JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 5, No. 3, DECEMBER 1986

A Relaxed Version of Bregman's Method for Convex Programming

A. R. DE PIERRO[1] AND A. N. IUSEM[2]

Communicated by O. L. Mangasarian

Abstract. A new type of relaxation for Bregman's method, an iterative primal-dual algorithm for linearly constrained convex programming, is presented. It is shown that the new relaxation procedure generalizes the usual concept of relaxation and preserves the convergence properties of Bregman's algorithm for a suitable choice of the relaxation parameters. For convergence, Bregman's method requires that the objective function satisfy certain conditions. A sufficient and easily checkable condition for these requirements to hold is also given.

Key Words. Linearly constrained convex programming, entropy optimization, large and sparse matrices, nonorthogonal projections.

1. Introduction

We consider the following linearly constrained convex programming problem:

$$\min f(x), \tag{1a}$$

$$\text{s.t. } Ax \leq b, \tag{1b}$$

where $f: \Lambda \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ is a continuously differentiable function, A is an $m \times n$-matrix, and b is an m-vector. Our relaxed version of Bregman's method is an algorithm for solving (1) when f belongs to a family B (Definition 2.1 below) which includes the Euclidean norm $\|x\|$ as well as the $x \ln x$ entropy function. Problem (1) with f in B arises in various fields of applications, like transport planning (Ref. 1) and image reconstruction from projections (Refs. 2 and 3). Generally, the matrices associated with such

[1] Associate Professor, Instituto de Matemática, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.
[2] Researcher, Instituto de Matemática Pura e Aplicada, Rio de Janeiro, Brazil.

0022-3239/86/1200-0421$05.00/0

applications are very large and sparse with no detectable structure pattern. In such a case, row-action methods (Ref. 4) have been shown to be successful even for matrices with dimensions exceeding 100,000 (Ref. 2).

Bregman's method (Refs. 5 and 6) as well as Hildreth's method (Ref. 8), of which it is a generalization, are row-action methods. Hildreth's algorithm solves (1) for the case $f(x) = \|x\|$, and can be described as follows: If x^k is the current iterate and H_k is the hyperplane associated with one of the constraints of (1) (i.e., the set of points which satisfy such constraint with equality), then x^{k+1} is the orthogonal projection of x^k on H_k if x^k does not satisfy that constraint, and the orthogonal projection of x^k on $\bar{H}_k$, a hyperplane parallel to H_k lying between x^k and H_k, otherwise. The location of $\bar{H}_k$ is determined by the current value of z^k, a sequence of dual variables generated simultaneously with the primal sequence x^k. The constraints are used in a cyclic way. Bregman's method extends Hildreth's method to the minimization of any function f belonging to B (which is called the set of Bregman functions). The only difference with respect to the previous description is that orthogonal projections are substituted by so-called Bregman projections (Definition 2.2) which are related to the function f.

The usual concept of relaxation, applicable to a wide class of iterative algorithms, can be summarized in the following way: If the original (unrelaxed) algorithm generates a sequence

$$x^{k+1} = f_k(x^k),$$

then the relaxed algorithm is defined as

$$x^{k+1} = (1-\alpha_k)x^k + \alpha_k f_k(x^k),$$

where α_k is a sequence of real numbers called relaxation parameters. Usually it is required that the relaxation parameters lie in the interval $(0, 2)$. The algorithm is said to be overrelaxed when α_k is constrained to belong to $(1, 2)$ and underrelaxed if it has to belong to $(0, 1)$. The option of using relaxation parameters has been shown to be a very important tool in practical implementation of row-action methods (see the discussions in Section 6 of Ref. 3, Section 6.2 of Ref. 6, and the results in Ref. 7). A relaxed version of Hildreth's algorithm has been shown to converge in Ref. 8, with the relaxation parameters belonging to a compact subset of $(0, 2)$, but no relaxation strategy retaining convergence properties has been proposed up to now for Bregman's algorithm.

The idea behind our relaxation strategy is the following: relax the constraints before computing the Bregman projection. More precisely, if at iteration k the ith constraint: $\langle a^i, x\rangle \le b_i$ is to be used, we substitute it by the relaxed constraint

$$\langle a^i, x\rangle \le \alpha_k b_i + (1-\alpha_k)\langle a^i, x^k\rangle,$$

and then we apply Bregman's procedure, where α_k is the relaxation parameter. When f is the Euclidean norm, this approach generates the same sequence as the usual relaxation scheme described above, and so it becomes the relaxed Hildreth's algorithm studied in Ref. 8. We prove convergence of the algorithm for any Bregman function when the sequence $\{\alpha_k\}$ is included in a compact subset of the interval $(0, 1]$ (underrelaxation). If the function f satisfies an additional condition [see (29) below], our convergence proof holds for $\{\alpha_k\}$ included in a compact subset of $(0, 2)$. Since quadratic positive-definite functions satisfy this condition, the convergence proof of Ref. 8 is a particular case of our results.

In Refs. 8 and 6, Hildreth's and Bregman's methods, respectively, are studied under almost cyclical control. This means that the m constraints are not used cyclically, but there exists a number $r \geq m$ such that all constrains are used in any block of r consecutive iterations. This extension becomes useful to prove convergence of special versions of the algorithms adapted to the interval problem (k.e., when the constraints are of the form $c \leq Ax \leq d$). Our proofs are given under almost cyclical control, although we have not been able to find a satisfactory convergent relaxation scheme for these special versions.

Finally, we give a sufficient and easily checkable condition for a function defined on all $\mathbb{R}^n$ to belong to the set B of Bregman functions.

In Section 2, Bregman functions and projections are introduced with some important properties. The relaxed algorithm is presented in Section 3, and its convergence is established in Section 4. Section 5 discusses the sufficient condition for Bregman functions, and Section 6 consists of some final remarks.

Here and below $\mathbb{R}^n$ stands for the Euclidean n-space and $\mathbb{R}^n_+$ is the orthant of nonnegative vectors. Vectors are assumed to be column vectors, and T denotes transpose. $\bar{S}$ is the closure of the set S, $\langle\ ,\ \rangle$ is the usual inner product in $\mathbb{R}^n$, and $\|\cdot\|$ is the Euclidean norm.

