. Censor, A. Lent: An iterative row-action method for interval convex programming, Journal of Optimization Theory and Applications 34 (1981) 321–353.

[13] Y. Censor, S. Reich: Iterations of paracontractions and firmly nonexpansive operators with applications to feasibility and optimization, Optimization 37 (1996) 323–339.

[14] Y. Censor, S. Reich: The Dykstra algorithm with Bregman projections, Communications in Applied Analysis 2 (1998) 407–419.

[15] Y. Censor, S. A. Zenios: Proximal minimization algorithm with D-functions, Journal of Optimization Theory and Applications 73 (1992) 451–464.

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

[17] P. L. Combettes: Signal recovery by best feasible approximation, IEEE Transactions on Image Processing IP-2 (1993) 269–271.

[18] P. L. Combettes: The foundations of set theoretic estimation, Proceedings of the IEEE 81 (1993) 182–208.

[19] P. L. Combettes: The convex feasibility problem in image recovery, Advances in Imaging and Electron Physics 95 (1996) 155–270.

[20] G. Crombez: Finding projections onto the intersection of convex sets in Hilbert spaces, Numerical Functional Analysis and Optimization 16 (1995) 637–652.

[21] I. Csiszár: Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems, The Annals of Statistics 19 (1991) 2032–2066.

[22] G. B. Dantzig: Linear Programming and Extensions, Princeton University Press, Princeton, NJ, USA, 1963.

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

[24] D. A. D'Esopo: A convex programming procedure, Naval Research Logistics Quarterly 6 (1959) 33–42.

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

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

[27] R. L. Dykstra: An iterative procedure for obtaining I-projections onto the intersection of convex sets, The Annals of Probability 13 (1985) 975–984.

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

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

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

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

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

[33] S. -P. Han, G. Lou: A parallel algorithm for a class of convex programs, SIAM Journal on Control and Optimization 26 (1988) 345–355.

[34] G. T. Herman, A Lent: A family of iterative quadratic optimization algorithms for pairs of inequalities with application in diagnostic radiology, Mathematical Programming Study 9 (1978) 15–29.

[35] C. Hildreth: A quadratic programming procedure, Naval Research Logistics Quarterly 4 (1957) 79–85. Erratum, ibid., p. 361.

[36] A. N. Iusem: On dual convergence and the rate of primal convergence of Bregman's convex programming method, SIAM Journal on Optimization 1 (1991) 401–423.

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

[38] A. N. Iusem, S. A. Zenios: Interval underrelaxed Bregman's method with an application, Optimization 35 (1995) 227–250.

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

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

[41] A. Lent, Y. Censor: Extensions of Hildreth's row-action method for quadratic programming, SIAM Journal on Control and Optimization 18 (1980) 444–454.

[42] A. L. Peressini, F. E. Sullivan, J. J. Uhl jun.: The Mathematics of Nonlinear Programming, Springer-Verlag, New York etc., 1988.

[43] T. Robertson, F. T. Wright, R. L. Dykstra: Order Restricted Statistical Inference, John Wiley & Sons, Chichester, UK, 1988.

[44] R. T. Rockafellar: Convex Analysis, Princeton University Press, Princeton, NJ, USA, 1970.

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

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