2. Bregman Functions and Projections

In this section, we follow closely Censor and Lent (Ref. 6). Let Λ be a subset of $\mathbb{R}^n$, and let $f: \Lambda \to \mathbb{R}$. Let S be a nonempty, convex set such that $\bar{S} \subseteq \Lambda$. Assume that $f(x)$ has continuous first partial derivatives at any $x \in S$, and denote by $\nabla f(x)$ its gradient at x.

From f, construct the function D: $\bar{S} \times S \subset \mathbb{R}^{2n} \to \mathbb{R}$ by

$$D(x, y) = f(x) - f(y) - \langle \nabla f(y), x - y \rangle. \tag{2}$$

Consider the partial level sets of D, for $\alpha \in \mathbb{R}$,

$$L_1(y, \alpha) = \{x \in \bar{S}: D(x, y) \leq \alpha\}, \tag{3a}$$

$$L_2(x, \alpha) = \{y \in S: D(x, y) \leq \alpha\}. \tag{3b}$$

Definition 2.1. A function $f: \Lambda \subseteq \mathbb{R}^n \to \mathbb{R}$ is called a Bregman function if there exists a nonempty, open convex set S, such that $\bar{S} \subseteq \Lambda$ and

(i) $f(x)$ is continuously differentiable at every $x \in S$;
(ii) $f(x)$ is strictly convex on $\bar{S}$;
(iii) $f(x)$ is continuous on $\bar{S}$;
(iv) for every $\alpha \in \mathbb{R}$, the partial level sets $L_1(y, \alpha)$ and $L_2(x, \alpha)$ are bounded for every $y \in S$, and every $x \in \bar{S}$, respectively;
(v) if $y^k \xrightarrow[k\to\infty]{} y^* \in \bar{S}$, then $D(y^*, y^k) \xrightarrow[k\to\infty]{} 0$;
(vi) if $D(x^k, y^k) \xrightarrow[k\to\infty]{} 0$, $y^k \xrightarrow[k\to\infty]{} y^* \in \bar{S}$, and $\{x^k\}$ is bounded, then $x^k \xrightarrow[k\to\infty]{} y^*$.

We denote the family of Bregman functions by B and refer to the set S as the zone of the function f.

Proposition 2.1. For every $f \in B$, $D(x, y) \geq 0$ and $D(x, y) = 0$, iff $x = y$.

Proof. See Ref. 6, Lemma 2.1. □

Let us now define Bregman projections.

Definition 2.2. Given a closed convex set $C \subset \mathbb{R}^n$ such that $C \cap \bar{S} \neq \emptyset$, let $P_C: S \to \mathbb{R}^n$ be defined as

$$P_C(y) = \underset{x \in C \cap \bar{S}}{\arg\min}\, D(x, y). \tag{4}$$

We call $P_c(y)$ the Bergman projection of y on C (with respect to f). Existence and uniqueness of Bregman projections when f is a Bregman function are guaranteed by Lemma 2.2 of Ref. 6.

A key role in Bregman's method is played by Bregman projections on hyperplanes. We take a closer look at such projections now.

Definition 2.3. (i) A function $f \in B$ with zone S is said to be zone consistent with respect to the hyperplane H, if $P_H(y) \in S$ for every $y \in S$.

(ii) A function $f \in B$ with zone S is strongly zone consistent with respect to the hyperplane H and the point $y \in S$, if it is zone consistent with respect to H as well as every hyperplane H' parallel to H which lies between y and H.

If H is given by

$$H=\{x\in\mathbb{R}^n, \langle a, x\rangle = b\}$$

and f is zone consistent with respect to H, then $P_H(y)$ is the unique solution of

$$\nabla f(x)=\nabla f(y)+\lambda a, \tag{5a}$$

$$\langle a, x\rangle = b, \tag{5b}$$

where λ is a real number uniquely determined when a and b are fixed, i.e., is the unique real number for which (5) has a solution. This statement is proved in Lemma 3.1 of Ref. 6. We will write

$$\lambda=\Pi_H(y) \tag{6}$$

and call λ the Bregman parameter associated with the Bregman projection of y on H (for the representation of H given by a and b).

Proposition 2.2. Let

$$H=\{x\in\mathbb{R}^n: \langle a, x\rangle = b\}.$$

If f is a Bregman function zone consistent with respect to H, then, for a and b fixed,

$$\Pi_H(y)(b-\langle a, y\rangle)>0, \qquad \text{if } \langle a, y\rangle \neq b.$$

Proof. See Lemma 3.2 in Ref. 6. □

3. Relaxed Bregman's Method

In this section, we describe the relaxed Bregman's method for solving problem (1). Let

$$A\in\mathbb{R}^{m\times n}, \qquad b\in\mathbb{R}^m, \qquad C=\{x\in\mathbb{R}^m: Ax\leq b\},$$

and let $f\colon \Lambda\subseteq\mathbb{R}^n\to\mathbb{R}$ be a Bregman function with zone S. Let a^i be rows of A, $a^i\neq 0$; and let e^i be the vector with components $e^i_j=\delta_{ij}$ (Kronecker's delta). Assume that f is strongly zone consistent with respect to the hyperplanes

$$H_i=\{x\in\mathbb{R}^n: \langle a^i, x\rangle = b_i\}, \qquad 1\leq i\leq m.$$

Let $\epsilon>0$ and $\{\alpha_k\}_{k=0}^{\infty}$ be a sequence of real numbefs such that $\epsilon\leq\alpha_k\leq 1$, which will be called the relaxation parameters.

Consider an almost cyclical control sequence $i(k)$, i.e., a sequence $i(k) \in \{1, 2, \ldots, m\}$ such that there is a constant r so that, for all $j \in \{1, 2, \ldots, m\}$ and all positive integer k, there exists $l \in \{1, \ldots, r\}$ with $j = i(k+l)$.

Given an arbitrary $z^0 \in \mathbb{R}_+^M$, define a sequence $\{(x^k, z^k)\} \subset \mathbb{R}^{n+m}$ by

$$\nabla f(x^0) = -A^T z^0, \tag{7}$$

$$\nabla f(x^{k+1}) = \nabla f(x^k) + c_k a^{i(k)}, \tag{8}$$

$$z^{k+1} = z^k - c_k e^{i(k)}, \tag{9}$$

where

$$c_k = \min\{z_{i(k)}^k, \beta_k\}, \tag{10}$$

$$\beta_k = \Pi_{H(k)}(x^k), \tag{11}$$

$$H(k) = \{x \in \mathbb{R}^n : \langle a^{i(k)}, x \rangle = \alpha_k b_{i(k)} + (1 - \alpha_k)\langle a^{i(k)}, x^k \rangle\}. \tag{12}$$

From now on, the preceding notation will be used with reference to the algorithm defined by (7)-(12). For further simplification, let

$$\bar{b}_{i(k)} = \alpha_k b_{i(k)} + (1 - \alpha_k)\langle a^{i(k)}, x^k \rangle.$$

Note that, when $c_k = \beta_k$, then

$$x^{k+1} = P_{H(k)}(x^k).$$

When $c_k \neq \beta_k$, x^{k+1} is also well defined, since Eq. (8) has a unique solution x^{k+1} for any value of c_k when f is a Bregman function, as can be easily checked. Also, by strong zone consistency, $x^{k+1} \in S$ when $c_k = \beta_k$, because $H(k)$ is parallel to $H_{i(k)}$ and lies between $H_{i(k)}$ and x^k, since $0 \leq \alpha_k \leq 1$. When $c_k \neq \beta_k$, x^{k+1} also belongs to S, since it will be shown later (Proposition 4.4) that, in that case, x^{k+1} is the Bregman projection of x^k on a hyperplane parallel to $H(k)$ hence to $H_{i(}0_{k)}$, lying between x^k and $H(k)$, and strong zone consistence applies again.

4. Convergence Results

We give now a complete proof of convergence of the relaxed Bregman's method under almost cyclical control. We follow the convergence proof of Censor and Lent (Ref. 6), which is a streamlined version (with the addtion of almost cyclical control) of Bregman's original convergence proof (Ref. 5). Steps 1, 3, 4 of Censor and Lent are kept virtually untouched and appear as Proposition 4.2, Proposition 4.7, and Corollary 4.3 in our proof. The remaining steps of Censor and Lent's proof are not applicable to the relaxed case. Our Proposition 4.10 substitutes for Proposition 4.1 and Lemmas 7.1, 7.2, 7.3 in Ref. 6.

Proposition 4.1. Assume that

$$H_i = \{x \in \mathbb{R}^n : \langle a, x \rangle = b_i\}, \qquad i = 1, 2,$$

are two parallel hyperplanes in $\mathbb{R}^n$. Then, if $y \in S$ and $f \in B$ is zone consistent with respect to both hyperplanes, we have

$$\Pi_{H_1}(y) < \Pi_{H_2}(y) \Leftrightarrow b_1 < b_2. \tag{13}$$

Proof. See Lemma 3.4 in Ref. 6, observing that $\Pi_{H_1}(y) = \Pi_{H_2}(y)$ iff $b_1 = b_2$. □

The following propositions describe the behavior of the sequence defined by the algorithm and lead to our main convergence result.

Proposition 4.2. $\nabla f(x^k) = -A^T z^k$, for all k.

Proof. It proceeds by induction. The result is true for $k = 0$, by (7). Assume that it is true for any k. Using (8) and (9),

$$\begin{aligned} \nabla f(x^{k+1}) &= \nabla f(x^k) + c_k a^{i(k)} \\ &= -A^T(z^k - c_k a^{i(k)}) = -A^T z^{k+1}. \end{aligned}$$

□

Proposition 4.3. $z^k \geq 0$, for all k.

Proof. It proceeds again by induction. The result is true for $k = 0$, by definition of z^0. Assume that it is true for any k. Then,

$$z_i^{k+1} = \begin{cases} z_i^k \geq 0, & \text{if } i \neq i(k), \\ z_i^k - c_k \geq 0, & \text{if } i = i(k), \end{cases}$$

because of (10). □

Proposition 4.4. The following results hold:

(i) $x^{k+1} = P_{\tilde{H}(k)}(x^k)$, $c_k = \Pi_{\tilde{H}(k)}(x^k)$, where

$$\tilde{H}(k) = \{x \in \mathbb{R}^n : \langle a^{i(k)}, x \rangle = \gamma_k b_{i(k)} + (1 - \gamma_k)\langle a^{i(k)}, x^k \rangle\},$$

for some γ_k such that $0 \leq \gamma_k \leq \alpha_k$;

(ii) if $c_k = \beta_k$, then $\gamma_k = \alpha_k$ and $\tilde{H}(k) = H(k)$;

(iii) if $c_k \neq \beta_k$, then $\gamma_k < \alpha_k$ and $\langle a^{i(k)}, x^{k+1} \rangle < \tilde{b}_{i(k)}$;

(iv) if $\gamma_k = 0$, then $x^{k+1} = x^k$.

Proof. Note that x^{k+1} is the Bregman projection of x^k on the hyperplane

$$\tilde{H}(k) = \{x \in \mathbb{R}^n : \langle a^{i(k)}, x \rangle = \langle a^{i(k)}, x^{k+1} \rangle\},$$

because x^{k+1} belongs to $\tilde{H}(k)$ and together with (8), this has the form prescribed by (5). So,

$$c_k = \Pi_{\tilde{H}(k)}(x^k).$$

It remains to be seen that the right-hand side in the definition of $\tilde{H}(k)$, i.e., $\langle a^{i(k)}, x^{k+1}\rangle$, has the desired form.

It is clear that, if $c_k = \beta_k$, then $\tilde{H}(k) = H(k)$, and we may take $\gamma_k = \alpha_k$, proving at the same time part (ii). On the other hand, if $c_k \neq \beta_k$, by (10) and Proposition 4.3,

$$0 \leq z^k_{i(k)} = c_k < \beta_k. \tag{14}$$

Now,

$$\begin{aligned} 0 &\leq D(x^{k+1}, x^k) + D(x^k, x^{k+1}) \\ &= \langle \nabla f(x^{k+1}) - \nabla f(x^k), x^{k+1} - x^k\rangle \\ &= c_k\langle a^{i(k)}, x^{k+1} - x^k\rangle. \end{aligned} \tag{15}$$

If $c_k = 0$, then $x^{k+1} = x^k$ by (15) and Proposition 2.1, and we may take $\gamma_k = 0$. From (14), the only remaining case is $0 < c_k < \beta_k$, i.e.,

$$0 < \Pi_{\tilde{H}(k)}(x^k) < \Pi_{H(k)}(x^k).$$

By Proposition 4.1,

$$\langle a^{i(k)}, x^k\rangle < \langle a^{i(k)}, x^{k+1}\rangle < \bar{b}_{i(k)},$$

and there is a $\gamma_k \in (0, \alpha_k)$ satisfying the lemma. This establishes parts (i) and (iii). For part (iv), observe that $\gamma_k = 0$ implies

$$\langle a^{i(k)}, x^{k+1} - x^k\rangle = 0.$$

From (15) and Proposition 2.1, get $x^{k+1} = x^k$. □

Let us now define the Lagrangian of problem (1),

$$L(x, z) = f(x) + \langle z, Ax - b\rangle,$$

and

$$d_k = L(x^{k+1}, z^{k+1}) - L(x^k, z^k). \tag{16}$$

Proposition 4.5. We have

$$D(x^{k+1}, x^k) + (1 - \gamma_k)D(x^k, x^{k+1}) = \gamma_k d_k.$$

Proof. Observe that

$$\begin{aligned}\gamma_k d_k &= \gamma_k[\langle A^T(z^{k+1}-z^k), x^{k+1}\rangle - \langle z^{k+1}-z^k, b\rangle + D(x^{k+1}, x^k)]\\ &= \gamma_k[\langle \nabla f(x^k) - \nabla f(x^{k+1}), x^{k+1}\rangle - \langle z^{k+1}-z^k, b\rangle + D(x^{k+1}, x^k)]\\ &= \gamma_k[c_k(b_{i(k)} - \langle a^{i(k)}, x^{k+1}\rangle) + D(x^{k+1}, x^k)]. \end{aligned} \tag{17}$$

By Proposition 4.4,

$$x^{k+1} = P_{\tilde{H}(k)}(x^k),$$

so

$$\begin{aligned}\langle a^{i(k)}, x^{k+1}\rangle &= \gamma_k b_{i(k)} + (1-\gamma_k)\langle a^{i(k)}, x^k\rangle \Rightarrow \gamma_k\langle a^{i(k)}, x^{k+1}\rangle\\ &= \gamma_k b_{i(k)} + (1-\gamma_k)\langle a^{i(k)}, x^k - x^{k+1}\rangle \Rightarrow \gamma_k(b_{i(k)} - \langle a^{i(k)}, x^{k+1}\rangle)\\ &= (1-\gamma_k)\langle a^{i(k)}, x^{k+1} - x^k\rangle. \end{aligned} \tag{18}$$

Substituting (18) into (17), we have

$$\begin{aligned}\gamma_k d_k &= (1-\gamma_k)c_k\langle a^{i(k)}, x^{k+1}-x^k\rangle + \gamma_k D(x^{k+1}, x^k)\\ &= (1-\gamma_k)\langle \nabla f(x^{k+1}) - \nabla f(x^k), x^{k+1}-x^k\rangle + \gamma_k D(x^{k+1}, x^k)\\ &= (1-\gamma_k)(D(x^{k+1}, x^k) + D(x^k, x^{k+1})) + \gamma_k D(x^{k+1}, x^k)\\ &= (1-\gamma_k)D(x^k, x^{k+1}) + D(x^{k+1}, x^k). \end{aligned}$$

□

Corollary 4.1. $d_k \geq 0$, for all k.

Proof. If $\gamma_k = 0$, from Proposition 4.4(iv) $x^{k+1} = x^k$, so $c_k = 0$ and $z^{k+1} = z^k$. So,

$$d_k = L(x^{k+1}, z^{k+1}) - L(x^k, z^k) = 0.$$

Otherwise, from Proposition 4.4(i), $0 < \gamma_k \leq \alpha_k \leq 1$, and the result follows from Propositions 4.5 and 2.1. □

We prove now that $L(x^k, z^k)$ is a bounded sequence.

Proposition 4.6. We have

$$L(x^k, z^k) \leq f(z) - D(z, x^k) \leq f(z), \qquad \text{for all } z \in C \cap \bar{S}.$$

Proof. Take $z \in C \cap \bar{S}$. Since $Az \leq b$ and $z^k \geq 0$,

$$\langle A^T z^k, z - x^k\rangle = \langle z^k, Az\rangle - \langle z^k, Ax^k\rangle \leq \langle z^k, b - Ax^k\rangle. \tag{19}$$

On the other hand,

$$\begin{aligned}D(z, x^k) &= f(z) - f(x^k) - \langle \nabla f(x^k), z - x^k\rangle = f(z) - f(x^k) + \langle A^T z^k, z - x^k\rangle\\ &\leq f(z) - f(x^k) + \langle z^k, b - Ax^k\rangle = f(z) - L(x^k, z^k),\end{aligned}$$

producing the left inequality. The right one follows from Proposition 2.1. □

Corollary 4.2. The sequence $L(x^k, z^k)$ converges.

Proof. By Corollary 4.1, the sequence is increasing. By Proposition 4.6, it is bounded above by $f(z)$ for any $z \in C \cap \bar{S}$. Then, it converges. □

Corollary 4.3. The sequence $\{x_k\}_{k=0}^{\infty}$ is bounded.

Proof. Take any $z \in C \cap \bar{S}$. Applying Corollary 4.1 recursively and Proposition 4.6, we have

$$D(z, x^k) \leq f(z) - L(x^0 z^0).$$

So

$$x^k \in \{x \in \mathbb{R}^n: D(z, y) \leq f(z) - L(x^0, z^0)\}.$$

Such set is bounded by Definition 2.1(iv). □

It follows from Corollary 4.3 that $\{x^k\}$ has a convergent subsequence. From now on, let $\{x^{j_k}\}_{k=1}^{\infty}$ be a convergent subsequence of $\{x^k\}$.

Proposition 4.7. $\lim_{k\to\infty} D(x^{k+1}, x^k) = 0$.

Proof. From Propositions 2.1 and 4.5, since $\gamma_k \leq \alpha_k \leq 1$, we have

$$0 \leq D(x^{k+1}, x^k) \leq \gamma_k d_k \leq d_k \xrightarrow[k\to\infty]{} 0,$$

by Corollary 4.2. □

Proposition 4.8. Assume that $x^{jk} \to_{k\to\infty} x^*$. Fix t and take a sequence $\{l_k\}_{k=1}^{\infty}$ with $l_k \in \{1, \ldots, t\}$. Then, $x^{j_k+l_k} \to_{k\to\infty} x^*$.

Proof. Consider first the t sequences $\{x^{j_k+l}\}_{k=1}^{\infty}$, with $1 \leq l \leq t$. Since they are subsequences of $\{x^k\}$, they are all bounded. Observe that, by Proposition 4.7,

$$D(x^{j_k+s+1}, x^{j_k+s}) \xrightarrow[k\to\infty]{} 0, \qquad \text{for all } s \in \{1, \ldots, t\}.$$

So, applying recursively Definition 2.1(vi), we conclude that

$$x^{j_k+l} \xrightarrow[k\to\infty]{} x^*, \qquad 0 \leq l \leq t.$$

Interlace these $t+1$ sequences, forming the sequence

$$x^{j_1}, x^{j_1+1}, \ldots, x^{j_1+t}, x^{j_2}, x^{j_2+1}, \ldots, x^{j_2+t}, \ldots, x^{j_k}, x^{j_k+1}, \ldots, x^{j_k+t}, \ldots. \tag{20}$$

This sequence converges to x^*, and $\{x^{j_k+l_k}\}$ is a subsequence of it. □

Proposition 4.9. If $x^{j_k} \xrightarrow[k\to\infty]{} x^*$, then $x^* \in C$.

Proof. Take $p \in \{1, \ldots, m\}$ and $l_k \in \{1, \ldots, r\}$, such that

$$i(j_k + l_k) = p \tag{21}$$

[r is the constant of almost cyclicality in the definition of the control sequence $i(k)$]. By Proposition 4.8 (with $r = t$),

$$x^{j_k + l_k} \xrightarrow[k\to\infty]{} x^*.$$

Take a subsequence $\{x^{s_k}\}$ of $\{x^{j_k+l_k}\}$ such that

$$\gamma_{s_k} \xrightarrow[k\to\infty]{} \gamma, \qquad \alpha_{s_k} \xrightarrow[k\to\infty]{} \alpha \geq \epsilon. \tag{22}$$

By (21), we have

$$\langle a^p, x^{s_k+1}\rangle = \gamma_{s_k} b_p + (1 - \gamma_{s_k})\langle a^p, x^{s_k}\rangle.$$

Taking limits as $k \to \infty$ above, we have

$$\langle a^p, x^*\rangle = \gamma b_p + (1-\gamma)\langle a^p, x^*\rangle \Rightarrow \gamma(\langle a^p, x^*\rangle - b_p) = 0.$$

If $\gamma \neq 0$,

$$\langle a^p, x^*\rangle = b_p. \tag{23}$$

If $\gamma = 0$, we have $\gamma_{s_k} \neq \alpha_{s_k}$ for big enough k, because of (22). So,

$$\langle a^p, x^{s_k+1}\rangle < \alpha_{s_k} b_p + (1 - \alpha_{s_k})\langle a^p, x^{s_k}\rangle, \tag{24}$$

because of Proposition 4.4(iii). Taking limits in (24) as $k \to \infty$,

$$\langle a^p, x^*\rangle \leq \alpha b_p + (1-\alpha)\langle a^p, x^*\rangle,$$

or

$$0 \leq \alpha(b_p - \langle a^p, x^*\rangle) \Rightarrow \langle a^p, x^*\rangle \leq b_p. \tag{25}$$

From (23) and (25), x^* satisfy the pth-constraint. Since p is an arbitrary index, we get $Ax^* \leq b$; so, $x^* \in C$. $\square$

For $x \in C$, define

$$I_1(x) = \{i : \langle a^i, x\rangle < b_i\},$$

$$I_2(x) = \{i : \langle a^i, x\rangle = b_i\}.$$

Proposition 4.10. Assume that $x^{j_k} \to x^*$. Then, $z_p^{j_k+r+1} = 0$, for all p in $I_1(x^*)$ (r is the constant of almost cyclicality).

Proof. Let

$$\rho = (\epsilon/5) \min_{p \in I_1(x^*)} \{[b_p - \langle a^p, x^* \rangle]/\|a^p\|\} > 0.$$

By Proposition 4.8, there exists k_0 such that

$$\|x^{j_k+l} - x^*\| < \rho, \qquad \text{for all } l \in \{0, \dots, r+1\} \text{ and } k \geq k_0.$$

Given $p \in I_1(x^*)$, define

$$l_k = \max_{0 \leq l \leq r} \{l : i(j_k + l) = p\}.$$

The existence of l_k is guaranteed by almost cyclicality. Let $s_k = l_k + j_k$. Assume $\beta_{s_k} = c_{s_k}$. Then,

$$\langle a^p, x^{s_k+1} \rangle = \alpha_{s_k} b_p + (1 - \alpha_{s_k}) \langle a^p, x^{s_k} \rangle$$
$$\Rightarrow \langle a^p, x^{s_k+1} - x^{s_k} \rangle = \alpha_{s_k}(b_p - \langle a^p, x^{s_k} \rangle).$$

Therefore,

$$\begin{aligned} \alpha_{s_k}(b_p - \langle a^p, x^{s_k} \rangle) &< \langle a^p, x^{s_k+1} - x^{s_k} \rangle + \alpha_{s_k} \langle a^p, x^{s_k} - x^* \rangle \\ &\leq \|a^p\| (\|x^{s_k+1} - x^{s_k}\| + \alpha_{s_k} \|x^{s_k} - x^*\|) \\ &\leq \|a^p\| (\|x^{s_k+1} - x^*\| + (1 + \alpha_{s_k}) \|x^{s_k} - x^*\|) \\ &\leq \|a^p\| (2 + \alpha_{s_k}) \rho \leq 4\rho \|a^p\|, \qquad \text{for } k \geq k_0. \end{aligned} \tag{26}$$

So

$$4\rho > \epsilon([b_p - \langle a^p, x^* \rangle]/\|a^p\|) \geq 5\rho,$$

a contradiction. It follows that

$$\beta_{s_k} \neq c_{s_k} \Rightarrow z_p^{s_k} = c_{s_k} \Rightarrow z_p^{s_k+1} = 0.$$

By the definition of l_k, the index p is not used in iteration $j_k + l$ for $l_k < l \leq r$, so $z_p^{j_k+l}$ remains unaffected. We conclude that

$$z_p^{j_k+r+1} = 0.$$

Observe that (26) holds even when $1 \leq \alpha_{s_k} \leq 2$, i.e., for the overrelaxed case. □

Proposition 4.11. If $x^{j_k} \to x^*$ and $z_p^{j_k} = 0$, for $p \in I_1(x^*)$ and all $k \geq 0$, then x^* is a solution of problem (1) with the additional constraint $x \in \bar{S}$.

Proof. Observe that

$$\langle z^{j_k}, Ax^{j_k} - b\rangle = \langle A^T z^{j_k}, x^{j_k} - x^*\rangle,$$

because

$$z_p^{j_k} = 0, \qquad p \in I_1(x^*),$$

$$\langle a^p, x^*\rangle = b_p, \qquad p \in I_2(x^*).$$

Then,

$$\begin{aligned}\langle z^{j_k}, Ax^{j_k} - b\rangle &= -\langle \nabla f(x^{j_k}), x^{j_k} - x^*\rangle \\ &= -D(x^*, x^{j_k}) + f(x^*) - f(x^{j_k}) \underset{k\to\infty}{\longrightarrow} 0,\end{aligned}$$

using Definition 2.1(i), (iii), (v). So, by continuity of f and Proposition 4.6, for any $z \in C \cap \bar{S}$, we have

$$\begin{aligned}f(x) &\geq \lim_{k\to\infty} L(x^{j_k}, z^{j_k}) \\ &= \lim_{k\to\infty} [f(x^{j_k}) + \langle z^{j_k}, Ax^{j_k} - b\rangle] = f(x^*).\end{aligned}$$

It follows that x^* is a minimizer of f on $\bar{S} \cap C$. □

Let now problem (1′) be

$$\min f(x), \tag{1'a}$$

$$\text{s.t. } Ax \leq b, \tag{1'b}$$

$$x \in \bar{S}. \tag{1'c}$$

We present our main convergence results as the following theorem.

Theorem 4.1. Let the following conditions hold:

(i) $f \in B$;
(ii) f is strongly zone consistent with respect to each H_i, $i \in \{1, \ldots, m\}$;
(iii) $\{i(k)\}_{k=0}^{\infty}$ is an almost cyclical control sequence on $\{1, \ldots, m\}$ with constant r;
(iv) $z^0 \in \mathbb{R}_+^m$;
(v) $C \cap \bar{S} \neq \varnothing$.

Then, any sequence generated by (7)–(12) convergex to a point $x^* \in \bar{S}$ which is the solution of problem (1′).

Proof. By Corollary 4.3, there is a convergent subsequence

$$x^{j_k} \xrightarrow[k\to\infty]{} x^*.$$

By Proposition 4.9, $x^* \in C \cap \bar{S}$. By Proposition 4.10, there is another convergent subsequence

$$x^{s_k} \xrightarrow[k\to\infty]{} x^*,$$

satisfying the hypothesis of Proposition 4.11. So, x^* is a solution of problem (1′). By the strict convexity of f [Definition 2.1(ii)], such a solution is unique. Hence, all convergent subsequences of $\{x^k\}$ have the same limit, namely x^*. Ift follows that

$$x^k \xrightarrow[k\to\infty]{} x^*. \qquad \square$$

Observation 4.1. The hypothesis that $\alpha_k \leq 1$ is used only in the proof of Proposition 4.7 and Corollary 4.1. Remove the condition $\alpha_k \leq 1$, but assume that, for all k:

$$D(x^k, x^{k+1}) \leq D(x^{k+1}, x^k), \tag{27}$$

$$0 < \epsilon_1 \leq \alpha_k \leq \epsilon_2 < 2, \qquad \epsilon_1, \epsilon_2 > 0. \tag{28}$$

By Proposition 4.5,

$$\begin{aligned}\gamma_k d_k &= D(x^{k+1}, x^k) - D(x^k, x^{k+1}) + (2-\gamma_k)D(x^k, x^{k+1})\\ &\geq (2-\gamma_k)D(x^k, x^{k+1}) \geq (2-\alpha_k)D(x^k, x^{k+1}) \geq 0,\end{aligned}$$

and Corollary 4.1 holds. As in the proof of Proposition 4.7,

$$\begin{aligned}0 &\leq (2-\epsilon_2)D(x^k, x^{k+1})\\ &\leq (2-\alpha_k)D(x^k, x^{k+1})\\ &\leq \gamma_k d_k \xrightarrow[k\to\infty]{} 0\\ &\Rightarrow D(x^k, x^{k+1}) \xrightarrow[k\to\infty]{} 0.\end{aligned}$$

So, if (27) holds, the algorithm converges with relaxation parameters α_k subject to (28), i.e., also in the overrelaxed case (disregarding the zone consistency issue).

A sufficient condition for (27) to hold, depending only on f, is that

$$D(x, y) = D(y, x), \qquad \text{for all } x, y \in S. \tag{29}$$

Condition (29) is true for

$$f(x) = x^T Q x + q^T x, \tag{30}$$

with $Q \in \mathbb{R}^{n \times n}$ symmetric, positive definite and $q \in \mathbb{R}^n$. In summary, the convergence of Hildreth's quadratic programming algorithm with almost cyclical control (Ref. 8) and relaxation parameters α_k satisfying (28) is a particular case of our proof (in this case, the zone is all $\mathbb{R}^n$).

We conjecture that (30) is the only solution of condition (29).

5. Characterization of a Family of Bregman Functions

Although Bregman's method (Ref. 5) was devised for solving problem (1) when f is a quadratic function or the maximum entropy function, it is important to provide easily checkable conditions guaranteeing that a function f belongs to the set B, which allow the extension of possible applications of the algorithm. In this section, we provide such a condition, for functions defined on all R^n.

Let $f: R^n \to R$ such that:

(I) f is twice continuously differentiable and strictly convex;

(II) $\lim_{\|x\| \to \infty}[f(x)/\|x\|] = \infty$.

Theorem 5.1. If f satisfies (I) and (II), then $f \in B$.

The proof of Theorem 5.1 requires four lemmas. Before presenting them, observe that, in Definition 2.1, conditions (v) and (vi) hold trivially in the interior of the domain as a consequence of condition (i) and need to be checked only on the boundary of S. So, when considrering functions defined on all $\mathbb{R}^n$, those conditions are void. Also, in this case, condition (ii) is a consequence of (I), which also implies conditions (i) and (iii) of Definition 2.1. The only trouble lies in condition (iv).

Consider a function $g: \mathbb{R}^n - \{0\} \to \mathbb{R}$ such that:

(III) g is twice continuously differentiable;

(IV) $\lim_{\|x\| \to \infty} g(x) = \infty$;

(V) the function $h: \mathbb{R}^n \to \mathbb{R}$ defined by $h(x) = \|x\| g(x)$, if $x \neq 0$, and $h(0) = 0$, is strictly convex.

For $x \neq 0$, define $\varphi_x: \mathbb{R}_{>0} \to \mathbb{R}$ as

$$\varphi_x(\lambda) = \langle x, \nabla f(\lambda x) \rangle,$$

where $\mathbb{R}_{>0}$ is the set of positive real numbers. Because of (III), φ_x is continuously differentiable.

Lemma 5.1. For any $x \neq 0$, $\varphi_x(1) \leq 0$ implies $\varphi_x(\lambda) \leq 0$, for all $\lambda \in (0, 1]$.

Proof. Assume that there exists a $\lambda_0 \in (0, 1)$ such that

$$\varphi_x(\lambda_0) > 0. \tag{31}$$

Let

$$\lambda_1 = \inf\{\lambda > \lambda_0 \colon \varphi_x(\lambda) = 0\}.$$

So,

$$\varphi_x(\lambda) > 0, \qquad \text{for } \lambda \in (\lambda_0, \lambda_1). \tag{32}$$

Since φ_x is continuous, λ_1 is well defined and

$$\varphi_x(\lambda_1) = 0; \tag{33}$$

also,

$$\varphi'_x(\lambda_1) \leq 0, \tag{34}$$

because otherwise

$$\varphi_x(\lambda) < \varphi_x(\lambda_1), \qquad \text{for } \lambda \in (\lambda_1 - \epsilon, \lambda_1),$$

in contradiction with (32). From (33), (34), we have

$$\begin{aligned} 0 &\geq 2\lambda_1 \varphi_x(\lambda_1) + \lambda_1^2 \varphi'_x(\lambda_1) = 2\lambda_1 x^T \nabla g(\lambda_1 x) + \lambda_1^2 x^T \nabla^2 g(\lambda_1 x) x \\ &= 2y^T \nabla g(y) + y^T \nabla^2 g(y) y, \qquad \text{with } y = \lambda_1 x, \end{aligned} \tag{35}$$

where $\nabla^2 g$ is the Hessian matrix of g. From condition (V), we have

$$0 < y^T \nabla^2 h(y) y = 2y^T \nabla g(y) + y^T \nabla^2 g(y) y,$$

in contradiction with (35). □

Lemma 5.2. For every $x \neq 0$, there exists $\mu > 1$ such that $\varphi_x(\mu) > 0$.

Proof. Otherwise,

$$0 \geq \varphi_x(\mu) = x^T \nabla g(\mu x), \qquad \text{for all } \mu > 1.$$

Then, by the mean-value theorem,

$$g(\mu x) \leq g(x), \qquad \text{for all } \mu > 1,$$

in contradiction with (IV). □

Let

$$V = \{x \in \mathbb{R}^n - \{0\} \colon x^T \nabla g(x) \leq 0\} = \{x \in \mathbb{R}^n - \{0\} \colon \varphi_x(1) \leq 0\},$$

and let

$$\hat{V} = V \cap \{x \in \mathbb{R}^n \colon \|x\| \geq 1\}.$$

By continuity of φ_x, $\hat{V}$ is closed.

Lemma 5.3. V is a bounded set.

Proof. Suppose that there exists a sequence $\{x^k\}_{k=1}^{\infty} \subset V$ such that $\|x^k\| \longrightarrow_{k\to\infty} \infty$. Let x^* be a cluster point of $\{x^k/\|x^k\|\}$. Take any $\mu \geq 1$. Given $\epsilon > 0$, take k such that

$$\|x^k/\|x^k\| - x^*\| < \epsilon/\mu \quad \text{and} \quad \|x^k\| > \mu.$$

Let

$$\lambda = \mu/\|x^k\| < 1.$$

So,

$$\|\lambda x^k - \mu x^*\| < \epsilon. \tag{36}$$

By Lemma 5.1, $\lambda x^k \in V$. Since $\mu \geq 1$, $\lambda x^k \in \hat{V}$. Since $\hat{V}$ is closed, (36) implies that $\mu x^* \in \hat{V} \subset V$. So,

$$\varphi_{x^*}(\mu) \leq 0 \qquad \text{for all } \mu \geq 1,$$

in contradiction with Lemma 5.2. □

Lemma 5.4. If f satisfies conditions (I) and (II), then

(i) $\lim_{\|x\|\to\infty}(f(x) - a^T x) = \infty$, for all $a \in \mathbb{R}^n$;
(ii) $\lim_{\|x\|\to\infty}[(x-a)^T \nabla f(x) - f(x)] = \infty$, for all $a \in \mathbb{R}^n$.

Note. Part (i) of this lemma is proposed as an exercise in Ref. 9, p. 110.

Proof. (i) Take M such that

$$f(x)/\|x\| \geq 2\|a\|, \qquad \text{for } \|x\| > M,$$

using (II). Then,

$$\|a\|\,\|x\| \leq f(x) - \|a\|\,\|x\| \leq f(x) - a^T x,$$

and the left-hand side tends to infinity as $\|x\| \to \infty$, if $a \neq 0$. If $a = 0$, (i) follows directly from (II).

(ii) For any $\rho > 0$, consider the function

$$\bar{f}(x) = f(x+a) + \rho.$$

Clearly, $\bar{f}$ satisfies (I) and (II). So, $g(x) = \bar{f}(x)/\|x\|$ satisfies (III), (IV), (V). From Lemma 5.3,

$$V = \{x \in \mathbb{R}^n - \{0\}: x^T \nabla g(x) \leq 0\}$$

is bounded. Take M such that, for $\|y\| > M$,

$$y^T \nabla g(y) \geq 0.$$

Then, for $\|y\| > M$,

$$0 \leq y^T \nabla g(y) = (1/\|y\|)[y^T \nabla \bar{f}(y) - \bar{f}(y)] \Rightarrow 0 \leq y^t \nabla \bar{f}(y) - \bar{f}(y). \tag{37}$$

Let $x = y + a$. It follows from (37) that, for $\|x\| \geq M + \|a\|$,

$$0 \leq (x-a)^T \nabla f(x) - f(x) - \rho \Rightarrow \rho \leq (x-a)^T \nabla f(x) - f(x).$$

Since ρ is arbitrary, (ii) holds. □

Proof of Theorem 5.1. As noted before, only condition (iv) in Definition 2.1 has to be checked. We have

$$L_1(y, \alpha) = \{x: f(x) - \nabla f(y)^T x \leq \alpha + f(y) - \nabla f(y)^T y\}.$$

Applying Lemma 5.4(i) with $a = \nabla f(y)$, we conclude that

$$f(x) - \nabla f(y)^T x \xrightarrow[\|x\| \to \infty]{} \infty.$$

So, $L_1(y, \alpha)$ is bounded for all $y \in R^n$. For

$$L_2(x, \alpha) = \{y: (y-x)^T \nabla f(y) - f(y) \leq \alpha - f(x)\},$$

apply Lemma 5.4 (ii) with $a = x$, and conclude that

$$(y-x)^T \nabla f(y) - f(y) \xrightarrow[\|y\| \to \infty]{} \infty.$$

So, $L_2(x, \alpha)$ is bounded for all $x \in \mathbb{R}^n$. □

Condition (II) is not a necessary condition for a function to belong to B, even for twice continuously differentiable functions defined in all $\mathbb{R}^n$, as the following example, with $n = 1$, shows:

$$f(x) = \begin{cases} \frac{1}{2}(x^2 - 4x + 3), & \text{if } x \leq 1, \\ -\log x, & \text{if } x \geq 1. \end{cases}$$

It is straightforward to verify that f is a twice continuously differentiable Bregman function; however,

$$\lim_{x \to +\infty} [f(x)/x] = 0.$$

6. Final Remarks

The question of convergence of algorithm (7)–(12) in the overrelaxed case (i.e., with $1 \leq \alpha_k \leq \epsilon_2 < 2$) remains open. Observe that an *a priori* restriction in such case is that zone consistency is required with respect to all hyperplanes parallel to any constraint lying between any point of S and its reflection on that constraint, rather than the constraint itself, just in order to have all the iterates within the domain of f. Such a condition seems almost infeasible unless $S = \mathbb{R}^n$ (it is not satisfied by the $x \ln x$ entropy function, for instance). But it may be the case that the overrelaxed algorithm converges when f is defined on all $\mathbb{R}^n$, or with an additional condition on f less restrictive than condition (29), which, as noted before, is conjectured to hold only in the quadratic case.

Another open problem is to find easily checkable conditions for f to be a Bregman function, even when $S \nsubseteq \mathbb{R}^n$, or at least for f to satisfy condition (iv) of Definition 2.1 in such a case. A possible useful observation in connection with this issue is the fact (rather immediate) that the set of Bregman functions is a positive cone, i.e., that a positive linear combination of Bregman functions, all with the same zone S, is a Bregman function with zone S.

Note that our convergence proof can be trivially adjusted to the problem of equality constraints, in which case Eq. (10) is replaced by

$$c_k = \beta_k,$$

and the dual variables [i.e., Eq. (9)] are eliminated. The sequence generated by this algorithm, however, is not the same as the sequence which results from converting each equality into two inequalities and then applying (7)-(12), if the relaxation parameters α_k are not all equal to one.

References

1. LAMOND, B., and STEWART, N. F., *Bregman's Balancing Method,* Transportation Research, Vol. 15B, pp. 239-248, 1981.
2. HERMAN, G. T., and LENT, A., *A Family of Iterative Quadratic Optimization Algorithms for Pairs of Inequalities, with Application in Diagnostic Radiology,* Mathematical Programming Study, Vol. 9, pp. 15-29, 1978.
3. HERMAN, G. T., and LENT, A., *Iterative Reconstruction Algorithms,* Computers in Biology and Medicine, Vol. 6, pp. 273-294, 1976.
4. CENSOR, Y., *Row-Action Methods for Huge and Sparse Systems and Their Applications,* SIAM Review, Vol. 23, pp. 444-464, 1981.
5. BREGMAN, L. M., *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, Vol. 7, pp. 200-217, 1967.
6. CENSOR, Y., and LENT, A., *An Iterative Row-Action Method for Interval Convex Programming,* Journal of Optimization Theory and Applications, Vol. 34, pp. 321-353, 1981.

7. HERMAN, G. T., LENT, A., and LUTZ, P. H., *Relaxation Methods for Image Reconstruction*, Communications of the Association for Computing Machinery, Vol. 21, pp. 152–158, 1978.
8. LENT, A., and CENSOR, Y., *Extensions of Hildreth's Row-Action Method for Quadratic Programming*, SIAM Journal on Control and Optimization, Vol. 18, pp. 444–454, 1980.
9. ORTEGA, J. M., and RHEINBOLDT, W. S., *Iterative Solutions of Nonlinear equations in Several Variables*, Academic Press, New York, New York, 1970.